A multivariate stochastic model for the generation of synthetic time series at multiple time scales reproducing long-term persistence

Andreas Efstratiadis[*1], Yannis G. Dialynas[2], Stefanos Kozanis[1] & Demetris Koutsoyiannis[1]*

1 Department of Water Resources & Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Greece

2 School of Civil & Environmental Engineering, Georgia Institute of Technology, Atlanta, GA, USA

(* Corresponding author; andreas@itia.ntua.gr)

Paper submitted to *Environmental Modelling & Software*

Revised version, July 2014

Abstract

A time series generator is presented, employing a robust three-level multivariate scheme for stochastic simulation of correlated processes. It preserves the essential statistical characteristics of historical data at three time scales (annual, monthly, daily), using a disaggregation approach. It also reproduces key properties of hydrometeorological and geophysical processes, namely the long-term persistence (Hurst-Kolmogorov behaviour), the periodicity and intermittency. Its efficiency is illustrated through two case studies in Greece. The first aims to generate monthly runoff and rainfall data at three reservoirs of the hydrosystem of Athens. The second involves the generation of daily rainfall for flood simulation at five rain gauges. In the first emphasis is given to long-term persistence – a dominant characteristic in the management of large-scale hydrosystems, comprising reservoirs with carry-over storage capacity. In the second we highlight to the consistent representation of intermittency and asymmetry of daily rainfall, and the distribution of annual daily maxima.

Keywords: stochastic simulation; hydrometeorological processes; disaggregation; long-term persistence; intermittency; hydrosystems

Software availability

Name of Software: Castalia

Developer: ITIA research team (www.itia.ntua.gr)

Contact: Demetris Koutsoyianns / Andreas Efstratiadis, Department of Water Resources & Environmental Engineering, National Technical University of Athens, Heroon Polytechneiou 5, 157 80 Zographou, Athens, Greece

Year first available: 2000 (version 1); 2004 (version 2); 2011 (version 3); 2014 (version 4 beta)

Hardware required: PC

Program language: CodeGear Delphi 2009

Availability: Castalia is freely provided upon request to the authors.

1 Introduction

Stochastic simulation is aimed to generate synthetic data that represent non-deterministic inputs to the system under study. This allows accounting for uncertainty and large variability of

input into the related processes. Particularly, design and management of water resources systems is a suitable field for the implementation of such approaches, due to the intrinsically uncertain nature of hydro-meteorological phenomena, which are often unpredictable for even short-term control horizons. Moreover, the use of synthetic time series instead of historical records is essential for providing sufficiently large samples (e.g., with length of hundreds or thousands of years) or ensembles of different time series of the same process, in order to evaluate a wide range of possible outcomes.

Probabilistic assessment through stochastic simulation is of high importance for all typical water-related problems. For instance, a major objective in the optimal planning and management of hydrosystems is the maximization of system reliability, namely the probability of satisfying the associated water uses and constraints. In this context, a hydrosystem operation model is driven by synthetic inflows of usually monthly time step, to evaluate the statistical regime of the regulated outflows (e.g. water withdrawals). For the representation of streamflows, finer time steps are also adopted (e.g., daily), in order to properly account for reservoir spills (Ilich, 2014) and small-scale regulations (e.g., through retention tanks). Another field of application of stochastic approaches involves the evaluation of flood risk, which requires even more detailed temporal resolutions (e.g., hourly). Although this problem has been traditionally tackled through semi-empirical methods, in particular by constructing "design storms" to be inputs to event-based rainfall-runoff models, during the last years much attention has been paid to continuous flood modelling, which make use of synthetic rainfall (Boughton and Droop, 2003). In this regard, there is an increasing demand for rainfall generators that properly represent not only the spatial and temporal variability of rainfall, but also the statistical properties of derived floods (Verhoest et al., 2010). Finally, synthetic meteorological (weather) data (i.e., temperature, potential evapotranspiration, solar radiation, wind velocity, etc.), can be important to a wide range of water, energy and environmental applications, including the design and management of renewable energy systems (Tsekouras and Koutsoyiannis, 2014).

Stochastic simulation constitutes a widely used methodology that extends over several disciplines, from signal processing to econometrics. Most of related time series analysis tools employ rather simplistic approaches, particularly ARMA-type models, which may only ensure fundamental statistical consistency, by means of reproducing the mean, variance and

autocorrelations for short lags of the parent historical data. However, hydrometeorological (and, more generally, geophysical) processes exhibit much more complex statistical behaviour, characterized by skewed rather than Gaussian distributions, as well as statistical inter-dependencies. The latter characteristic is important since hydrometeorological variables are correlated either due to cause-effect relationships (e.g., rainfall-runoff), or due to common hydroclimatic regimes (e.g., point rainfall at neighbouring stations). This sets the application of multivariate schemes that enable the preservation of cross-correlations, a necessity.

Finally, hydrometeorological processes exhibit several characteristic properties that are closely related to their temporal evolution, particularly: (a) long-term persistence, i.e. the tendency of wet years to cluster into multi-year wet periods or of dry years to cluster into multi-year drought periods, which is dominant property of the annual and over-annual processes; (b) periodicity, which appears at the sub-annual scale (e.g., monthly) and is due to the Earth motion; (c) intermittency, which is a key feature of several processes at fine temporal scales (e.g., daily rainfall) and is quantified by the probability that the value of the process within a time interval is zero (often referred to as probability dry). Intermittency also results in significant variability and high positive skewness, which are difficult to reproduce by most generators.

Since general-purpose approaches for time series analysis (summarized in the classic book by Box and Jenkins, 1970) fail to represent the characteristic properties of hydrometeorological processes, several specialised methodologies have been developed for hydrological applications. As mentioned by Koutsoyiannis (2000; see also the comprehensive review by Grygier and Stedinger, 1990), early efforts on stochastic hydrological modelling are found in the works of Barnes (1954), Maass et al. (1962), Thomas and Fiering (1962), Beard (1965) and Matalas (1967). The decades of 1970 and 1980 provided significant progress, including the implementation of cyclo-stationary and multivariate schemes, the preservation of skewness, the representation of long-term persistence, the effective handling of numerical problems related to parameter estimation, etc. The advances of this period are summarized in the classic works of Matalas and Wallis (1976), Salas et al. (1980), Bras and Rodriguez-Iturbe (1985) and Salas (1993). Such methodologies were implemented within specialized computer tools, including HEC-4 (USACE, 1971), WASIM (McLeod and Hipel, 1978), WGEN (Richardson, 1981; Richardson and Wright 1984), LAST (Lane and Frevert, 1990), SPIGOT (Grygier and Stedinger,

1990), CSUPAC1 (Salas, 1993), SAMS (Sveinsson et al., 2003; Salas et al., 2006), NSRP (Kilsby et al., 2007) and RainSim (Burton et al., 2008). Yet, most of the known modelling tools have important shortcomings, which mainly involve parameter estimation drawbacks, the preservation of narrow type of autocorrelation functions, and the inability to perform in multivariate problems (Koutsoyiannis, 2000), particularly in fine (i.e., sub-monthly) time scales.

Another deficiency of many of the widely used stochastic packages is the fact that they merely preserve statistical characteristics at a specific temporal scale, which coincide with the time resolution of simulation. Yet, given that hydrometeorological processes exhibit different behaviour at different temporal scales, a fully consistent approach should imply the generation of synthetic time series that reproduce the statistical characteristics of the parent historical samples, not only at the time scale of simulation, but also at coarser ones. Traditionally, this problem is tackled by disaggregation techniques, which follow the general scheme proposed by Valencia and Schaake (1973). In this scheme, which major advantage is simplicity, disaggregation is implemented in two or more steps, where in the first step higher-level (e.g., annual) time series are generated that are next disaggregated to finer scales (e.g. monthly), in subsequent steps. However, most of well-known disaggregation approaches exhibit difficulties in parameter estimation, inaccuracies in preserving skewness and cross-correlations, and computational inefficiency (Langousis and Koutsoyiannis, 2005). For this reason, some researchers have proposed non-parametric approaches, in an attempt to preserve statistical correlations without having to resort to disaggregation (e.g., Srinivas and Srinivasan, 2005; Ilich, 2014; Srivastav and Simonovic, 2014).

An original three-level stochastic simulation framework, implemented within Castalia computer package, is presented in this paper. Castalia reproduces all essential characteristics and peculiarities of hydrometeorological processes at the annual, monthly and daily time scale. The whole modelling concept is unique, in terms of handling all aforementioned challenges, through effective and statistically consistent techniques. Next, we briefly review the key features of the methodological framework, which synthesizes several individual techniques that are described in detail in a number of research articles (Koutsoyiannis 1994, 1999, 2000, 2001; Koutsoyiannis and Manetas, 1996; Koutsoyiannis et al., 2003a). Castalia has been used, mostly in its early implementation (Efstratiadis and Koutsoyiannis, 2004), in several applications in the last years

(e.g., Koutsoyiannis et al., 2003b, Efstratiadis et al., 2004; Nalbantis et al., 2011; Tsekouras and Koutsoyiannis, 2014; Efstratiadis et al., 2014). However, a comprehensive presentation of the software and the methodology on which is based was never made and therefore it is the subject of this paper. In addition to the methodological elements, the paper illustrates the advantages of the modelling procedure and the software features through two case studies, involving the generation of synthetic monthly and daily time series, to be inputs in water management and flood modelling studies, respectively.

2 Software description and model overview

2.1 Key features

Castalia is free software, developed by the research team ITIA, in the National Technical University of Athens. The initial version of the program for monthly stochastic simulations (Efstratiadis and Koutsoyiannis, 2004), was implemented as component of a decision support system for the management of the water supply system of Athens (Koutsoyiannis et al., 2003b). The current version also supports daily simulations, through a three-level multivariate disaggregation scheme (Dialynas, 2011; Dialynas et al., 2011). For intermediate time scales (e.g., seasonal, weekly), synthetic data is straightforwardly provided by aggregating from the closest finer scale. Next, the key steps of the upgraded scheme are outlined for the generation of daily time series; more details are provided in sections 3 to 5, dealing with each specific time scale (annual, monthly and daily, respectively).

Castalia implements an original multivariate stochastic simulation scheme, in which each variable refers to a specific hydrometeorological process, at a specific location. All variables are assumed to be mutually correlated. The generating procedure preserves the marginal statistics up to third order (mean, standard deviation, skewness) as well as the joint second order statistics, particularly the first order autocorrelations and lag zero cross-correlations, at the daily, monthly and annual time scales. These are generally assumed as the essential statistical properties that should be preserved by stochastic hydrological models (Matalas and Wallis, 1976). Moreover, the model reproduces the long-term persistence (LTP) at the annual and over-annual scales, the periodicity at the monthly scale, and the intermittency at the daily scale (in terms of preserving the probability dry of the process of interest). We remark that the lagged cross-correlations (for lags higher than zero) are not explicitly preserved, to avoid complex schemes with many

parameters, whose estimation can be highly uncertain. On the other hand, the model already preserves the autocorrelations at all scales, which is indirectly transferred in approximating lagged cross-correlations.

