

# Simple stochastic simulation of time irreversible and reversible processes

Demetris Koutsoyiannis<sup>1,2</sup>

<sup>1</sup> School of Civil Engineering, National Technical University of Athens, Zographou, Greece

<sup>2</sup> Visiting Professor, Department of Civil, Constructional and Environmental Engineering, Sapienza University of Rome  
(dk@itia.ntua.gr)

**Citation:** Koutsoyiannis, D., 2020. Simple stochastic simulation of time irreversible and reversible processes, *Hydrological Sciences Journal*, DOI: [10.1080/02626667.2019.1705302](https://doi.org/10.1080/02626667.2019.1705302).

**Abstract** As time irreversibility of streamflow is marked for time scales up to several days, while common stochastic generation methods are good only for time symmetric processes, the need for new methods to handle irreversibility, particularly in flood simulations, has been recently highlighted. From an investigation of the historical evolution of existing stochastic generation methods, which is a useful step before proposing new methods, the strengths and weaknesses of current approaches are located. Following this investigation, a generic solution to the stochastic generation problem is proposed. This is an analytical exact method based on an asymmetric moving average scheme, capable of handling time irreversibility in addition to preserving the second-order stochastic structure, as well as higher-order marginal statistics, of a process. The method is studied theoretically in its general setting, as well as in its most interesting special cases, and is successfully applied to streamflow generation at hourly scale.

**Keywords** Time's arrow; Irreversibility; Stochastic simulation; Time series; Monte Carlo method

## 1 Introduction

*Simplicity is the final achievement.*

Frédéric Chopin

A recent study (Koutsoyiannis 2019b) has investigated the possible existence of time irreversibility in hydrometeorological processes and the related theoretical framework. According to the latter, a simple definition of time reversibility within stochastics is the following, where underlined symbols denote stochastic (random) variables and non-underlined ones denote values thereof or regular variables.

A stochastic process  $\underline{x}(t)$  at continuous time  $t$ , with  $n$ th order distribution function:

$$F(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) := P\{\underline{x}(t_1) \leq x_1, \underline{x}(t_2) \leq x_2, \dots, \underline{x}(t_n) \leq x_n\} \quad (1)$$

is time-symmetric or time-reversible if its joint distribution does not change after reflection of time about the origin, i.e., if for any  $n, t_1, t_2, \dots, t_n$ ,

$$F(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = F(x_1, x_2, \dots, x_n; -t_1, -t_2, \dots, -t_n) \quad (2)$$

If times  $t_i$  are equidistant, i.e.  $t_i - t_{i-1} = D$ , the definition can be also written by reflecting the order of points in time, i.e.:

$$F(x_1, x_2, \dots, x_{n-1}, x_n; t_1, t_2, \dots, t_{n-1}, t_n) = F(x_1, x_2, \dots, x_{n-1}, x_n; t_n, t_{n-1}, \dots, t_2, t_1) \quad (3)$$

A process that is not time-reversible is called time-asymmetric, time-irreversible or time-directional. Important results related to time (ir)reversibility are the following:

- A time reversible process is also stationary (Lawrance, 1991).
- If a scalar process  $\underline{x}(t)$  is Gaussian (i.e., all its finite dimensional distributions are multivariate normal) then it is reversible (Weiss, 1975). The consequences are: (a) a directional process cannot be Gaussian; (b) a discrete-time ARMA process (and a continuous-time Markov process) is reversible if and only if it is Gaussian.
- However, a vector (multivariate) process can be Gaussian and irreversible at the same time. A multivariate Gaussian linear process is reversible if and only if its autocovariance matrices are all symmetric (Tong and Zhang, 2005).

