Journal of Hydrology 575 (2019) 354–380

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Building a puzzle to solve a riddle: A multi-scale disaggregation approach for multivariate stochastic processes with any marginal distribution and correlation structure

Ioannis Tsoukalas*, Andreas Efstratiadis, Christos Makropoulos

Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Heroon Polytechneiou 5, 15780 Zographou, Greece

ARTICLE INFO

This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Uwe Haberlandt, Associate Editor

Keywords:
Stochastic processes
Nataf-based models
Non-Gaussian distributions
Correlation structures
Multivariate simulation
Disaggregation

ABSTRACT

The generation of hydrometeorological time series that exhibit a given probabilistic and stochastic behavior across multiple temporal levels, traditionally expressed in terms of specific statistical characteristics of the observed data, is a crucial task for risk-based water resources studies, and simultaneously a *puzzle* for the community of stochastics. The main challenge stems from the fact that the reproduction of a specific behavior at a certain temporal level does not imply the reproduction of the desirable behavior at any other level of aggregation. In this respect, we first introduce a pairwise coupling of Nataf-based stochastic models within a disaggregation scheme, and next we propose their *puzzle-type* configuration to provide a generic stochastic simulation framework for multivariate processes exhibiting any distribution and any correlation structure. Within case studies we demonstrate two characteristic configurations, i.e., a three-level one, operating at daily, monthly and annual basis, and a two-level one to disaggregate daily to hourly data. The first configuration is applied to generate correlated daily rainfall and runoff data at the river basin of Achelous, Western Greece, which preserves the stochastic behavior of the two processes at the three temporal levels. The second configuration disaggregates daily rainfall, obtained from a meteorological station at Germany, to hourly. The two studies reveal the ability of the proposed framework to represent the peculiar behavior of hydrometeorological processes at multiple temporal resolutions, as well as its flexibility on formulating generic simulation schemes.

1. Introduction

Today, most water recourse studies employ Monte Carlo simulations, by running deterministic models that are driven by synthetic inputs, which are typically generated by stochastic models. In this context, the key requirement for extracting statistically consistent outcomes is the concise representation of the probabilistic behavior and stochastic structure of the input hydrometeorological processes (e.g., rainfall, runoff, temperature). It is well-known that these exhibit a significantly complex regime, the most prominent aspects of which are non-Gaussianity, intermittency, auto- and cross-dependence, as well as periodicity (Koutsoyiannis, 2005; Moran, 1970; Salas et al., 1980). All the above peculiarities dictate the specifications of a *good* simulation model.

Actually, the first two characteristics (non-Gaussianity and intermittency) are directly associated with the marginal properties of the process, and imply the need for a suitable probability distribution model. On the other hand, auto- and cross-dependencies are associated with the stochastic (joint) properties of the process, both in time and space, and point out the need for stochastic simulation models *per se*. In fact, if the physical processes to simulate were not (auto- or cross-) correlated, the problem would be substantially simpler, as the production of synthetic data would be trivially made by generating uniform numbers and then employing probability integral transformations. Finally, periodicity introduces further complexity, since it implies representing the processes as cyclostationary, thus differentiating their marginal and joint properties not only across different temporal scales but also across seasons (or systematically repeated time intervals, in general).

During more than a half century, the need for *good* synthetic data generators, to be used within risk-aware decision-making frameworks for design, assessment and operation of water resource systems (Celeste & Billib, 2009; Feng et al., 2017; Giuliani et al., 2014; Haberlandt et al., 2011; Koutsoyiannis & Economou, 2003; Paschalis et al., 2014;

* Corresponding author.
E-mail address: itsoukal@mail.ntua.gr (I. Tsoukalas).

https://doi.org/10.1016/j.jhydrol.2019.05.017
Received 21 November 2018; Received in revised form 26 March 2019; Accepted 5 May 2019
Available online 08 May 2019

Tsoukalas et al., 2016; Tsoukalas & Makropoulos, 2015a,b) has triggered numerous researchers for developing a plethora of stochastic approaches and associated modelling tools. These can be primarily classified into two broad categories, i.e., single-scale and multi-scale. The former ensure the reproduction of a set of statistical and stochastic properties at a unique time scale of interest, i.e., the time interval of simulation, while the latter attempt to simultaneously represent the desirable properties of the simulated data, as well as the properties of the aggregated data at coarser temporal scales.

The numerous single-scale simulation schemes that have been developed so far can be further classified into: 1) linear stochastic models, also known as time series generators (e.g., Bras & Rodríguez-Iturbe, 1985; Koutsoyiannis, 1999, 2000; Matalas & Wallis, 1976; Matalas, 1967; Salas et al., 1980; Thomas & Fiering, 1962); 2) point process models (Bo et al., 1994; Burton et al., 2008; Evin & Favre, 2008; Kaczmarska et al., 2014; Kilsby et al., 2007; Onof et al., 2000; Tarpanelli et al., 2012); 3) two-part models, i.e., product models of occurrence and amount that are represented as discrete and continuous processes, respectively (e.g., Ailliot et al., 2015; Baigorria & Jones, 2010; Breinl et al., 2013, 2015; Katz, 1977; Khalili et al., 2009; Richardson & Wright, 1984; Srikanthan & Pegram, 2009; Todorovic & Woolhiser, 1975; Wilks, 1998); 4) resampling methods (Caraway et al., 2014; Clark et al., 2004; Ilich, 2014; Ilich & Despotovic, 2008; Lall & Sharma, 1996; Mehrotra et al., 2006; Mehrotra & Sharma, 2007; Rajagopalan & Lall, 1999; Wójcik & Buishand, 2003); and 5) copula-based models (e.g., Bárdossy & Pegram, 2009; Chen et al., 2015; Hao & Singh, 2011, 2013; Jeong & Lee, 2015; Lee, 2017; Lee & Salas, 2011; Serinaldi, 2009).

By design, single-scale simulation models attempt to reproduce the desirable statistical and stochastic behavior within the synthetic data at the scale of simulation, yet they provide limited control to the properties of the same process, when aggregated at higher (coarser) time scales. It is well-known that the reproduction of the probabilistic and stochastic behavior of a process, expressed either in terms of a distribution function or a set of statistical properties, at a certain time scale does not ensure the reproduction of the associated characteristics of the aggregated process at any other time scale.

The necessity for *multi-scale consistency* has been early recognized by the hydrological community, through the pioneering work by Harms and Campbell (1967). Actually, from the first steps of Monte Carlo approaches in water resources it has been accepted that that the outcomes of stochastic analyses are associated with the overall statistical and stochastic behavior of the input hydrometeorological processes, which may extend far beyond the time interval of the underlying (deterministic) simulation model (see, Klemeš, 1981; Koutsoyiannis, 2005). For instance, the design and operation of large reservoir systems that employ overyear regulation, which are typically modelled in monthly intervals, is strongly dictated by the probabilistic and stochastic properties of the aggregated inflows, at the annual and even over-annual scales. Similarly, the outputs of continuous flood simulation models, driven by fine-time (e.g., hourly) rainfall series, are substantially affected by the sequence of accumulated rainfall, as the runoff production strongly depends on the antecedent soil moisture conditions. In this respect, multi-scale consistency in stochastic simulation can be regarded as an operational *sine qua non*.

Furthermore, multi-scale consistency is directly linked with the so-called issue of low-frequency variability or over-dispersion (i.e., the deficiency to reproduce the process' variance at higher time scales), which is encountered in many popular daily *weather-generation* models (e.g., Baigorria & Jones, 2010; Breinl et al., 2013, 2015; Brissette et al., 2007; Katz & Parlange, 1998; Khalili et al., 2009; Lee, 2017; Mehrotra et al., 2006; Mhanna & Bauwens, 2012; Serinaldi, 2009; Srikanthan & Pegram, 2009; Wilks, 1998; Wilks & Wilby, 1999).

Multi-scale simulation schemes, with the exception of few specifically designed models (e.g., Langousis & Koutsoyiannis, 2006; Rodriguez-Iturbe et al., 1987), is typically built upon the disaggregation paradigm. The essential element of disaggregation is the *additive property*, which enables the generation of multi-scale consistent time series via the transfer of information among different temporal scales. This implies that the sum of the generated variables at the lower level (e.g., monthly) at any period should add to the corresponding value at the higher level (e.g., annual), which is assumed known, either from observed or synthetic (simulated) data. This property distinguishes disaggregation from downscaling (e.g., Cannon, 2008; Lombardo et al., 2012; Wilks & Wilby, 1999), which focus on generating lower level time series that statistically resemble the properties of higher level ones, and not necessarily honor the additive constraint.

As already mentioned, the beginning of the quest (at least in hydrological domain) for multi-scale simulation models can be attributed to Harms and Campbell (1967), who developed a two-level version of the classical stochastic model by Thomas and Fiering (1962) that preserves some key statistical properties of the observed data at both the annual and monthly scale. A little later, the interest on such methods reinforced with the theoretical research on disaggregation by Valencia and Schakke (1973) and Mejia and Rousselle (1976). However, the proposed methods were fully general only for normally distributed variables, thus limiting their applicability to a relatively narrow range of processes and scales.

The next generation approaches offered multi-scale schemes that utilized the notion of the so-called *adjusting procedures* (Grygier & Stedinger, 1988; Harms & Campbell, 1967; Koutsoyiannis, 2001; Koutsoyiannis & Manetas, 1996; Stedinger & Vogel, 1984). These aimed at coupling single-scale simulation models of any type, operating independently at different time scales. The rationale is generating low-level synthetic data as auxiliary information, and next adjusting them to the known higher-level values, by using relatively simple algebraic transformations, such as the partial sums at the low level equal the values of the higher level. Koutsoyiannis and Manetas (1996) and Koutsoyiannis (2001) investigated several adjusting procedures, and also standardized the concept of *repetitive sampling* (kind of Monte Carlo approach), to ensure that the partial sums are close to the given values. This can be regarded as an informal method of conditional sampling, that can significantly improve the efficiency of such schemes (for an early formulation of this idea see Glasbey et al., 1995).

Adjusting procedures of varying complexity have been implemented within a number of disaggregation-based schemes, in order to couple single-scale simulation models (such as the ones described above) across various time scales. In particular, they were used within linear stochastic models (e.g., Allard & Bourotte, 2015; Efstratiadis et al., 2014; Koutsoyiannis et al., 2003; Lombardo et al., 2012; Segond et al., 2006; Tsoukalas et al., 2018b), point processes (e.g., Glasbey et al., 1995; Kossieris et al., 2015, 2016; Koutsoyiannis & Onof, 2001; Onof et al., 2005), two-part models (Evin et al., 2018; Shao et al., 2016), resampling methods (e.g., Breinl & Di Baldassarre, 2019; Lee et al., 2010; Li et al., 2018) and copula-based models (e.g., Gyasi-Agyei, 2011; Gyasi-Agyei & Melching, 2012). It is highlighted that the overall simulation capabilities of adjusting-based schemes are determined by the underlying simulation models, which consist the core data generation mechanism.

In addition, several modern schemes for establishing multi-scale consistency are built upon the concepts of scaling and multifractality (Kantelhardt et al., 2006; Tessier et al., 1996; Veneziano et al., 2006). Typically, these employ multiplicative random cascade models (Gupta & Waymire, 1990, 1993) to generate multi-scale consistent (in terms of typically high-order moments) realizations (Deidda et al., 1999; Menabde et al., 1997; Molnar & Burlando, 2005; Müller & Haberlandt, 2015, 2018; Olsson, 1998; Rupp et al., 2009). Recent works by Lombardo et al. (2012) and Pui et al. (2012) provide comparative studies involving such models, as well as alternative downscaling or disaggregation methods.

Besides the vast effort made so far, the quest for full generality and full consistency across multiple scales (herein the focus is on temporal scales, but may also concern spatial ones) still remains a *puzzle*. Recently, Tsoukalas et al. (2018a) highlighted that many of widespread schemes, including linear stochastic models with non-Gaussian innovations, point-process models and resampling techniques, emphasize the reproduction of a specific set of summary statistical characteristics, which arguably cannot capture the full behavior of a random process. As also shown, under some common conditions these may lead to bounded dependence patterns, which are not realistic (Tsoukalas et al., 2018d). On the other hand, two-part and copula-based models are actually able to explicitly account for the distributional properties of simulated processes, yet they are mainly designed to represent specific correlation structures. For instance, two-part models often neglect temporal dependencies, while copula-based schemes typically account for temporal dependencies spanning over only few time lags.

In this work, our focus is not on disaggregation *per se*, rather than we employ the flexibility provided by the concepts of repetitive sampling and adjusting procedures to link individual multivariate (or univariate) stochastic models, in order to represent the varying regime of hydrometeorological processes across multiple temporal scales. Our emphasis is to shift from the classical paradigm of describing a process in terms of few summary statistics (in particular, moments up to third order and low order correlation coefficients), to the explicit representation of its marginal and stochastic properties, in terms of distribution functions and theoretical correlation structures, respectively. This is accomplished by building upon a recently introduced (in hydrology) class of stochastic models, the so-called *Nataf-based* models, named after the contribution of Nataf (1962), and the further developments by Tsoukalas et al. (2017, 2018a, c), as well as by Papalexiou (2018). These are able to simulate multivariate, stationary and cyclostationary processes with any marginal distributions and any correlation structures. These properties allow for characterizing Nataf-based models as *good* single-scale stochastic simulators, and thus appropriate data generators within multi-scale adjusting-based schemes. Taking advantage of the above concepts, we propose a scale-free disaggregation approach for the pairwise coupling of Nataf-based models, aiming to preserve the desirable marginal distributions and correlation structures. The proposed approach is next referred to as *Nataf-based Disaggregation to Anything* (NDA). Eventually, a chain configuration of NDA allows for developing *puzzle-type*, i.e., modular, simulation schemes that ensure consistent simulations across any sequence of temporal scales. A sequence may involve any combination of stochastic models, stationary or cyclostationary.

This paper is organized as follows: Section 2 discusses the core generation procedure, through Nataf-based stochastic models, Section 3 describes the disaggregation-based coupling approach, i.e., NDA, designed to maintain consistency across pairwise scales, and Section 4 presents the puzzle-type multi-level simulation scheme. Next two sections demonstrate two characteristic configurations of this puzzle. Particularly, Section 5 presents a three-level multivariate scheme, applied for generating synthetic rainfall and runoff data at the river basin of Achelous, Greece. On the other hand, the configuration of Section 6 aims at synthesizing hourly rainfall data from a given (i.e., observed) daily record, thus illustrating the efficiency of the method against challenging disaggregation problems. Finally, Section 7 summarizes the overall modelling framework and discusses its potential applications and future improvements.

2. Nataf-based stochastic models

In a recent work, Tsoukalas et al. (2018a) highlighted the need for generalized generation schemes, which are able to represent processes from *any* distribution and *any* correlation structure. This has been also regarded as a shift in classical stochastic modelling, emphasizing on the reproduction of a finite set of *essential* statistical characteristics (cf. Matalas & Wallis, 1976), estimated from the historical data.

An effective and efficient handling of this requirement is offered by the so-called Nataf-based models (Tsoukalas, 2019; Tsoukalas et al., 2017, 2018a, 2018c) that use the notion of parent-Gaussian processes (Papalexiou, 2018). As the name suggests, these are built upon a quite old idea by Nataf (1962), who proposed the (non-linear) mapping of an auxiliary multivariate Gaussian distribution in order to construct the joint distribution of random variables with arbitrary marginal distributions (whose variance exists; an assumption implied throughout this paper). Using a similar rationale, it is possible to establish stochastic processes with any target marginal distribution and correlation structure (expressed in terms of Pearson's correlation coefficient) through the mapping of an appropriately specified auxiliary (stationary or cyclostationary) Gaussian process (Gp) with zero mean and unit variance, to which an *equivalent* correlation structure is assigned (see details below). The mapping operation is typically a non-linear function, often implemented through the inverse cumulative distribution function (ICDF). These approaches can be viewed as Gaussian copula-based schemes (since they rely on the mapping of a Gaussian process) or non-linear versions of the classic (i.e., Gaussian) linear stochastic schemes (Tsoukalas et al., 2018c). Nataf-based models have been widely employed within the domain of operations research (e.g., Biller & Nelson, 2003; Cario & Nelson, 1996) and probabilistic engineering mechanics (e.g., Deodatis & Micaletti, 2001; Grigoriu, 1998), while their application in hydrological modelling was, until recently, unknown. Beyond the aforementioned works, see the work of Kossieris et al. (2019), who for the first time, employed such models for the simulation of water demand processes at fine time scales. Interestingly, similar ideas can be revealed in several classical hydrological papers; cf. review by Tsoukalas et al. (2018a) and Tsoukalas (2019). Beyond the hydrological domain, more details about the Nataf's model, as well as its relationship with the Gaussian copula, are provided by Lebrun & Dutfoy (2009).

Herein we briefly discuss the theoretical background and key implementation steps of the proposed, Nataf-based, simulation approach, also providing guidelines for its *optimal* use. For convenience, we first present the most involved modelling case of multivariate cyclostationary processes, and next deal with the simpler case of stationary processes. For demonstration, we also present a hypothetical study involving the simulation of contemporaneous cross-correlated processes. We remark that throughout the paper, the underbar notation (e.g., $\underline{x}$) is used to denote a random variable (RV), while the italic typeface (e.g., x) denotes a realization of it.

2.1. Cyclostationary processes

In general, cyclostationarity is regarded as a special type of non-stationarity that implies a cyclic switching on the marginal and joint characteristics of the process over a period (e.g., year). To elaborate, let $\{\underline{\boldsymbol{x}}_{s,n}\}$ be an m-dimensional multivariate cyclostationary process. Each individual process $\{\underline{x}^i_{s,n}\}$ is consisted of $s = 1, \cdots, S$ sub-periods (e.g., months), while $n \in \mathbb{Z}^>$, denotes the time index. The sub-period (i.e., season) that corresponds to a time step n may be recovered by $s = n \bmod(S)$, while when $n \bmod(S) = 0$ we get $s = S$. This process can also be written as $\underline{x}_{s,t}$, where now $t \in \mathbb{Z}^>$, denotes the period (e.g., year). In this formulation, the period t is related with the time index n and sub-period s by, $t = 1 + (n - s)/S$. For convenience, the first formulation will be employed in the following paragraphs. Due to cyclostationarity, each process $\underline{x}^i_{s,n}$ is characterized by seasonally varying (herein referred to as target) marginal distributions $F_{\underline{x}^i_s} = P(\underline{x}^i_s \le x)$, while their correlation structure is expressed through the Pearson's correlation coefficient $\rho^{i,j}_{s,s-\tau}$: =Corr$[\underline{x}^i_s \underline{x}^j_{s-\tau}]$, where τ denotes the time lag (the index n is omitted for simplicity). Also let $\{\underline{\boldsymbol{z}}_{s,n}\}$ denote an auxiliary m-dimensional cyclostationary standard Gp with $\underline{z}^i_s \sim \mathcal{N}(0, 1)$.

Due to cyclostationarity, the Gp is completely defined by its correlation structure, which is expressed through the so-called equivalent correlation coefficients $\widetilde{\rho}^{i,j}_{s,s-\tau}$: =Corr[$\underline{z}^i_s \underline{z}^j_{s-\tau}$]. The target process $\{\underline{\boldsymbol{x}}_{s,n}\}$ can be established through the auxiliary process $\{\underline{\boldsymbol{z}}_{s,n}\}$ via the mapping function:

$$\underline{x}^i_{s,n} = F^{-1}_{\underline{x}^i_s}(\Phi(\underline{z}^i_{s,n})) \tag{1}$$

where $F^{-1}_{\underline{x}^i_s}$ denotes the ICDF of $F_{\underline{x}^i_s}$ and $\Phi(\bullet)$ denotes the cumulative density function (CDF) of the standard Gaussian distribution. This mapping allow us to relate the target correlation coefficients $\rho^{i,j}_{s,s-\tau}$ with the equivalent correlation coefficients $\widetilde{\rho}^{i,j}_{s,s-\tau}$ of the auxiliary Gp (Tsoukalas et al., 2017, 2018a). Specifically, they are related by:

$$\rho^{i,j}_{s,s-\tau} = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F^{-1}_{\underline{x}^i_s}(\Phi(z^i_s)) F^{-1}_{\underline{x}^j_{s-\tau}}(\Phi(z^j_{s-\tau}))\varphi_2(z^i_s, z^j_{s-\tau}; \widetilde{\rho}^{i,j}_{s,s-\tau})\mathrm{d}z^i_s \mathrm{d}z^j_{s-\tau} - \mathrm{E}[\underline{x}^i_s]\mathrm{E}[\underline{x}^j_{s-\tau}]}{\sqrt{\mathrm{Var}[\underline{x}^i_s]\mathrm{Var}[\underline{x}^j_{s-\tau}]}} \tag{2}$$

where E[•], and Var[•], denote the corresponding mean and variance (known from the target distributions), while $\varphi_2(z^i_s, z^j_{s-\tau}; \widetilde{\rho}^{i,j}_{s,s-\tau})$ stands for the bivariate standard normal probability density function (PDF). The relationship of Eq. (2) is abbreviated as:

$$\rho^{i,j}_{s,s-\tau} = \mathcal{F}(\widetilde{\rho}^{i,j}_{s,s-\tau} | F_{\underline{x}^i_s}, F_{\underline{x}^j_{s-\tau}}) \tag{3}$$

2.2. Stationary processes

A similar relationship can be established between a target multivariate stationary process $\{\underline{\boldsymbol{x}}_t\}$ and an auxiliary multivariate stationary standard Gp $\{\underline{\boldsymbol{z}}_t\}$. Particularly, let $\{\underline{\boldsymbol{x}}_t\}$ be comprised of m univariate stationary processes $\{\underline{x}^i_t\}$, indexed using $t \in \mathbb{Z}^>$. Furthermore, let each one be described by a target CDF, $F_{\underline{x}^i} = P(\underline{x}^i \le x)$ and let their correlation structure be expressed by $\rho^{i,j}_\tau$: =Corr[$\underline{x}^i_t \underline{x}^j_{t+\tau}$]. Similarly, the process $\{\underline{\boldsymbol{z}}_t\}$ is a m-dimensional stationary standard Gp, with equivalent correlation structure, $\widetilde{\rho}^{i,j}_\tau$: =Corr[$\underline{z}^i_t \underline{z}^j_{t+\tau}$]. Using a similar rationale to the cyclostationary case, each target process is established by, $\underline{x}^i_t = F^{-1}_{\underline{x}^i}(\Phi(\underline{z}^i_t))$. In this case, the relationship between the target and equivalent correlation coefficients reads (e.g., Biller & Nelson, 2003; Tsoukalas et al., 2018c):

$$\rho^{i,j}_\tau = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F^{-1}_{\underline{x}^i}(\Phi(z^i_t)) F^{-1}_{\underline{x}^j}(\Phi(z^j_{t+\tau}))\varphi_2(z^i_t, z^j_{t-\tau}; \widetilde{\rho}^{i,j}_\tau)\mathrm{d}z^i_t \mathrm{d}z^j_{t+\tau} - \mathrm{E}[\underline{x}^i]\mathrm{E}[\underline{x}^j]}{\sqrt{\mathrm{Var}[\underline{x}^i]\mathrm{Var}[\underline{x}^j]}} \tag{4}$$

which is abbreviated as:

$$\rho^{i,j}_\tau = \mathcal{F}(\widetilde{\rho}^{i,j}_\tau | F_{\underline{x}^i}, F_{\underline{x}^j}) \tag{5}$$

Both Eqs. (3) and (5) imply that the correlation structure of the target process depends on the target distributions and the equivalent correlation structure of the auxiliary Gp. We underline that the term *equivalent* is used to highlight the fact that the correlation coefficients of the target process and those of the auxiliary Gp, rarely coincide (due to the non-linear mapping operation; see Embrechts et al. (1999)). Particularly, it can be shown (Liu & Der Kiureghian, 1986) that for any pair of correlated RVs $\underline{x}^i$, $\underline{x}^j$ established through a pair of correlated Gaussian RVs $\underline{z}^i$, $\underline{z}^j$ using a mapping operation analogous to Eq. (1), i.e., $\underline{x}^i = F^{-1}_{\underline{x}^i}(\Phi(\underline{z}^i))$, we get $|\rho^{i,j}| \le |\widetilde{\rho}^{i,j}|$, where $\rho^{i,j}$ and $\widetilde{\rho}^{i,j}$ denote the target and equivalent (i.e., in the Gaussian domain) correlation coefficients, respectively (cf. Fig. 1 in Tsoukalas et al. (2018a)).

2.3. Estimation of equivalent correlation coefficients

Given Eqs. (3) and (5), it is possible to establish a variety of approaches to simulate processes with a priori specified (target) marginal distributions and (target) correlation structures. This specification is subject to several criteria, which are discussed in Sections 2.4–2.6.

Nevertheless, an essential step for ensuring the reproduction of target correlations through Eq. (3) or Eq. (5) (for cyclostationary and stationary processes, respectively) is the identification of the equivalent correlation coefficients. This is implemented by establishing and then inverting the corresponding $\mathcal{F}(\bullet)$ relationships.

Among the available methods (e.g., Biller & Nelson, 2003; Cario & Nelson, 1996, 1997; Liu & Der Kiureghian, 1986), herein we will employ the Monte-Carlo procedure by Tsoukalas et al. (2017, 2018a), due to its efficiency and straightforward implementation, since it avoids the use of integration methods (for a method comparison see, Tsoukalas et al., 2018c).