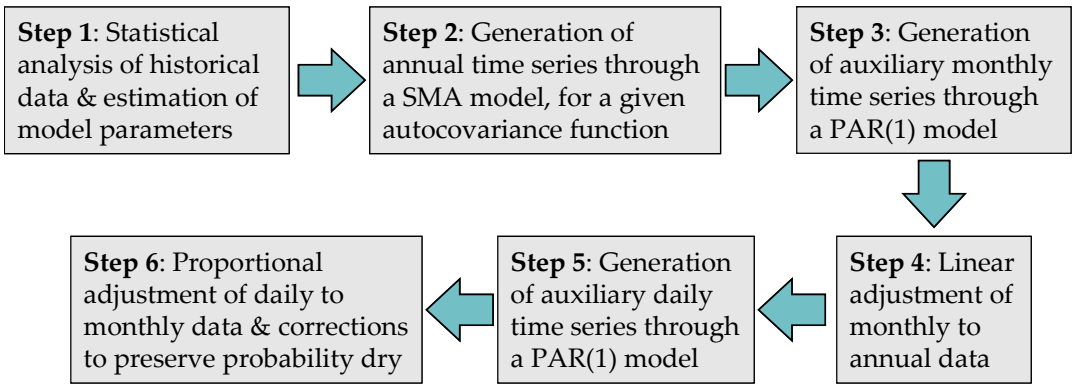

Figure 1 Outline of computational procedures in Castalia.

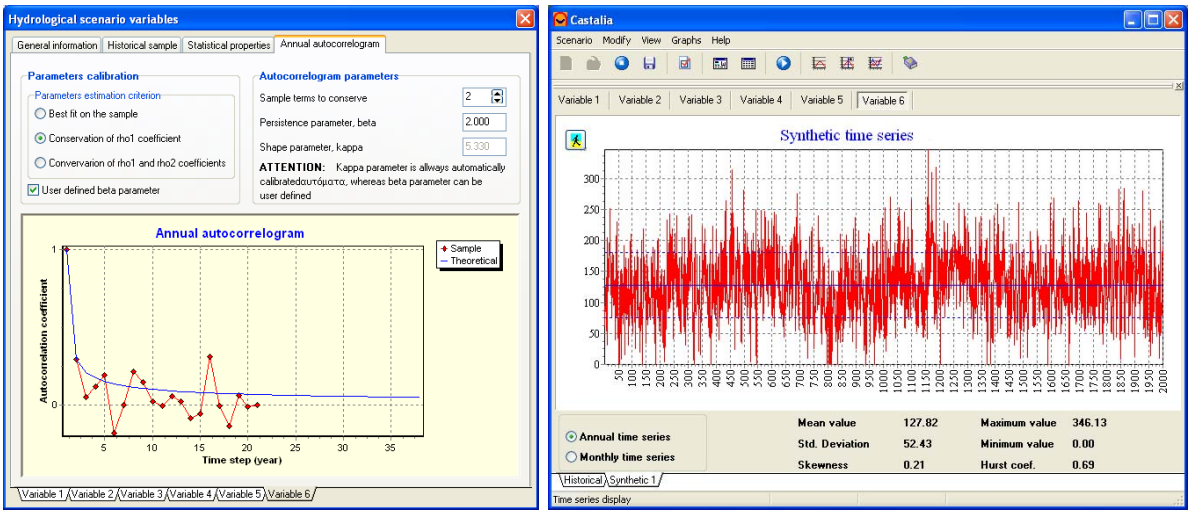

Figure 2 Characteristic screenshots of Castalia: (left) determination of autocovariance function for annual simulations and (right) plot of synthetic annual time series.

Fig. 1 illustrates the flow diagram for daily simulations, which follows a typical two-phase disaggregation scheme. First, the statistical characteristics of the parent historical data are computed, through which all model parameters are estimated. At the annual time scale, LTP is reproduced through a symmetric moving average scheme that implements a user-defined autocovariance function, which enables the representation of a wide range of stochastic structures, i.e. from ARMA-type, which are characterized by short-term persistence, to Hurst-

Kolmogorov behaviour, with as high long-term persistence as needed. For the monthly and daily time scales, auxiliary time series are initially provided by a multivariate periodic autoregression scheme. Next, a disaggregation procedure is employed to establish statistical consistency between the three temporal scales; first the monthly series are adjusted to the known annual ones. Finally, the daily time series are adjusted to the disaggregated monthly data, using a multivariate coupling scheme. Technical details are provided in sections 3, 4 and 5, describing the annual, monthly and daily generation schemes, respectively. The model novelties are also highlighted in section 2.2.

The above procedure can be formulated in two alternative modes. In steady-state simulations long time series are generated to estimate long-term performance characteristics, such as the reliability or safe yield through a hydrosystem. The length of simulations may reach several thousands of years, in order to represent statistically rare events and evaluate extreme probabilities. Apparently, even for much shorter time horizons, the outcomes of simulations become practically independent of the initial conditions. The other mode refers to terminating simulations, in which the present and past states of the system under study must be considered, thus the observed values of the present and past must condition the hydrological time series of the future. In terminating simulations, the model runs in forecast mode for a time horizon of, typically, few years, where the observed past records of the hydrological variables are introduced to the generation scheme, in order to obtain statistical predictions of their future values. In this context, numerous "ensemble" time series of short length are generated, which represent multiple hydrological scenarios for relative small time horizons.

Castalia operates on a windows environment with several graphical capabilities, comprising charts, tables and specific tools for adjusting the parameters of the modelling procedure (e.g., Fig. 2). The synthetic time series, either individually or by means of hydrological scenarios, can be exported in text file formats.

2.2 Comparison with other packages

Table 1 illustrates a comparison of Castalia's technical characteristics with two widespread stochastic hydrology packages, i.e. SAMS (Sveinsson et al., 2003; Salas et al., 2006), and SPIGOT (Grygier and Stedinger, 1990), which implement multivariate disaggregation schemes and temporal disaggregation at different scales (i.e., up to monthly scale for SAMS, and up to weekly or daily for SPIGOT). The two packages also support further spatial disaggregation to

generate, for instance, consistent annual flows at different stations, which at present is not the case in Castalia. In Castalia, the concept of "location" is extended to any kind of correlated variables (not particularly spatially correlated).

The great advantage of Castalia is the preservation of long-term persistence (LTP). On the contrary, SAMS and SPIGOT only represent processes with short-term persistence (AR, MA and ARMA), which cannot reproduce the Hurst phenomenon, as discussed by Koutsoyiannis (2011). In fact, Castalia is capable of handling arbitrary annual autocorrelation functions, through the implemented generalized autocorrelation structure (section 3.1), which is also applicable for multivariate simulations. Moreover, Castalia has the additional advantage of simultaneous preservation of all essential statistical characteristics at the annual, monthly, and daily scale, with emphasis to characteristic peculiarities, such as skewness and intermittency that are difficult to handle through analytical models. Another original feature of Castalia is the use of a multicriteria optimization approach for the decomposition of covariance matrices, which is essential for preserving the observed cross-correlations at all temporal scales. Finally, it supports terminating simulations (i.e. generation of ensemble time series for stochastic forecast), apart from steady-state ones. These points constitute unique elements of Castalia compared to popular packages. Note that SAMS and SPIGOT also include additional advantages not listed in Table 1 and thus may be preferable for particular types of problems.

Table 1 Comparison of stochastic simulation packages Castalia, SAMS, and SPIGOT

	Castalia	SAMS	SPIGOT
Multivariate analysis	Yes	Yes	Yes
Time scales of temporal disaggregation (A: annual; M: monthly; W: weekly; D: daily)	A → M → D	A → M	A → M, A → W→ D
Preservation of all essentail statistical characteristics at the annual, monthly, and daily scales	Yes	No	No
Preservation of LTP	Yes	No	No
User-defined annual autocorrelation function	Yes	No	No
Preservation of seasonality	Yes	Yes	Yes
Preservation of probability dry, at the daily scale	Yes	No	Yes
Spatial disaggregation	No	Yes	Yes
Decomposition of covariance matrices through optimization	Yes	No	No
Applicable for terminating simulations, conditioned on past data	Yes	No	Yes

3 Generation of annual time series

3.1 The symmetric moving average (SMA) generating scheme

In annual simulations, a key requirement is the reproduction of the long-term persistence, also referred to as scaling behaviour or Hurst-Kolmogorov dynamics, which is an omnipresent property of hydrometeorological (and, generally, geophysical) processes (Koutsoyiannis 2002, 2003, 2011; Koutsoyiannis and Montanari, 2007). This behaviour has major effects on the management of water resource systems and the design of all related infrastructures (Koutsoyiannis, 2011), as dry periods tend to follow dry ones, while wet periods also tend to follow wet ones. In this context, long lasting droughts or wet periods can be regarded as the result of large-scale random fluctuations of climate. These can be represented by means of stationary stochastic processes with a generalized autocorrelation structure, such as the one proposed by Koutsoyiannis (2000):

$$\gamma_j = \gamma_0 \left[1 + \kappa \beta j\right]^{-1/\beta} \tag{1}$$

where γ_j is the autocovariance of the annual stochastic process for lag j, γ_0 is the variance and κ, β are shape and scale parameters, respectively, that are related to the persistence of the process. By adjusting the values of κ and β, one can take a wide range of feasible autocovariance structures. In particular, for $\beta = 0$ we obtain an ARMA-type structure, corresponding to a Hurst coefficient $H = 0.50$; in that case (by applying l'Hôpital's rule) eq. (1) is written as $\gamma_j = \gamma_0 \exp(-\kappa j)$. Any other positive value of parameter β represents a persistent process, with $H > 0.50$. We remark that the estimation of the Hurst coefficient of the historical data is quite uncertain, due to the inadequate length of data records. For this reason, we recommend manually setting a plausible value of parameter β, and estimating κ by fitting eq. (1) to the observed (empirical) lag-one autocorrelation coefficient, $\rho_1 := \gamma_1 / \gamma_0$. For the estimation of parameters κ and β, the program also offers alternative options, particularly: (a) analytical computation of κ and β by fitting eq. (1) to the first two autocorrelation coefficients ρ_1 and ρ_2, (b) calibration of eq. (1) against the $N/2$ first terms of the autocorrelogram, where N is the length of historical data, and (c) manual setting of parameter β and estimation of κ similarly to case (b). It is noted that in a typical Hurst-Kolmogorov process the autocovariance decays with lag j according to a power law with exponent $-2 - 2H$; therefore the parameter β is related to H by $H = 1 - 1/2\beta$. Thus $\beta = 2$ results in $H = 0.75$, which is a common value for hydrological processes (see also section 6.1). Note that

eq. (1) is an expression more general than that of the Hurst-Kolmogorov process and offers a great number of possibilities.

Castalia implements the autocovariance function (1) within a symmetric moving average (SMA) scheme introduced by Koutsoyiannis (2000), which is used to generate synthetic annual time series through the formula:

$$\underline{z}_i = \sum_{j=-n}^{n} \alpha_{|j|}\, \underline{v}_{i+j} = \alpha_s\, \underline{v}_{i-s} + \ldots + \alpha_1\, \underline{v}_{i-1} + \alpha_0\, \underline{v}_i + \alpha_1\, \underline{v}_{i+1} + \ldots\ \alpha_s\, \underline{v}_{i+s} \qquad (2)$$

where $\underline{z}_i$ denotes the annual stochastic process for year i, $\underline{v}_i$ are independent identically distributed innovations, and α_j are numerical coefficients that can be analytically determined from the sequence of γ_j. (Notice that underlined symbols denote random variables according to the so-called Dutch convention; cf. Hemelrijk, 1966) Koutsoyiannis (2000) has shown that the inverse finite Fourier transform $s_\alpha(\omega)$ of the coefficients a_j is related to that of the coefficients γ_j by:

$$s_\alpha(\omega) = \sqrt{2 s_\gamma(\omega)} \qquad (3)$$

Finally, the auxiliary variables (also referred to as noise variables or innovations) $\underline{v}_i$ are generated through a three-parameter Gamma distribution, which ensures the preservation of the mean value and the coefficient of skewness of the observed annual data. This distribution, which is generally used for the generation of noise variables at the three time scales of interest, is quite flexible since it can represent from exponentially to normally-distributed variables (we note that at the annual scale, most of hydrometeorological variables are close to normal). Heavy-tailed distributions (e.g. Pareto) are not supported by the current version of the program but are scheduled for the future versions. The variance and lag-one autocorrelation are explicitly preserved through the proper evaluation of coefficients α_j.