Time asymmetry of a process can be studied through the differenced process, i.e.:

$$\tilde{\underline{x}}_\tau := \underline{x}_\tau - \underline{x}_{\tau-1} \quad (4)$$

where  $\underline{x}_\tau$  is the process representation in discrete time  $\tau$  (denoting the continuous-time interval  $[(\tau - 1)D, \tau D]$ ), i.e.:

$$\underline{x}_\tau := \frac{1}{D} \int_{(\tau-1)D}^{\tau D} \underline{x}(u) du \quad (5)$$

For a stationary stochastic process  $\underline{x}(t)$ , the differenced process  $\tilde{\underline{x}}_\tau$  has mean zero and variance:

$$\tilde{\gamma}_1 := \text{var}[\tilde{\underline{x}}_\tau] = 2(\gamma_1 - c_1) \quad (6)$$

where  $\gamma_1$  and  $c_1$  are the variance and lag-one autocovariance, respectively, of  $\underline{x}_\tau$ . Then the time asymmetry is quantified by the skewness coefficient:

$$\tilde{C}_S := \frac{\mu_3[\tilde{\underline{x}}_\tau]}{(\text{var}[\tilde{\underline{x}}_\tau])^{3/2}} \quad (7)$$

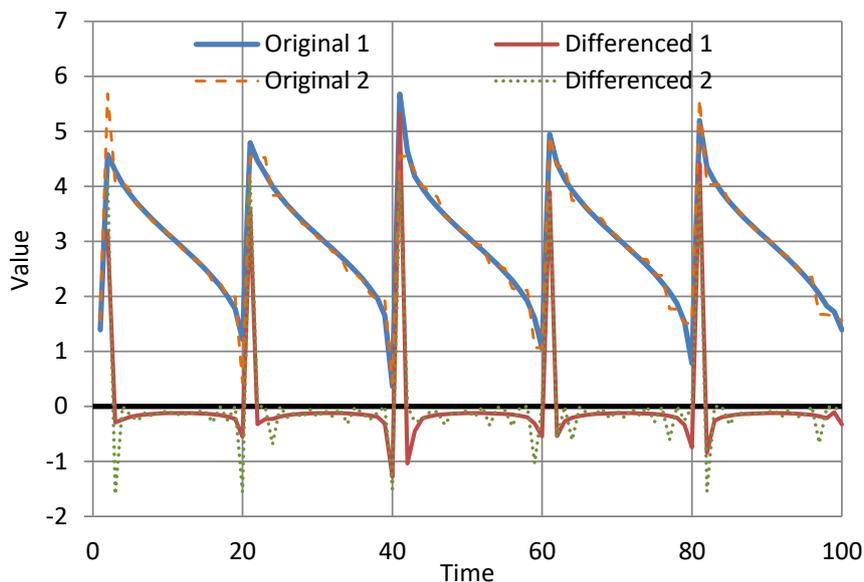
where  $\mu_3[\ ]$  denotes the third moment. Alternatively, it is quantified by the probability of  $\tilde{\underline{x}}_\tau$  being negative:

$$\tilde{P}_N := P\{\tilde{\underline{x}}_\tau \leq 0\} = \tilde{F}_{\tilde{\underline{x}}_\tau}(0) \quad (8)$$

Processes with large  $\tilde{C}_S$  or large departure of  $\tilde{P}_N$  from the value 1/2 signify high (positive) time irreversibility. A sketch of a realization of such a process is given in Figure 1. The graph is characterized by rapid increases followed by gradual decreases. The entire setting in constructing the time series of Figure 1 is nonlinear (the values are obtained by optimization rather than generated by a linear stochastic model) and while the marginal distribution is by construction Gaussian, the multivariate one is not, so there is no conflict with Weiss's (1975) result mentioned above.

The study by Koutsoyiannis (2019b) has provided evidence that time irreversibility may be profound in the streamflow process at scales of several days or lower. In atmospheric processes

irreversibility appears as well, but at much finer time scales. On the other hand, most stochastic simulation methods do not account for irreversibility. In the same study two methods have been proposed for a stochastic generation that respects irreversibility and reproduces it in a controlled manner, quantifying it by the skewness of the differenced process. After exploring them, one of the two methods proved successful. A disadvantage of that method is that it needs iterations to converge to a solution, in which vector derivatives need to be evaluated.