2.4. Admissible marginal distributions

As already explained, Nataf-based models can be used for the simulation of processes with arbitrary (continuous, discrete or mixed-type) marginal distributions and valid correlation structures, provided that their combination is feasible (i.e., leads to a positive definite correlation structure) and the variance of the distributions is finite (which is true when modelling hydrometeorological processes; Koutsoyiannis (2016)). Regarding the marginal distributions, and in contrast to the classical working paradigm of stochastic hydrology, it is stressed that by design, Nataf-based models do not aim at resembling the process's moments; in fact, they aim to simulate processes with target, *a priori* specified, distributions, in order to fully describe their marginal properties (cf. discussion by Tsoukalas et al. (2018a)). In this respect, questions about skewness handling or "*how many moments should be reproduced for approximating the distribution of a specific process?*" are now out of interest.

For instance, within Nataf-based schemes, simulating a process following a Gamma or Log-Normal distribution requires the identification of just two parameters (shape and scale), which can be easily determined by straightforward methods. Even the classical method of product moments would ensure reliable estimations, since in these specific cases it only requires computations up to second order (Lombardo et al., 2014).

In a more general context, the Nataf-based approach offers the flexibility to employ robust fitting methods for parameter estimation, that rely on alternative notions, such as maximum-likelihood, probability weighted moments (Greenwood et al., 1979) or L-moments (Hosking, 1990). In our view, this is a major advantage, since it can avoid the data-driven estimation of high-order moments (e.g., kurtosis or higher), since it is well known that they are prone to sample uncertainties and bias (Lombardo et al., 2014; Matalas, 1967p. 945). Appendix A summarizes all distribution models employed in this work, which were generally fitted using the L-moments method.

2.5. Specific case: mixed distributions

Mixed distributions are often advocated within hydrological applications, either to better represent the tails of the understudy hydrometeorological variable (e.g., Evin et al., 2018; Foufoula-Georgiou & Lettenmaier, 1987; Furrer & Katz, 2008; Li et al., 2013; Wilks, 1998), or to simultaneously represent the dual character of intermittent processes (e.g., Bárdossy & Pegram, 2016; Cannon, 2008; Papalexiou, 2018; Serinaldi, 2009; Tsoukalas et al., 2018c; Williams, 1998). Herein we briefly describe the second case, which can be accomplished using a *zero-inflated* (also referred to as *zero-augmented* or *discrete-continuous*) distribution model. This model is composed of both a discrete and a continuous part, and its CDF is given by:

$$F_{\underline{x}}(x) = \begin{cases} p_D, & x = 0 \\ p_D + (1 - p_D)G_{\underline{x}}(x), & x > 0 \end{cases} \tag{6}$$

The discrete part is represented by p_D: =$P(\underline{x} = 0)$, and denotes the probability of a zero value. The continuous part is given by

$G_{\underline{x}}: = F_{\underline{x}|\underline{x}>0} = P(\underline{x} \leqslant x | \underline{x} > 0)$, which denotes a continuous distribution function for the non-zero data. For instance, within the context of intermittent hydrometeorological processes (e.g., rainfall), p_D stands for the probability of a dry interval (i.e., probability dry), and $G_{\underline{x}}$ represents the distribution of positive amounts. In real-world situations, the most straightforward way to specify p_D and $G_{\underline{x}}$ is through the available data. Specifically, p_D is estimated as the ratio of dry occurrences to the total number of observations, while $G_{\underline{x}}$ can be identified by fitting a continuous distribution function to the positive amounts. For completeness, the ICDF of the zero-inflated model, which can be used for RVs generation, is given by:

$$F_{\underline{x}}^{-1}(u) = \begin{cases} 0, & 0 \leq u \leq p_D \\ G_{\underline{x}}^{-1}\left(\frac{(u-p_D)}{(1-p_D)}\right), & p_D < u \leq 1 \end{cases} \tag{7}$$

The implementation of Eq. (6) for simulating intermittent process, in a context similar to Nataf-based models, has been recently justified by Papalexiou (2018). See also the similarly constructed three-component distribution function employed by Kossieris et al. (2019), aiming to simultaneously describe both the discrete-continuous character (i.e., intermittent nature) and tail behavior of water demand processes at fine time scales (up to 1 minute).

2.6. Admissible correlation structures

Classical stochastic modelling strategies are designed to reproduce a limited number of low-order dependence metrics in space and time, typically expressed in terms of Pearson's correlation coefficients. Actually, most of them still follow the specifications posed by Matalas and Wallis (1976), thus aiming to reproduce just two dependencies, i.e., lag-1 autocorrelations and lag-0 cross-correlations. We remark that herein, the term spatial correlation will denote any dependence between different processes, either referring to different geographical locations or not.

More modern approaches suggest the use of theoretical models for the mathematical description of the auto- and cross-dependence structures that span over any lag (e.g., Gneiting, 2000; Gneiting & Schlather, 2004; Koutsoyiannis, 2000, 2016; Papalexiou, 2018). These typically concern stationary processes, and are based on the notions of correlation, spectrum or variance over aggregated time scales, which are all interrelated (see, Beran, 1994; Koutsoyiannis, 2016). The use of theoretical dependence models instead of sample statistics is mostly implied from the significant uncertainties and biases of data-driven estimates.

Arguably, the most popular type of theoretical dependence models are correlation-based ones. These can be further classified to full spatiotemporal models (Chilès & Delfiner, 1999; Genton & Kleiber, 2015; Gneiting et al., 2010), which simultaneously model the auto- and cross-correlation structure of the process, and separable (e.g., Genton, 2007; Mardia & Goodall, 1993; Rodríguez-Iturbe & Mejía, 1974), which describe the two correlation structures independently, as the product of two functions (i.e., one for the spatial and one for the temporal component).

Throughout this work, and without loss of generality (since alternative models can be used), we will employ the separable approach. Specifically, we model directly the lag-0 contemporaneous cross-correlations of the processes, while the auto-dependence structure of each individual stationary process is modelled using the two-parameter Cauchy-type autocorrelation structure (CAS), introduced by Koutsoyiannis (2000):

$$\rho_\tau^{\mathrm{CAS}}(\kappa, \beta) = (1 + \kappa\beta\tau)^{-1/\beta}, \quad \tau \geq 0 \tag{8}$$

where $\beta > 0$ and $\kappa > 0$ are model parameters. The specific case $\beta = 0$ is also feasible, through the use of L' Hôpital's rule, and represents ARMA-type processes (see, Koutsoyiannis, 2000). By construction, CAS can resemble a wide spectrum of processes, characterized by both short- and long-range dependence, i.e., SRD and LRD (e.g., Efstratiadis et al., 2014; Tsoukalas et al., 2018c). SRD refers to a stochastic process with a weak autocorrelation structure (e.g., exponential) that decays rapidly, while LRD implies the exact opposite. In this case, the autocorrelation structure is a slowly decreasing function (typically power-type) of the time lag (Beran, 1994; Beran et al., 2013; Koutsoyiannis, 2002; O'Connell et al., 2016). This type of auto-dependence is also associated with the widely studied Hurst phenomenon (Hurst, 1951), nowadays also referred to as Hurst-Kolmogorov (HK) dynamics (Koutsoyiannis, 2011; Koutsoyiannis & Montanari, 2007), as well as with the fractional Gaussian noise process (e.g., Mandelbrot & Wallis, 1969). These properties and its parsimonious character (as the model has only two parameters), make CAS a good candidate model for modelling hydrometeorological processes. Regarding parameter identification, the most straightforward option is to fit CAS to the empirical estimates of autocorrelation coefficients. However, this simple approach neglects the estimator's biases (e.g., Beran, 1994; Koutsoyiannis, 2016; Marriott & Pope, 1954), which are considered to be significant in the presence of LRD and for large time lags (due to small sample sizes). In such cases, it may be advantageous to explicitly account for bias by using alternative robust parameter identification procedures, such as the climacogram (e.g., Dimitriadis & Koutsoyiannis, 2015; Koutsoyiannis, 2016), or even through empirical approaches, accounting for regional information and user expertise (Efstratiadis et al., 2014).

In summary, the combined use of Nataf-based models along with theoretical distribution functions and theoretical correlation structures (e.g., CAS), offers several advantages, such as:

- easy exploration of alternative scenarios (by perturbing the models parameters);
- regional transferability (through spatial interpolation);
- improved model stability (since a valid correlation structure owes to be positive definite; a fact guaranteed by a proper theoretical model);
- decoupling of parameter identification (involving the parameters of the distribution model and the theoretical correlation structure) and generation mechanism.

2.7. The auxiliary Gaussian processes

In order to deploy a Nataf-based stochastic simulation scheme, it is finally required to simulate realizations from an auxiliary Gaussian process (Gp) model that preserves the equivalent correlation coefficients, which in turn, after the mapping procedure, reproduces the target stochastic structure.

A convenient option is the use of Gaussian linear stochastic models (often called time series models). Characteristic examples, adapted from operations research, are, the works of Cario and Nelson (1996) and Biller and Nelson (2003), who used as an auxiliary Gp, univariate and multivariate stationary AutoRegressive (AR) processes, respectively. The resulting Nataf-based models are termed AutoRegressive To anything (ARTA) and Vector AutoRegressive To Anything (VARTA). A notable difference of these works compared to our approach lies in the fact that the previous works did not employ the notion of theoretical correlation structures. This implies that the order p of the associated AR model dictates the correlation structure of the process to simulate. This may be also the reason for the typical use of low order models. On the other hand, if the auto-correlation structure has been *a priori* specified (e.g., using CAS), it is possible to employ high-order models (e.g., AR (p)) without sacrificing parsimony. In this case, the order of the Gp model solely controls the degree of resemblance of the correlation structure up to the desired lag τ (since a higher order model provides more flexibility), while the associated model's parameters can be viewed as internal coefficients (for a bivariate example, see Section 2.9).

In the water resources domain, a comprehensive treatment of

multivariate and univariate Nataf-based schemes, based on stationary and cyclostationary Gaussian linear stochastic models, is presented by Tsoukalas et al. (2018c) and Tsoukalas et al. (2017, 2018a) respectively, as well as by Papalexiou (2018). In this work, aiming to develop a multi-scale consistent simulation scheme (see Section 4), we employ (and couple in a pairwise basis) several Nataf-based models, which use different Gp models (all of them also capable of multivariate simulations; see the summary and the abbreviations of Table 1).

From the aforementioned models, SMARTA and CMARTA are designed for stationary processes, while SPARTA for cyclostationary ones (i.e., accounting for the season-to-season correlations). A common characteristic of the three models is the direct reproduction of lag-0 cross-correlations coefficient among multiple contemporaneous processes. SMARTA and SPARTA are fully described in the works cited in Table 1, while a detailed description of the auxiliary Gaussian CMAR(p) model is given in Appendix B. It is stressed that, regardless of the choice of the auxiliary Gp model, in order to generate realizations with the equivalent correlation structure, the model parameters have to be estimated using the equivalent correlation coefficients. For this reason, within the description of CMAR(p) we employ the tilde notation (see Appendix B).

2.8. Brief overview via a step-by-step procedure

For a given stochastic process (univariate case) or a set of processes (multivariate case) to simulate, the required methodological steps of any Nataf-based model are:

Step 1. Identify the type (i.e., stationary or cyclostationary) of the processes, accounting for process properties and the time scale of simulation.
Step 2. Based on the available information (e.g., historical data), as well as the user expertise, assign appropriate target marginal distributions to all processes and identify the target correlation structure, in time and (case of multivariate simulation) space.
Step 3. Select a suitable linear stochastic model to simulate the auxiliary Gp.
Step 4. Estimate the equivalent correlation coefficients for all pairs of variables that are required by the parameter estimation procedure of the auxiliary model, i.e., Gp.
Step 5. Estimate the parameters of the Gp model through the equivalent correlation coefficients.
Step 6. Generate a synthetic time series by employing the Gp (i.e., $\underline{\boldsymbol{z}}_t$ or $\underline{\boldsymbol{z}}_{s,n}$).
Step 7. Map the auxiliary (i.e., Gaussian) time series to the actual domain in order to attain a realization of the target process (i.e., $\underline{\boldsymbol{x}}_t$, or $\underline{\boldsymbol{x}}_{s,n}$).

The methodology and simulation models described in section 2, as briefly summarized by the above-described step-by-step procedure, has been implemented in the *anySim* R package (Tsoukalas and Kossieris, 2019) which enables the straightforward and easy generation of synthetic time series with any distribution and correlation structure.

2.9. A hypothetical simulation example

To enhance readability, we employ CMARTA(p) and setup three bivariate hypothetical simulation experiments (hereafter termed A, B and C) that regard the simulation of two contemporaneously cross-correlated stationary processes, $\underline{x}_t^1$ and $\underline{x}_t^2$, with either zero-inflated or discrete marginal distributions.

For convenience, the three experiments share some common assumptions which are: 1) the order of the auxiliary Gaussian CMAR(p) model, which is set to be $p = 100$; 2), the length of the synthetic time series to simulate, i.e., 100 000 steps; 3), the target lag-0 cross-correlation coefficient, i.e., $\rho_0^{1,2} = 0.70$; and 4) the target auto-dependence structure of each process, which is provided by CAS (i.e., Eq. (8)), i.e., $\underline{x}_t^1 \sim \rho_\tau^{CAS;1}(\beta = 0, \kappa = 0.67)$ and $\underline{x}_t^2 \sim \rho_\tau^{CAS;2}(\beta = 1.2, \kappa = 1)$. As briefly mentioned earlier, and since the autocorrelation structure of the processes is already specified, the use of a high-order Gp model (i.e., CMAR(p) in this example) does not introduces additional parameters, but solely controls the degree of resemblance of the target correlation structure. Particularly, by setting $p = 100$, the model will resemble the target CAS up to time lag 100, while for $\tau > 100$ it will reduce according to its theoretical properties. Similarly, if we employed a higher-order model, e.g., $p = 1000$, we would resemble the target CAS up to time lag $\tau = 1000$, without needing more parameters for the description of the autocorrelation structure.

The three simulation studies differ in terms of target marginal distributions of the individual processes. More specifically, in case A it is assumed that the marginal distribution is provided by a zero-inflated model (i.e., Eq. (6)) with $p_D = 0.80$ and $p_D = 0.75$ for $\underline{x}_t^1$ and $\underline{x}_t^2$ respectively; while the continuous part is given by the Gamma ($\mathcal{G}$) and Log-Normal ($\mathcal{LN}$) distribution (Eq. (A.1) and Eq. (A.2)) respectively. Particularly, they are: for $\underline{x}_t^1$ by $\mathcal{G}(a = 0.20, b = 0.15)$ and for $\underline{x}_t^2$ by $\mathcal{LN}(a = 1, b = 1)$. In case B the target distributions are regarded to be discrete, and given by the Poisson distribution ($\mathcal{P}oi$; see also, Eq. (A.5)). Particularly we assume that, $\underline{x}_t^1 \sim \mathcal{P}oi(\lambda = 1)$ and $\underline{x}_t^2 \sim \mathcal{P}oi(\lambda = 2)$. Finally, in case C we assigned a discrete-type Bernoulli distribution ($\mathcal{B}ern$; see Eq. (A.6)). Specifically, $\underline{x}_t^1 \sim \mathcal{B}ern(p = 0.8)$ and $\underline{x}_t^2 \sim \mathcal{B}ern(p = 0.75)$.

Case A can be considered as the most common simulation scenario, since it involves the multivariate simulation of intermittent processes with zero-inflated (non-Gaussian) marginal distributions. For instance, it could represent fine-time rainfall processes at two locations. On the other hand, cases B and C can arise in practice when aiming to model counting (e.g., number of drought events in a given year) or occurrence (e.g., sequences of wet and dry transitions) processes, respectively.

Regarding CMARTA(p) implementation and evaluation for case studies A-C, let us focus on the first case, since the simulation procedure, as well as the results, are similar for all three cases. Given that the target distributions (red line in Fig. 1c,d), auto-correlation structure (red line in Fig. 1h,i) and lag-0 cross-correlation (red line for $\tau = 0$ in Fig. 1j), are already known, we begin by estimating the required $\mathcal{F}(\bullet)$ relationships (Fig. 1e–g). These are subsequently inverted to estimate the corresponding equivalent correlation coefficients, and the parameters of the auxiliary Gp model, i.e., CMAR(p). Next a bivariate realization of the Gp process, i.e., $\underline{\boldsymbol{z}}_t = [\underline{z}_t^1, \underline{z}_t^2]^T$, with the desired length (in this case, 100 000), is generated. The final step involves the

Table 1
Summary of employed Nataf-based models of order p.

Auxiliary Gp model	Associated Nataf-based model	Type	References
SMA(p)	SMARTA(p)	Stationary	(Tsoukalas et al., 2018c)
CMAR(p)	CMARTA(p)	Stationary	This work (Appendix B)
PAR(p)	SPARTA(p)	Cyclostationary	(Tsoukalas et al., 2017, 2018a)

Abbreviations: SMA (Symmetric Moving Average), CMAR (Contemporaneous Multivariate AutoRegressive), PAR (Periodic AutoRegressive), SMARTA (Symmetric Moving Average neaRly To Anything), CMARTA (Contemporaneous Multivariate AutoRegressive neaRly To Anything), SPARTA (Stochastic Periodic AutoRegressive To Anything).

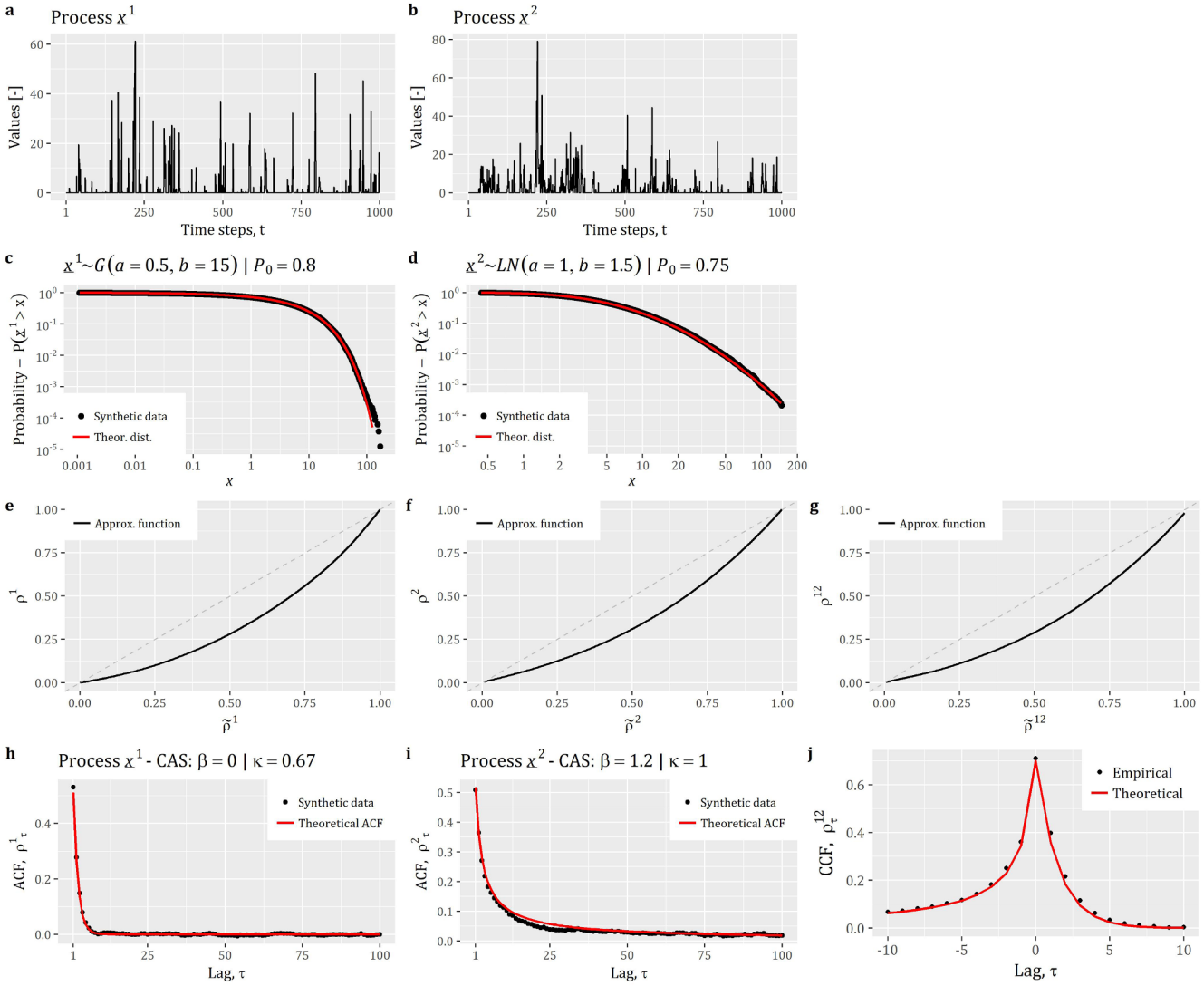

Fig. 1. Case A – Zero-inflated marginal distributions. Simulated realization of process a) $\underline{x}_t^1$ and b) $\underline{x}_t^2$. Comparison of non-zero, simulated and theoretical distribution function for process c) $\underline{x}_t^1$ and d) $\underline{x}_t^2$. The established relationships between the equivalent, $\tilde{\rho}$, and target, ρ, correlation coefficients given the marginal distribution of each process; e) $\underline{x}_t^1$, f) $\underline{x}_t^2$, as well as their g) interaction. Simulated and theoretical autocorrelation function (ACF) for process h) $\underline{x}_t^1$ and i) $\underline{x}_t^2$, as well as j) the cross-correlation function (CCF) among $\underline{x}_t^1$ and $\underline{x}_t^2$, i.e., $\rho_\tau^{1,2} = \mathrm{Corr}[\underline{x}_t^1, \underline{x}_{t+\tau}^2]$.

mapping of the Gaussian process, to attain a realization of the target one, i.e., $\underline{\boldsymbol{x}}_t = [\underline{x}_t^1, \underline{x}_t^2]^T$, using the corresponding ICDF, i.e., $\underline{x}_t^i = F_{\underline{x}^i}^{-1}(\Phi(\underline{z}_t^i))$. The validity of the model is verified by the plots of Fig. 1 (case A), Fig. C.1 (case B) and Fig. C.2 (case C), which illustrate that CMARTA is able to accurately reproduce the probabilistic and stochastic structure of the target processes, regardless if the marginal distributions are zero-inflated or discrete.

3. Addressing multi-scale consistency

The above simulation framework, although fulfilling the requirements of a *good* stochastic model, i.e., the explicit reproduction of any distribution and any correlation structure, it does not account for multi-scale consistency. Since the problem is independent of the generation procedure and the time scale of simulation, we first provide a global overview and then propose a generic solution for Nataf-based models, herein referred to as Nataf-based Disaggregation To Anything (NDA).

3.1. Problem description

Let us begin from the univariate case, denoting by $\{\underline{\omega}_n\}_{n\in\mathbb{Z}^>}$ a discrete-time, stationary or cyclostationary (the season indicator s is omitted for simplicity), stochastic process at time scale $k = 1$, where n is a time index. Let also define the aggregated process $\underline{\omega}_l^{(k)}$ at a higher time scale $k \in \mathbb{Z}^{\geq 2}$, obtained by:

$$\underline{\omega}_l^{(k)} = \sum_{n=(l-1)k+1}^{kl} \underline{\omega}_n \tag{9}$$

where l is the time index of the aggregated process. Alternatively (e.g., if $\underline{\omega}_n$ refers to an instantaneous quantity), we can define the averaged process, also denoted by $\underline{\omega}_l^{(k)}$, by, $\underline{\omega}_l^{(k)} = \sum_{n=(l-1)k+1}^{kl} \underline{\omega}_n / k$. Apparently, the properties of $\{\underline{\omega}_n\}$ at scale $k = 1$ are related with those of the aggregated (or averaged) process at a higher time scale $k \in \mathbb{Z}^{\geq 2}$.

Herein, without loss of generality, we focus on the aggregated case. To simplify, we first remark that the operations implied by Eq. (9), can

be viewed as a sum of k RVs. Thus, if we were interested in the distribution of $\{\underline{\omega}_l^{(k)}\}$, it would be the same as solving an aggregated distribution problem. If the process $\{\underline{\omega}_n\}$ is stationary at $k = 1$, then at any higher scale k we would have the sum of k identical RVs. On the other hand, if $\{\underline{\omega}_n\}$ is cyclostationary at the lower scale $k = 1$, at any higher scale k we would have the sum of k non-identical RVs (their marginal and dependence properties depend on the season $s = 1, \cdots, S$, implied by the time index n; see Section 2.1).

Arguably, the problem of identifying the distribution of $\underline{\omega}_l^{(k)}$ at $k > 1$ is particularly challenging, since there is no general method (without resorting to simulation) to identify the distribution of the sum of k RVs, especially in the presence of dependence, which is typical for hydrometeorological processes. Furthermore, apart from some low order moments (i.e., mean, variance, autocovariance and autocorrelation), higher order moments of the aggregated process are also particularly difficult to estimate, either analytically or theoretically. Analogously, it is also challenging to specify a process $\{\underline{\omega}_n\}$ that has the desirable (for this time scale) marginal and stochastic properties, when it is aggregated at a higher scale $k > 1$.