3.2 Multivariate formulation

The SMA scheme is easily generalized for multivariate simulations, thus also preserving the cross-correlations of the historical variables. Let a set m of correlated variables, with known covariance matrix $\boldsymbol{C}$, which is an $m \times m$ matrix representing the historical variances, in the diagonal elements, and the lag zero cross-correlation coefficients, in the off-diagonal ones. At

each time step (i.e. year) i, correlated innovation variables are generated, in terms of an m-dimensional vector $\boldsymbol{v}_i := [\underline{v}_i^1, \ldots, \underline{v}_i^m]^T$ given by:

$$\underline{\boldsymbol{v}}_i = \boldsymbol{B}\, \underline{\boldsymbol{w}}_i \tag{4}$$

where $\underline{\boldsymbol{w}}_i := [\underline{w}_i^1, \ldots, \underline{w}_i^m]^T$ is a vector of gamma-distributed noise variables with unit variance, independent both in time and location, and $\boldsymbol{B}$ is a matrix with size $m \times m$ which is obtained by decomposing the covariance matrix $\boldsymbol{C}$, such that:

$$\boldsymbol{B}\, \boldsymbol{B}^{\mathrm{T}} = \boldsymbol{C} \tag{5}$$

The methodology for solving (5) is briefly discussed in section 3.3 below. The remaining parameters required to define model (4) are the vector of mean values and coefficients of skewness of $\boldsymbol{w}_i$, which are analytically derived from the associated statistical characteristics of the historical data (Koutsoyiannis, 2000).

3.3 Decomposition of covariance matrices

The decomposition of covariance matrices is one of the most challenging numerical problems of operational stochastics, which appears in all multivariate stochastic schemes. This problem has several peculiarities. Specifically, eq. (5) has infinite number of solutions when $\boldsymbol{C}$ is positive definite and no (real) solution otherwise. The latter case appears very often and is due to inconsistencies of statistical estimation, particularly when different items of the covariance matrices are estimated using records of different lengths (Grygier and Stedinger, 1990). Another drawback is encountered when attempting to preserve the coefficients of skewness of the historical data, since the innovation variables associated with the stochastic model may potentially have too high coefficients of skewness, which are practically impossible to reproduce by random number generators (Todini, 1980).

In Castalia, the above issues are effectively handled through an optimization approach proposed by Koutsoyiannis (1999), whether the matrix $\boldsymbol{C}$ is positive definite or not. In this respect, a weighted multicriteria function is formulated that comprises three components aiming at (a) accurate preservation of the observed variances, (b) satisfactory approximation of the observed covariances, and (c) minimization of the skewness coefficients of the innovation variables, which are proportional to the inverse of a matrix whose elements are the cubes of $\boldsymbol{B}$.

Through a suitable formulation of $\boldsymbol{B}$, one can restrict the skewness of innovations up to reasonable limits, which allows, in turn, preserving the skewness of the actual variables, i.e. the observed data. Koutsoyiannis (1999) provides analytical expressions of the objective function and its derivatives, which strongly facilitate the optimization procedure. Castalia employs a hybrid scheme, in which a conjugate gradient local search technique, i.e. the Fletcher-Reeeves algorithm (Press et al., 1992), runs from multiple, randomly generated initial points in the feasible space. This approach allows avoiding an early trapping of the algorithm to local minima, thus ensuring both effectiveness (i.e., satisfactory approximation of a good-compromise solution) and efficiency (i.e., computational speed). In this procedure, the user has to specify the maximum number of local searches (default value 100) as well as the convergence criterion, in terms of a minimum desirable value of the norm $||\boldsymbol{B}\,\boldsymbol{B}^{\mathrm{T}} - \boldsymbol{C}||$. Usually, even a single trial suffices to obtain an acceptable solution, with the exception of highly skewed variables, which may require several iterations (i.e. local optimizations) to converge.

The above procedure can be used regardless of the model's autocorrelation structure, which makes it suitable for the three-level simulation scheme. Yet, given that in most applications the method is approximate (when the variance-covariance matrix $\boldsymbol{C}$ is not positive semi-definite), the cross-correlations of the historical data that are contained in $\boldsymbol{C}$ cannot be preserved with perfect accuracy in the synthetic time series. In fact, the generating scheme reproduces the cross-correlations that are derived by performing inverse calculations of the decomposition algorithm, i.e. by using the resulting matrices $\boldsymbol{B}'$ as true ones, and thus estimating a new set of theoretical cross-correlations on the basis of the approximated variance-covariance matrix $\boldsymbol{C}' = \boldsymbol{B}'\boldsymbol{B}'^{\mathrm{T}}$.

4 Generation of monthly time series

4.1 Generation of auxiliary monthly series

To construct the monthly synthetic time series, we initially generate auxiliary series without any reference to the known annual ones. In this temporal scale, a key specification is the preservation of periodicity, which is achieved by employing a cyclostationary model. In particular, we use a periodic autoregressive scheme of first order, PAR(1), which is the most parsimonious among linear stochastic models. In multivariate terms, it is given by the recursive equation:

$$\underline{\tilde{\boldsymbol{x}}}_{i,s} = \boldsymbol{A}_s \, \underline{\tilde{\boldsymbol{x}}}_{i\,s-1} + \boldsymbol{B}_s \, \underline{\boldsymbol{v}}_{i,s} \qquad (6)$$

where $\underline{\tilde{\boldsymbol{x}}}_{i,s} := [\underline{\tilde{x}}_{i,s}^{\,1}, \ldots, \underline{\tilde{x}}_{i,s}^{\,m}]^T$ represents a vector of m stochastic processes in year i and month s (s = 1, …, 12), which represent auxiliary variables, to be next adjusted to annual synthetic data (see section 4.2); $\boldsymbol{A}_s$ and $\boldsymbol{B}_s$ are $m \times m$ parameter matrices; and $\underline{\boldsymbol{v}}_{i,s}$ is an m-dimensional vector of innovations, namely independent, in time and space, random variables, with unit variance. In the generating scheme (6) we assume diagonal matrices $\boldsymbol{A}_s$, thus formulating the so-called contemporaneous PAR(1) model (Matalas and Walis, 1976; Salas, 1993, p. 19.31), which is mathematically convenient, and also suffices for preserving the essential statistical properties of the historical samples (Koutsoyiannis, 1999).

For each month s, the model parameters $\boldsymbol{A}_s$ and $\boldsymbol{B}_s$ are determined from the joint second order statistics of the monthly historical samples. Specifically, the diagonal matrix $\boldsymbol{A}_s$ contains the monthly lag one autocorrelations, while matrix $\boldsymbol{B}_s$ derives through decomposing the variance-covariance matrix of the historical data, following the optimization approach that was discussed in 3.3. Finally, innovations $\underline{\boldsymbol{v}}_{i,s}$ are generated through a three-parameter Gamma distribution, the parameters of which are estimated from the monthly means and skewness coefficients of the historical samples. Analytical equations are given by Koutsoyiannis (1999).

4.2 Adjusting monthly to annual time series

The model defined by (6) is proper for sequential generation of correlated monthly series $\underline{\tilde{\boldsymbol{x}}}_{i,s}$ but it cannot account for the annual values $\underline{\boldsymbol{z}}_i$, which are already generated through the multivariate SMA model. Apparently, the two data sets are not consistent, since for any year i, the annual sum of $\underline{\tilde{\boldsymbol{x}}}_{i,s}$, denoted as $\underline{\tilde{\boldsymbol{z}}}_i$, is not equal to the corresponding vector of annual variables, $\underline{\boldsymbol{z}}_i$. To establish consistency, we employ an adjusting procedure, introduced by Koutsoyiannis and Manetas (1996) and generalized by Koutsoyiannis (2001), in terms of the transformation:

$$\underline{\boldsymbol{x}}_{i,s} = \underline{\tilde{\boldsymbol{x}}}_{i,s} + \boldsymbol{H}_s \, (\underline{\boldsymbol{z}}_i - \underline{\tilde{\boldsymbol{z}}}_i) \qquad (7)$$

where $\boldsymbol{H}_s$ is a matrix of monthly parameters, estimated by:

$$\boldsymbol{H}_s = \mathrm{Cov}[\underline{\boldsymbol{x}}_{i,s}, \underline{\boldsymbol{z}}_i] \, \{\mathrm{Cov}[\underline{\boldsymbol{z}}_i, \underline{\boldsymbol{z}}_i]\}^{-1} \qquad (8)$$

In the case of a single variable, a linear transformation is employed that distributes the departure $\Delta\underline{\mathbf{z}}_i = (\underline{\mathbf{z}}_i - \underline{\widetilde{\mathbf{z}}}_i)$ of the additive property to each lower-level (i.e. monthly) variable proportionally to the covariance of this lower-level variable with the higher-level (i.e. annual) variable; at the multivariate case the definition of $\boldsymbol{H}_s$ is still provided by (8) but there is no easy interpretation (see details in Koutsoyiannis, 2001). It is proved that this adjusting procedure, defined by (7) and (8), preserves the vectors of means, the variance-covariance matrix and any linear relationship that holds among $\underline{\boldsymbol{x}}_{i,s}$ and $\underline{\mathbf{z}}_i$, including correlations between annual and monthly variables.

The above transformation has two disadvantages. First, skewness is hard to preserve in an analytical manner, yet such preservation is of great importance, as most of hydrometeorological processes, particularly at small time scales, exhibit non-symmetric distributions. Moreover, highly negative departures $\Delta\underline{\mathbf{z}}_i$ may result in negative values of the adjusted variables. To remedy these problems, we employ a simple repetitive procedure based on conditional sampling, as proposed by Koutsoyiannis and Manetas (1996). This procedure, which is a type of Monte Carlo simulation, aims at minimizing the departures $\Delta\mathbf{z}_i$, by repeating the generation process for the variables of each year (rather than performing a single generation for the entire simulation horizon), until the distance $\Delta\mathbf{z}_i$ becomes lower that an accepted limit, which is expressed as percentage (default value, 1%) of the annual standard deviation of the associated variable.

5 Generation of daily time series

5.1 Generation of auxiliary daily series

The general scheme for generating synthetic daily data resembles the case of monthly data, since auxiliary time series are produced through a PAR(1) model initially, which are then adjusted to the known monthly ones. Yet, the computational procedure is somewhat more complicated, given that, apart from the essential statistical characteristics that are, similar to monthly simulations, periodic functions of time, it is also necessary to reproduce intermittency, i.e. the proportions of intervals with zero values of the modelled variables. In the case of rainfall, this characteristic is often referred to as probability dry.