**Figure 1** Plot of two synthetic time series generated by maximizing time irreversibility properties of a process restricted to be marginally Gaussian ( $N(3, 1)$ ) with lag-one autocorrelation 0.5, so that the variance of the differenced process is also 1 (equal to that of the original process). Solution 1 maximizes the skewness of the differenced process. Solution 2 maximizes the frequency that the differenced process has a negative value, without taking into account the skewness. In both series the frequency that the differenced process has negative values is 0.94. The coefficients of skewness of the differenced processes for series 1 and 2 are 4.10 and 3.34, respectively. (Source: Koutsoyiannis 2019b.)

Here we present a simple method that is exact, analytical and direct, and its application is fast avoiding the need of iterations. As far as the general framework related to time's arrow and time irreversibility of atmospheric and hydrological processes is concerned, the interested reader is referred to the study by Koutsoyiannis (2019b) and its extended list of references. However, in terms of the stochastic synthesis per se, the current paper is a stand-alone study with a detailed presentation, starting from main historical landmarks and ending with proposing a generic method of stochastic synthesis which is entirely new (to the author's knowledge). In this manner, an earlier study for a generic and parsimonious stochastic modelling scheme (Koutsoyiannis 2016) is complemented and extended to time irreversible processes.

## 2 Approaches in stochastic synthesis

It is common knowledge that when we want to model the (marginal) distribution of a stochastic variable, we usually fit a model (a distribution such as Gaussian, gamma, generalized extreme value, etc.) using the available data, typically given as a time series  $x_n, n = 1, \dots, N$ . But let us

consider for a minute a different tactic: We estimate from the data sufficiently many, say  $J$ , moments, and instead of fitting a customary distribution, we approximate the distribution through moments. These moments are  $\mu'_j := E[\underline{x}^j], j = 1, 2, \dots, J$ , and their unbiased estimates from the data are  $\hat{\mu}'_j := \sum_{n=1}^N x_n^j / N$ . From the moments we could recover the density function  $f(x)$  of the stochastic variable of interest  $\underline{x}$ , in at least two ways:

1. We could exploit the inversion formula (e.g. Papoulis, 1991, p. 116):

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi(\omega) e^{-i\omega x} d\omega \quad (9)$$

where  $\varphi(\omega)$  is the characteristic function of the distribution and  $i := \sqrt{-1}$  is the imaginary unit, combine it with the moment theorem, according to which the  $j$ th derivative of  $\varphi(\omega)$  is:

$$\varphi^{(j)}(\omega) = i^j \mu'_j \quad (10)$$

and then use some type of expansion to approximate  $\varphi(\omega)$  and eventually  $f(x)$  based on a number  $J$  of moment estimates.

2. We could use as an approximation of  $f(x)$  the maximum entropy distribution conditional on the specified moments, which is (e.g. Papoulis 1991, p. 571):

$$f(x) = \exp\left(\sum_{j=0}^J \lambda_j x^j\right) \quad (11)$$

where  $\lambda_j$  are constants determined so that the distribution have the specified moments.

This tactic, if implemented in either variant, would have problems, including the following:

- Assuming that the number  $J$  is not very small, the resulting models would not be parsimonious, as they would include  $J$  moments and an equal number of parameters.
- The moment estimates beyond second are unreliable for typical sample sizes, despite the fact that unbiasedness is theoretically guaranteed (see the work by Lombardo et al. 2014, with the characteristic title “Just two moments”.)
- In particular, heavy tailed distributions would never appear as models, because in these distributions all moments beyond a threshold are infinite, while their sample estimates are finite.

The origin of problems in such a tactic is not actually the approximate character of the specified distribution—after all, any model is an approximation of some real phenomenon. Rather, it is the lack of parsimony and the direct link of the parameters (such as  $\lambda_j$  in (11)) with the data, on which their estimation is based. In contrast, the same approximation of the distribution would be valid and not problematic if coefficients  $\lambda_j$  were not model parameters and if they were determined theoretically rather than estimated from the data.