The problem becomes even harder when multiple processes are involved, in the context of multivariate simulation problems. Let $\underline{\boldsymbol{\xi}}_t = [\underline{\xi}_t^1, \cdots, \underline{\xi}_t^m]^{\mathrm{T}}$ and $\underline{\boldsymbol{\omega}}_n = [\underline{\omega}_n^1, \cdots, \underline{\omega}_n^m]^{\mathrm{T}}$ be two m-dimensional vectors of two discrete-time processes $\underline{\xi}_t^i$ and $\underline{\omega}_n^i$, indexed using $t \in \mathbb{Z}^>$ and $n \in \mathbb{Z}^>$, respectively. Furthermore, let assume that $\underline{\xi}_t^i$ and $\underline{\omega}_n^i$ represent the same process at two different temporal scales, higher and lower, respectively, with time units denoted by $\delta_{\underline{\xi}}$ and $\delta_{\underline{\omega}}$, respectively (i.e., $\delta_{\underline{\xi}} > \delta_{\underline{\omega}}$).

Similarly to Eq. (9), when $k^*: = k = \delta_{\underline{\xi}}/\delta_{\underline{\omega}}$ (e.g., 1 year/1 month = 12, or 1 month/1 hour = 28 × 24, 30 × 24, 31 × 24; depending on the number of days of the month), we obtain an aggregated process at the same temporal level of $\underline{\xi}_t^i$, i.e.,

$$\widehat{\underline{\xi}}_t^i: = \underline{\omega}_l^{i;(k^*)} = \sum_{n=(l-1)k^*+1}^{lk^*} \underline{\omega}_n^i, \ (l = t) \tag{10}$$

Evidently, when $\underline{\omega}_n^i$ is simulated without reference to the higher-level process $\underline{\xi}_t^i$, then $\underline{\xi}_t^i \neq \widehat{\underline{\xi}}_t^i$. Hence, for each process $i = 1, ..,m$, our target is to generate a k^*-dimensional random sequence, $\underline{\boldsymbol{\omega}}_{t;(k^*)}^i = [\underline{\omega}_{(t-1)k^*+1}^i, \cdots, \underline{\omega}_{tk^*}^i]$, of the low-level process ($k = 1$), with the desirable properties, which honors the equality, $\underline{\xi}_t^i = \widehat{\underline{\xi}}_t^i$, when aggregated to the time scale k^*. The multivariate formulation of the problem is written as:

$$\underline{\boldsymbol{\Omega}}_{t;(k^*)} = [\underline{\boldsymbol{\omega}}_{t;(k^*)}^1{}^{\mathrm{T}}, \cdots, \underline{\boldsymbol{\omega}}_{t;(k^*)}^m{}^{\mathrm{T}}]^{\mathrm{T}} = \begin{bmatrix} \underline{\omega}_{(t-1)k^*+1}^1 & \cdots & \underline{\omega}_{tk^*}^1 \\ \vdots & \ddots & \vdots \\ \underline{\omega}_{(t-1)k^*+1}^m & \cdots & \underline{\omega}_{tk^*}^m \end{bmatrix}, \text{ and}$$

$$\widehat{\underline{\boldsymbol{\xi}}}_t = [\widehat{\underline{\xi}}_t^1, ..., \widehat{\underline{\xi}}_t^m]^{\mathrm{T}} = \left[\sum_{n=(t-1)k^*+1}^{tk^*} \underline{\omega}_n^1, ..., \sum_{n=(t-1)k^*+1}^{tk^*} \underline{\omega}_n^m \right]^{\mathrm{T}} \tag{11}$$

3.2. The NDA approach: step-by-step implementation

In order to address the problem, we develop the Nataf-based Disaggregation To Anything (NDA) approach, which combines Nataf-based models, considered as data generation mechanisms, with a coupling procedure that encompasses the notions of repetitive sampling and adjusting procedures. These two key notions are thoroughly discussed by Koutsoyiannis and Manetas (1996).

The NDA procedure starts from a given realization, $\boldsymbol{\xi}_t$, of a process $\underline{\boldsymbol{\xi}}_t$, at a specific time scale, aiming to produce a consistent realization, $\boldsymbol{\omega}_n$, at a lower scale. The given realization $\boldsymbol{\xi}_t$ is known either from observations or already generated by another model (deterministic or stochastic). In the second case, if a Nataf-based model is employed, the synthesized higher-level realization would have the desirable marginal distributions and correlation structure, hence the problem would reduce to generating a lower-level realization with the target properties, which when aggregated to the higher-level honors the additive property. Fulfilling both conditions allows preserving the properties of the process at both temporal levels, given that the realization at the higher level is kept as is.

Therefore, given the realization $\boldsymbol{\xi}_t$, and assuming a *temporary* Nataf-based lower-level process, denoted by $\widetilde{\boldsymbol{\omega}}_n$, with properties identical to those of the target process $\widetilde{\boldsymbol{\omega}}_n$ (i.e., $\widetilde{\boldsymbol{\omega}}_n = \underline{\boldsymbol{\omega}}_n$), the following steps are applied for all time indices t.

1) Using a Nataf-based model (cf. Table 1), generate $N_{\xi/\omega}$ temporary realizations $\widetilde{\boldsymbol{\omega}}_n$ of the lower level process $\widetilde{\boldsymbol{\omega}}_n$, of length k^*, thus obtaining $N_{\xi/\omega}$ sets of matrices $\{\widetilde{\boldsymbol{\Omega}}_{t;(k^*)}(\nu);\ \nu = 1, ..., N_{\xi/\omega}\}$.
2) For each of the $N_{\xi/\omega}$ matrices $\widetilde{\boldsymbol{\Omega}}_{t;(k^*)}$, estimate the corresponding vector $\widetilde{\boldsymbol{\xi}}_t$ and obtain a set of vectors $\{\widetilde{\boldsymbol{\xi}}_t(\nu);\ \nu = 1, ..., N_{\xi/\omega}\}$.
3) Calculate the difference between $\widetilde{\boldsymbol{\xi}}_t(\nu)$ and the known $\boldsymbol{\xi}_t$ using a distance metric, $e_t(\nu) = D(\widetilde{\boldsymbol{\xi}}_t(\nu), \boldsymbol{\xi}_t)$.
4) Formulate the set $\{e_t(\nu);\ \nu = 1, \cdots, N_{\xi/\omega}\}$ and select the realization $\widetilde{\boldsymbol{\Omega}}_{t;(k^*)}(\nu)$ with the minimum value of $e_t(\nu)$, hereafter denoted $\boldsymbol{\Omega}_{t;(k^*)}^{'}$ (the breve notation has been omitted for simplicity). Under this premise, by aggregating $\boldsymbol{\Omega}_{t;(k^*)}^{'}$ to time scale k^*, thus obtaining the corresponding sum $\widetilde{\boldsymbol{\xi}}_t'$, its difference with the target values of $\boldsymbol{\xi}_t$ will be the minimum over the simulated set.
5) Produce the final values of $\boldsymbol{\Omega}_{t;(k^*)}$ by adjusting the remaining difference between $\widetilde{\boldsymbol{\xi}}_t'$ and $\boldsymbol{\xi}_t$, by employing a specific adjusting procedure.

We remark that since we employ Nataf-based models, in order to ensure a proper sequential generation procedure, it is essential to maintain an archive of the realizations generated by the auxiliary Gp model. These are needed to condition the generation mechanism on the required number of previous values. For instance, if we employ CMARTA(p) for generating the temporary realizations $\widetilde{\boldsymbol{\omega}}_n$, p previous values of the auxiliary Gaussian realization are needed to condition the generation of $\widetilde{\boldsymbol{\omega}}_{n+1}$.

3.3. Computational details

For convenience, within repetitive sampling (step 3), we employ as a distance metric the following quantity, also used by Koutsoyiannis and Manetas (1996):

$$e_t = D(\widetilde{\boldsymbol{\xi}}_t, \boldsymbol{\xi}_t) = \frac{1}{m} \sum_{i=1}^{m} |\xi_t^i - \xi_t^i| / \mathrm{Var}[\underline{\xi}_t^i] \tag{12}$$

On the other hand, all available adjusting procedures (APs) that are found in the literature (see, Grygier & Stedinger, 1988; Harms & Campbell, 1967; Koutsoyiannis, 2001) are compatible with the proposed approach. Here we employ the so-called proportional AP that can be implemented independently for each $\omega'^i_{t;(k^*)}$ and reads as:

$$\omega_{t;(k^*)}^i = \omega'^i_{t;(k^*)} \xi_t^i \Big/ \widetilde{\xi}_t'^i \tag{13}$$

Apart from its simplicity, key advantage of this AP is the preservation of the sign of each realization $\omega'^i_{t;(k^*)}$. For instance, in case of rainfall, where the underlying Nataf-based model is combined with a mixed-type distribution to represent intermittency (see Section 2.5), the proportional adjustment not only prohibits the generation of negative rainfall values but also preserves the sequence of zero and non-zero values, as explicitly foreseen by the auxiliary Nataf model.

A final technical issue involves the termination criteria for repetitive sampling. Here, we consider that the iterative procedure terminates

when reaching a maximum number of allowable iterations, $N_{\xi/\omega}$. An alternative option would imply the use of a convergence criterion, by means of a similarity metric between ξ_t' and ξ_t. Nevertheless, the stopping criteria should be carefully assigned, since they control both the accuracy and computational efficiency of NDA, which are inherently conflicting. In our examples, we set $N_{\xi/\omega} = 250$ to 350, which was heuristically identified as a fair conciliation for multivariate problems involving up to five processes. We remark that for the univariate case *per se*, Papalexiou et al. (2018) addressed the challenge of a priori determining the required number of iterations within disaggregation through Bernoulli trials.

We remark that in contrast to other disaggregation schemes, where repetitive sampling had an optional role (cf. Koutsoyiannis & Manetas, 1996), in our approach its role is pivotal, since it allows the preservation of the advantages of Nataf-based models, and hence generate lower-level realizations with the target probabilistic and stochastic properties.

4. Modular framework for developing multi-temporal simulation schemes

4.1. Multi-temporal stochastic simulation as a puzzle game

As already discussed, there does not exist a general, *bottom-up* solution to the problem of multi-scale consistency, by means of a generation procedure that provides consistent synthetic data at a time scale of interest, and simultaneously captures the scale-varying stochastic-probabilistic behavior of the aggregated process at higher time scales. In a practical context, the generally accepted requirement for a *good* stochastic model is to reproduce the desirable probabilistic and dependence properties across specific temporal scales that have operational interest. Typically, these follow the standard resolutions of hydrometeorological time series, i.e., annual, monthly, daily, hourly, etc.

In this context, we propose a *puzzle-type* implementation of NDA, to address multi-scale simulation problems of any complexity. Essentially, this can be done by coupling, in a pairwise manner, multiple Nataf-based models, which operate independently of each other. Thereby, one can establish a modular, top-down approach, starting from the first level, which corresponds to the highest time scale of interest, and subsequently moving to next levels, until reaching the lowest scale, which is dictated by the simulation problem at hand. As shown in Fig. 2, each individual coupling of subsequent scales through NDA can be considered as the pieces of a puzzle. The generic design of NDA ensures flexibility regarding the combination of temporal scales, while at the same time, the robustness of the underlying Nataf-based approach ensures the preservation of the desirable process properties.

For demonstration, we next present a typical configuration of this puzzle, by means of a three-level scheme for annual to daily simulation, which is of significant interest for a wide range of operational hydrological problems. In Section 5, we explore the capacities of this configuration, in the context of a real-world case study, involving the generation of synthetic daily rainfall and runoff series. Moreover, in Section 6, we present another useful configuration, this time for handling a classical disaggregation problem, i.e., the generation of hourly rainfall from a given daily time series.

4.2. Three-level configuration for annual to daily simulation

In this configuration we couple three Nataf-based models, shown in Table 1, to provide a multivariate three-level simulation scheme. This modular scheme (i.e., puzzle) aims to preserve the probabilistic and dependence properties of typical hydrometeorological processes at the annual, monthly, and daily scales.

Let $\underline{\boldsymbol{y}}_t = [\underline{y}_t^1, \cdots, \underline{y}_t^m]^{\mathrm{T}}$ be a vector of m stationary stochastic process at the annual time scale (where $t \in T_{\underline{y}}$ denotes the time index, i.e., year, over the set $T_{\underline{y}}$). In the context of this configuration we model the annual processes using SMARTA, in order to preserve:

- the distribution function of $\underline{y}_t^i$, i.e., $F_{\underline{y}^i}(y)$;
- its autocorrelation structure, $\rho_{\underline{y};\tau}^i = \mathrm{Corr}[\underline{y}_t^i, \underline{y}_{t+\tau}^i]$;
- the lag-0 cross-correlations among processes $\underline{y}_t^i$ and $\underline{y}_t^j$, i.e., $\rho_{\underline{y}}^{i,j} = \mathrm{Corr}[\underline{y}_t^i, \underline{y}_t^j]$.

On the other hand, the standard hypothesis for the monthly time scale is cyclostationarity. Let the monthly process be represented by an m dimensional vector $\underline{\boldsymbol{x}}_{s,n} = [\underline{x}_{s,n}^1, \cdots, \underline{x}_{s,n}^m]^{\mathrm{T}}$, where $s(=1, \cdots, 12, 1 \cdots, 12, \cdots)$ denotes the month and $n \in T_{\underline{x}}$ is the time index. The index t of the annual process (i.e., the year) may be recovered by $t = 1 + (n - s)/12$. For monthly simulation we employ SPARTA in order to resemble:

- the seasonally-varying marginal distribution of $\underline{x}_{s,n}^i$, i.e., $F_{\underline{x}s}(x) = F_{\underline{x}s+12}(x)$;
- the lag-1 month-to-month correlation coefficients $\rho_{\underline{x};s,s-1}^i = \mathrm{Corr}[\underline{x}_s^i, \underline{x}_{s-1}^i]$;
- the lag-0 cross-correlations among processes $\underline{x}_s^i$ and $\underline{x}_s^j$ for each season s, i.e., $\rho_{\underline{x}s}^{i,j} = \mathrm{Corr}[\underline{x}_s^i, \underline{x}_s^j]$.

Finally, the hydrometeorological processes at sub-monthly time scales (e.g., daily) are typically regarded to be stationary within in each month s. In this respect, let $\underline{\boldsymbol{w}}_{s;d} = [\underline{w}_{s;d}^1, \cdots, \underline{w}_{s;d}^m]^{\mathrm{T}}$ be an m-dimensional vector of stationary processes at month s, where $d \in T_{\underline{w}s}$, denotes the time index. We remark that in this case, $k^* = \delta_{\underline{x}s}/\delta_{\underline{w}s}$, where $\delta_{\underline{x}s}$ and $\delta_{\underline{w}s}$denote the time units of $\underline{x}_{s,n}^i$ and $\underline{w}_{s;d}^i$ respectively. For instance, if $\underline{w}_{s;d}^i$ represents the process of month s, at the daily temporal level, $k^* = D_s$, where D_s stands for the days of a month s (i.e., 28, 30 or 31, excluding leap years; similarly, if $\underline{w}_{s;d}^i$ denotes an hourly process, then $k^* = D_s \times 24$). Nonetheless, for the simulation of daily temporal level, we employ CMARTA model, and aim to reproduce:

- the seasonally varying marginal distribution of $\underline{w}_{s;n}^i$, i.e., $F_{\underline{w}s}(w) = F_{\underline{w}s;n}(w)$;
- the within-month autocorrelation structure $\rho_{\underline{w}s;\tau}^i$ =Corr$[\underline{w}_{s;d}^i, \underline{w}_{s;d+\tau}^i]$;
- the lag-0 cross-correlation coefficients among processes $\underline{w}_s^i$ and $\underline{w}_s^j$ for each season s, i.e., $\rho_{\underline{w}S}^{i,j} = \mathrm{Corr}[\underline{w}_s^i, \underline{w}_s^j]$.

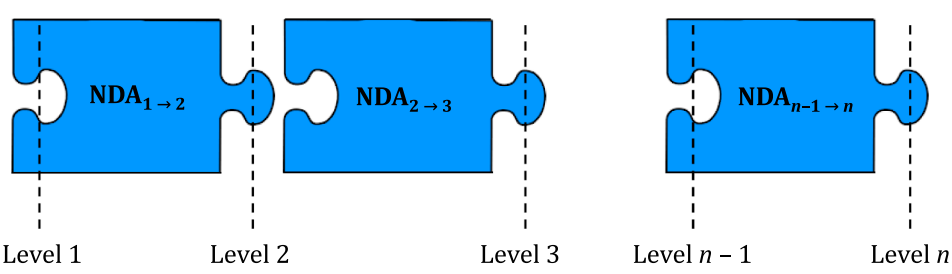

Fig. 2. The stochastic simulation framework as a puzzle, involving a chain implementation of individual NDA *pieces*.

Provided that the parameters of the individual models have been identified (see Section 2.8 for a general overview, as well as Appendix D for a model-specific description), the simulation procedure starts with generating a realization of the annual processes, using the SMARTA model, and subsequently, moves to the monthly and daily level, through the NDA approach. The overall procedure is organized as follows:

4.2.1. Generation of annual synthetic time series

Using SMARTA synthesize an m-dimensional realization of the annual process $\underline{\boldsymbol{y}}_t$ with $t = 1, \cdots, T$, where T denotes the desirable simulation length. The synthesized realization is represented by a $m \times T$ matrix $\boldsymbol{Y}$, i.e.,

$$\boldsymbol{Y} = \begin{bmatrix} y_1^1 & \cdots & y_T^1 \\ \vdots & \ddots & \vdots \\ y_1^m & \cdots & y_T^m \end{bmatrix}$$

4.2.2. Generation and adjustment of monthly synthetic time series

By construction, the realization $\boldsymbol{y}_t$ fulfils the specifications of the annual level, hence the next step is to generate T realizations of the monthly multivariate process $\underline{\boldsymbol{x}}_{s,n}$, each of length 12 (i.e., equal to the number of months). These realizations have to reproduce the specifications implied for the monthly time scale, and, additionally, when aggregated to the annual temporal level, to honor the additive property, i.e., $y_t^i = \sum_{n=(t-1)12+1}^{12t} x_{s,n}^i$.

Therefore, for each year $t = 1, \cdots T$, we employ NDA with the SPARTA model as the generation mechanism (by setting $\boldsymbol{\xi}_t = \boldsymbol{y}_t$ and $\boldsymbol{\omega}_n = \boldsymbol{x}_{s,n}$), and obtain T matrices $\boldsymbol{X}_1, \cdots \boldsymbol{X}_t, \ldots, \boldsymbol{X}_T$, which contain the final adjusted monthly realizations. Each matrix has the form:

$$\boldsymbol{X}_t = \begin{bmatrix} x_{1,(t-1)12+1}^1 & \cdots & x_{12,t12}^1 \\ \vdots & \ddots & \vdots \\ x_{1,(t-1)12+1}^m & \cdots & x_{12,t12}^m \end{bmatrix}$$

Finally, the matrices are concatenated in $\boldsymbol{X} = [\boldsymbol{X}_1, \cdots, \boldsymbol{X}_t, \cdots, \boldsymbol{X}_T]$.

4.2.3. Generation and adjustment of daily synthetic time series

For the disaggregation of monthly generated values to the daily temporal level, and given the previous matrix organization, it is convenient to refer to the obtained, adjusted, monthly realization with reference to season s and year t (not time index n), i.e., $\boldsymbol{x}_{s,t}$, where $s = 1, \cdots, 12$ and $t = 1, \cdots, T$. For instance, in this notation, $\boldsymbol{x}_{3,2}$, refers to the third month of the second year. At this point we have at our disposal, a realization at the monthly level of length $12 \times T$, and seek to generate an equal number of realizations of the daily time scale, each one with length D_s (i.e., 30, 31 or 28). Similarly, to the previous level and for the same reasons, we want the realizations of $\underline{\boldsymbol{w}}_{s;d}$ to resemble the specifications of the sub-monthly time scale, and fulfil the additive property, i.e., $x_{s,t}^i = \sum_{d=(t-1)D_s+1}^{D_s t} w_{s;d}^i$. In this vein, for each month $s = 1, \cdots, 12$ and year $t = 1, \cdots T$, employ NDA using CMARTA for data generation (by setting $\boldsymbol{\xi}_t = \boldsymbol{x}_{s,t}$ and $\boldsymbol{\omega}_n = \boldsymbol{w}_{s;d}$), and obtain $12 \times T$ matrices $\boldsymbol{W}_{s,t}$, which contain the final adjusted daily realizations, i.e.,

$$\boldsymbol{W}_{s,t} = \begin{bmatrix} w_{s;(t-1)D_s+1}^1 & \cdots & w_{s;tD_s}^1 \\ \vdots & \ddots & \vdots \\ w_{s;(t-1)D_s+1}^m & \cdots & w_{s;tD_s}^m \end{bmatrix}$$

Finally, the matrices are concatenated in $\boldsymbol{W} = [\boldsymbol{W}_{1,1}, \cdots, \boldsymbol{W}_{12,1}, \cdots, \boldsymbol{W}_{1,T}, \cdots, \boldsymbol{W}_{12,T}]$, which contains the complete sequence of the daily realization.

5. Case study A: multi-temporal simulation of daily rainfall and runoff processes

To assess the performance of the aforementioned three-level configuration scheme, we selected a case study that regards the contemporaneous synthesis of daily rainfall and runoff data, at the river basin of Achelous, Western Greece, upstream of Kremasta dam, draining an area of 3570 km^2 (Fig. 3a and b). The evaluation of the model is performed at all time scales of interest (daily, monthly, annual), by comparing the empirical, simulated and theoretical (i.e., target) marginal and stochastic characteristics of both the daily and aggregated data. The time horizon of simulation is 2 000 years (i.e., ~730 000 days; Fig. 3c and d).

Regarding the model parameterization, we employed a theoretical autocorrelation model, i.e., Cauchy-type (CAS; Eq. (8)) for describing the auto-dependence structure of the processes, at the annual and daily time scales. It is noted that at the daily scale, the parameters of CAS were varied on a monthly basis. Furthermore, the target distribution functions were varied according to the time scale of simulation, the season and the type of processes (i.e., runoff or rainfall). In all cases, the parameters of the distribution functions have been identified on the basis of historical data, using the L-moments method. Particularly, in the case of runoff, we modeled the data using either the three-parameter Log-Normal ($\mathcal{LN}$; Eq. (A.2)), the Generalized Gamma ($\mathcal{GG}$; Eq.

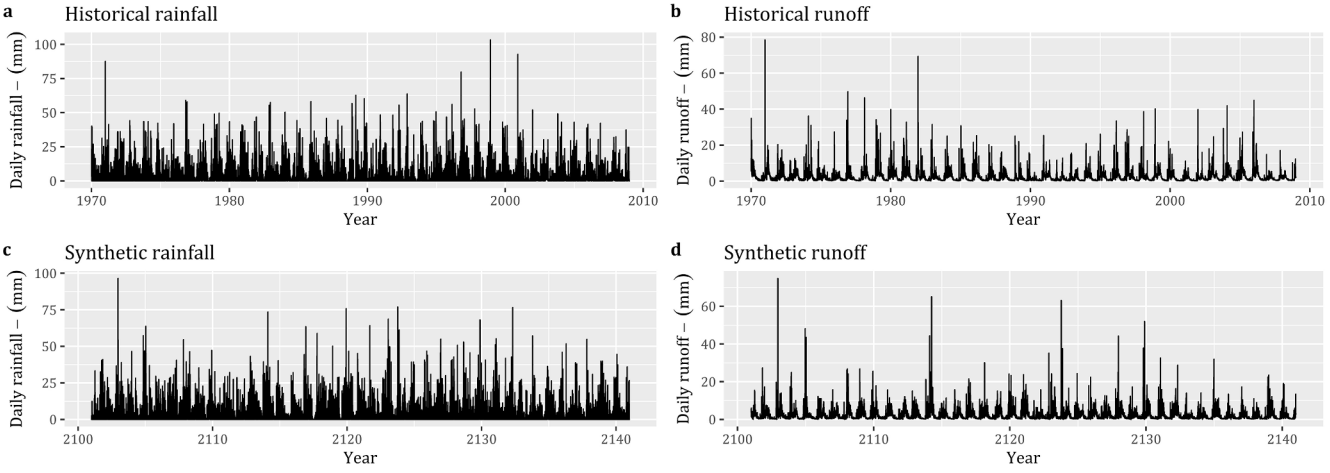

Fig. 3. a-b) Historical daily rainfall-runoff time series (1 January 1970 to 31 December 2008). c-d) Synthetically generated time series (randomly selected window of 40 years).