The PAR(1) model for multivariate daily simulations is formulated as:

$$\underline{\widetilde{\boldsymbol{y}}}_{s,\tau} = \boldsymbol{A}_s \, \underline{\widetilde{\boldsymbol{y}}}_{s\,\tau-1} + \boldsymbol{B}_s \, \underline{\boldsymbol{v}}_{s,\tau} \qquad (9)$$

where $\underline{\widetilde{\boldsymbol{y}}}_{s,\tau} := [\widetilde{\boldsymbol{y}}_{s,\tau}^{\,1}, \ldots, \widetilde{\boldsymbol{y}}_{s,\tau}^{\,m}]^T$ represents a vector of m stochastic processes with indices denoting month s and day τ (s = 1, …, 12; τ = 1, …, 30 or 31), $\boldsymbol{A}_s$ is an $m \times m$ diagonal matrix containing the lag-1 autocorrelations of historical data, $\boldsymbol{B}_s$ is an $m \times m$ matrix of parameters, which are estimated by decomposing the variance-covariance matrix, and $\underline{\boldsymbol{v}}_{s,\tau}$ is an m-dimensional vector of innovations, independent in time and space, which are generated through a Gamma distribution that preserves the mean values and the skewness coefficients of historical data. For convenience, in eq. (9), annual indices for $\widetilde{\boldsymbol{y}}_{s,\tau}$ and $\underline{\boldsymbol{v}}_{s,\tau}$ are omitted.

A key assumption of (9) is the homoscedasticity of $\widetilde{\boldsymbol{y}}_{s,\tau}$ and hence of innovations $\underline{\boldsymbol{v}}_{s,\tau}$, namely the hypothesis of constant variance of $\widetilde{\boldsymbol{y}}_{s,\tau}$ regardless of the value $\widetilde{\boldsymbol{y}}_{s,\tau-1}$. However, this prohibits from properly representing the high variability and asymmetry of historical data, which becomes more significant as the time scale of simulation decreases. Koutsoyiannis et al. (2003a) studied this problem within a simplified multivariate rainfall model, which resulted in synthetic hyetographs characterized by unrealistically similar peaks. One of the suggested methods was the power transformation of daily variables, such as:

$$\widetilde{\boldsymbol{y}}'_{s,\tau} = \widetilde{\boldsymbol{y}}_{s,\tau}^{\,(n)} \qquad (10)$$

where (n) denotes that all items of $\widetilde{\boldsymbol{y}}_{s,\tau}$ are raised to a common power n, where $0 < n < 1$ (n is assumed to be the same at all locations). Preserving the statistical characteristics of the transformed variables does not necessarily ensure that the characteristics of the original (i.e. untransformed) variables will also be preserved. However, Koutsoyiannis et al. (2003a) showed that for relatively high values of n (e.g., $n \geq 0.5$), the discrepancies are insignificant. Moreover, thanks to the power transformation it is much easier to reproduce the (usually) particularly high coefficients of skewness of the daily historical data. In this respect, for the generation of auxiliary daily time series, Castalia employs a modified expression of the PAR(1) model, where the auxiliary variables $\widetilde{\boldsymbol{y}}_{s,\tau}$ are replaced by $\widetilde{\boldsymbol{y}}'_{s,\tau}$.

5.2 Adjusting daily to monthly time series

In order to establish consistency between the monthly and daily synthetic data, an adjusting procedure is applied to the auxiliary time series that are generated by (9), to add-up to the known monthly values. Yet, in daily time scales, linear transformations, such as the one used for adjusting monthly to annual time series (section 4.2), are not appropriate, because they fail to preserve the probability dry, and may also result in negative values (Valencia and Schaake, 1973). In this context, for daily disaggregation, instead of (7), we employ a proportional adjusting scheme (Lane and Frevert, 1990; Grygier and Stedinger, 1990; Koutsoyiannis, 1988, 1994):

$$y'_{s,\tau} = \tilde{y}'_{s,\tau}\, x_s / \tilde{x}_s \qquad (11)$$

where $y'_{s,\tau}$ and $\tilde{y}'_{s,\tau}$ denote the initially generated and adjusted daily series, respectively, $\tilde{x}_s$ is the sum of $\tilde{y}'_{s,\tau}$ for month s, and x_s is the known monthly value. The above scheme, which is implemented for each individual location (i.e. simulated process), never results in negative values of $y_{s,\tau}$ and does not affect the preservation of probability dry, as zero values of the auxiliary variables remain zero after the adjusting. Moreover, whenever $\tilde{y}'_{s,\tau}$ are independent and two-parameter Gamma distributed, with common scale parameter, the procedure ensures accurate preservation of the entire distribution function (Koutsoyiannis, 1994). Numerical applications showed that the same procedure provides satisfactory approximations for variables with distributions approaching the two-parameter Gamma distribution (e.g. the three-parameter Gamma, which is generally employed in Castalia) even if the variables are correlated.

Similarly to monthly time series, a Monte Carlo repetitive procedure is applied to ensure a minimal departure between x_s and $\tilde{x}_s$. This aims to improve the approximations of the characteristics of historical daily data that are not explicitly preserved by (11), namely skewness and cross-correlation coefficients. Here the reproduction of skewness coefficients is much easier, as all calculations refer to power-transformed data (eq. 10).

5.3 Preservation of probability dry

The proportions of dry intervals or, equivalently, the probability dry of the parent time series, constitute major information of hydrometeorological processes at fine time scales. Since this characteristic cannot be explicitly preserved by single-state linear stochastic models, such as

PAR(1), we follow a hybrid procedure, involving the sequential application of three rules, as explained below.

5.3.1 Truncation of negative values

In order to preserve the usually high coefficients of variation in the daily time scale, the linear stochastic models unavoidably generate some negative values. Negative values may also appear in monthly simulations, e.g. in the case of summer rainfall, which in fact makes essential to employ the same truncation rule. Given that most hydrometeorological variables are by definition non-negative, all simulated negative values should be truncated to zero.

5.3.2 Rounding off rule for small positive values

The daily time series generation scheme often underestimates the historical probability dry, although the statistical characteristics that are related, to some extent, to this probability, i.e. the variance, skewness, and lag-1 autocorrelation, are satisfactory approximated. In particular, it cannot generate sequences of dry (zero) values, since there is no explicit distinction between the two states of the modelled process (i.e., the dry and the wet one).

This problem was investigated by Koutsoyiannis et al. (2003a), who suggested that by applying a rounding off rule to the stochastic process is preferable over modelling rainfall as a two-state process, which is much more complicated. Thus, they argued that the rounding off rule, according to which small values (e.g., < 0.10 mm) are set to zero values is more convenient and equally precise to two-state rainfall modelling, in terms of periods with very small rainfall depths that are handled as dry ones.

Castalia implements the rounding off rule suggested by Koutsoyiannis et al. (2003a), particularly for multivariate simulations. According to this rule, a proportion π_0 of the days with very small positive values, which are randomly chosen among all values that are smaller than a threshold l_0, are set to zero. The two arguments of the rounding off rule (i.e. π_0 and l_0) are constants, defined by the user. Note that this rule does not overlap with the truncation of negative values, because the former constitutes a probabilistic rule, and it clearly does not ensure truncation of all generated negative values.

5.3.3 Markov-based approach accounting for dry conditions in time and space

The application of the rounding off rule significantly increases the number of dry periods, which is added to the number of dry periods emerging from the truncation of negative values. Yet, as the total proportion of dry intervals may still be smaller than the historical one, we also use a Markov-based approach, considering the temporal and spatial distribution of dry periods.

Specifically, for a dry value $y_{\tau\text{-}1}{}^{l} = 0$ generated in Castalia in day $\tau - 1$ and location l, there is a probability μ_s to be followed by another dry value, thus $y_{\tau}{}^{l} = 0$. The conditional probability $\mu_s{}^{l} = P\{y_{\tau}{}^{l} = 0 \mid y_{\tau\text{-}1}{}^{l} = 0\}$ is defined for every month s as constant proportion of the corresponding probability dry $p_s{}^{l}$, i.e. $\mu_j{}^{l} = \lambda\, p_s{}^{l}$, where λ is an input parameter. On the other hand, for every dry value at location l, i.e. $y_{\tau}{}^{l} = 0$, there is also a conditional probability ξ that dry periods are forced to the rest of $m - 1$ simulated locations, in the same day τ. This is a reasonable assumption, particularly when dealing with rain gauges at close distances. Through appropriate selection of parameters λ and ξ, as explained in next section, this approach can generate extra dry periods, thus preserving the historical probability dry.

The combined use of the three aforementioned procedures (i.e., the truncation and rounding off rules, as well as the Markov-based approach), allow for preserving even very high values of probability dry, which may be typical in several processes (e.g., summer rainfall in dry climates). Thus, long sequences of zero daily values can be provided, at individual locations. Moreover, the generation of such sequences, both in space and time, is also achieved through the preservation of cross-correlation coefficients by the multivariate daily stochastic model.

5.3.4 Potential sources of bias

A negative outcome of the above procedure is the introduction of bias in some key statistical characteristics of the historical data. For instance, the truncation of negative values may result in overestimation of cross-correlations, since negative values are often contemporary. Furthermore, forcing dry periods in space may also overestimate cross-correlations, since for several dry days the same (i.e., zero) value is manually assigned to all modelled variables. Nevertheless, a slight overestimation of cross-correlations could counterbalance the underestimation resulting from the adjusting procedure of section 5.2. Also, this bias only depends on the value of k, so it can be adjusted to be negligible, through a careful adjustment of this parameter.

The application of the procedures outlined in section 5.3.3 may also affect the autocorrelation structure of the simulated variables. In general, by setting high values to parameters λ and ξ, the lag-1 autocorrelations are underestimated, and this may be unavoidable in cases of historical data with high proportions of dry periods (e.g., see case study in section 6.2). To counterbalance this, an autocorrelation adjusting factor is applied, which introduces positive prior bias to daily autocorrelation coefficients. The adjusting factor can be estimated by a trial and error procedure through inspection of the model outputs.

In terms of parameter sensitivity, λ and ξ have much greater impact than π_0 and l_0. A suitable range for λ and ξ cannot be specified a priori, since the effect of the method directly depends on specific characteristics of each particular case study, such as the number of simulated variables, the actual proportions of dry intervals, etc. In general, high values of λ and ξ should be avoided, as they may introduce significant bias to the statistical characteristics to be preserved. We recommend employing a trial and error approach to determine the aforementioned parameters empirically, i.e., by evaluating the statistical characteristics of the synthetic time series. Preliminary investigations showed that such a procedure requires at most two or three trial runs.

6 Case studies

6.1 Generation of monthly inflows for hydrosystem simulation

The first case study aims at the generation of simultaneous monthly inflows, i.e. rainfall and runoff into three major reservoirs (Evinos, Mornos, Hylike) of the water supply system of Athens (Koutsoyiannis et al., 2003b). The map of Fig. 3 shows the three reservoirs and their upstream catchments. This constitutes a multivariate generation problem with six variables (i.e. two processes at three basins), for a simulation length of 2000 years. We remark that the historical rainfall data have been obtained from rain gauges that are located close to each of the three reservoirs (not shown in the map), while the runoff data have been estimated by solving the monthly water balance equation for the unknown naturalized inflows. Apart from the rainfall sample at Hylike, the rest of historical records cover a period of around 40 years (1970-2008).