Fortunately, however, this tactic is not commonly implemented. Rather, a parametric model is assumed based, on the one hand, on some theoretical considerations (e.g. central limit

theorem, maximum entropy principle, asymptotic theory of distribution of extremes) and on the other hand, on some exploration of the available data.

Strangely enough though, this tactic has been the rule in fitting and applying so-called time series models—and it is still the norm even today—despite the fact that fitting a stochastic process is more demanding a task than fitting a marginal distribution and, at the same time, the presence of autocovariance increases uncertainty in model fitting. Indeed, time series models are typically fitted based on estimates of the autocovariance (or autocorrelation) function and even the model formulation (in terms of the multivariate distribution) is determined by the number of the autocorrelation coefficients they are used—similar to what equation (11) suggests for the marginal distribution. Apparently, the alternative of fitting a parametric model to the multivariate distribution is valid (Koutsoyiannis 2000, 2016) but it has not been the norm. The reasons for following such a weird practice in most disciplines whose modelling relies on data, including hydrology, are historical. An attempt to trace their evolution is contained in the next section.

### 3 An historical overview of stochastic modelling

Current stochastic modelling practices have been influenced by at least three Schools of thought, which will be referred to as (a) the *Stochastic School*, (b) the *Time Series School* and (c) the *Monte Carlo School*. We attempt to trace out the evolution in each of them in the next subsections, while an individual subsection is devoted to hydrology.

#### 3.1 The Stochastic School

Perhaps Ludwig Boltzmann (1844 –1906, Universities of Graz and Vienna, Austria, and Munich, Germany) was the first to introduce fundamental concepts related to stochastic process. In particular, Boltzmann (1877) explained the concept of entropy in probability theoretic context and Boltzmann (1884/85) introduced the notion of ergodic systems<sup>1</sup>, which is central in stochastics. Later George D. Birkhoff (1884 – 1944; Princeton, Harvard, USA) discovered the ergodic theorem (Birkhoff 1931) also known as Birkhoff–Khinchin theorem to give credit to Aleksandr Khinchin (also spelled Khintchine, 1894 – 1959; Moscow State University, Russia) who gave a purely measure-theoretic proof of the theorem (Khintchine 1933). In addition, Khintchine (1934) defined the stationary stochastic processes and introduced a probabilistic setting of the Wiener-Khinchin theorem relating autocovariance and power spectrum.

But it was Andrey N. Kolmogorov (1903 – 1987; Moscow State University, Russia) the great scientist who founded the stochastic theory. Kolmogorov (1931) introduced the term *stochastic*

---

<sup>1</sup> Boltzmann coined the terms *ergode* and *isodic*, both of which are etymologized from Greek words but which ones exactly is uncertain. Most probably, *ergodic* comes from the Greek *ἔργον* (ergon = work) and *ὁδός* (hodos = pathway). According to another interpretation, the second noun is *εἶδος* (eidos = form, kind, nature), or the whole word is a transliteration of the Greek adjective *ἐργώδης* (ergodes = laborious, troublesome; see Mathieu 1988).

*process*<sup>2</sup>, identifying *process* with *change*. He also used the term *stationary* to describe a probability density function that is unchanged in time. Kolmogorov (1933) defined *probability* (founded on measure theory) in an axiomatic manner based on three fundamental concepts (a triplet called probability space) and four axioms (non-negativity: normalization, additivity and continuity at zero). Kolmogorov (1937, 1938) gave a concise presentation of the concept of a stationary stochastic process and Kolmogorov (1947) defined wide-sense stationarity.