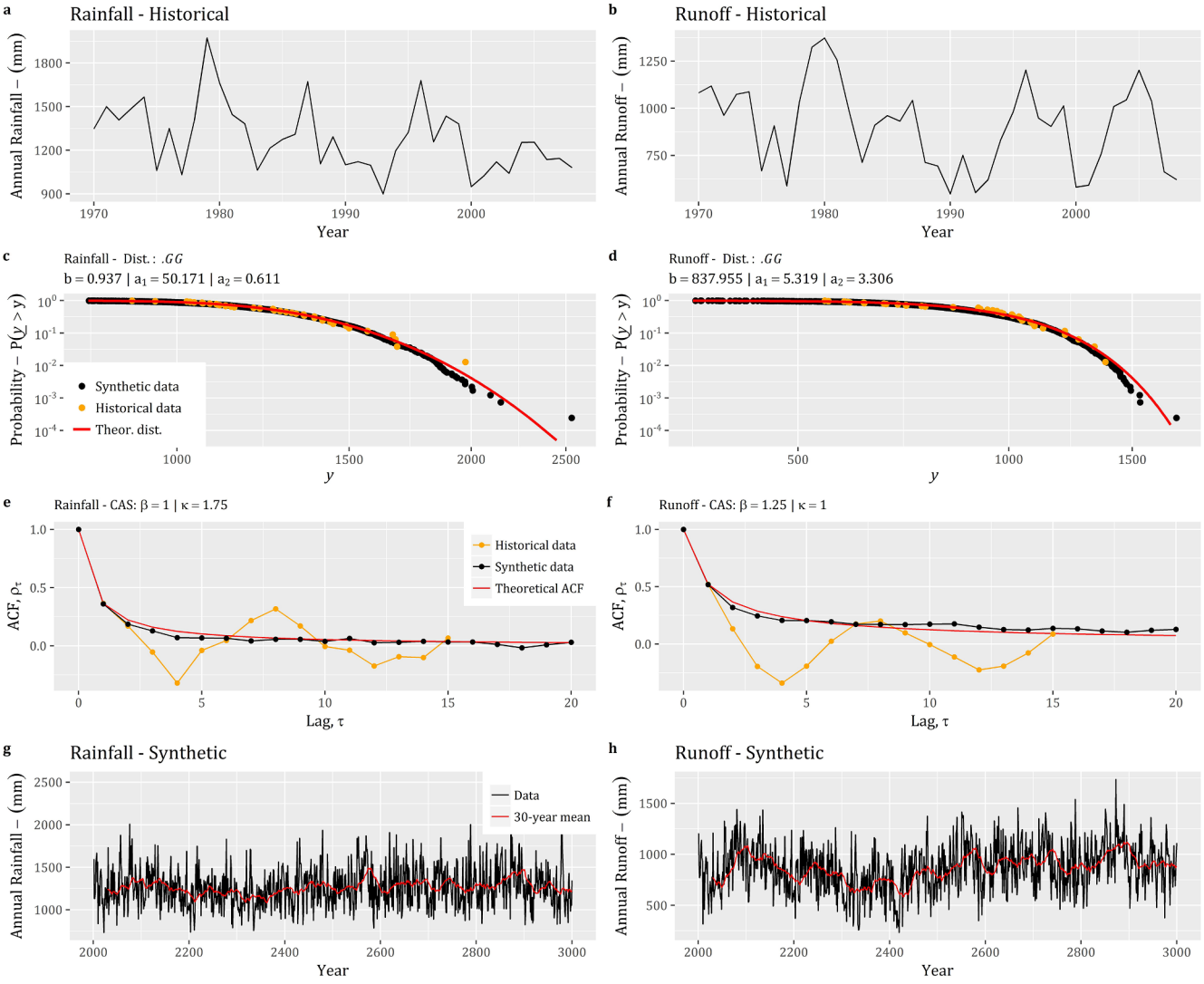

Fig. 4. Rainfall-runoff series: (a-b) Historical annual time series. (c-d) Empirical, simulated and theoretical distribution functions (using the Weibull's plotting position). (e-f) Empirical, simulated and theoretical ACFs. (g-h) Synthetic annual time series (randomly selected window of 1 000 years). Abbreviations: Generalized Gamma distribution (GG), Cauchy-type autocorrelation structure (CAS).

(A.3)) or the Burr type-XII ($\mathcal{B}r$XII; Eq. (A.4)) distribution. On the hand, for the daily rainfall process, which is characterized by intermittent behavior, we employed the zero-inflated distribution model of Eq. (6), using for the continuous component one of the aforementioned distributions.

Starting from the annual temporal level, Fig. 4, summarizes the ability of the highest-level model to preserve both the target distribution function and the autocorrelation structure of each process. Furthermore, the model resembled the lag-0 cross-correlation among the two processes with high accuracy (the historical and simulated values are 0.813 and 0.815 respectively; the high values are physically explained by the significantly rich hydrological conditions of the basins). It is noted that the parameters of CAS have been manually fine-tuned in order to increase the *degree* of annual long-range dependence and stress-test the capabilities of the associated simulation scheme.

Figs. 5 and 6a provides a quick outlook of the results obtained at the monthly time scale, preserving with high accuracy, the empirical L-moments, the seasonality, expressed by means of month-to-month correlation coefficients, as well as the lag-0 cross-correlations.

Beyond summary statistics, a more challenging test is the reproduction of the monthly target marginal distributions. Figs. 7 and 8, compare the empirical distribution of the historical and synthetic data with the target theoretical model (the fitted distribution, as well as its parameters are shown in the title of each sub-plot). In all cases, the model resembled the target distribution with notable accuracy.

The previous figures, illustrate the ability of the integrated model, to generate cyclostationary realizations that are also consistent with the specifications of the annual temporal level. As an additional diagnostic, and to test the model for *envelope behavior* we employed scatter plots, and depicted the established dependence patterns. An example is given in Fig. 9, which depicts the lag-1 month-to-month dependence patterns of runoff series. The scheme does not exhibit the aforementioned behavior, yet more interestingly, it was found capable of creating a variety of dependence forms, which are also in accordance with the historical ones. The results obtained for other time scales (or rainfall) are similar, hence not shown herein.

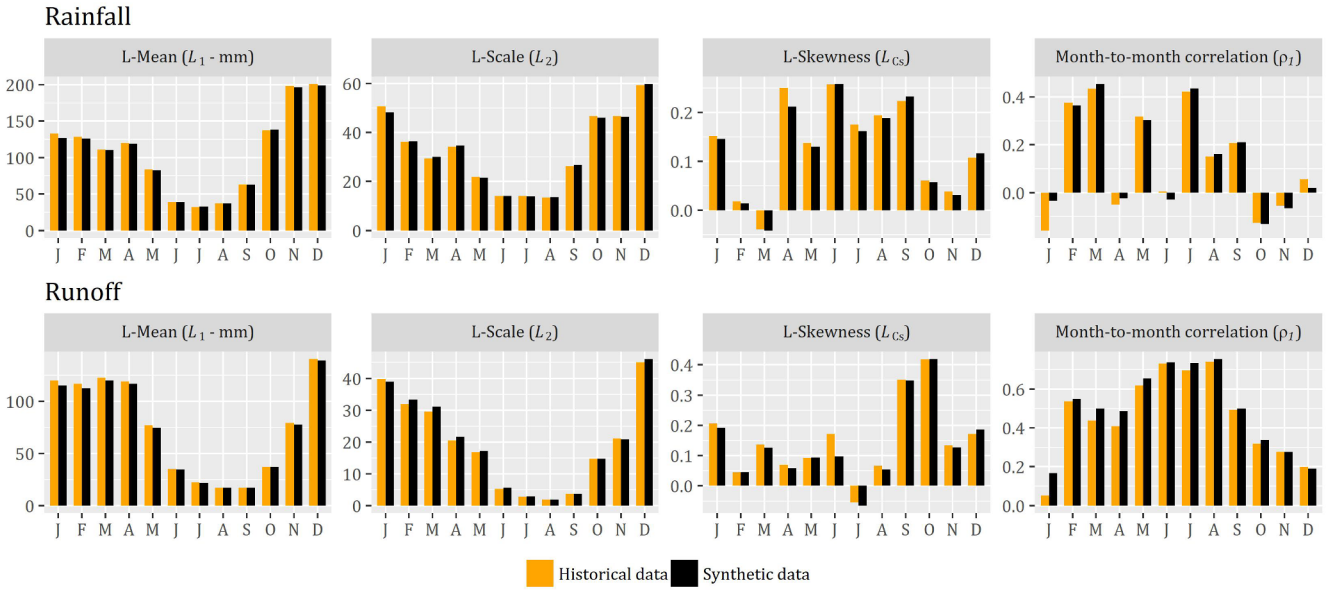

Fig. 5. Comparison of monthly empirical and simulated L-Mean, L-Scale and L-Skewness, as well as historical and simulated lag-1 month-to-month correlations.

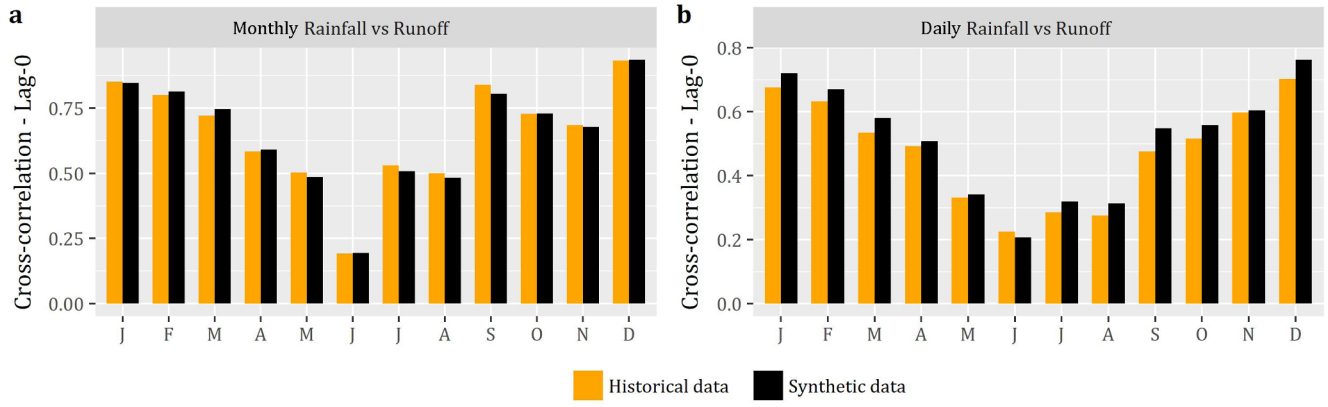

Fig. 6. Comparison of historical and simulated lag-0 cross-correlations at (a) monthly and (b) daily scale.

Regarding the lowest level of simulation, that is the daily time scale, the comparison among summary statistics of Figs. 10 and 6b (depicting the lag-0 cross-correlations at daily scale), as well as the empirical, simulated and theoretical distribution functions depicted in Figs. 11 and 12, underline the ability of the model to generate consistent realizations with the higher levels, and also preserve the target distribution functions of the daily process, which at this time scale, are characterized by considerably heavier tails. Notice that for daily runoff, and for the months, February to May, we selected the $\mathcal{B}r$XII model, which is a heavy-tailed distribution with power-type tail. Recall, that the r^{th}-moment of the $\mathcal{B}r$XII exist only if $a_1 a_2 < r$. Remarkably, NDA accurately simulated even February's daily runoff, which is characterized by $a_1 a_2 < 2.90$; implying that it only has finite mean and variance.

Furthermore to this, Figs. 13 and 14 depict a monthly-based comparison of the empirical, simulated and theoretical autocorrelation function (ACF) of the daily process, which in most cases deviates from the typical AR(1) ACF, that most daily stochastic models are capable of simulating. Inspection of this figure, reveals that the integrated model can resemble the theoretical auto-dependence structure with high precision. This result stems from the combination, within NDA, of two modelling components; the CMARTA and the use of theoretical auto-dependence structure (e.g., CAS).

An additional assessment of model's performance concerns its capabilities regarding the reproduction of statistical characteristics that are not explicitly modelled by the method. These may involve marginal properties at intermediate temporal scales, as well as a plethora of other features, such as transition probabilities, dry/wet spells, and asymptotic properties, associated with the reproduction of extremes.

In this context, in order to evaluate the model's behavior at the intermediate time scales between daily and monthly, we aggregated the historical and synthetic daily series at several scales $k \in \{2, ..., D_s\}$ and compared, on a monthly basis, the L-mean ($L_1^{(k)}$), L-scale ($L_2^{(k)}$), L-Skewness ($L_{Cs}^{(k)}$) coefficients and the probability dry ($P_D^{(k)}$) at each scale k. This analysis is presented in the Supplementary material (Figs.

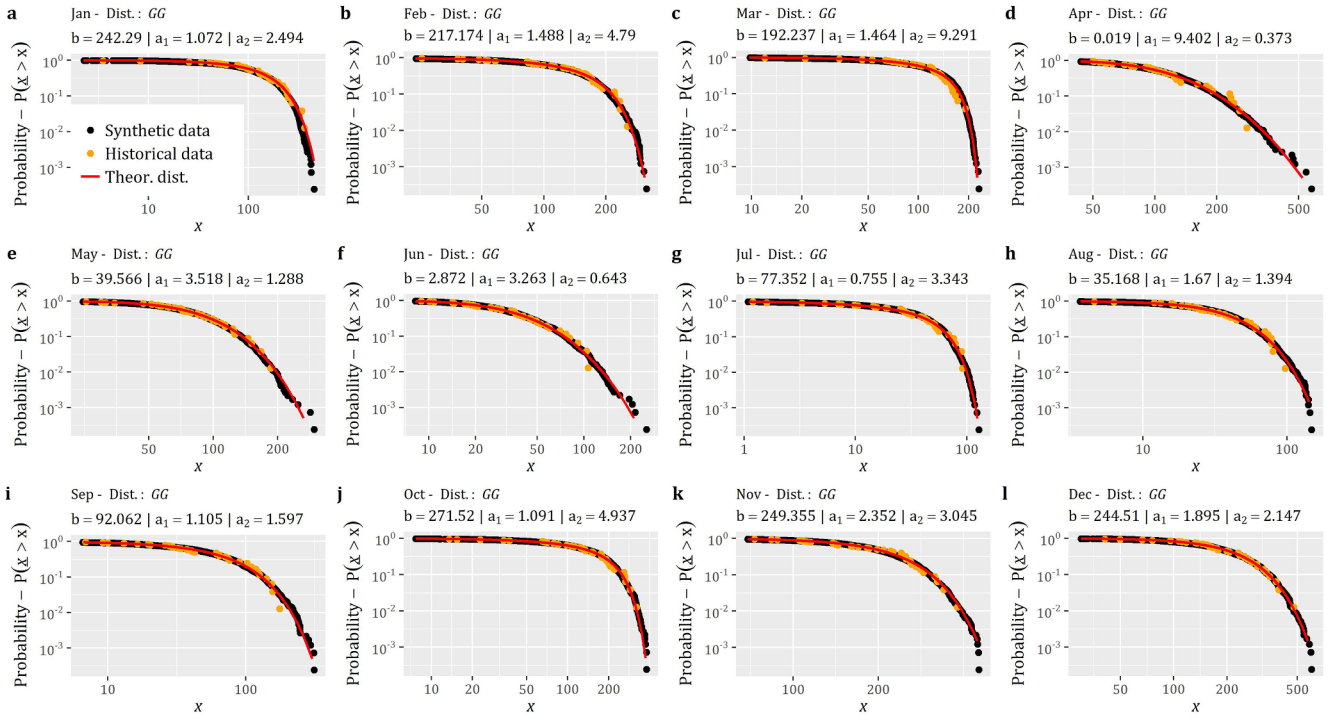

Fig. 7. Monthly rainfall - monthly-based comparison of empirical, simulated and theoretical distribution functions (using the Weibull's plotting position). The title of each subplot provides the selected distribution and its parameters, as well as the historical (p_D) and simulated ($\hat{p}_D$) values of probability dry. Abbreviations: Generalized Gamma distribution (GG).

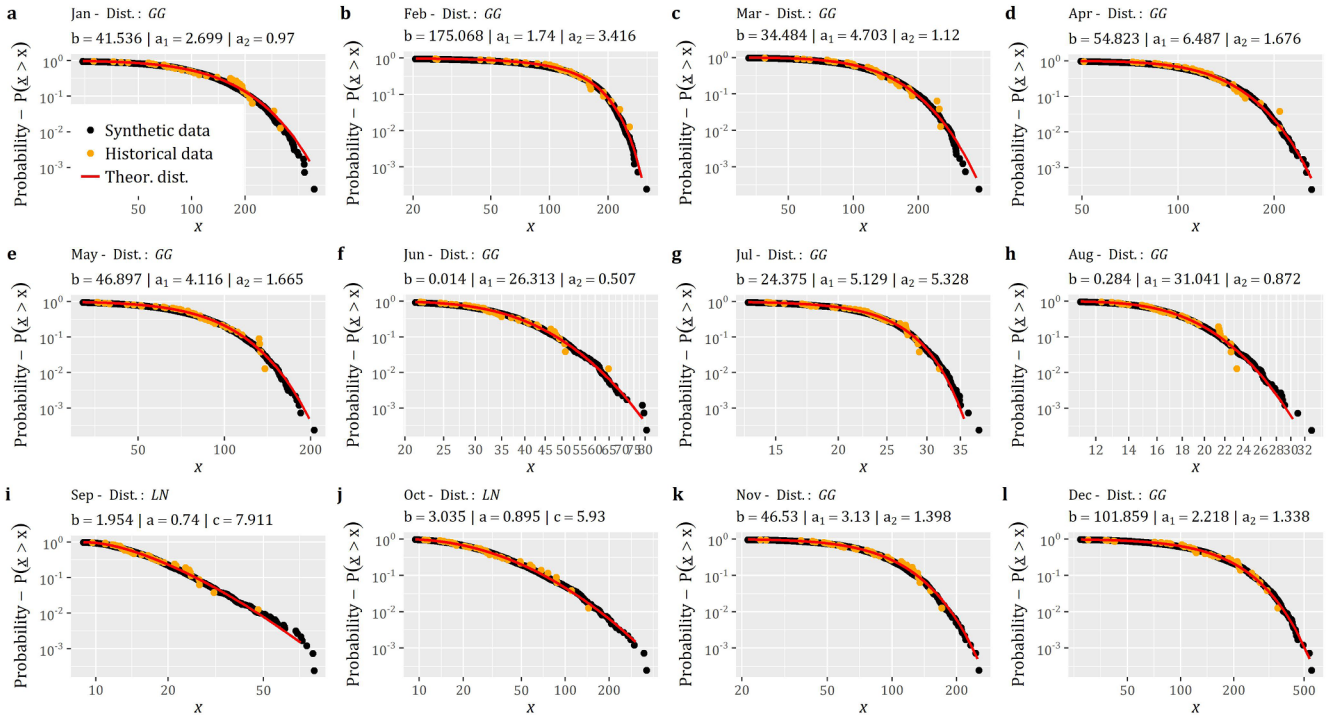

Fig. 8. Monthly runoff - monthly-based comparison of empirical, simulated and theoretical distribution functions (using the Weibull's plotting position). The title of each subplot provides the selected distribution and its parameters, as well as the historical (p_D) and simulated ($\hat{p}_D$) values of probability dry. Abbreviations: Generalized Gamma distribution (GG), Log-Normal distribution (LN).

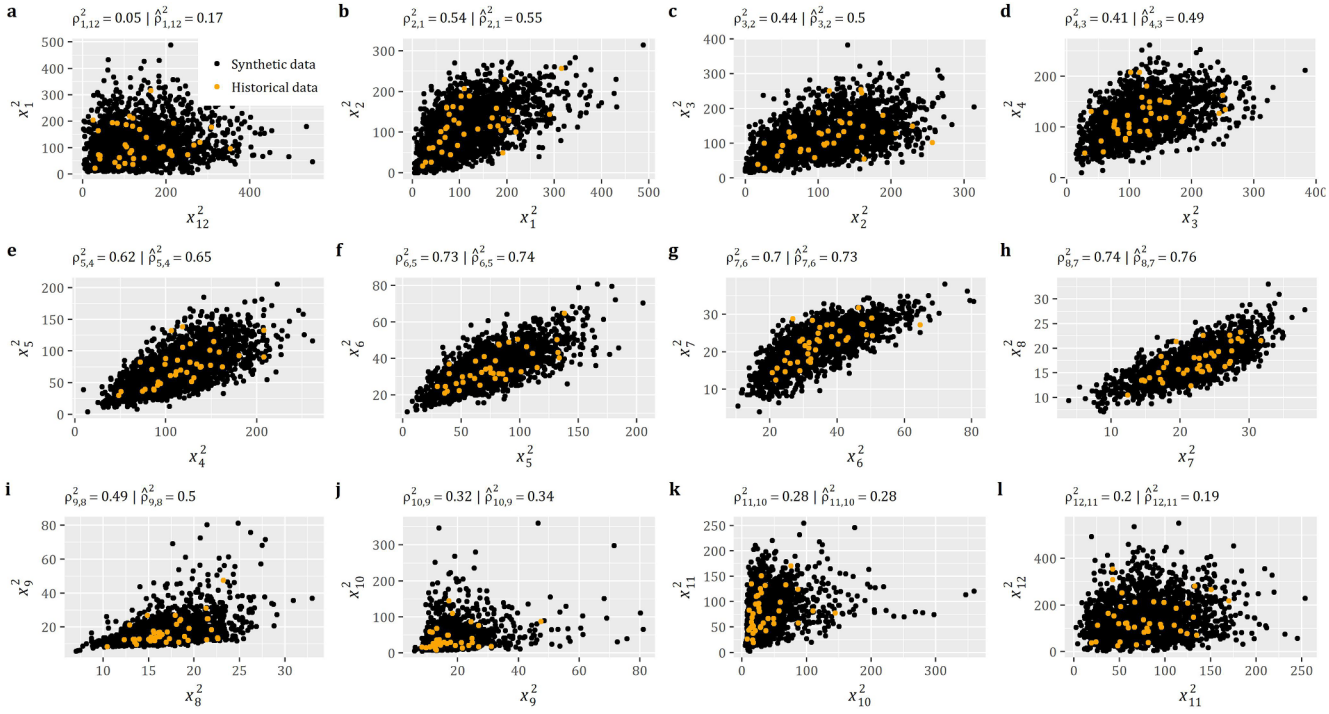

Fig. 9. Monthly runoff (mm) month-to-month scatter plots of historical and simulated series. The title of each subplot provides the lag-1 month-to-month target ($\rho_{s,s-1}$) and simulated ($\hat{\rho}_{s,s-1}$) correlation coefficients.

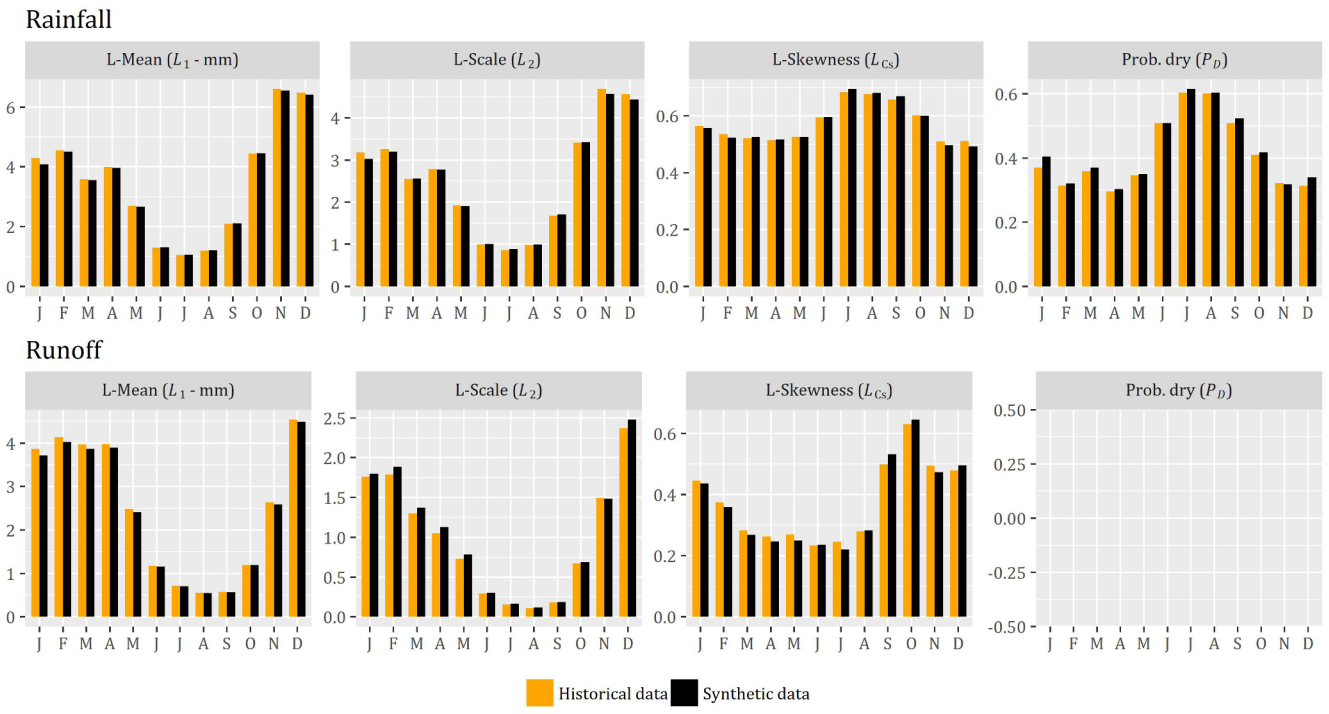

Fig. 10. Comparison of daily empirical and simulated L-Mean, L-Scale, L-Skewness, as well as probability dry.

S1–S8). As shown, although the intermediate time scales (i.e., $k \neq \{1, D_s\}$) are not explicitly modelled neither by the three-level scheme or NDA, the arguably good agreement with historical data, can be attributed to the accurate simulation of the process at daily and monthly time scales.