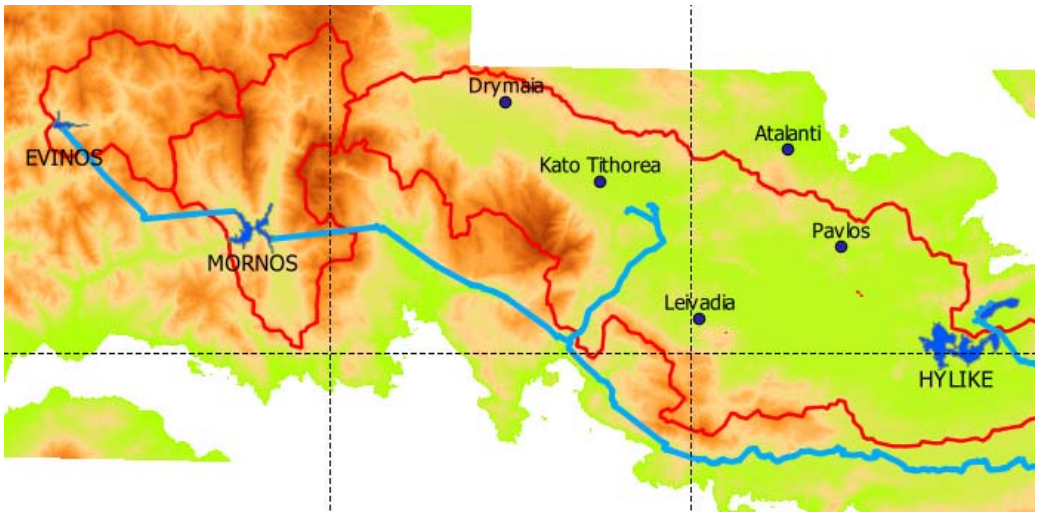

Figure 3 Part the water resource system of Athens, in which are illustrrated the three reservoirs, their upstream basins, the conveyance network, and the five rain gauges used in case study 2.

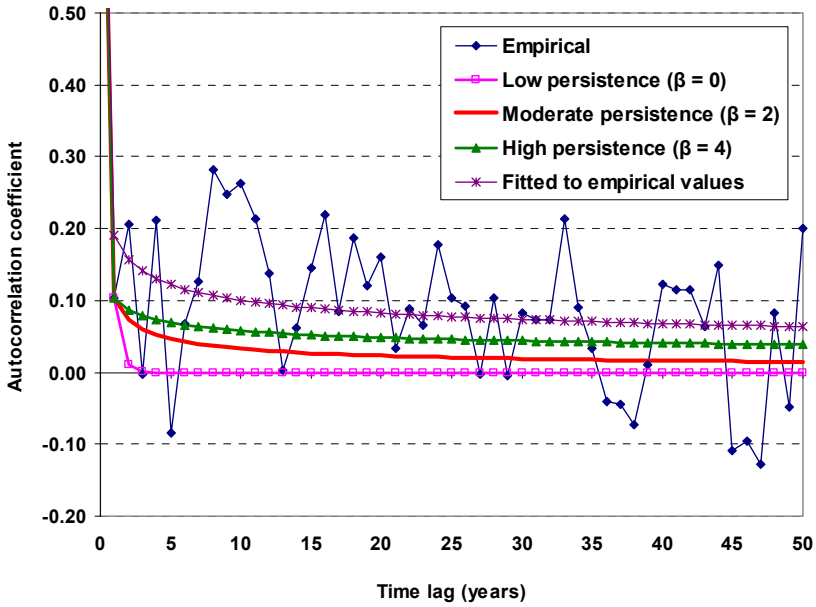

Figure 4 Empirical and theoretical autocorrelograms of the annual rainfall at Hylike, for different parameters of eq. (1).

The small length of all but one time series makes rather unreliable the estimation of the long-term persistence characteristics of the associated processes, which are mathematically expressed by the generalized autocovariance function (section 3.1). Fortunately, safer conclusions can be obtained from the annual rainfall record at Hylike, which extends over a 100-year period (1907-

2008). Fig. 4 illustrates the corresponding empirical autocorrelogram, i.e. the annual autocorrelation coefficients ρ_j for time lags up to j = 50 years. This exhibits a significantly long tail, since most of the empirical autocorrelation coefficients retain particularly high values in the long run (with many of them being higher than the lag-1 value, $\rho_1 = 0.103$). Also, four theoretical autocorrelograms are presented, derived by different formulations of eq. (1). The first three were estimated by setting the scale parameter of eq. (1) equal to $\beta = 0.0$, 2.0, and 4.0, thus representing low (ARMA-type), moderate and high persistence, respectively, while the shape parameter κ was analytically computed to preserve the lag-1 autocorrelation of the observed rainfall. For $\beta = 0.0$, 2.0, and 4.0, the corresponding values of κ were 2.3, 46.9, and 2248.0, respectively. The last theoretical autocorrelogram was estimated via calibration, i.e. by fitting the theoretical against the empirical autocorrelation coefficients; is that case we obtained $\beta = 3.55$ and $\kappa = 100.8$. Following a similar approach for all variables, we examined the relationship between the parameters of the generalized autocovariance function (1) and the resulting Hurst coefficient, H. The outcomes of this analysis are summarized in Table 2.

Table 2 Simulated Hurst coefficients, estimated from synthetic series by the algorithm given by Koutsoyiannis (2003), at the six locations of interest, for different formulations of the generalized autocovariance function.

	Low persistence ($\beta = 0.0$)	Moderate persistence ($\beta = 2.0$)	High persistence ($\beta = 4.0$)	Fitted to empirical autocorrelograms of historical data
Evinos rain	0.55	0.63	0.69	0.56
Evinos runoff	0.59	0.72	0.76	0.62
Mornos rain	0.58	0.68	0.69	0.66
Mornos runoff	0.59	0.70	0.73	0.65
Hylike rain	0.56	0.57	0.67	0.74
Hylike runoff	0.59	0.69	0.76	0.65

Table 3 Comparison of annual statistical characteristics for all modelled variables..

		Mean (mm)	St. deviation (mm)	Skewness	Lag-1 autocorrelation
Evinos rain	Historical	1220.8	283.9	-0.487	0.172
	Synthetic	1231.0	272.4	-0.470	0.122
Evinos runoff	Historical	785.6	230.4	-0.267	0.315
	Synthetic	810.6	223.8	-0.359	0.272
Mornos rain	Historical	934.5	205.6	0.694	0.247
	Synthetic	944.9	195.5	0.533	0.201
Mornos runoff	Historical	408.2	144.7	-0.063	0.320
	Synthetic	418.4	137.0	-0.164	0.250
Hylike rain	Historical	653.6	158.3	0.509	0.095
	Synthetic	647.7	157.1	0.547	0.062
Hylike runoff	Historical	126.0	55.6	0.194	0.297
	Synthetic	128.1	52.1	0.249	0.248

Accepting that the empirical autocorrelogram of the annual rainfall at Hylike is relatively reliable, and thus representative of the scaling behaviour of the associated process, a suitable value of parameter β should be around 4.0. However, similarly safe conclusions cannot be extracted for the remaining processes, as the corresponding historical records are not sufficiently long. On the other hand, employing such a high value of β would probably result in too conservative estimations, with respect to the performance of the water resource system under study (in terms of reliability, cost, etc.). Therefore, in the following simulations we decided to assign a moderate value of $\beta = 2.0$ to all variables and fit parameter κ to the corresponding lag-1 autocorrelation, the estimation of which is relatively safer. We remark that in the case of small samples, significant bias and uncertainty is introduced in the estimation of autocorrelations as the lag increases, manifested in random fluctuations of the empirical autocorrelation coefficients (e.g., alternations between negative and positive values). For this reason, for lags greater than one, the annual synthetic data are forced to reproduce the theoretical autocorrelations derived by eq. (1), which are statistically consistent (their values decrease monotonically according to an appropriate model), and not the historical ones. The corresponding Hurst values are around 0.60 for rainfall and 0.70 for runoff. In all cases but one (Hylike rainfall) these are somewhat greater than the ones obtained when the theoretical autocorrelogram is fitted to the empirical one (Table 2, last column). However, due to the presense of negative coefficients in the empirical

autocorrelogram, this fitting approach is expected to provide underestimated values of parameter β, which in turn leads to underestimated Hurst coefficients. This further justifies the proposed formulation of the theoretical autocorrelation function.

The performance of the monthly module of Castalia is evaluated by comparing the statistical characteristics of simulated data to the historical ones. We remark that in the case of stochastic models, evaluations can only be made on statistical grounds, in contrast to typical quantitative evaluations employed in environmental modelling, which are based on comparisons of simulated data against observations (cf. Bennett et al., 2013). We also note that even the accuracy of the simulated statistical characteristics depends on the length of synthetic data. For infinite simulation horizons, the statistical characteristics of the synthetic time series should be identical to the historical ones, provided that the theoretical equations of the model are built to preserve the desirable statistics of the parent time series. Some inaccuracies may also emerge when the estimation of some parameters is made through numerical approaches, in the absence of analytical ones. The unavoidable errors of numerical methods are reflected in the representation of the statistical characteristics.

Under this premise, we compare the observed and simulated annual mean, standard deviation, skewness coefficient and lag-1 autocorrelation at every location, which are given at Table 3. The same characteristics are also illustrated, by means of monthly diagrams, for the rainfall at Hylike (the station with the largest sample) and the runoff at the wetest basin, i.e. upstream of Evinos dam (Figs. 5 and 6, respectively). In Fig. 7 the monthly cross-correlations are also compared, for several pairs of variables. It is demonstrated that the modelling scheme preserves with very satisfactory accuracy all essential statistical properties for both the annual and the monthly time scales; this also involves characteristics that are not explicitly represented in the theoretical equations of the model, such as the skewness and cross-correlations. Yet, in few cases, the preservation of the above statistics is less satisfactory. To remedy this, we should go back to the numerical routines embedded in Castalia (e.g., the optimization procedure for covariance matrix decomposition) and assign more strict convergence and termination criteria.

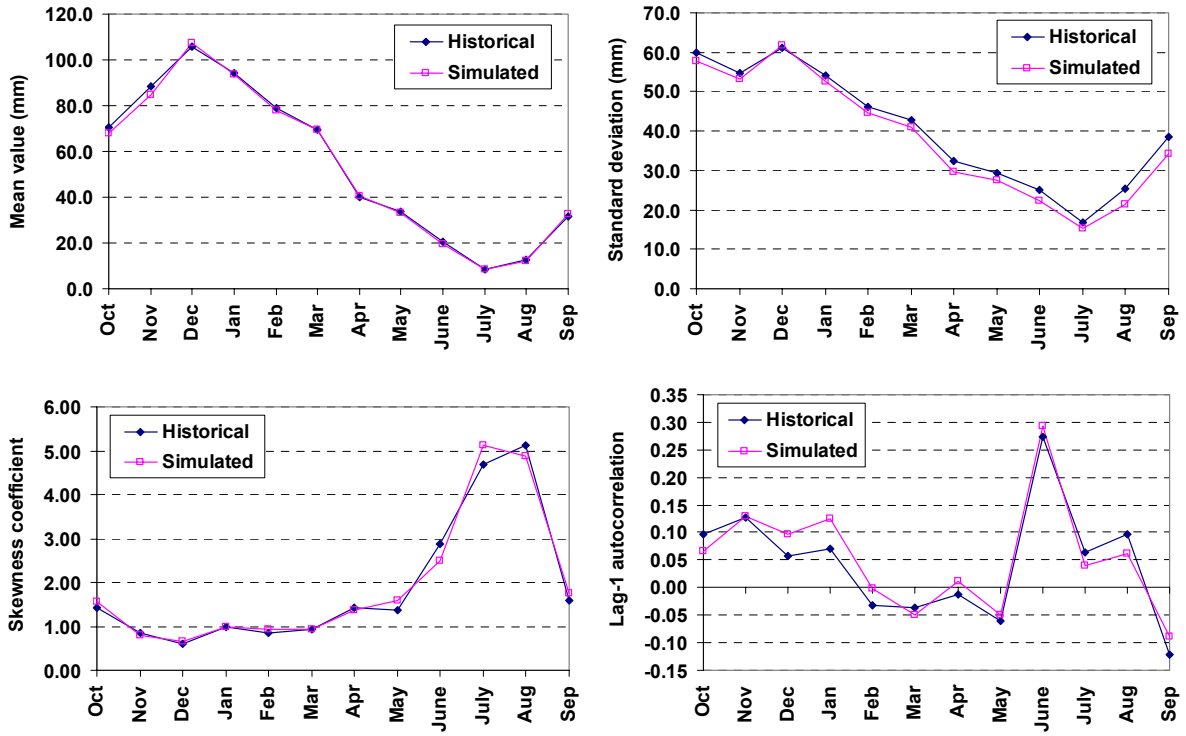

Figure 5 Comparison of monthly statistical characteristics of rainfall at Hylike.