The concept of stationarity (and its negation, nonstationarity), has a central role in stochastics yet is widely misunderstood and broadly misused (Montanari and Koutsoyiannis, 2014; Koutsoyiannis and Montanari, 2015). Stationarity is tightly connected to ergodicity, which in turn is a prerequisite to make inference from data, that is, *induction*. If a system that is modelled in a stochastic framework has deterministic dynamics then it is stationary if it is ergodic and vice versa (Mackey 2003, p. 52). Likewise, a nonstationary system is also non-ergodic and vice versa. If the system dynamics is stochastic, then ergodicity and stationarity do not necessarily coincide. However, recalling that a stochastic process is a model (not part of the real world), we can always devise a stationary stochastic process that is ergodic.

It may be useful to include here a reminder of the definitions of the two concepts. Following Kolmogorov (1931, 1938) and Khintchine (1934), a process is stationary if its statistical properties are invariant to a shift of time origin, i.e. the processes  $\underline{x}(t)$  and  $\underline{x}(t')$  have the same (multivariate) distribution for any  $t$  and  $t'$ . Based on the ergodic theorem (Birkhoff, 1931; Khintchine, 1933; see also Mackey, 2003, p. 54), a stochastic process  $\underline{x}(t)$  is ergodic if the time average of any (integrable) function  $g(\underline{x}(t))$ , as time tends to infinity, equals the true (ensemble) expectation, i.e.:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T g(\underline{x}(t)) dt = E[g(\underline{x}(t))] \quad (12)$$

### 3.2 The Time Series School

Contrary to the theoretical Stochastic School, the Time Series School is more intuitive and empirical and less rigorous. Its history was written more by economists and less by theoretical mathematicians. Perhaps the first who introduced time series analysis was W.M. Persons (American economist). In studying the problem “*When to buy or sell*”, Persons (1919) introduced the study of time series, which he called *statistical series*, and asserted that they:

---

<sup>2</sup> The Greek adjective “stochastic[os/e]” was used by Greek philosophers, including Plato and Aristotle, and was transplanted to the international scientific vocabulary by Jacob Bernoulli (evidently aware of the Greek literature and in particular of Plato’s texts) in his famous book *Ars Conjectandi* (written in Latin in 1684-89 but published after his death, in 1713). The term was revived by Bortkiewicz (1917; Russian economist and statistician of Polish ancestry) and also by Slutsky (1925, 1928a,b, 1929; Ukrainian/Russian/Soviet mathematical statistician and economist). It appears that the prevalence in USSR of the more sophisticated term “stochastic” (over the rather equivalent term “random”) must have been related to political and ideological reasons (incongruence of randomness with the *dialectic materialism*: models beyond strict deterministic were considered with a priori suspicion; see Mazliak 2018).

*result from the combination of four elements: secular trend, seasonal variation, cyclical fluctuation, and a residual factor.*

He also proposed methods for “*Eliminating secular trends*” and “*Eliminating seasonal variation*”. A definition of a time series was given ten years later by Bailey (1929) (American statistician):

*A time series is a series of observations taken at different times and recorded with the time at which they were taken.*

Interestingly, Slutsky (1927) demonstrated that what Persons (and other economists) regarded as cyclical component is nothing but a statistical artefact with no essential meaning (see e.g. Kyun and Kim 2006; Barnett, 2006). Subsequently, the notion of a cyclical component was abandoned but the decomposition of a time series to the remaining three components, trends, seasonal variation and residuals is popular even today.

Yule (1927) and Walker (1931) (both British statisticians), starting from an analysis of sunspot numbers, studied autoregressive processes and in particular their periodogram and autocorrelation properties. But the biggest progress in the Time Series School was made in Uppsala by H.O.A. Wold (Norwegian-born econometrician and statistician with career in Sweden) and P. Whittle (New-Zealand-born mathematician and statistician) who in their doctoral theses provided the stochastic foundation of time series analysis. Wold (1938, 1948) proved that a stochastic process (even though he referred to it as a time series) can be decomposed into a regular process (i.e., a process linearly equivalent to a white noise process) and a predictable process (i.e., a process that can be expressed in terms of its past values). Whittle (1951, 1952, 1953) laid the mathematical foundation of autoregressive and moving average models in univariate and multivariate setting. Later, in their influential book, Box and Jenkins (1970) named these models with acronyms such as  $AR(p)$  (standing for autoregressive models of order  $p$ ),  $MA(q)$  (standing for moving average models of order  $q$ ),  $ARMA(p,q)$  (standing for autoregressive – moving average models) and  $ARIMA(p,d,q)$  (standing for autoregressive integrated moving average models). These became popular with these names and have also been known as Box-Jenkins models (cf. Stigler’s law of eponymy; Stigler, 2002). A useful extension of these models to apply to processes with long-range dependence was proposed by Hosking (1981). This is done by replacing the integer parameter  $d$  in  $ARIMA(p,d,q)$  with a real one (fractional differencing) and the models are usually termed  $ARFIMA(p,d,q)$ .