In the Supplementary material, we further explore the intermittency-related features of daily rainfall, by comparing the dry and wet spell distributions (Fig. S9), as well as the transition probabilities between dry-dry and wet-wet sequences (Fig. S10). In this case, also, the synthetic data exhibit a satisfactory agreement with the historical

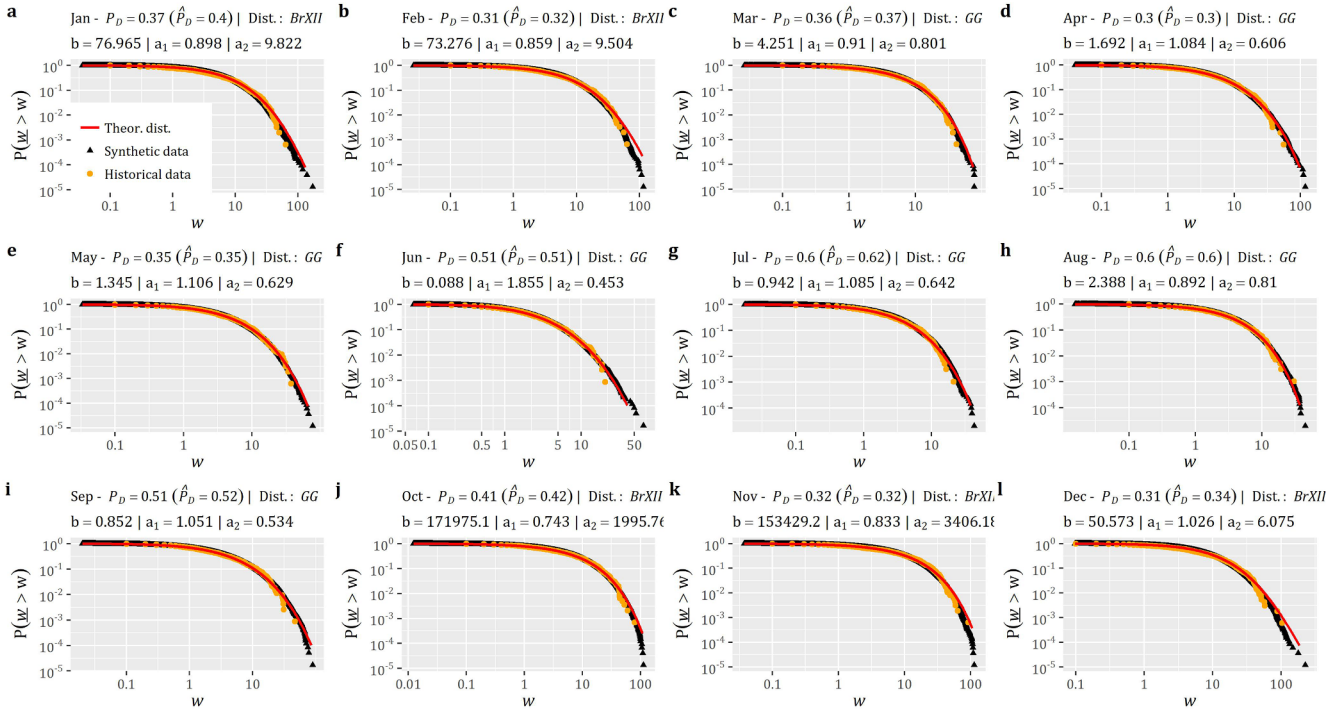

Fig. 11. Daily non-zero rainfall - monthly-based comparison of empirical, simulated and theoretical distribution functions (using the Weibull's plotting position). The title of each subplot provides the selected distribution and its parameters, as well as the historical (p_D) and simulated ($\hat{p}_D$) values of probability dry. Abbreviations: Generalized Gamma distribution (GG), Burr type-XII distribution (BrXII).

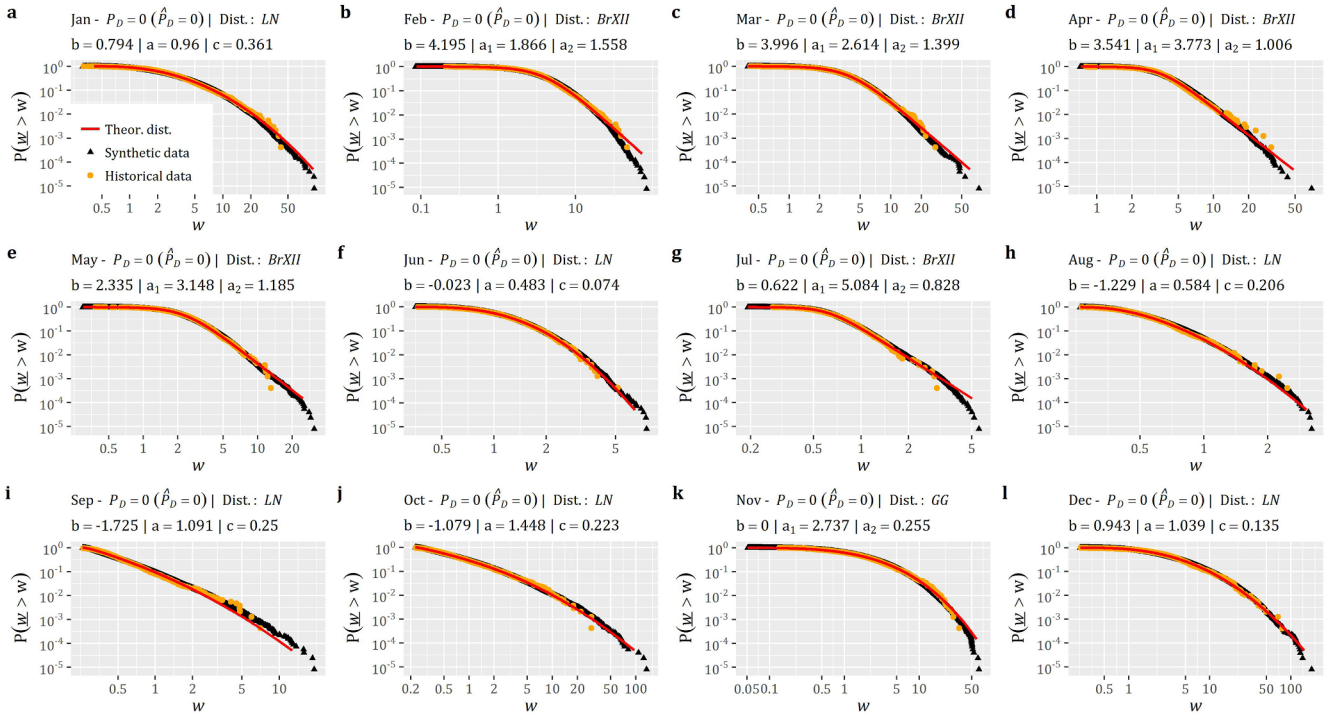

Fig. 12. Daily non-zero runoff - monthly-based comparison of empirical, simulated and theoretical distribution functions (using the Weibull's plotting position). The title of each subplot provides the selected distribution and its parameters, as well as the historical (p_D) and simulated ($\hat{p}_D$) values of probability dry. Abbreviations: Generalized Gamma distribution (GG), Log-Normal distribution (LN), Burr type-XII distribution (BrXII).

ones, which is mainly attributed to the reproduction of the auto-dependencies, in combination with the flexibility provided by the mixed-type distributions.

However, in our view, an even more important evidence of the model's success involves the reproduction of the extremes. In this vein, Fig. 15 depicts the empirical and simulated daily annual maxima, as well as the fitted (using the L-moments method), to the historical data, Generalized Extreme Value ($\mathscr{GEV}$) distribution (i.e., Eq. (A.7)). As

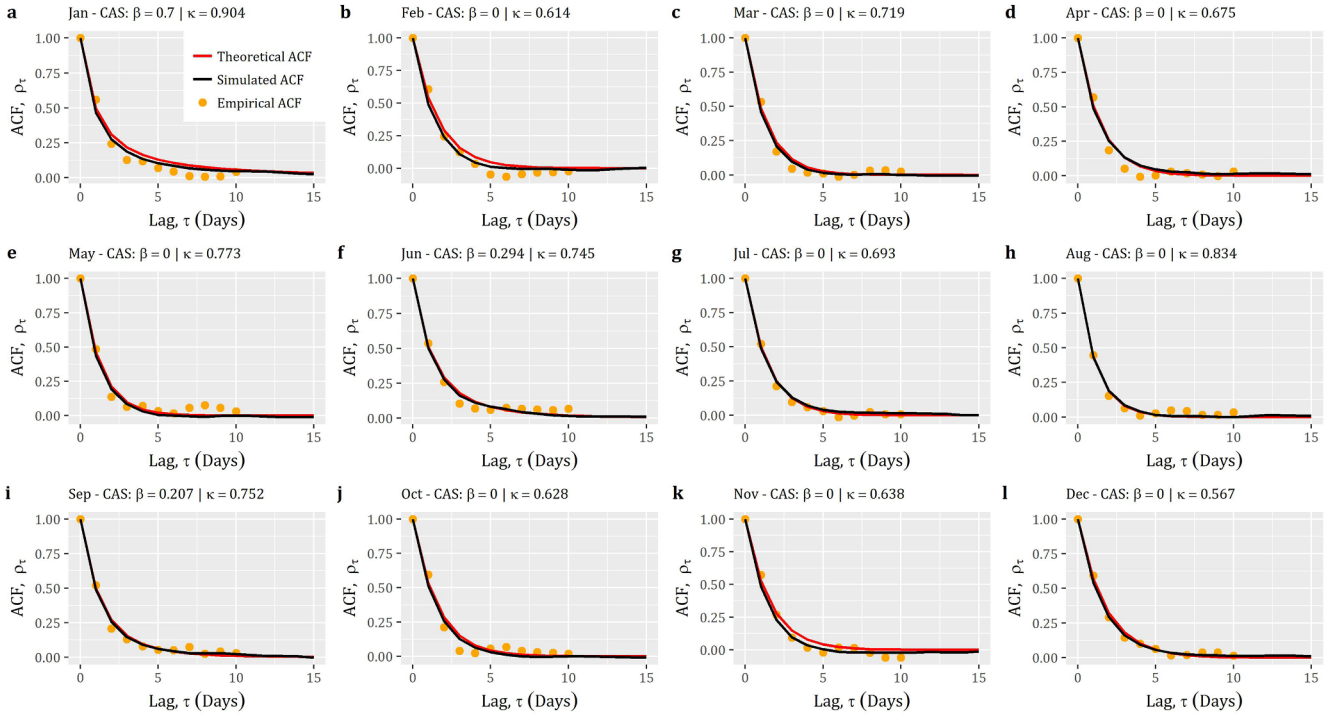

Fig. 13. Daily rainfall - monthly-based comparison of empirical, simulated and theoretical autocorrelation function (ACF); the parameters of the Cauchy-type autocorrelation structure (CAS) are given on the title of each subplot.

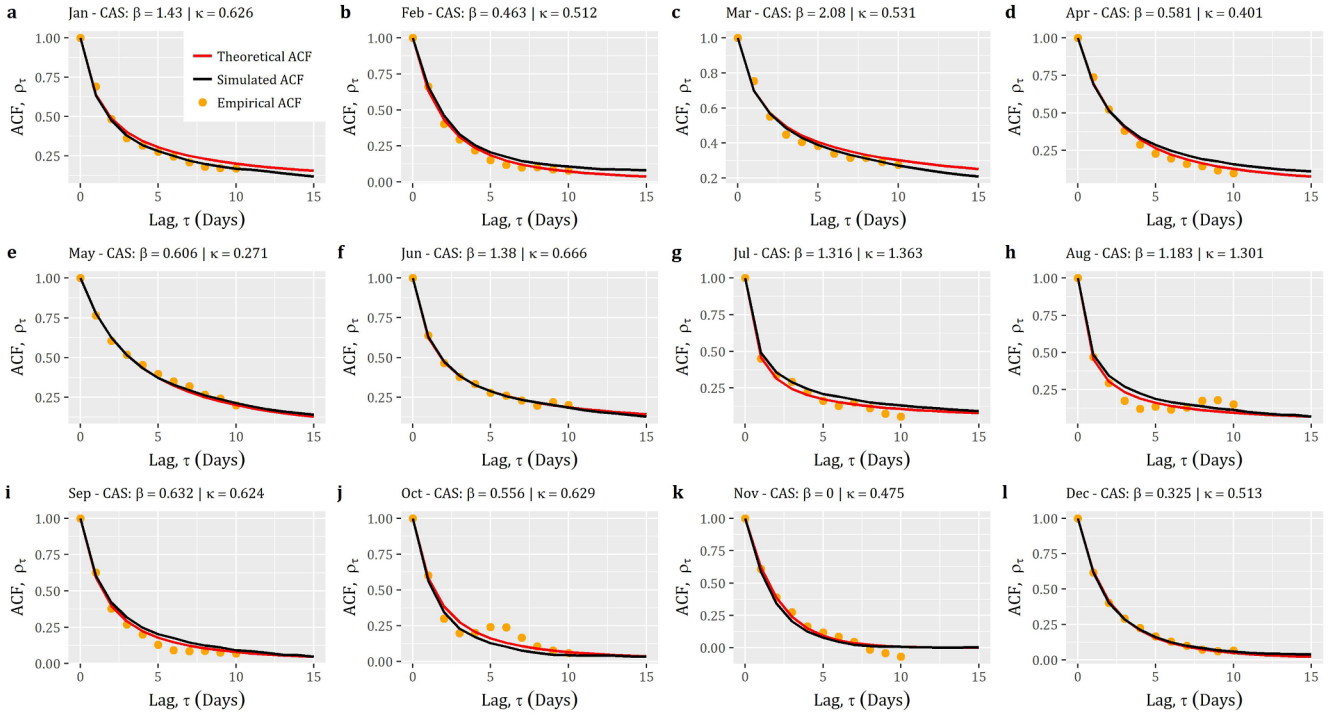

Fig. 14. Daily runoff - monthly-based comparison of empirical, simulated and theoretical autocorrelation function (ACF); the parameters of the Cauchy-type autocorrelation structure (CAS) are given on the title of each subplot.

shown, in both cases the model manages to resemble the distributional form of the identified $\mathscr{GEV}$ distribution, characterized by an arguably heavy-tailed behavior, which is expressed through the Fréchet distribution (since $a > 0$). Reasonably, this behavior is attributed to the concise reproduction of the distributions at the daily time scale, which was modelled using either the power-type $\mathcal{B}r$XII or the $\mathcal{LN}$ distribution (we recall that at each month we applied different distribution models, based on empirical criteria and hydrological evidence).

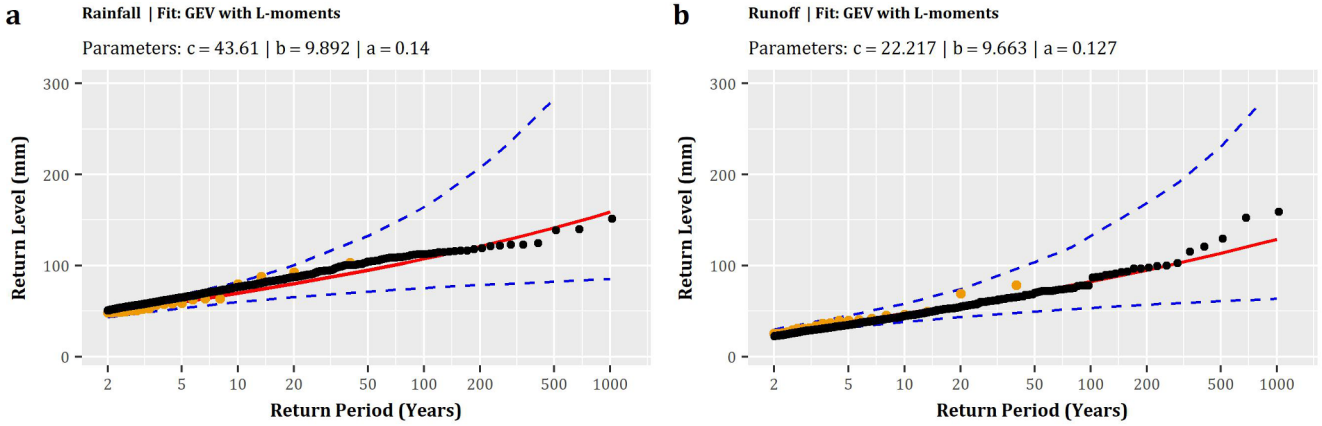

Fig. 15. Empirical (●) and simulated (●) daily annual rainfall-runoff maxima, as a function of the return period. The solid red line (—) depicts the fitted to historical data Generalized Extreme Value ($\mathscr{GEV}$) distribution (parameters: location (c), scale (b) and shape (a)). The dashed blue lines (– – –) represent the 95% confidence intervals (estimated using the parametric bootstrap method). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

6. Case study B: disaggregation of daily rainfall to hourly scale

To demonstrate the flexibility provided by NDA, as well the potential to extend the three-level scheme of Section 4 for even lower temporal levels, we now provide a two-level configuration for disaggregating a univariate daily sequence to the hourly scale. This scheme is applied to an hourly rainfall dataset at Oberstdorf, Germany, provided from the German Weather Service (Deutscher Wetterdienst; station ID 3730). The historical data extend over the period 01/09/1995 to 31/12/2017 (Fig. 16).

In this example, we do not aim to generate synthetic data that represents the actual process across multiple time scales of interest (such as in case study A). In contrast, our goal is to provide a synthetic hourly realization, under the following requirements:

- the synthetic data at the hourly scale reproduces the probabilistic and stochastic properties of the historical sample;
- the additive property is preserved between the aggregated hourly ($k = 1$) synthetic data and the corresponding historical ones ($k^* = 24$).

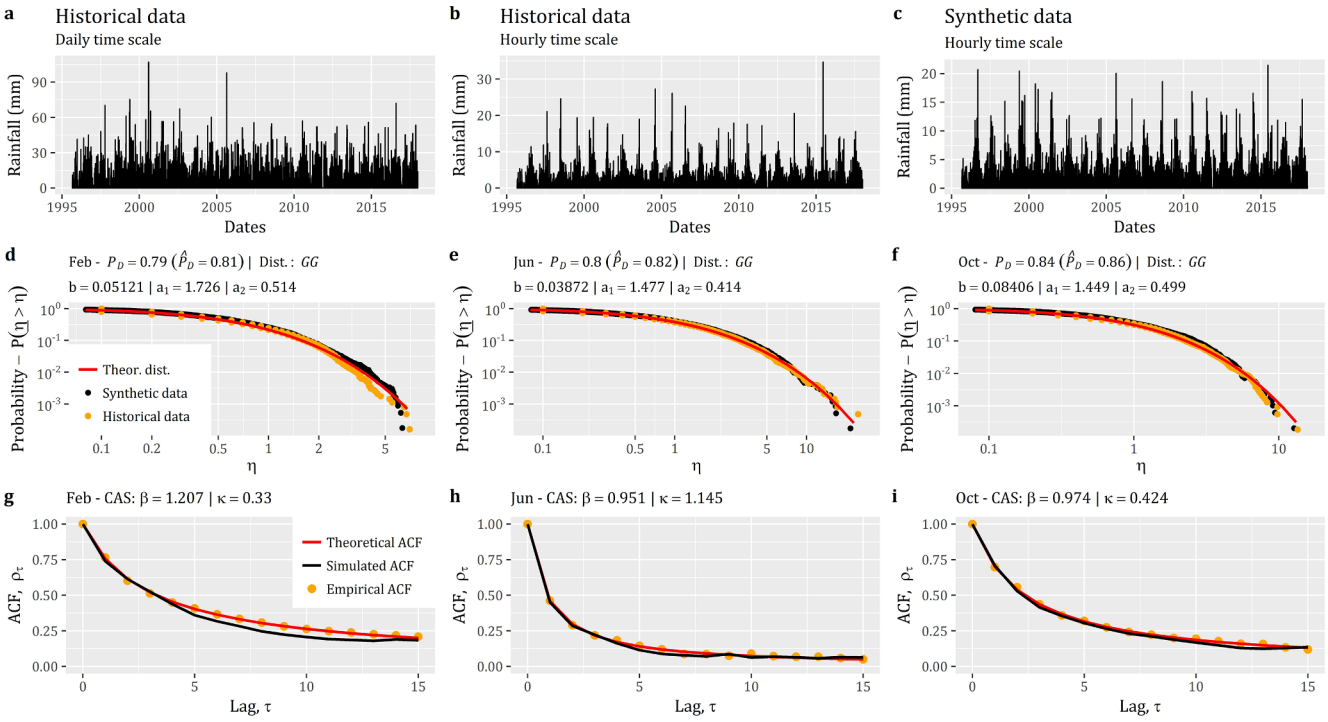

Fig. 16. Historical a) daily and b) hourly rainfall series. c) Synthetic (disaggregated) hourly rainfall realization. d-f) Comparison of distribution function of non-zero amounts for hourly historical and disaggregated series for February, June and October respectively (the fitted theoretical model is shown with red line). g-i) Comparison of autocorrelation function (ACF) for hourly historical and disaggregated series for February, June and October respectively (the fitted theoretical model is shown with the red line). Abbreviations: Generalized Gamma distribution (GG), Cauchy-type autocorrelation structure (CAS).

By definition, in disaggregation problems, the synthetic sequence has the same length as the given data.

To cope with the effect of seasonality, we employ the typical assumption for fine-time scale rainfall processes (e.g., daily, hourly or finer), that of cyclical stationarity with annual period and monthly subperiod (see also Section 4). Assuming that the sequence $\{w_{s;d}\}$ denotes the observed daily records for month s, we simulate an hourly process, e.g., $\{\underline{\eta}_{s;h}\}_{h\in\mathbb{Z}^>}$, which is also considered stationary within month s. This implies that the distribution function, $F_{\underline{\eta}_s}$, of the process, as well as its auto-correlation structure, i.e., $\rho_{\underline{\eta}_s;\tau} = \mathrm{Corr}[\underline{\eta}_{s;h}, \underline{\eta}_{s;h+\tau}]$ remain invariant within the month s. Furthermore, to account for temporal consistency we impose the requirement of generating realizations of the process $\{\underline{\eta}_{s;h}\}$ constrained by, $w_{s;d} = \overline{w}_{s;d}$, where $\overline{W}_{s;d}: = \eta^{(24)}_{s;l} = \sum_{h=(l-1)24+1}^{l24} \eta_{s;h}$ (analogous to Eq. (10)).

In order to simulate the hourly rainfall, we employ as generation mechanism the univariate version of CMARTA, which is known as ARTA (Cario & Nelson, 1996). We recall that this model uses, as an auxiliary Gp, a Gaussian AR process (see Appendix B.1). The generation scheme is employed on a monthly basis, since the hourly process properties are reasonably considered seasonally varying.

Regarding the parameterization of ARTA, the marginal distribution of hourly rainfall of each month is modelled using the zero-inflated model of Eq. (6). In this case, for the continuous part we fitted (using L-moments) the $\mathcal{GG}$ distribution (i.e., Eq. (A.3); the parameters of the model are shown in Fig. 16d–f). For the autocorrelation structure of the hourly rainfall, we fitted monthly-varying CAS models (Eq. (8)) to the corresponding empirical autocorrelation coefficients (red line in Fig. 16g–i). Eventually, each individual hourly process is modeled using five parameters (three for the marginal distribution and two for the

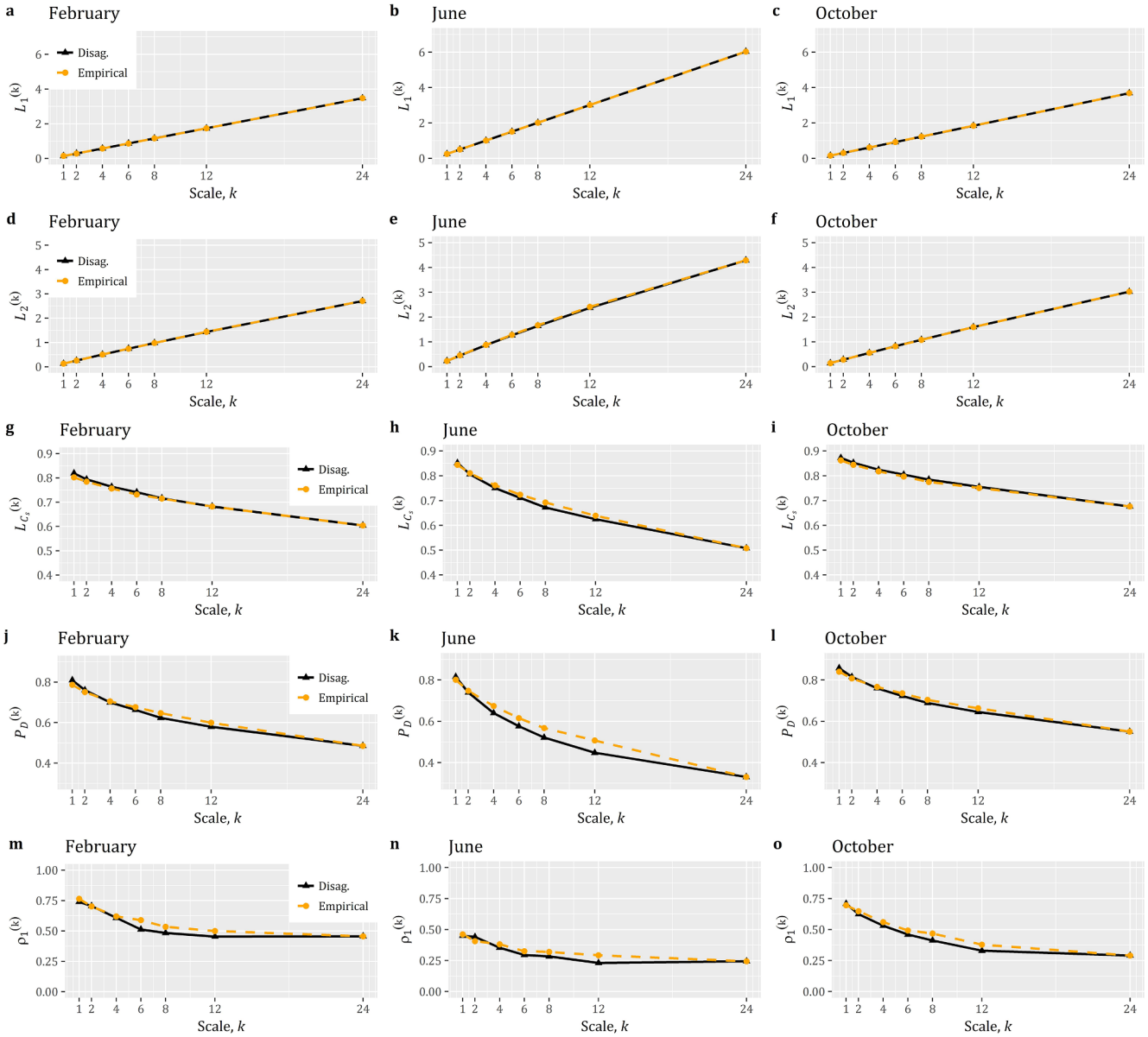

Fig. 17. Comparison of empirical and disaggregated, a-c) L-mean ($L_1^{(k)}$), d-f) L-scale ($L_2^{(k)}$), g-i) L-skewness ($L_{Cs}^{(k)}$), j-l) probability dry ($p_0^{(k)}$) and m-o) lag-1 autocorrelation coefficient ($\rho_1^{(k)}$), as a function of aggregation scale k, for February, June and October.

autocorrelation structure).