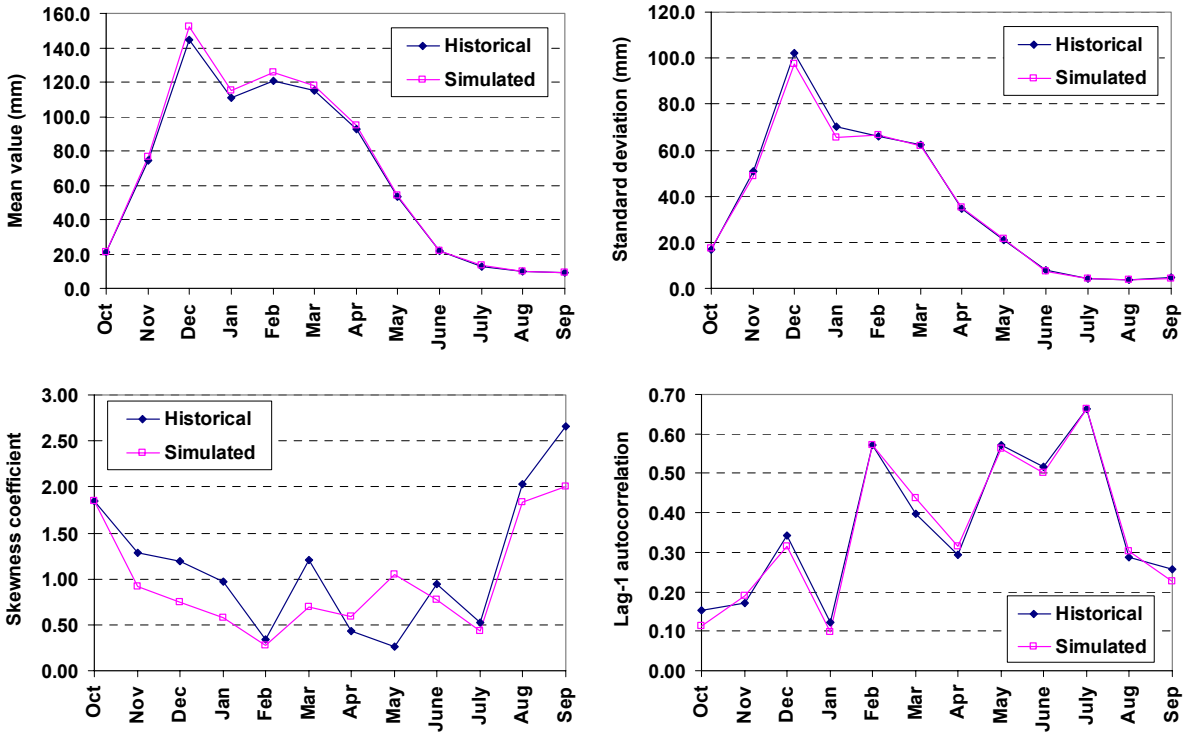

Figure 6 Comparison of monthly statistical characteristics of runoff at Evinos.

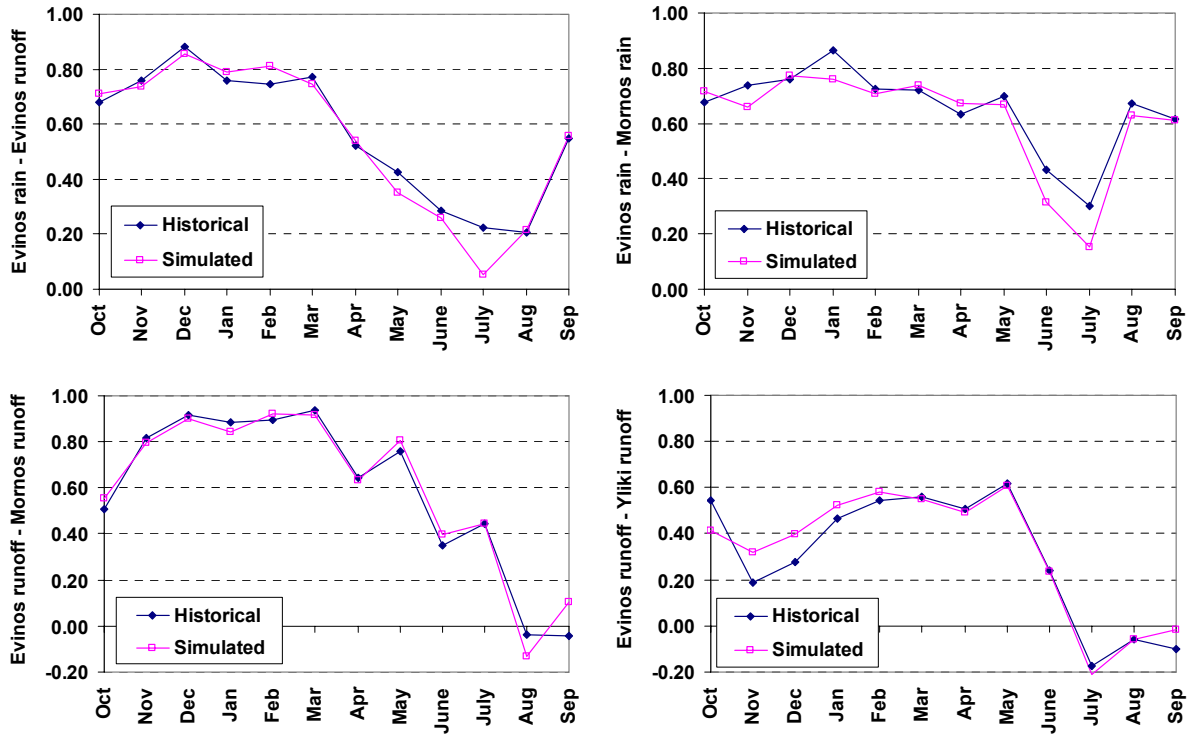

Figure 7 Comparison of monthly cross-correlations for several pairs of variables.

6.2 Generation of daily rainfall for flood simulation

In the second case study we tested the newer version of Castalia for the stochastic simulation of daily rainfall at five (out of the 13 available) gauges located at the Boeoticos Kephisos river basin, in Eastern Greece. The basin, which extends over an area of 1930 km^2, is also part of the water supply system of Athens, since it discharges into Lake Hylike. The locations of the stations are shown Fig. 3. The historical records cover a common observation period of 42 years (1964-2006). The mean annual rainfall over all stations ranges from 500 to 750 mm. Synthetic data are necessary for the estimation of the proper representation of the areal rainfall at the sub-basin scale, which can constitute inputs to semi-distributed flood models at the specific basin.

The three-level multivariate disaggregation scheme was applied for generating 1000 years of daily rainfall, at the five gauges (Dialynas, 2011). In the context of model configurations, we assigned the following values: for the power transformation of daily variables (eq. 10), we applied the exponent $n = 0.8$; for the rounding off rule of section 5.3.2, we assumed $\pi_0 = 0.90$ and $l_0 = 0.30$; finally, for the Markovian model of section 5.3.3, we set $\lambda = 0.23$ and $\xi = 0.55$ (the

basis on parameter selection is explained in section 5.3.4), and we also applied an autocorrelation adjusting factor equal to 1.25.

Table 4 Comparison of daily statistical characteristics of the five rain gauges in December.

		Pavlos	Drymea	Atalanti	Livadia	Tithorea
Mean (mm)	Historical	2.53	2.78	2.57	4.05	3.31
	Synthetic	2.49	2.75	2.54	3.96	3.29
St. deviation (mm)	Historical	7.68	8.18	7.16	9.89	8.7
	Synthetic	8.08	9.01	7.5	10.73	9.28
Skewness	Historical	5.6	5.2	4.6	3.9	4.3
	Synthetic	7.4	6.4	5.9	5.2	5.3
Lag 1 autocorrelation	Historical	0.15	0.28	0.22	0.27	0.24
	Synthetic	0.22	0.29	0.21	0.27	0.26
Probability dry	Historical	0.74	0.73	0.74	0.65	0.67
	Synthetic	0.69	0.71	0.69	0.65	0.66

Table 5 Comparison of daily statistical characteristics of the five rain gauges in July..

		Pavlos	Drymea	Atalanti	Livadia	Tithorea
Mean (mm)	Historical	0.29	0.63	0.46	0.58	0.62
	Synthetic	0.33	0.65	0.48	0.64	0.62
St. deviation (mm)	Historical	2.52	3.85	3.17	4.4	3.76
	Synthetic	2.41	4.25	3.33	4.68	3.62
Skewness	Historical	13.8	9.4	9.5	14	9.2
	Synthetic	12.5	10.3	12.7	12.7	9.7
Lag 1 autocorrelation	Historical	0.077	0.039	0.08	0.022	0.086
	Synthetic	0.056	0.016	0.045	0.033	0.066
Probability dry	Historical	0.96	0.94	0.95	0.93	0.92
	Synthetic	0.94	0.93	0.93	0.94	0.91

The evaluation of the model performance is implemented for the lowest level of simulation, i.e. the daily scale. In this scale, apart from the reproduction of the essential statistical charateristics of the synthetic time series (moments up to third order, auto- and cross-correlations, and probability dry), emphasis is also given to the statistical regime of the extremes. In this context, we first compare the statistical characteristics of the simulated time series against the historical ones, for the wettest (December) and driest (July) month of the year, which are shown

in Tables 4 and 5, respectively. In general, Castalia reproduces with considerable accuracy the statistical behaviour of the observed data, even during summer months that are characterized by particularly high coefficients of skewness (~10 to 15) and significantly high percentages of dry days (~0.90 to 0.95), In Tables 6 and 7 the cross-correlation coefficients of the two months of interest are compared, which are estimated both on the basis of the power-transformed (through eq. 10), and the raw data series, respectively. Even though the model is built to merely preserve the "theoretical" cross-correlations (the estimation of which is explained in section 3), the deviations appearing in Tables 6 and 7 are rather minor, which also indicates the suitability of the implemented decomposition approach for the variance-covariance matrices.

Table 6 Cross-correlation coefficients of power-transformed daily data for December (upper-diagonal table) and July (lower-diagonal table); in each triplet of rows the first shows the historical values, the second the "theoretical" (see definition in section 3.3) and the third the synthetic ones.