In Fig. 16c are shown the synthetic (i.e., disaggregated to hourly scale) rainfall data. In Fig. 16d-e and g–i we show for three representative months (February, June, October) that the model resembles with good accuracy the target distributions and autocorrelation structures, respectively. Similarly good performance is achieved for the rest of the months (see Supplementary material; Figs. S11 and S12).

Furthermore, in order to investigate the behavior of the model at the intermediate time scales ($1 < k < 24$), in Fig. 17, we depict for three characteristics months (February, June and October), both the historical and synthetic series, the L-mean, $L_1^{(k)}$, L-scale, $L_2^{(k)}$, L-skewness, $L_{Cs}^{(k)}$, probability dry, $p_D^{(k)}$, and lag-1 autocorrelation coefficient, $\rho_1^{(k)}$, as a function of the aggregation level k (the rest of the months are shown in Figs. S13–S17 of Supplementary material). Inspection of these plots reveals the potential of the NDA approach to preserve the empirical scaling properties of rainfall, without requiring the use of cascading of techniques and direct simulation of rainfall at intermediate temporal levels.

7. Conclusions

In order to address the puzzle of multi-temporal simulation of hydrometeorological processes, we developed a puzzle-type approach, employing chain implementation of a novel generation procedure, called Nataf-based Disaggregation to Anything (NDA). This is built upon recent advances in stochastics by means of Nataf-based models, coupled with the concepts of repetitive sampling and adjusting.

This coupling allows taking advantage of the primary ability of Nataf-based models to represent stationary processes that exhibit any distribution and any correlation structure. The recent extension of Nataf-based models to simulate cyclostationary as well as multivariate processes, offered the essential *generality* to handle challenging single-scale hydrometeorological simulation problems.

However, as widely discussed, the reproduction of a target's probabilistic and stochastic behavior at a single temporal scale does not guarantee similarly consistent performance at higher temporal scales. In this work, the issue of *consistency across any pair of scales* is handled via the NDA approach, while the general puzzle-type framework enables the transition to multi-scale simulations. We recall that NDA uses Nataf-based models at two independent scales as underlying data generators, and coupling mechanisms to adjust the lower-level data to the higher one.

The above approach ensures significant *flexibility*, since it allows establishing any configuration of scale-consistent simulators, through pairwise link of NDAs. This flexibility and the advantages of NDA itself have been mainly revealed by configuring a multivariate simulation scheme that reproduces the probabilistic and stochastic properties of the processes of interest at three characteristic temporal scales (i.e., annual, monthly and daily). In this configuration, we integrated different Nataf-based models for each scale, i.e., SMARTA for the annual, SPARTA for the monthly, and CMARTA (and incidental contribution of this work) for the daily one.

The multi-temporal simulation capabilities of the integrated scheme were evaluated on the basis of a long-term bi-variate simulation study, aiming at the generation of synthetic rainfall and runoff data. As shown, the model reproduced with accuracy the characteristics of the underlying hydrometeorological processes, which exhibit substantial differences among the two processes and across scales and seasons. Key requirements in this study were:

- the reproduction of a wide range of *target* distribution functions, varying across processes, scales and seasons;
- the simultaneous simulation of intermittent and continuous processes (i.e., daily rainfall and runoff), exhibiting significant correlations;
- the preservation of *target* short-term and long-term auto-dependence structures, at the annual scale, as well as the daily scale, on seasonal basis;
- the preservation of *target* season-to-season correlations at the monthly scale;
- the preservation of *target* lag-0 cross-correlations at all scales.

One can observe that in the above *bucket list* we make repeated use of term *target*, in order to highlight the multidimensional role of the user. Actually, before employing simulations, there are several critical modelling decisions to make, regarding the assignment of suitable distribution functions and correlation structures to the processes of interest (this also involves the selection of time scales to represent, thus the configuration of the puzzle). This flexibility may offer significant advantages. For instance, in this specific study, the careful selection of the daily distribution models resulted in reproducing the heavy-tailed behavior of the observed daily extremes. We recall that the reproduction of extremes was not set as an explicit requirement of the model, thus making this surprising outcome a promising topic for further research.

The model performance at even finer temporal scales (i.e., hourly) was demonstrated through a disaggregation example, where we employed NDA for the synthesis of hourly rainfall realizations that are consistent with the observed daily data. Similarly to the first study, the model faithfully reproduced the target behavior of the hourly process, simultaneously ensuring consistency with the daily scale. Moreover, it reproduced with accuracy important statistical properties of rainfall (expressed in terms of L-moments) at intermediate scales. Above all, this study highlighted the scale-free character of NDA, as well as its ability to handle hydrological disaggregation problems.

Arguably, the potential applications of our puzzle-type approach extend beyond the realm of hydrometeorological time series generation (or disaggregation). Essentially, it is a general-purpose stochastic simulation scheme. Depending on the synthesis of the puzzle pieces (i.e., chain of NDAs), as well as the underlying decisions of each NDA (in terms of target marginal distributions and correlation structures), it is possible to apply the method for the simulation of a widely extended range of processes, geophysical and socioeconomic.

Beyond simulation, other applications of NDA may concern downscaling or disaggregation problems, which require: a) replacing the corresponding higher-level simulation model with the realizations provided by global or regional climate models, and b) identifying the marginal and stochastic properties of the lower-level model, using, e.g., in-situ gauging stations, regional information, and/or scaling laws.

Eventually, the proposed approach can be employed within broader Monte Carlo experiments, to provide long synthetic input data to deterministic simulation models. Given that the type and number of processes to simulate, as well as their temporal resolution, is dictated by the deterministic model, a major computational challenge arises. In particular, the repetitive sampling within NDA imposes a bottleneck, when applied to high-dimensional multivariate problems and/or long-term simulations at fine time scales. Potential remediation to this technical problem may be the use of parallel computing or the model implementation in low-level programming languages.

Regarding the modelling framework per se, potential future

research may focus on two interesting aspects that have been revealed in the two case studies. The first involves the reproduction of non-explicitly preserved statistical features, with emphasis to extremes, while the second is the validation of the model behavior at intermediate time scales. Another interesting research target is the comparison with alternative synthetic generation schemes, to assess the practical impacts of stochastic model selection in problems that typically rely on synthetic inputs, such as water resources planning and management, hydrological simulations, and flood risk assessment. Since each problem is influenced by different statistical aspects of the input processes, the generality and flexibility of the proposed method makes it favorable for a wide range of applications.

Declaration of Competing Interest

None.

Acknowledgments

The authors would like to thank the Associate Editor and the three reviewers, Prof. Geoffrey Pegram, Prof. Elena Volpi, and an anonymous one, for their constructive comments and suggestions, which helped us improve the manuscript.

Data availability: The historical dataset of rainfall and runoff of Achelous river basin upstream of Kremasta dam in Western Greece (employed in section 5) is available at: www.itia.ntua.gr/1914/. The hourly rainfall series used in section 6, were retrieved from the German Weather Service (Deutscher Wetterdienst; DWD), available at: https://www.dwd.de/EN/climate_environment/cdc/cdc.html.

Code availability: The methodology, as well as the stochastic models presented herein have been implemented in an R package named *anySim* (Tsoukalas and Kossieris, 2019), available at: http://www.itia.ntua.gr/en/softinfo/33/.

Appendix A:. Distribution functions

The probability density function (PDF) of the Gamma distribution ($\mathcal{G}$) is given by,

$$f_{\mathcal{G}}(x;\, a,\, b) = \frac{1}{|b|\ \Gamma(a)}\left(\frac{x}{b}\right)^{a-1}\exp\left(-\frac{x}{b}\right),\ x > 0 \tag{A.1}$$

where $a > 0$, $b \neq 0$, are shape and scale parameters respectively, while $\Gamma(\bullet)$ stands for the gamma function.

The PDF of the 3-parameter Log-Normal distribution ($\mathcal{LN}$) is given by,

$$f_{\mathcal{LN}}(x;\, a,\, b,\, c) = \frac{1}{(x-c)a\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{\log(x-c)-b}{a}\right)^2\right),\ x > c \tag{A.2}$$

where $a > 0$, $b \in \mathbb{R}$, and $c \in \mathbb{R}$ denote the shape, scale and location parameters respectively.

The PDF of the Generalized Gamma ($\mathcal{GG}$) distribution is given by (Stacy, 1962),

$$f_{\mathcal{GG}}(x;\, a_1,\, a_2,\, b) = \frac{a_2}{b\Gamma(a_1/a_2)}\left(\frac{x}{b}\right)^{a_1-1}\exp\left(-\left(\frac{x}{b}\right)^{a_2}\right),\ x > 0 \tag{A.3}$$

where $\Gamma(\bullet)$ denotes the gamma function, while, $a_1 > 0$, $a_2 > 0$ are shape parameters and $b > 0$ is a scale parameter.

The PDF of the Burr type-XII distribution ($\mathcal{B}r$XII) is (Burr, 1942; Tadikamalla, 1980),

$$f_{\mathcal{B}rXII}(x;\, a_1,\, a_2,\, b) = \left(\frac{a_1 a_2}{b}\right)\left(\frac{x}{b}\right)^{a_1-1}\left(1+\left(\frac{x}{b}\right)^{a_1}\right)^{-a_2-1},\ x > 0 \tag{A.4}$$

where a_1, $a_2 > 0$ are shape parameters and $b > 0$ is a scale parameter. It is noted that, the r^{th} moment of the $\mathcal{B}r$XII distribution is finite, if and only if, $a_1 a_2 < r$.

The probability mass function (PMF) of the Poisson distribution ($\mathcal{P}oi$) is given by,

$$P_{\mathcal{P}oi}(x;\, \lambda) = (\exp(-\lambda)\lambda^x)/x!, \quad x = 0,\, 1,\, 2,\, \cdots \tag{A.5}$$

where $\lambda > 0$ is the distribution parameter; and has the meaning of average number of occurrences within a time interval.

The PMF of the Bernoulli distribution ($\mathcal{B}ern$) is given by,

$$f_{\mathcal{B}ern}(x;\, p) = \begin{cases} 1-p, & x = 0 \\ p, & x = 1 \end{cases} \tag{A.6}$$

where $p \in [0,\, 1]$.

The CDF of the Generalized Extreme Value ($\mathcal{GEV}$) distribution is given by,

$$F_{\mathcal{GEV}}(x;\, a,\, b,\, c) = \begin{cases} \exp\left(-\left(1+a\frac{x-c}{b}\right)^{-\frac{1}{a}}\right), & a \neq 0 \\ \exp\left(-\exp\left(-\frac{x-c}{b}\right)\right), & a = 0 \end{cases} \tag{A.7}$$

where $a, c \in \mathbb{R}$ and $b > 0$ are shape, location and scale parameters respectively. $\mathcal{GEV}$ encompasses three distributions, the Fréchet ($a > 0$ with $\underline{x} \in [c - b/a, +\infty)$), the Gumbel ($a = 0$ with $\underline{x} \in (-\infty, +\infty)$) and the reversed Weibull ($a < 0$ with $\underline{x} \in (-\infty, c - b/a]$); the latter case is not considered herein, since it regards upper bounded RVs.

Appendix B:. The auxiliary Gaussian AutoRegressive (AR) model

B.1 The univariate AR model

A particularly popular model for stationary processes is the autoregressive model of order p (i.e., AR(p)). A standard Gaussian AR(p) process with zero mean and unit variance can be simulated by,

$$\underline{z}_t = \sum_{l=1}^{p} \widetilde{a}_l \underline{z}_{t-l} + \underline{\varepsilon}_t \tag{B.1}$$

where p denotes the order of the model, and $\widetilde{a}_l$ are the model's parameters, while $\underline{\varepsilon}_t \sim \mathcal{N}(0, \sigma^2_{\underline{\varepsilon}})$. The parameters $\widetilde{\alpha}_l$ can be obtained by solving the Yule-Walker system. Specifically, given a p-dimensional vector of correlation coefficients, $\widetilde{\boldsymbol{\rho}}_p = [\widetilde{\rho}_1, \cdots, \widetilde{\rho}_p]^{\mathrm{T}}$, the parameter vector $\widetilde{\boldsymbol{a}}_p = [\widetilde{a}_1, \cdots, \widetilde{a}_p]^{\mathrm{T}}$, can be obtained by,

$$\widetilde{\boldsymbol{a}}_p = \widetilde{\boldsymbol{P}}_p^{-1} \widetilde{\boldsymbol{\rho}}_p \tag{B.2}$$

where, $\widetilde{\boldsymbol{P}}_p^{-1}$ denotes the inverse of $(p \times p)$ matrix $\widetilde{\boldsymbol{P}}_p$ whose i^{th} and j^{th} element are $[\widetilde{P}_p]_{i,j} = \widetilde{\rho}_{|i-j|}$. After the specification of $\widetilde{a}_l$, $\sigma^2_{\underline{\varepsilon}}$ is obtained by, $\sigma^2_{\underline{\varepsilon}} = 1 - \sum_{l=1}^{p} \widetilde{a}_l \widetilde{\rho}_l$. A stationary AR($p$) process reproduces the autocorrelation structure of the process up to lag p, while for $\tau \geq p + 1$ its correlation structure is given by,

$$\widetilde{\rho}_\tau = \widetilde{\alpha}_1 \widetilde{\rho}_{\tau-1} + \widetilde{\alpha}_2 \widetilde{\rho}_{\tau-2} + \cdots + \widetilde{\alpha}_p \widetilde{\rho}_{\tau-p} = \sum_{l=1}^{p} \widetilde{\alpha}_l \widetilde{\rho}_{\tau-l} \tag{B.3}$$

As a side note, let us provide an additional relationship that will be subsequently used within the parameter estimation procedure of the auxiliary Gaussian Contemporaneous Multivariate AutoRegressive (CMAR) model (Appendix B.2). According to Wold's representation theorem, any covariance stationary causal process can be written as a general linear process, i.e., as a weighted linear combination of past and present i.i.d. random variables $\underline{w}_t$.

$$\underline{z}_t = \psi_0 \underline{w}_t + \psi_1 \underline{w}_{t-1} + \psi_2 \underline{w}_{t-2} + \cdots = \sum_{\zeta=0}^{\infty} \psi_\zeta \underline{w}_{t-\zeta} \tag{B.4}$$

where ψ_ζ are weight coefficients. This representation is also known as infinite moving average representation, i.e., MA(∞). It can be shown that ψ_ζ are related with the coefficients $\widetilde{\alpha}_\zeta$ of AR(p) model by (e.g., Cryer & Chan, 2008; Shumway & Stoffer, 2017),

$$\begin{aligned} \psi_0 &= 1 \\ \psi_1 &= \widetilde{\alpha}_1 \\ \psi_2 &= \widetilde{\alpha}_2 + \widetilde{\alpha}_1 \psi_1 \\ &\vdots \\ \psi_\zeta &= \widetilde{\alpha}_p \psi_{\zeta-p} + \widetilde{\alpha}_{p-1} \psi_{\zeta-p+1} + \cdots + \widetilde{\alpha}_1 \psi_{\zeta-1} \end{aligned} \tag{B.5}$$

or more compactly,

$$\psi_\zeta = \sum_{l=1}^{\zeta} \widetilde{\alpha}_l \psi_{\zeta-l}, \qquad \text{for } \zeta = 1, 2, \cdots \tag{B.6}$$

where $\psi_0 = 1$ and $\widetilde{\alpha}_\zeta = 0$ for $\zeta > p$. It is also noted that a similar relationship exists for ARMA-type models. Nevertheless, since ψ_j decay with increasing ζ and approach zero after some large value of ζ we can truncate Eq. (B.4) at some q to read,

$$\underline{z}_t = \sum_{\zeta=0}^{q} \psi_\zeta \underline{w}_{t-\zeta} \tag{B.7}$$

B.2 The multivariate AR model

The univariate AR(p) model can been extended for multivariate processes (e.g., Bras & Rodríguez-Iturbe, 1985; Cryer & Chan, 2008; Kottegoda, 1980; Pegram & James, 1972; Shumway & Stoffer, 2017), and it is often referred to as Multivariate or Vector Autoregressive (MAR(p) or VAR(p)) model. Assuming that we wish to model an m-dimension vector of Gaussian processes $\underline{\boldsymbol{z}}_t = [\underline{z}_t^1, .., \underline{z}_t^m]^{\mathrm{T}}$ with zero and unit variance, its generating equation is given by,

$$\underline{\boldsymbol{z}}_t = \sum_{l=1}^{p} \widetilde{\boldsymbol{A}}_l \underline{\boldsymbol{z}}_{t-l} + \underline{\boldsymbol{\varepsilon}}_t \tag{B.8}$$

where p denotes the order of the model, $\widetilde{\boldsymbol{A}}_l$ are $(m \times m)$ parameter matrices and $\underline{\boldsymbol{\varepsilon}}_t = [\underline{\varepsilon}_t^1, .., \underline{\varepsilon}_t^m]^{\mathrm{T}}$ is a vector of m Gaussian variates with zero mean and covariance matrix $\widetilde{G}$: =Cov[$\underline{\varepsilon}_t$, $\underline{\varepsilon}_t$] (whose i^{th} j^{th} element is denoted by $\widetilde{g}^{i,j}$). The correlation (since we assume a standard Gaussian model) matrix of time lag τ, is denoted by $\widetilde{R}_\tau$: =Corr[$\underline{z}_t$, $\underline{z}_{t-\tau}$], and is related with the parameter matrices $\widetilde{A}_l$ by,

$$\widetilde{\boldsymbol{R}}_\tau - \widetilde{\boldsymbol{A}}_1 \widetilde{\boldsymbol{R}}_{\tau-1} - \cdots - \widetilde{\boldsymbol{A}}_p \widetilde{\boldsymbol{R}}_{\tau-p} = \begin{cases} \widetilde{\boldsymbol{G}} & \text{if} \quad \tau = 0 \\ 0 & \text{if} \quad \tau > 0 \end{cases} \tag{B.9}$$

Specifically, for $\tau = 0$, the system reads,

$$\widetilde{\boldsymbol{G}} = \widetilde{\boldsymbol{R}}_0 - \widetilde{\boldsymbol{A}}_1\widetilde{\boldsymbol{R}}_1^{\mathrm{T}} - \cdots - \widetilde{\boldsymbol{A}}_p\widetilde{\boldsymbol{R}}_p^{\mathrm{T}} = \widetilde{\boldsymbol{R}}_0 - \sum_{l=1}^{p}\widetilde{\boldsymbol{A}}_l\widetilde{\boldsymbol{R}}_l^{\mathrm{T}} \tag{B.10}$$

Furthermore, Eq. (B.9) can be written in matrix notation as follows (for $\tau = 1, \cdots, p$),

$$[\boldsymbol{R}_1, \boldsymbol{R}_2, \cdots, \boldsymbol{R}_p] = [\widetilde{\boldsymbol{A}}_1, \widetilde{\boldsymbol{A}}_2, \cdots, \widetilde{\boldsymbol{A}}_p]\begin{bmatrix} \widetilde{\boldsymbol{R}}_0 & \widetilde{\boldsymbol{R}}_1 & \cdots & \widetilde{\boldsymbol{R}}_{p-1} \\ \widetilde{\boldsymbol{R}}_1^{\mathrm{T}} & \widetilde{\boldsymbol{R}}_0 & \cdots & \widetilde{\boldsymbol{R}}_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \widetilde{\boldsymbol{R}}_{p-1}^{\mathrm{T}} & \widetilde{\boldsymbol{R}}_{p-2}^{\mathrm{T}} & \cdots & \widetilde{\boldsymbol{R}}_0 \end{bmatrix} \tag{B.11}$$

where $\widetilde{\boldsymbol{R}}_{-\tau} = \widetilde{\boldsymbol{R}}_{\tau}^{\mathrm{T}}$. Eq. (B.11) is also known as the multivariate Yule-Walker system of the MAR(p) model. Provided that the matrices $\widetilde{\boldsymbol{R}}_1, \widetilde{\boldsymbol{R}}_2, \cdots, \widetilde{\boldsymbol{R}}_p$ are known and non-signular, Eq. (B.11) can be solved for $\widetilde{\boldsymbol{A}}_1, \widetilde{\boldsymbol{A}}_2, \cdots, \widetilde{\boldsymbol{A}}_p$, i.e.,

$$[\widetilde{\boldsymbol{A}}_1, \widetilde{\boldsymbol{A}}_2, \cdots, \widetilde{\boldsymbol{A}}_p] = [\boldsymbol{R}_1, \boldsymbol{R}_2, \cdots, \boldsymbol{R}_p]\begin{bmatrix} \widetilde{\boldsymbol{R}}_0 & \widetilde{\boldsymbol{R}}_1 & \cdots & \widetilde{\boldsymbol{R}}_{p-1} \\ \widetilde{\boldsymbol{R}}_1^{\mathrm{T}} & \widetilde{\boldsymbol{R}}_0 & \cdots & \widetilde{\boldsymbol{R}}_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \widetilde{\boldsymbol{R}}_{p-1}^{\mathrm{T}} & \widetilde{\boldsymbol{R}}_{p-2}^{\mathrm{T}} & \cdots & \widetilde{\boldsymbol{R}}_0 \end{bmatrix}^{-1} \tag{B.12}$$

Arguably, this is a complex system of equations that requires the specification of p matrices $\widetilde{\boldsymbol{R}}_p$. The overall parameter estimation procedure can be significantly simplified if we assume that the parameter matrices $\widetilde{\boldsymbol{A}}_1, \widetilde{\boldsymbol{A}}_2, \cdots, \widetilde{\boldsymbol{A}}_p$ are diagonal, i.e.,

$$\widetilde{\boldsymbol{A}}_l = \begin{bmatrix} \widetilde{a}_{l[1,1]} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \widetilde{a}_{l[m,m]} \end{bmatrix} = [\widetilde{\boldsymbol{A}}_l]_{i,j} \tag{B.13}$$

Thereby formulating the so-called Contemporaneous Multivariate Autoregressive model of order p (i.e., CMAR(p); see, the work of Pegram and James (1972)). This simplification apart from the reproduction of the autocorrelation structure of the process up to time lag p (as in the case of full matrices $\widetilde{\boldsymbol{A}}_l$), implies the direct reproduction of the lag-0 cross-correlation coefficients, i.e., correlation matrix $\widetilde{\boldsymbol{R}}_0$. Using the contemporaneous formulation, the model can be decomposed into m univariate AR(p) models, which are contemporaneously cross-correlated at lag 0, i.e.,

$$\begin{aligned} \underline{z}_t^{\,1} &= \widetilde{a}_{1[1,1]}\underline{z}_{t-1}^{\,1} + \widetilde{a}_{2[1,1]}\underline{z}_{t-2}^{\,1} + \cdots + \widetilde{a}_{l[1,1]}\underline{z}_{t-l}^{\,1} + \cdots + \widetilde{a}_{p[1,1]}\underline{z}_{t-p}^{\,1} + \underline{\varepsilon}_t^{\,1} \\ &\vdots \\ \underline{z}_t^{\,m} &= \widetilde{a}_{1[m,m]}\underline{z}_{t-1}^{\,m} + \widetilde{a}_{2[m,m]}\underline{z}_{t-2}^{\,m} + \cdots + \widetilde{a}_{l[m,m]}\underline{z}_{t-l}^{\,m} + \cdots + \widetilde{a}_{p[m,m]}\underline{z}_{t-p}^{\,m} + \underline{\varepsilon}_t^{\,m} \end{aligned} \tag{B.14}$$

Alternatively, and assuming that, $\widetilde{\boldsymbol{B}}\widetilde{\boldsymbol{B}}^{\mathrm{T}} = \widetilde{\boldsymbol{G}}$, where $\widetilde{\boldsymbol{B}}$ is a $m \times m$ matrix that denotes the square root matrix of $\widetilde{\boldsymbol{G}}$ (typically identified using standard matrix decomposition methods; e.g., Cholesky), then, Eq. (B.8) can be rewritten as,

$$\underline{z}_t^{\,i} = \sum_{l=1}^{p}\widetilde{a}_{l[i,i]}\underline{z}_{t-l}^{\,i} + \sum_{j=1}^{m}\widetilde{b}_{[i,j]}\underline{w}_t^{\,j} \tag{B.15}$$

where $\underline{w}_t^{\,j}$ are i.i.d. standard Gaussian variates, i.e., $\underline{w}_t^{\,j} \sim \mathcal{N}(0, 1)$.