	Pavlos	Drymea	Atalanti	Livadia	Tithorea
Pavlos		0.40 0.40 0.39	0.39 0.38 0.43	0.47 0.47 0.48	0.54 0.52 0.54
Drymea	0.38 0.38 0.39		0.38 0.39 0.43	0.46 0.47 0.48	0.57 0.58 0.56
Atalanti	0.31 0.31 0.35	0.40 0.40 0.57		0.42 0.43 0.46	0.43 0.43 0.46
Livadia	0.53 0.52 0.46	0.40 0.41 0.45	0.32 0.33 0.37		0.64 0.64 0.60
Tithorea	0.37 0.38 0.35	0.46 0.44 0.50	0.45 0.46 0.50	0.50 0.48 0.45	

Table 7 Cross-correlation coefficients of daily data for December (upper-diagonal table) and July (lower-diagonal table); in each row the first value is estimated from the the historical data and the second from the simulated ones.

	Pavlos	Drymea	Atalanti	Livadia	Tithorea
Pavlos		0.39/ 0.33	0.37 / 0.40	0.46 / 0.48	0.52 / 0.48
Drymea	0.40 / 0.34		0.34 / 0.33	0.45 / 0.45	0.56 / 0.53
Atalanti	0.28 / 0.33	0.37 / 0.35		0.39 / 0.46	0.40 / 0.41
Livadia	0.54 / 0.52	0.42 / 0.38	0.27 / 0.27		0.63 / 0.58
Tithorea	0.34 / 0.39	0.43 / 0.42	0.47 / 0.43	0.48 / 0.38	

The average values of historical and synthetic annual daily maxima are also reported in Table 8 (even though these values are not directly comparable), where there seems to be a slight yet systematic overestimation of the average synthetic maxima. Moreover, in order to investigate the statistical behaviour of the simulated annual daily maxima and evaluate the ability of Castalia to reproduce extreme rainfall events, we fit the Generalized Extreme Value (GEV) distribution to daily maxima, as illustrated in the example of Fig. 8 (rainfall at Pavlos station), which constitutes direct graphical output of Castalia. The GEV distribution has three parameters, i.e. k (shape), λ (scale) and ψ (position), which are estimated via the L-moments method. The parameter values for the historical and synthetic samples of daily maxima are compared in Table 8. Once again we emphasize that the accuracy of the estimated parameters substantially differs for historical and synthetic data (larger samples offer more confident estimations, as the variance of estimators decreases with sample size), and thus comparisons should be carefully interpreted. Nevertheless, as shown in Table 8, the values of the three parameters are quite close for most stations, although the stochastic model does not explicitly reproduce the statistical characteristics of the extremes, as quantified through the GEV parameters. Interestingly, the shape parameter k of synthetic maxima, which determines the tail of the distribution, is always positive (suggesting distribution unbounded from above), while for two stations (Drymea and Atalanti) its analytical computation on the grounds of historical daily maxima results in negative values. However, negative sample estimates of k can be expected, as indicated by Papalexiou and Koutsoyiannis (2013), who analyzed numerous large-size samples of rainfall data globally. Therefore, the statistical regimes of the synthetic daily maxima seem to be consistent with the historical ones, which constitutes another important advantage of the daily generation scheme.

Table 8 Mean annual maxima and parameters k, λ, ψ of the GEV distribution.

		Pavlos	Drymea	Atalanti	Livadia	Tithorea
Mean annual maximum (mm)	Historical	58.8	65.0	54.6	69.3	59.4
	Synthetic	62.5	74.1	61.1	77.6	68.0
Shape parameter k	Historical	0.084	-0.389	-0.042	0.144	0.082
	Synthetic	0.104	0.036	0.030	0.030	0.081
Scale parameter λ	Historical	13.409	22.565	16.751	17.743	15.034
	Synthetic	18.308	21.556	18.491	22.367	19.009
Position parameter ψ	Historical	3.721	2.594	2.721	3.162	3.288
	Synthetic	2.723	2.826	2.696	2.860	2.913

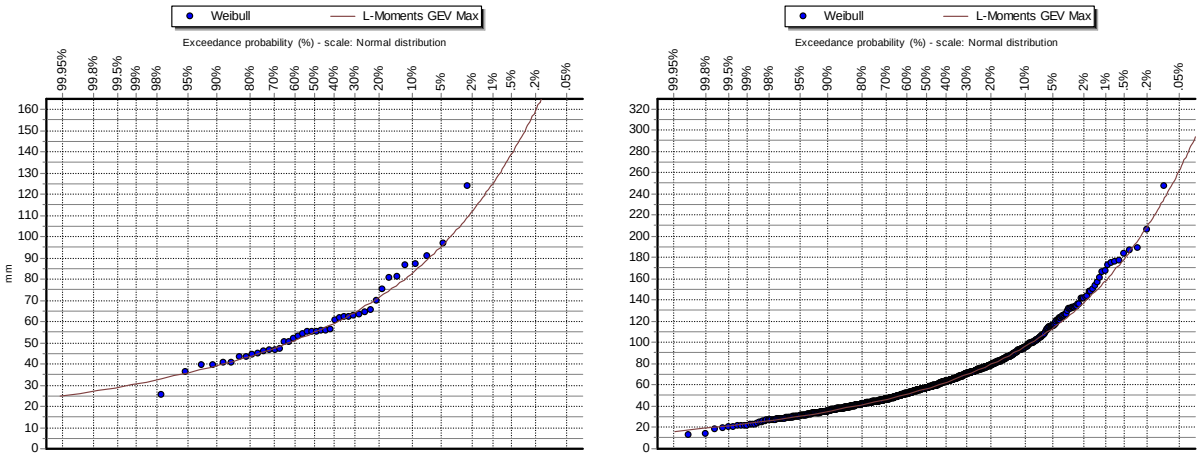

Figure 8 Fitting of GEV distribution to historical (left) and synthetic (right) daily rainfall maxima at Pavlos.

7 Summary and discussion

Castalia employs a stochastic simulation framework for generating synthetic time series at multiple locations and at three time scales (daily, monthly, and annual). This constitutes a synthesis of several individual methodologies, combining analytical and numerical procedures that allow preserving simultaneously all important statistical characteristics of the observed data, at all time scales. The two case studies dealing with different simulation problems, illustrated the model advantages, as well as the software capacities, e.g. its flexibility on representing a wide range of autocovariance structures, and thus respecting the Hurst-Kolmogorov behaviour of the annual stochastic processes, in addition to preserving high proportions of dry intervals and

representing GEV-distributed maxima at the daily scale. However, they also revealed some issues that require further investigation, such as: (a) the ability to handle variables with fat (subexponential) distributions (i.e., tails approaching zero less rapidly than an exponential tail); (b) the improvement of the parameter estimation procedures by taking into account the statistical bias and uncertainty in estimation; and (c) the automatic tuning of several algorithmic parameters, particularly within the daily simulation procedures. With respect to the latter issue, it is recalled that Castalia, in addition to autocovariance and cross-covariance parameters, also requires the estimation of several other parameters and algorithmic constants, to be implemented within hybrid or numerical procedures. All these are directly or indirectly associated with the preservation of the statistical characteristics that are not explicitly reproduced by stochastic models. Apparently, the performance of these parameters should be systematically evaluated in a wide range of problems, in order to provide generalized guidance for their estimation. In our opinion, a full automatization is not possible, since most of associated routines are iterative, thus it is impossible to make *a priori* estimations on the basis of the available information. Yet, some of these routines could be improved by embedding hybrid, self-tuning rules that are based on the additional information gained during the trials.

Another possible improvement is related to the number of variables that can be represented in multivariate stochastic simulations. If this number is large, numerical problems that emerge are related to a large number of parameters to be estimated, combined with the error accumulation issues (e.g., Kottegoda et al., 2003). For instance, at the decomposition of the variance-covariance matrix, which is a common procedure at all time scales, a nonlinear optimization problem of $m \times m$ control variables is to be solved, which is characterized by significant sensitivity of the objective function against its parameters as well as the existence of many local optima. Efstratiadis (2001) found that well-recognized evolutionary optimization algorithms were easily trapped to local optima, which in turn resulted to improper representation of the associated statistical characteristics (i.e., skewness and cross-correlations). At the daily time scale, additional perplexity arises due to the implementation of hybrid approaches to preserve the probability of dry historical data, which may introduce significant bias to the rest of statistical properties. Nevertheless, it is not possible to set a specific upper limit for the number of modelled variables, since this depends on each specific hydrological problem. In our limited experience with daily rainfall simulation, the use of more than 7 or 8 variables may not ensure satisfactory

preservation of the statistical characteristics of historical data, especially when the latter exhibit considerably high proportions of dry intervals; nonetheless up to 20 variables have been simulated by Ilich (2014) using a different approach. At any rate, if the number of variables of interest is quite high, the multivariate module can be individually applied to each subset, and thus indirectly represent the total set of variables, as practiced e.g., in LAST (Lane and Frevert, 1990), SPIGOT (Grygier and Stedinger, 1990), and SAMS (Sveinsson et al., 2003; Salas et al., 2006). Even though this implies that cross-correlations among different datasets are ignored, this is preferable over employing single-site simulations, since simultaneous preservation of cross-correlations at the three time scales of interest leads to more sufficient representation of hydrological reality. In this case, cross-correlations between variable subsets could either be considered unimportant and thus, be neglected, or be treated by alternative approaches. For instance, Koutsoyiannis et al. (2008) applied a method based on maximization of joint entropy to optimally estimate unknown cross-correlation values for a multivariate stochastic model.

The first case study (section 6.1) highlights the significant uncertainty on the estimation of parameters β and κ, which are associated with the generalized autocovariance function and identify, in turn, the Hurst coefficient of the simulated time series. Theoretically, if large data records were available, of 100 years as an order of magnitude, the obvious and most reliable practice would be the calibration of these parameters against the empirical autocorrelation coefficients. However, the majority of hydrometeorological data sets do not exceed few decades, thus making the estimation of historical autocorrelations (apart from few time lags) unreliable; as already mentioned, the fewer the historical data the more uncertain the autocorrelation estimation is. In this context, it is essential to employ systematic analyses using a few in number, yet large in size samples, in order to provide representative regional values for the Hurst coefficients, over broader regions, for all typical hydrometeorological variables. The outcomes of such analysis would allow formulating realistic autocovariance functions, and consequently generate synthetic time series that properly reproduce the Hurst-Kolmogorov behaviour of the associated processes.

Until now, Castalia was mainly applied for generating monthly rainfall and runoff, as well as daily rainfall, in the context of water resources management and flood modelling studies, respectively. However, a wider spectrum of hydrometeorological variables can be simulated by Castalia (Venediki et al., 2013). Tsekouras and Koutsoyiannis (2014) produced synthetic time

series of daily wind speed and sunshine duration time series, which are essential in renewable energy studies. Next research steps can be the investigation of the model performance against other hydrometeorological variables, particularly at the daily time scale. This may require some improvements in order to efficiently represent the peculiarities of each individual process (e.g., baseflow characteristics and snowmenlt effects of daily discharge, negative values of daily temperature, etc.). Regarding streamflow generation, our preliminary investigations in wet basins with permanent runoff showed that Castalia generated realistic patterns of daily discharge, due to the preservation of lag-1 autocorrelations and cross-correlations with rainfall. However, to ensure the generation of physically consistent streamflow data under any hydroclimatic regime, it is essential to account for complex interactions throughout the rainfall-runoff transformation, including both physical regulations (due to soil moisture storage, snow accumulation, flow routing, etc.) and man-made interventions (e.g., abstractions). In this context, Efstratiadis et al. (2014) proposed a nonlinear stochastic framework, comprising effective coupling of Castalia, for the generation of meteorological inputs, deterministic hydrological models, as well as stochastic error models to represent structural and parameter uncertainties.

A further research objective is the implementation of a fourth temporal level, e.g. hourly, which is more convenient for flood simulations. We expect that, by employing the robust disaggregation methodology of section 4.2, it is rather straightforward to couple fine-scale stochastic models, such as the one proposed by Koutsoyiannis et al. (2003a), within higher-level simulation schemes. While some models include weekly and seasonal time scales, following a disaggregation methodology, in order to explicitly preserve weekly or seasonal statistical properties, we do not see the reason to incorporate such scales or intermediate ones; by choosing key time scales, a disaggregation framework ensures preservation of statistics at these scales and one can expect that the statistics at intermediate scales (in between two explicitly considered in the disaggregation) would be appropriately interpolated if the finer scale data are aggregated to any desired time scale.

Acknowledgments

The authors would like to thank Dr. N. Ilich and three anonymous reviewers for their constructive comments. The research leading to this paper was partly funded by the Greek General Secretariat

for Research and Technology through the research project Combined REnewable Systems for Sustainable ENergy DevelOpment (CRESSENDO; grant number 5145).

References

Barnes, F.B., 1954. Storage required for a city water supply, 1954. J. Inst. Eng. Australia 26(9), 198–203.

Beard, L.R., 1965. Use of interrelated records to simulate streamflow, Proc. ASCE, J. Hydraul. Div. 91(HY5), 13–22.

Bennett N.D., Croke B.F.W., Guariso G., Guillaume J.H.A., Hamilton S.H., Jakeman A.J., Marsili-Libelli S., Newham L.T.H., Norton J.P., Perrin C., Pierce S.A., Robson B., Seppelt R., Voinov A.A., Fath B.D. and Andreassian V., 2013. Characterising performance of environmental models. Environmental Modelling and Software 40, 1–20.

Boughton, W.C., Droop, O.P., 2003. Continuous simulation for design flood estimation – a review. Environmental Modelling and Software 18(4), 309–318.

Box, G. E., Jenkins, G. M., 1970. Time Series Analysis: Forecasting and Control. Holden Day.

Bras, R.L., Rodriguez-Iturbe, I., 1985. Random Functions in Hydrology. Addison-Wesley, Reading, MA.

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. Environmental Modelling and Software 23(12), 1356–1369.

Dialynas, Y., 2011. A computer system for the multivariate stochastic disaggregation of monthly into daily hydrological time series. Diploma thesis, 337 pp., Dept. of Water Resources & Environmental Engineering, National Technical University of Athens, Athens (in Greek; http://itia.ntua.gr/1142/).

Dialynas, Y., Kozanis, S., Koutsoyiannis, D., 2011. A computer system for the stochastic disaggregation of monthly into daily hydrological time series as part of a three–level multivariate scheme. European Geosciences Union General Assembly 2011, Geophysical Research Abstracts, Vol. 13. Vienna (http://itia.ntua.gr/1137/).

Efstratiadis, A., 2001. Investigation of global optimum seeking methods in water resources problems, MSc thesis, 139 pages, Dept. of Water Resources, Hydraulic & Maritime Engineering, National Technical University of Athens, Athens (in Greek; extended abstract in English available at http://itia.ntua.gr/446/).

Efstratiadis, A., Koutsoyiannis, D., 2004. Castalia (version 2.0) – A system for stochastic simulation of hydrological variables. Modernisation of the supervision and management of the water resource system of Athens, Report 23, Dept. of Water Resources, Hydraulic & Maritime Engineering, National Technical University of Athens, Athens (in Greek; http://itia.ntua.gr/619/).

Efstratiadis, A., Koutsoyiannis, D., Xenos, D., 2004. Minimising water cost in the water resource management of Athens. Urban Water Journal 1(1), 3–15.

Efstratiadis, A., Nalbantis, I., Koutsoyiannis, D., 2014. Hydrological modelling of temporally-varying catchments: Facets of change and the value of information, Hydrol. Sci. J. (submitted).

Grygier, J.C., Stedinger, J.R., 1990. SPIGOT: A synthetic streamflow generation software package, Version 2.5, Technical description. Cornell Univ., Ithaca, New York.

Hemelrijk, J., 1966. Underlining random variables. Statistica Neerlandica, 20(1).

Ilich, N., 2014. An effective three-step algorithm for multi-site generation of weekly stochastic hydrologic time series. Hydrol. Sci. J., 59(1), 85–98.

Kilsby, C.G., Jones, P.D., Burton, A., Ford, A.C., Fowler, H.J., Harpham, C., James, P., Smith, A., Wilby, R.L., 2007. A daily weather generator for use in climate change studies, Environmental Modelling & Software, 22(12), 1705–1719.

Kottegoda, N. T., Natale, L., Raiteri, E., 2003. A parsimonious approach to stochastic multisite modelling and disaggregation of daily rainfall. J. Hydrol. 274, 47–61.

Koutsoyiannis, D., 1988. A disaggregation model of point rainfall. PhD thesis, 310 p., National Technical University of Athens, Athens (in Greek).

Koutsoyiannis, D., 1994. A stochastic disaggregation method for design storm and flood synthesis. J. Hydrol. 156, 193–225.

Koutsoyiannis, D., 1999. Optimal decomposition of covariance matrices for multivariate stochastic models in hydrology. Water Resour. Res. 35(4), 1219–1229.

Koutsoyiannis, D., 2000. A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series, Water Resour. Res. 36(6), 1519–1534.

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

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

Koutsoyiannis, D., 2003. Climate change, the Hurst phenomenon, and hydrological statistics. Hydrol. Sci. J. 48(1), 3–24.

Koutsoyiannis, D., 2011. Hurst-Kolmogorov dynamics and uncertainty. J. Amer. Water Res. Assoc. 47(3), 481–495.

Koutsoyiannis, D., Manetas, A., 1996. Simple disaggregation by accurate adjusting procedures. Water Resour. Res. 32(7), 2105–2117.

Koutsoyiannis, D. Montanari, A., 2007. Statistical analysis of hydroclimatic time series: uncertainty and insights. Water Resour. Res. 43(5), W05429.

Koutsoyiannis, D., Onof, C., Wheater, H.S., 2003a. Multivariate rainfall disaggregation at a fine time scale. Water Resour. Res. 39(7), 1173.

Koutsoyiannis, D., Karavokiros, G., Efstratiadis, A., Mamassis, N., Koukouvinos, A., Christofides, A., 2003b. A decision support system for the management of the water resource system of Athens. Phys. Chem. Earth 28 (14-15), 599–609.

Koutsoyiannis, D., Yao, H., Georgakakos, A., 2008. Medium-range flow prediction for the Nile: a comparison of stochastic and deterministic methods. Hydrol. Sic. J. 53(1), 142–164.

Lane, W.L., Frevert, D.K., 1990. Applied stochastic techniques, personal computer version 5.2, user's manual. Earth Sciences Division, United States Bureau of Reclamation. Denver.

Langousis, A., Koutsoyiannis, D., 2005. A stochastic methodology for generation of seasonal time series reproducing overyear scaling behaviour. J. Hydrol. 322, 138–154.

Maass, A., Hufschmidt, M.M., Dorfman, R., Thomas, H.A., Marglin, S.A., Fair, J.M., 1962. Design of Water Resource Systems. Harvard University Press, Cambridge, Mass.

Matalas, N.C., 1967. Mathematical assessment of synthetic hydrology. Water Resour. Res. 3(4), 937–945.

Matalas, N.C., Wallis, J.R., 1976. Generation of synthetic flow sequences. Systems Approach to Water Management, edited by Biswas, A.K., McGraw-Hill, New York.

McLeod, A.I., Hipel, K.W., 1978. Simulation procedures for Box-Jenkins models. Water Resour. Res. 14, 969–975.

Nalbantis, I., Efstratiadis, A., Rozos, E., Kopsiafti, M., Koutsoyiannis, D., 2011. Holistic versus monomeric strategies for hydrological modelling of human-modified hydrosystems. Hydrol. Earth Sys. Sci. 15, 743–758.

Papalexiou, S.M., Koutsoyiannis, D., 2013. Battle of extreme value distributions: A global survey on extreme daily rainfall, Water Resour. Res. 49(1), 187–201.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. Numerical Recipes in C. Cambridge Univ. Press, New York.

Richardson, C.W., 1981. Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resour. Res. 17, 182–190.

Richardson, C.W., Wright, D.A., 1984. WGEN: A model for generating daily weather variables. USDA Publication ARS-8.

Salas, J.D., 1993. Analysis and modeling of hydrologic time series. Chapter 19, Handbook of Hydrology, edited by Maidment, D., McGraw-Hill, New York.

Salas, J.D., Delleur, J.W., Yevjevich, V., Lane, W.L., 1980. Applied Modeling of Hydrologic Time Series. Water Resources Publications. Littleton, Colorado.

Salas J.D., Sveinsson O.G., Lane W.L., Frevert D.K., 2006. Stochastic streamflow simulation using SAMS-2003. J Irrig. Drain. Eng. 132(2), 112–123.

Srinivas, V.V., Srinivasan, K., 2005. Hybrid moving block bootstrap for stochastic simulation of multi-site multi-season streamflows. J. Hydrol., 302(1-4), 307–330.

Srivastav, R.K., Simonovic, S.P., 2014. An analytical procedure for multi-site, multi-season streamflow generation using maximum entropy bootstrapping. Environmental Modelling & Software, 59, 59–75.

Sveinsson O.G., Salas J.D., Lane W.L., Frevert D.K., 2003. Progress in Stochastic Analysis Modeling and Simulation: SAMS-2003. 23rd Annual American Geophysical Union Hydrology Days, Fort Collins, Colorado State University.

Thomas, H.A., Fiering, M.B., 1962. Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. Design of Water Resource Systems, edited by Maass A., Hufschmidt M.M., Dorfman R., Thomas H.A., Marglin S.A., Fair J.M. Harvard University Press, Cambridge, Mass.

Todini, E., 1980. The preservation of skewness in linear disaggregation schemes. J. Hydrol., 47, 199–214.

Tsekouras, G., Koutsoyiannis, D., 2014. Stochastic analysis and simulation of hydrometeorological processes associated with wind and solar energy. Renewable Energy, 63, 624–633.

USACE (U.S. Army Corp of Engineers), 1971. HEC-4 Monthly Streamflow Simulation. Hydrologic Engineering Center. Davis, California.

Valencia, D., Schaake, J.C., 1973. Disaggregation processes in stochastic hydrology. Water Resour. Res. 9(3), 211–219.

Venediki, A., Giannoulis, S., Ioannou, C., Malatesta, L., Theodoropoulos, G., Tsekouras, G., Dialynas, Y., Papalexiou, S.M., Efstratiadis, A., Koutsoyiannis, D., 2013. The Castalia stochastic generator and its applications to multivariate disaggregation of hydro-meteorological processes. European Geosciences Union General Assembly 2013, Geophysical Research Abstracts, Vol. 15. Vienna (http://itia.ntua.gr/1325/).

Verhoest, N.E.C., Vandenberghe, S., Cabus, P., Onof, C., Meca-Figueras, T., Jameleddine, S., 2010. Are stochastic point rainfall models able to preserve extreme flood statistics? Hydrol. Processes 24(23), 3439–3445.

Wilks, D.S., 1998. Multisite generalization of a daily stochastic precipitation generation model. J. Hydrol. 210, 178–191.