In this form, assuming that the autocorrelation structure of each process is known (e.g., specified by a theoretical model such as CAS) the parameters α_l ($l = 1, ..,p$) as well as the variance ($\sigma_{\underline{\varepsilon}}^2$) of $\underline{\varepsilon}_t$, can be easily computed through the univariate Yule-Walker system. Hence it is possible to fully estimate the matrices $\widetilde{\boldsymbol{A}}_1, \widetilde{\boldsymbol{A}}_2, \cdots, \widetilde{\boldsymbol{A}}_p$ as well as the diagonal elements of $\widetilde{\boldsymbol{G}}$, which are, $\widetilde{g}^{i,i} = \mathrm{Var}[\underline{\varepsilon}_t^{\,i}, \underline{\varepsilon}_t^{\,i}] = \sigma_{\underline{\varepsilon}^i}^2$.

According to Pegram and James (1972), in order to estimate the off-diagonal elements of $\widetilde{\boldsymbol{G}}$ one can resort to iterative methods or solve a complicated system of equations. Both solutions experience significant difficulties, especially when implemented in a computer software. Herein, we propose an alternative technique. It is recalled that according to Eq. (B.4) each individual process $\underline{z}_t^{\,i}$ can be represented in terms of an MA(∞) process, which can be truncated at some high value of q, i.e.,

$$\underline{z}_t^{\,i} = \sum_{\zeta=0}^{q}\psi_{\zeta}^{i}\underline{w}_{t-\zeta}^{\,i} \tag{B.16}$$

The elements ψ_{ζ}^{i} can be easily computed for each process $\underline{z}_t^{\,i}$ using Eq. (B.5) or (B.6). Provided that the ψ_{ζ}^{i} quantities have been estimated, the off-diagonal i^{th} j^{th} elements (for $i, j = 1, \cdots, m$ and $i \neq j$; since the diagonal elements are known) of matrix $\widetilde{\boldsymbol{G}}$ are identified as follows,

$$\widetilde{g}^{i,j} = \frac{[\widetilde{R}_o]_{i,j}}{\sum_{\zeta=0}^{q}\psi_{\zeta}^{i}\psi_{\zeta}^{j}} = \frac{\widetilde{\rho}_0^{i,j}}{\sum_{\zeta=0}^{q}\psi_{\zeta}^{i}\psi_{\zeta}^{j}} \tag{B.17}$$

It is also noted that the elements ψ_ς^i can be used for the estimation of any cross-correlation value for lag $\tau = 0, 1, 2\cdots$ through,

$$\text{Corr}[\underline{z}_t^i, \underline{z}_{t+\tau}^j] = [\widetilde{R}_o]_{i,j} \frac{\sum_{\varsigma=0}^{q-\tau} \psi_\varsigma^i \psi_{\varsigma+\tau}^j}{\sum_{\varsigma=0}^{q} \psi_\varsigma^i \psi_\varsigma^j} = \widetilde{\rho}_0^{i,j} \frac{\sum_{\varsigma=0}^{q-\tau} \psi_\varsigma^i \psi_{\varsigma+\tau}^j}{\sum_{\varsigma=0}^{q} \psi_\varsigma^i \psi_\varsigma^j} = \widetilde{g}^{i,j} \sum_{\varsigma=0}^{q-\tau} \psi_\varsigma^i \psi_{\varsigma+\tau}^j \tag{B.18}$$

Appendix C:. Additional figures for the simulation examples of Section 2.7

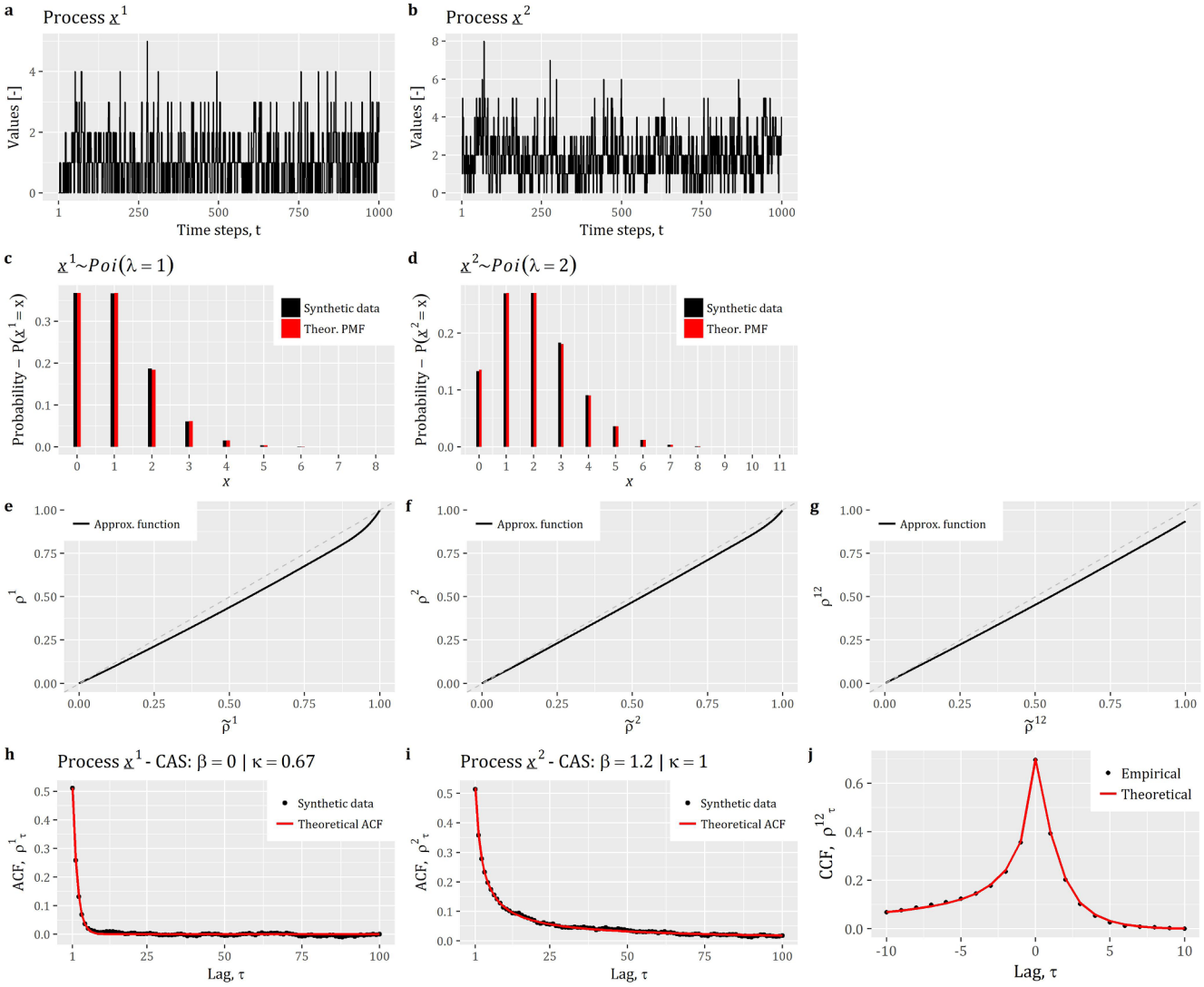

Fig. 18. Case B – Poisson marginal distributions. Simulated realization of process a) $\underline{x}_t^1$ and b) $\underline{x}_t^2$. Comparison of simulated and theoretical distribution function for process c) $\underline{x}_t^1$ and d) $\underline{x}_t^2$. The established relationships between the equivalent, $\widetilde{\rho}$, and target, ρ, correlation coefficients given the marginal distribution of each process; e) $\underline{x}_t^1$, f) $\underline{x}_t^2$, as well as their g) interaction. Simulated and theoretical autocorrelation function (ACF) for process h) $\underline{x}_t^1$ and i) $\underline{x}_t^2$, as well as j) the cross-correlation function (CCF) among $\underline{x}_t^1$ and $\underline{x}_t^2$, i.e., $\rho_\tau^{1,2} = \text{Corr}[\underline{x}_t^1, \underline{x}_{t+\tau}^2]$.

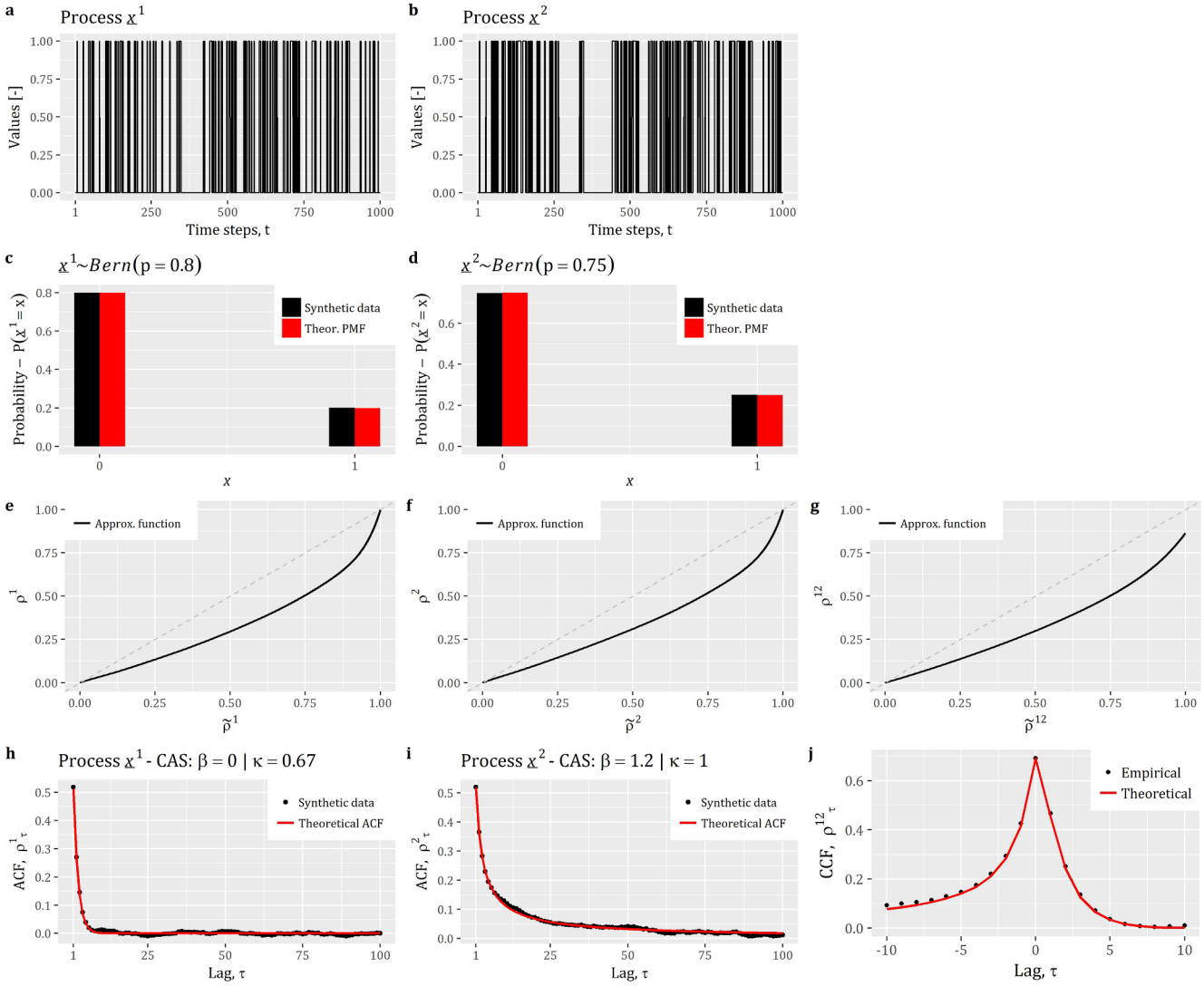

Fig. 19. Case C – Bernoulli marginal distributions. Simulated realization of process a) $\underline{x}_t^1$ and b) $\underline{x}_t^2$. Comparison of simulated and theoretical distribution function for process c) $\underline{x}_t^1$ and d) $\underline{x}_t^2$. The established relationships between the equivalent, $\tilde{\rho}$, and target, ρ, correlation coefficients given the marginal distribution of each process; e) $\underline{x}_t^1$, f) $\underline{x}_t^2$, as well as their g) interaction. Simulated and theoretical autocorrelation function (ACF) for process h) $\underline{x}_t^1$ and i) $\underline{x}_t^2$, as well as j) the cross-correlation function (CCF) among $\underline{x}_t^1$ and $\underline{x}_t^2$, i.e., $\rho_\tau^{1,2} = \text{Corr}[\underline{x}_t^1, \underline{x}_{t+\tau}^2]$.

Appendix D:. Parameter estimation of individual Nataf-based models

The parameter estimation procedure is organized according to the temporal scale of simulation.

Annual scale (SMARTA model)

Step1: Specify a target distribution function $F_{\underline{y}^i}$ for each process $\underline{y}_t^i$ $(i = 1, ..,m)$.
Step 2: Identify a target theoretical auto-dependence structure for each process $\underline{y}_t^i$ $(i = 1, ..,m)$. For instance, by fitting the Cauchy autocorrelation structure (CAS; Eq. (8)) to the empirical estimates of autocorrelation coefficients, i.e., specify the target $\rho_{\underline{y};\tau}^i = \text{Corr}[\underline{y}_t^i, \underline{y}_{t+\tau}^i]$.
Step 3: Specify the target annual lag-0 cross-correlation coefficients ($\rho_{\underline{y}}^{i,j} = \text{Corr}[\underline{y}_t^i, \underline{y}_t^j]$; $i \neq j = 1, \cdots, m$). For instance, using the empirical estimates of historical annual data.
Step 4: On the basis of the information provided by step 3, estimate the equivalent correlation coefficients, as well as the parameters of SMARTA model (Tsoukalas et al., 2018c).

Monthly scale (SPARTA model)

Step 1: Specify a target distribution function $F_{\underline{x}_s^i}$ for each season and process $\underline{x}_{s,n}^i$ $(i = 1, ..,m;\ s = 1, \cdots, 12)$.
Step 2: Specify the target lag-1 month-to-month-correlation coefficients of each process ($\rho_{\underline{x};s,s-1}^i = \text{Corr}[\underline{x}_s^i, \underline{x}_{s-1}^i]$; $i = 1, \cdots, m;\ s = 1, \cdots, 12$). For instance, using the empirical estimates of historical monthly data.
Step 3: Specify the target monthly lag-0 cross-correlation coefficients ($\rho_{\underline{x};s}^{i,j} = \text{Corr}[\underline{x}_s^i, \underline{x}_s^j]$; $i \neq j = 1, \cdots, m;\ s = 1, \cdots, 12$). For instance, using the empirical estimates of historical monthly data.

Step 4: On the basis of the information provided by step 3, estimate the equivalent correlation coefficients, and the parameters of SPARTA model (Tsoukalas et al., 2017, 2018a).

Daily scale (CMARTA model)

Step 1: Specify a target distribution function $F_{\underline{w}_s^i}$ for each season and process $\underline{w}_{s;d}^i$ ($i = 1, ..,m$; $s = 1, \cdots, 12$).

Step 2: Identify a target theoretical auto-dependence structure each process $\underline{w}_{s;d}^i$ ($i = 1, ..,m$). For instance, by fitting CAS (i.e., Eq. (8)), on a monthly basis, to the daily empirical estimates of autocorrelation coefficients, i.e., specify the target $\rho_{\underline{w}_{s;\tau}^i} = \mathrm{Corr}[\underline{w}_{s;d}^i, \underline{w}_{s;d+\tau}^i]$.

Step 3: Specify the target daily lag-0 cross-correlation coefficients ($\rho_{\underline{w},s}^{i,j} = \mathrm{Corr}[\underline{w}_{s;d}^i, \underline{w}_{s;d}^j]$; $i \neq j = 1, \cdots, m$; $s = 1, \cdots, 12$). For instance, using the empirical estimates of historical daily data.

Step 4: On the basis of the information provided by step 3, estimate the equivalent correlation coefficients, and the parameters of CMARTA model (see Section 2.8, and Appendix B).

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2019.05.017.

References

Ailliot, P., Allard, D., Monbet, V., Naveau, P., 2015. Stochastic weather generators: an overview of weather type models. Journal de La Société Française de Statistique 156 (1), 101–113.

Allard, D., Bourotte, M., 2015. Disaggregating daily precipitations into hourly values with a transformed censored latent Gaussian process. Stoch. Env. Res. Risk Assess. 29 (2), 453–462. https://doi.org/10.1007/s00477-014-0913-4.

Baigorria, G.A., Jones, J.W., 2010. GiST: a stochastic model for generating spatially and temporally correlated daily rainfall data. J. Clim. 23 (22), 5990–6008. https://doi.org/10.1175/2010JCLI3537.1.

Bárdossy, A., Pegram, G., 2009. Copula based multisite model for daily precipitation simulation. Hydrol. Earth Syst. Sci. Discuss. 6 (3), 4485–4534. https://doi.org/10.5194/hessd-6-4485-2009.

Bárdossy, A., Pegram, G.G.S., 2016. Space-time conditional disaggregation of precipitation at high resolution via simulation. Water Resour. Res. 52 (2), 920–937. https://doi.org/10.1002/2015WR018037.

Beran, J., 1994. Statistics for Long-memory Processes. CRC Press.

Beran, J., Feng, Y., Ghosh, S., Kulik, R., 2013. In: Long-Memory Processes. Springer Berlin Heidelberg, Berlin, Heidelberg, Heidelberg. https://doi.org/10.1007/978-3-642-35512-7.

Biller, B., Nelson, B.L., 2003. Modeling and generating multivariate time-series input processes using a vector autoregressive technique. ACM Trans. Model. Comput. Simul. 13 (3), 211–237. https://doi.org/10.1145/937332.937333.

Bo, Z., Islam, S., Eltahir, E.A.B., 1994. Aggregation-disaggregation properties of a stochastic rainfall model. Water Resour. Res. 30 (12), 3423–3435. https://doi.org/10.1029/94WR02026.

Bras, R.L., Rodríguez-Iturbe, I., 1985. Random Functions and Hydrology. Addison-Wesley, Reading, Mass.

Breinl, K., Di Baldassarre, G., 2019. Space-time disaggregation of precipitation and temperature across different climates and spatial scales. J. Hydrol.: Reg. Stud. 21, 126–146. https://doi.org/10.1016/j.ejrh.2018.12.002.

Breinl, K., Turkington, T., Stowasser, M., 2013. Stochastic generation of multi-site daily precipitation for applications in risk management. J. Hydrol. 498, 23–35. https://doi.org/10.1016/j.jhydrol.2013.06.015.

Breinl, K., Turkington, T., Stowasser, M., 2015. Simulating daily precipitation and temperature: a weather generation framework for assessing hydrometeorological hazards. Meteorol. Appl. 22 (3), 334–347. https://doi.org/10.1002/met.1459.

Brissette, F.P., Khalili, M., Leconte, R., 2007. Efficient stochastic generation of multi-site synthetic precipitation data. J. Hydrol. 345 (3–4), 121–133. https://doi.org/10.1016/j.jhydrol.2007.06.035.

Burr, I.W., 1942. Cumulative frequency functions. Ann. Math. Stat. 13 (2), 215–232. https://doi.org/10.1214/aoms/1177731607.

Burton, A., Kilsby, C.G., Fowler, H.J., Cowpertwait, P.S.P., O'Connell, P.E., 2008. RainSim: a spatial–temporal stochastic rainfall modelling system. Environ. Modell. Software 23 (12), 1356–1369. https://doi.org/10.1016/j.envsoft.2008.04.003.

Cannon, A.J., 2008. Probabilistic multisite precipitation downscaling by an expanded bernoulli-gamma density network. J. Hydrometeorol. 9 (6), 1284–1300. https://doi.org/10.1175/2008JHM960.1.

Caraway, N.M., McCreight, J.L., Rajagopalan, B., 2014. Multisite stochastic weather generation using cluster analysis and k-nearest neighbor time series resampling. J. Hydrol. 508, 197–213. https://doi.org/10.1016/j.jhydrol.2013.10.054.

Cario, M.C., Nelson, B.L., 1996. Autoregressive to anything: time-series input processes for simulation. Operations Res. Lett. 19 (2), 51–58. https://doi.org/10.1016/0167-6377(96)00017-X.

Cario, M.C., Nelson, B.L., 1997. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Industrial Engineering. Retrieved from. .Technical Report, Department of Industrial Engineering and Management Sciences. Northwestern University, Evanston, Illinois. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.48.281&rep=rep1&type=pdf.

Celeste, A.B., Billib, M., 2009. Evaluation of stochastic reservoir operation optimization models. Adv. Water Resour. 32 (9), 1429–1443. https://doi.org/10.1016/j.advwatres.2009.06.008.

Chen, L., Singh, V.P., Guo, S., Zhou, J., Zhang, J., 2015. Copula-based method for multisite monthly and daily streamflow simulation. J. Hydrol. 528, 369–384. https://doi.org/10.1016/j.jhydrol.2015.05.018.

Chilès, J.-P., Delfiner, P., 1999. Geostatistics: Modeling Spatial Uncertainty. Jhon Wiley & Sons Inc., New York, pp. 695.

Clark, M.P., Gangopadhyay, S., Brandon, D., Werner, K., Hay, L., Rajagopalan, B., Yates, D., 2004. A resampling procedure for generating conditioned daily weather sequences. Water Resour. Res. 40 (4), 1–15. https://doi.org/10.1029/2003WR002747.

Cryer, J. D., Chan, K.-S, 2008. Time Series Analysis. With Applications to R. Design. https://doi.org/10.1007/978-0-387-75959-3.

Deidda, R., Benzi, R., Siccardi, F., 1999. Multifractal modeling of anomalous scaling laws in rainfall. Water Resour. Res. 35 (6), 1853–1867. https://doi.org/10.1029/1999WR900036.

Deodatis, G., Micaletti, R.C., 2001. Simulation of highly skewed non-Gaussian stochastic processes. J. Eng. Mech. 127 (12), 1284–1295.

Dimitriadis, P., Koutsoyiannis, D., 2015. Climacogram versus autocovariance and power spectrum in stochastic modelling for Markovian and Hurst-Kolmogorov processes. Stoch. Env. Res. Risk Assess. 29 (6), 1649–1669. https://doi.org/10.1007/s00477-015-1023-7.

Efstratiadis, A., Dialynas, Y.G., Kozanis, S., Koutsoyiannis, D., 2014. A multivariate stochastic model for the generation of synthetic time series at multiple time scales reproducing long-term persistence. Environ. Modell. Software 62 (July), 139–152. https://doi.org/10.1016/j.envsoft.2014.08.017.

Embrechts, P., McNeil, A.J., Straumann, D., 1999. Correlation and dependence in risk management: properties and pitfalls. In: Dempster, M.A.H. (Ed.), Risk Management. Cambridge University Press, Cambridge, pp. 176–223. https://doi.org/10.1017/CBO9780511615337.008.

Evin, G., Favre, A.-C., 2008. A new rainfall model based on the Neyman-Scott process using cubic copulas. Water Resour. Res. 44 (3), 1–18. https://doi.org/10.1029/2007WR006054.

Evin, G., Favre, A., Hingray, B., 2018. Stochastic generation of multi-site daily precipitation focusing on extreme events. Hydrol. Earth Syst. Sci. 22 (1), 655–672. https://doi.org/10.5194/hess-22-655-2018.

Feng, M., Liu, P., Guo, S., Gui, Z., Zhang, X., Zhang, W., Xiong, L., 2017. Identifying changing patterns of reservoir operating rules under various inflow alteration scenarios. Adv. Water Resour. 104, 23–36. https://doi.org/10.1016/j.advwatres.2017.03.003.

Foufoula-Georgiou, E., Lettenmaier, D.P., 1987. A Markov Renewal Model for rainfall occurrences. Water Resour. Res. 23 (5), 875–884. https://doi.org/10.1029/WR023i005p00875.

Furrer, E.M., Katz, R.W., 2008. Improving the simulation of extreme precipitation events by stochastic weather generators. Water Resour. Res. 44 (12), 1–13. https://doi.org/10.1029/2008WR007316.

Genton, M.G., 2007. Separable approximations of space-time covariance matrices. Environmetrics 18 (7), 681–695. https://doi.org/10.1002/env.854.

Genton, M.G., Kleiber, W., 2015. Cross-covariance functions for multivariate geostatistics. Statist. Sci. 30 (2), 147–163. https://doi.org/10.1214/14-STS487.

Giuliani, M., Herman, J.D., Castelletti, A., Reed, P., 2014. Many-objective reservoir policy identification and refinement to reduce policy inertia and myopia in water management. Water Resour. Res. 50 (4), 3355–3377. https://doi.org/10.1002/2013WR014700.

Glasbey, C.A., Cooper, G., McGechan, M.B., 1995. Disaggregation of daily rainfall by conditional simulation from a point-process model. J. Hydrol. 165 (1–4), 1–9. https://doi.org/10.1016/0022-1694(94)02598-6.

Gneiting, T., 2000. Power-law correlations, related models for long-range dependence and their simulation. J. Appl. Probability 37 (4), 1104–1109.

Gneiting, T., Kleiber, W., Schlather, M., 2010. Matérn cross-covariance functions for multivariate random fields. J. Am. Stat. Assoc. 105 (491), 1167–1177. https://doi.org/10.1198/jasa.2010.tm09420.

Gneiting, T., Schlather, M., 2004. Stochastic models that separate fractal dimension and

the Hurst effect. Retrieved from. SIAM Rev. 46 (2), 269–282. http://www.jstor.org/stable/20453506.
Greenwood, J.A., Landwehr, J.M., Matalas, N.C., Wallis, J.R., 1979. Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form. Water Resour. Res. 15 (5), 1049–1054. https://doi.org/10.1029/WR015i005p01049.
Grigoriu, M., 1998. Simulation of stationary non-Gaussian translation processes. J. Eng. Mech. 124 (2), 121–126.
Grygier, J.C., Stedinger, J.R., 1988. Condensed disaggregation procedures and conservation corrections for stochastic hydrology. Water Resour. Res. 24 (10), 1574–1584. https://doi.org/10.1029/WR024i010p01574.
Gupta, V.K., Waymire, E., 1990. Multiscaling properties of spatial rainfall and river flow distributions. J. Geophys. Res. https://doi.org/10.1029/JD095iD03p01999.
Gupta, V.K., Waymire, E.C., 1993. A statistical analysis of mesoscale rainfall as a random cascade. J. Appl. Meteorol. 32 (2), 251–267. https://doi.org/10.1175/1520-0450(1993) 032<0251:ASAOMR>2.0.CO;2.
Gyasi-Agyei, Y., 2011. Copula-based daily rainfall disaggregation model. Water Resour. Res. 47 (7), 1–17. https://doi.org/10.1029/2011WR010519.
Gyasi-Agyei, Y., Melching, C.S., 2012. Modelling the dependence and internal structure of storm events for continuous rainfall simulation. J. Hydrol. 464–465, 249–261. https://doi.org/10.1016/j.jhydrol.2012.07.014.
Haberlandt, U., Hundecha, Y., Pahlow, M., Schumann, A.H., 2011. Rainfall generators for application in flood studies. In: Flood Risk Assessment and Management. Springer, pp. 117–147.
Hao, Z., Singh, V.P., 2011. Single-site monthly streamflow simulation using entropy theory. Water Resour. Res. 47 (9). https://doi.org/10.1029/2010WR010208. n/a-n/a.
Hao, Z., Singh, V.P., 2013. Modeling multisite streamflow dependence with maximum entropy copula. Water Resour. Res. 49 (10), 7139–7143. https://doi.org/10.1002/wrcr.20523.
Harms, A.A., Campbell, T.H., 1967. An extension to the Thomas-Fiering Model for the sequential generation of streamflow. Water Resour. Res. 3 (3), 653–661. https://doi.org/10.1029/WR003i003p00653.
Hosking, J.R.M., 1990. L-moments: analysis and estimation of distributions using linear combinations of order statistics. Retrieved from. J. Roy. Stat. Soc.: Ser. B (Methodol.) 52 (1), 105–124.
Hurst, H.E., 1951. Long-term storage capacity of reservoirs. Trans. Amer. Soc. Civil Eng. 116, 770–808.
Ilich, N., 2014. An effective three-step algorithm for multi-site generation of stochastic weekly hydrological time series. Hydrol. Sci. J. 59 (1), 85–98. https://doi.org/10.1080/02626667.2013.822643.
Ilich, N., Despotovic, J., 2008. A simple method for effective multi-site generation of stochastic hydrologic time series. Stoch. Env. Res. Risk Assess. 22 (2), 265–279. https://doi.org/10.1007/s00477-007-0113-6.
Jeong, C., Lee, T., 2015. Copula-based modeling and stochastic simulation of seasonal intermittent streamflows for arid regions. J. Hydro-Environ. Res. 9 (4), 604–613. https://doi.org/10.1016/j.jher.2014.06.001.
Kaczmarska, J., Isham, V., Onof, C., 2014. Point process models for fine-resolution rainfall. Hydrol. Sci. J. 59 (11), 1972–1991. https://doi.org/10.1080/02626667.2014.925558.
Kantelhardt, J.W., Koscielny-Bunde, E., Rybski, D., Braun, P., Bunde, A., Havlin, S., 2006. Long-term persistence and multifractality of precipitation and river runoff records. J. Geophys. Res. 111 (D1), D01106. https://doi.org/10.1029/2005JD005881.
Katz, R.W., 1977. Precipitation as a chain-dependent process. J. Appl. Meteorol. 16 (7), 671–676. https://doi.org/10.1175/1520-0450(1977) 016<0671:PAACDP>2.0.CO;2.
Katz, R.W., Parlange, M.B., 1998. Overdispersion phenomenon in stochastic modeling of precipitation. J. Clim. 11 (4), 591–601. https://doi.org/10.1175/1520-0442(1998)011<0591:OPISMO>2.0.CO;2.
Khalili, M., Brissette, F., Leconte, R., 2009. Stochastic multi-site generation of daily weather data. Stoch. Env. Res. Risk Assess. 23 (6), 837–849. https://doi.org/10.1007/s00477-008-0275-x.
Kilsby, C.G., Jones, P.D., Burton, A., Ford, A.C., Fowler, H.J., Harpham, C., Wilby, R.L., 2007. A daily weather generator for use in climate change studies. Environ. Modell. Software 22 (12), 1705–1719. https://doi.org/10.1016/j.envsoft.2007.02.005.
Klemeš, V., 1981. Applied stochastic theory of storage in evolution. In: Advances in Hydroscience. Elsevier, pp. 79–141.
Kossieris, P., Efstratiadis, A., Tsoukalas, I., Koutsoyiannis, D., 2015. Assessing the performance of Bartlett-Lewis model on the simulation of Athens rainfall. EGU General Assembly Conference Abstracts.
Kossieris, P., Makropoulos, C., Onof, C., Koutsoyiannis, D., 2016. A rainfall disaggregation scheme for sub-hourly time scales: coupling a Bartlett-Lewis based model with adjusting procedures. J. Hydrol. https://doi.org/10.1016/j.jhydrol.2016.07.015.
Kossieris, P., Tsoukalas, I., Makropoulos, C., Savic, D., 2019. Simulating marginal and dependence behaviour of water demand processes at any fine time scale. Water 11 (5), 885. https://doi.org/10.3390/w11050885.
Kottegoda, N.T., 1980. Stochastic Water Resources Technology. Springer.
Koutsoyiannis, D., 1999. Optimal decomposition of covariance matrices for multivariate stochastic models in hydrology. Water Resour. Res. 35 (4), 1219–1229. https://doi.org/10.1029/1998WR900093.
Koutsoyiannis, D., 2000. A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series. Water Resour. Res. 36 (6), 1519–1533. https://doi.org/10.1029/2000WR900044.
Koutsoyiannis, D., 2001. Coupling stochastic models of different timescales. Water Resour. Res. 37 (2), 379–391. https://doi.org/10.1029/2000WR900200.
Koutsoyiannis, D., 2002. The Hurst phenomenon and fractional Gaussian noise made easy. Hydrol. Sci. J. 47 (4), 573–595. https://doi.org/10.1080/02626660209492961.
Koutsoyiannis, D., 2005. Stochastic simulation of hydrosystems. In: Water Encyclopedia. John Wiley & Sons, Inc.. https://doi.org/10.1002/047147844X.sw913.
Koutsoyiannis, D., 2011. Hurst-kolmogorov dynamics and uncertainty. JAWRA J. Am. Water Resour. Assoc. 47 (3), 481–495. https://doi.org/10.1111/j.1752-1688.2011.00543.x.
Koutsoyiannis, D., 2016. Generic and parsimonious stochastic modelling for hydrology and beyond. Hydrol. Sci. J. 61 (2), 225–244. https://doi.org/10.1080/02626667.2015.1016950.
Koutsoyiannis, D., Economou, A., 2003. Evaluation of the parameterization-simulation-optimization approach for the control of reservoir systems. Water Resour. Res. 39 (6). https://doi.org/10.1029/2003WR002148. n/a-n/a.
Koutsoyiannis, D., Manetas, A., 1996. Simple disaggregation by accurate adjusting procedures. Water Resour. Res. 32 (7), 2105–2117. https://doi.org/10.1029/96WR00488.
Koutsoyiannis, D., Montanari, A., 2007. Statistical analysis of hydroclimatic time series: uncertainty and insights. Water Resour. Res. 43 (5), 1–9. https://doi.org/10.1029/2006WR005592.
Koutsoyiannis, D., Onof, C., 2001. Rainfall disaggregation using adjusting procedures on a Poisson cluster model. J. Hydrol. 246 (1–4), 109–122. https://doi.org/10.1016/S0022-1694(01)00363-8.
Koutsoyiannis, D., Onof, C., Wheater, H.S., 2003. Multivariate rainfall disaggregation at a fine timescale. Water Resour. Res. 39 (7), 1–62. https://doi.org/10.1029/2002WR001600.
Lall, U., Sharma, A., 1996. A nearest neighbor bootstrap for resampling hydrologic time series. Water Resour. Res. 32 (3), 679–693. https://doi.org/10.1029/95WR02966.
Langousis, A., Koutsoyiannis, D., 2006. A stochastic methodology for generation of seasonal time series reproducing overyear scaling behaviour. J. Hydrol. 322 (1–4), 138–154. https://doi.org/10.1016/j.jhydrol.2005.02.037.
Lebrun, R., Dutfoy, A., 2009. An innovating analysis of the Nataf transformation from the copula viewpoint. Probab. Eng. Mech. 24 (3), 312–320. https://doi.org/10.1016/j.probengmech.2008.08.001.
Lee, T., 2017. Multisite stochastic simulation of daily precipitation from copula modeling with a gamma marginal distribution. Theor. Appl. Climatol. https://doi.org/10.1007/s00704-017-2147-0.
Lee, T., Salas, J.D., 2011. Copula-based stochastic simulation of hydrological data applied to Nile River flows. Hydrol. Res. 42 (4), 318–330. https://doi.org/10.2166/nh.2011.085.
Lee, T., Salas, J.D., Prairie, J., 2010. An enhanced nonparametric streamflow disaggregation model with genetic algorithm. Water Resour. Res. 46 (8), 1–14. https://doi.org/10.1029/2009WR007761.
Li, X., Meshgi, A., Wang, X., Zhang, J., Tay, S.H.X., Pijcke, G., Babovic, V., 2018. Three resampling approaches based on method of fragments for daily-to-subdaily precipitation disaggregation. Int. J. Climatol. 38, e1119–e1138. https://doi.org/10.1002/joc.5438.
Li, C., Singh, V.P., Mishra, A.K., 2013. A bivariate mixed distribution with a heavy-tailed component and its application to single-site daily rainfall simulation. Water Resour. Res. 49 (2), 767–789. https://doi.org/10.1002/wrcr.20063.
Liu, P.L., Der Kiureghian, A., 1986. Multivariate distribution models with prescribed marginals and covariances. Probab. Eng. Mech. 1 (2), 105–112. https://doi.org/10.1016/0266-8920(86)90033-0.
Lombardo, F., Volpi, E., Koutsoyiannis, D., 2012. Rainfall downscaling in time: theoretical and empirical comparison between multifractal and Hurst-Kolmogorov discrete random cascades. Hydrol. Sci. J. 57 (6), 1052–1066. https://doi.org/10.1080/02626667.2012.695872.
Lombardo, F., Volpi, E., Koutsoyiannis, D., Papalexiou, S.M., 2014. Just two moments! A cautionary note against use of high-order moments in multifractal models in hydrology. Hydrol. Earth Syst. Sci. 18 (1), 243–255. https://doi.org/10.5194/hess-18-243-2014.
Mandelbrot, B., Wallis, J.R., 1969. Computer experiments with fractional gaussian noises: part 1, averages and variances. Water Resour. Res. 5 (1), 228–241. https://doi.org/10.1029/WR005i001p00228.
Mardia, K.V., Goodall, C.R., 1993. Spatial-temporal analysis of multivariate environmental monitoring data. Multivariate Environ. Statist. 6 (76), 347–385.
Marriott, F.H.C., Pope, J.A., 1954. Bias in the estimation of autocorrelations. Biometrika 41 (3/4), 390. https://doi.org/10.2307/2332719.
Matalas, N.C., 1967. Mathematical assessment of synthetic hydrology. Water Resour. Res. 3 (4), 937–945. https://doi.org/10.1029/WR003i004p00937.
Matalas, N.C., Wallis, J.R., 1976. Generation of Synthetic Flow Sequences, Systems Approach to Water Management. McGraw-Hill, New York, New York.
Mehrotra, R., Sharma, A., 2007. A semi-parametric model for stochastic generation of multi-site daily rainfall exhibiting low-frequency variability. J. Hydrol. 335 (1–2), 180–193. https://doi.org/10.1016/j.jhydrol.2006.11.011.
Mehrotra, R., Srikanthan, R., Sharma, A., 2006. A comparison of three stochastic multi-site precipitation occurrence generators. J. Hydrol. 331 (1–2), 280–292. https://doi.org/10.1016/j.jhydrol.2006.05.016.
Mejia, J.M., Rousselle, J., 1976. Disaggregation models in hydrology revisited. Water Resour. Res. 12 (2), 185–186. https://doi.org/10.1029/WR012i002p00185.
Menabde, M., Harris, D., Seed, A., Austin, G., Stow, D., 1997. Multiscaling properties of rainfall and bounded random cascades. Water Resour. Res. https://doi.org/10.1029/97WR02006.
Mhanna, M., Bauwens, W., 2012. A stochastic space-time model for the generation of daily rainfall in the Gaza Strip. Int. J. Climatol. 32 (7), 1098–1112. https://doi.org/10.1002/joc.2305.
Molnar, P., Burlando, P., 2005. Preservation of rainfall properties in stochastic

disaggregation by a simple random cascade model. Atmospheric Res. 77 (1–4 SPEC. ISS.), 137–151. https://doi.org/10.1016/j.atmosres.2004.10.024.
Moran, P.A.P., 1970. Simulation and evaluation of complex water systems operations. Water Resour. Res. 6 (6), 1737–1742. https://doi.org/10.1029/WR006i006p01737.
Müller, H., Haberlandt, U., 2015. Temporal rainfall disaggregation: from point disaggregation to spatial rainfall. J. Hydrol. Eng. 20 (11), 3973. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001195.
Müller, H., Haberlandt, U., 2018. Temporal rainfall disaggregation using a multiplicative cascade model for spatial application in urban hydrology. J. Hydrol. 556, 847–864. https://doi.org/10.1016/j.jhydrol.2016.01.031.
Nataf, A., 1962. Statistique mathematique-determination des distributions de probabilites dont les marges sont donnees. C. R. Acad. Sci. Paris 255 (1), 42–43.
O'Connell, P.E., Koutsoyiannis, D., Lins, H.F., Markonis, Y., Montanari, A., Cohn, T., 2016. The scientific legacy of Harold Edwin Hurst (1880–1978). Hydrol. Sci. J. 61 (9), 1571–1590. https://doi.org/10.1080/02626667.2015.1125998.
Olsson, J., 1998. Evaluation of a scaling cascade model for temporal rainfall disaggregation. Hydrol. Earth Syst. Sci. https://doi.org/10.5194/hess-2-19-1998.
Onof, C., Chandler, R.E., Kakou, A., Northrop, P., Wheater, H.S., Isham, V., 2000. Rainfall modelling using Poisson-cluster processes: a review of developments. Stoch. Env. Res. Risk Assess. 14 (6), 0384–0411. https://doi.org/10.1007/s004770000043.
Onof, C., Townend, J., Kee, R., 2005. Comparison of two hourly to 5-min rainfall disaggregators. Atmos. Res. 77 (1–4), 176–187. https://doi.org/10.1016/j.atmosres.2004.10.022.
Papalexiou, S.M., 2018. Unified theory for stochastic modelling of hydroclimatic processes: preserving marginal distributions, correlation structures, and intermittency. Adv. Water Resour. https://doi.org/10.1016/j.advwatres.2018.02.013.
Papalexiou, S.M., Markonis, Y., Lombardo, F., AghaKouchak, A., Foufoula-Georgiou, E., 2018. Precise Temporal Disaggregation Preserving Marginals and Correlations (DiPMaC) for Stationary and Nonstationary Processes. Water Resour. Res. https://doi.org/10.1029/2018WR022726.
Paschalis, A., Fatichi, S., Molnar, P., Rimkus, S., Burlando, P., 2014. On the effects of small scale space–time variability of rainfall on basin flood response. J. Hydrol. 514, 313–327. https://doi.org/10.1016/j.jhydrol.2014.04.014.
Pegram, G.G.S., James, W., 1972. Multilag multivariate autoregressive model for the generation of operational hydrology. Water Resour. Res. 8 (4), 1074–1076. https://doi.org/10.1029/WR008i004p01074.
Pui, A., Sharma, A., Mehrotra, R., Sivakumar, B., Jeremiah, E., 2012. A comparison of alternatives for daily to sub-daily rainfall disaggregation. J. Hydrol. 470–471, 138–157. https://doi.org/10.1016/j.jhydrol.2012.08.041.
Rajagopalan, B., Lall, U., 1999. A k-nearest-neighbor simulator for daily precipitation and other weather variables. Water Resour. Res. 35 (10), 3089–3101. https://doi.org/10.1029/1999WR900028.
Richardson, C.W., Wright, D.A., 1984. WGEN: A Model for Generating Daily Weather Variables. US Department of Agriculture, Agricultural Research Service Washington, DC.
Rodriguez-Iturbe, I., Cox, D.R., Isham, V., 1987. Some models for rainfall based on stochastic point processes. Proc. R. Soc. A 410 (1839), 269–288. https://doi.org/10.1098/rspa.1987.0039.
Rodríguez-Iturbe, I., Mejía, J.M., 1974. The design of rainfall networks in time and space. Water Resour. Res. 10 (4), 713–728. https://doi.org/10.1029/WR010i004p00713.
Rupp, D.E., Keim, R.F., Ossiander, M., Brugnach, M., Selker, J.S., 2009. Time scale and intensity dependency in multiplicative cascades for temporal rainfall disaggregation. Water Resour. Res. https://doi.org/10.1029/2008WR007321.
Salas, J.D., Delleur, J.W., Yevjevich, V., Lane, W.L., 1980. Applied Modeling of Hydrologic Time Series (2nd Print). Water Resources Publication, Littleton, Colorado.
Segond, M.-L., Onof, C., Wheater, H.S., 2006. Spatial–temporal disaggregation of daily rainfall from a generalized linear model. J. Hydrol. 331 (3–4), 674–689. https://doi.org/10.1016/j.jhydrol.2006.06.019.
Serinaldi, F., 2009. A multisite daily rainfall generator driven by bivariate copula-based mixed distributions. J. Geophys. Res. 114 (D10), D10103. https://doi.org/10.1029/2008JD011258.
Shao, Q., Zhang, L., Wang, Q.J., 2016. A hybrid stochastic-weather-generation method for temporal disaggregation of precipitation with consideration of seasonality and within-month variations. Stoch. Env. Res. Risk Assess. 30 (6), 1705–1724. https://doi.org/10.1007/s00477-015-1177-3.
Shumway, R.H., Stoffer, D.S., 2017. In: Time Series Analysis and Its Applications. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-52452-8.
Srikanthan, R., Pegram, G.G.S., 2009. A nested multisite daily rainfall stochastic generation model. J. Hydrol. 371 (1–4), 142–153. https://doi.org/10.1016/j.jhydrol.2009.03.025.
Stacy, E.W., 1962. A generalization of the gamma distribution. Ann. Math. Stat. 33 (3), 1187–1192. https://doi.org/10.1214/aoms/1177704481.
Stedinger, J.R., Vogel, R.M., 1984. Disaggregation procedures for generating serially correlated flow vectors. Water Resour. Res. 20 (1), 47–56. https://doi.org/10.1029/WR020i001p00047.
Tadikamalla, P.R., 1980. A look at the Burr and related distributions. Int. Statist. Rev. 337–344.
Tarpanelli, A., Franchini, M., Brocca, L., Camici, S., Melone, F., Moramarco, T., 2012. A simple approach for stochastic generation of spatial rainfall patterns. J. Hydrol. 472–473, 63–76. https://doi.org/10.1016/j.jhydrol.2012.09.010.
Tessier, Y., Lovejoy, S., Hubert, P., Schertzer, D., Pecknold, S., 1996. Multifractal analysis and modeling of rainfall and river flows and scaling, causal transfer functions. J. Geophys. Res.: Atmospheres 101 (D21), 26427–26440. https://doi.org/10.1029/96JD01799.
Thomas, H.A., Fiering, M.B., 1962. Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. Des. Water Resour. Syst. 459–493.
Todorovic, P., Woolhiser, D.A., 1975. A stochastic model of n-day precipitation. J. Appl. Meteorol. 14 (1), 17–24.
Tsoukalas, I., 2019. Modelling and simulation of non-Gaussian stochastic processes for optimization of water-systems under uncertainty (PhD Thesis). Department of Water Resources and Environmental Engineering. National Technical University of Athens.
Tsoukalas, I.Kossieris, P. anySim: Stochastic simulation of processes with any marginal distribution and correlation structure. R package. Retrieved from. http://www.itia.ntua.gr/en/softinfo/33/.
Tsoukalas, I., Kossieris, P., Efstratiadis, A., Makropoulos, C., 2016. Surrogate-enhanced evolutionary annealing simplex algorithm for effective and efficient optimization of water resources problems on a budget. Environ. Modell. Software 77, 122–142. https://doi.org/10.1016/j.envsoft.2015.12.008.
Tsoukalas, I., Makropoulos, C., 2015a. A surrogate based optimization approach for the development of uncertainty-aware reservoir operational rules: the case of nestos hydrosystem. Water Resour. Manage. 29 (13), 4719–4734. https://doi.org/10.1007/s11269-015-1086-8.
Tsoukalas, I., Efstratiadis, A., Makropoulos, C., 2017. Stochastic simulation of periodic processes with arbitrary marginal distributions. 15th International Conference on Environmental Science and Technology. CEST 2017. Rhodes, Greece.
Tsoukalas, I., Makropoulos, C., 2015b. Multiobjective optimisation on a budget: exploring surrogate modelling for robust multi-reservoir rules generation under hydrological uncertainty. Environ. Modell. Software 69, 396–413. https://doi.org/10.1016/j.envsoft.2014.09.023.
Tsoukalas, I., Efstratiadis, A., Makropoulos, C., 2018a. Stochastic Periodic Autoregressive to Anything (SPARTA): modeling and simulation of cyclostationary processes with arbitrary marginal distributions. Water Resour. Res. 54 (1), 161–185. https://doi.org/10.1002/2017WR021394.
Tsoukalas, I., Makropoulos, C., Koutsoyiannis, D., 2018c. Simulation of stochastic processes exhibiting any-range dependence and arbitrary marginal distributions. Water Resour. Res. https://doi.org/10.1029/2017WR022462.
Tsoukalas, I., Papalexiou, S., Efstratiadis, A., Makropoulos, C., 2018d. A cautionary note on the reproduction of dependencies through linear stochastic models with non-gaussian white noise. Water 10 (6), 771. https://doi.org/10.3390/w10060771.
Tsoukalas, I., Kossieris, P., Efstratiadis, A., Makropoulos, C., Koutsoyiannis, D., 2018b. CastaliaR: an R package for multivariate stochastic simulation at multiple temporal scales. European Geosciences Union General Assembly 2018, Geophysical Research Abstracts, Vol. 20, Vienna, EGU2018-18433. European Geosciences Union.
Valencia, R.D.V., Schakke, J.C., 1973. Disaggregation processes in stochastic hydrology. Water Resour. Res. 9 (3), 580–585. https://doi.org/10.1029/WR009i003p00580.
Veneziano, D., Langousis, A., Furcolo, P., 2006. Multifractality and rainfall extremes: a review. Water Resour. Res. https://doi.org/10.1029/2005WR004716.
Wilks, D.S., 1998. Multisite generalization of a daily stochastic precipitation generation model. J. Hydrol. 210 (1–4), 178–191. https://doi.org/10.1016/S0022-1694(98)00186-3.
Wilks, D.S., Wilby, R.L., 1999. The weather generation game: a review of stochastic weather models. Prog. Phys. Geogr. 23 (3), 329–357. https://doi.org/10.1191/030913399666525256.
Williams, P., 1998. Modelling seasonality and trends in daily rainfall data. Adv. Neural Inf. Process. Syst. 10, 985–991.
Wójcik, R., Buishand, T.A., 2003. Simulation of 6-hourly rainfall and temperature by two resampling schemes. J. Hydrol. 273 (1–4), 69–80. https://doi.org/10.1016/S0022-1694(02)00355-4.