Despite the wider influence of the Time Series School over the Stochastic School, there are several problems with the former. First, the term *time series* is ambiguous, sometimes denoting a series of observations as in the original definition of Bailey (1929) (or, equivalently, a realization of a stochastic process), and other times denoting the stochastic process *per se* as in the aforementioned use by Wold. We emphasize that here the term *time series* is used with the first meaning, a series of numbers, while for a series of random variables we use the term stochastic process. Second, with the exception of the simplest models of these families, such as the  $AR(1)$  and  $ARMA(1,1)$ , time series models are too artificial because, being complicated discrete-time models, they do not necessarily correspond to a continuous time process, while natural

processes typically evolve in continuous time. Furthermore, their identification, typically based on the estimation of the autocorrelation function from data, usually neglects estimation bias and uncertainty, which in stochastic processes (as opposed to purely random processes) are often tremendous (Lombardo et al., 2014).

Indeed, from their onset, time series models have been tightly associated with a large number of parameters and they usually become over-parameterized and thus not parsimonious. These parameters are estimated from data, which usually are too few to support a reliable estimation. In Whittle's (1952) words (emphasis added):

*There is, of course, nothing special with the autoregressive scheme; we could equally well graduate with a high-order moving average, and there are many other possibilities. [...] In practice, however, the autoregressive graduation has the advantage that the estimated residual **sum of squares can be written down directly in terms of the observations** [...] without the need to solve explicitly for the estimates of the 'a' coefficients.*

...

*It is, of course, not possible to estimate an infinite number of parameters from a finite sample, but the series of a coefficients must converge, and by **considering sufficiently many coefficients we should be able to obtain an arbitrarily good approximation to the real process.***

This is similar to fitting a distribution function using observations to estimate many moments as described in section 2.

The decomposition of a time series to components, trends, seasonal variation and residuals, is fundamentally problematic, despite being popular. (In particular trend analysis of hydroclimatic processes is more fashionable today than ever before.) However, it should be noted that a meaningful definition of a trend has never been given. Also, for some, including this author, it is hardly conceivable how *time per se* could be regarded as an explanatory variable in a complex process and what the logical basis is in expressing the statistics of a process as a deterministic function of time. Accumulation of data series with long time spans is revealing that, what have been regarded as trends, are mostly parts of long term fluctuations (and in accord to Slutsky's work, they could also be regarded as statistical artefacts). Hurst's (1951) and Kolmogorov's (1940) works provide a scientific basis to model what has been regarded as trends in the context of stationary stochastic processes. Finally, "deseasonalization" (in Persons's original terminology "*Eliminating seasonal variation*") is a delusion; we can hardly remove seasonality in the multivariate distribution of a stochastic process (what we typically do is in the marginal distribution).

### 3.3 The Monte Carlo School

The Monte Carlo method is essentially a numerical technique for solving deterministic integro-differential equations using random sampling. Ideas that could be classified as implementation of the Monte Carlo method are old. Georges Louis LeClerc (Comte de Buffon, French scientist;

1707-1788) became famous for “Buffon’s needle,” a method using needle tosses onto a lined background to estimate  $\pi$  (where, if the line distance is equal to needle length,  $\pi$  is found as twice the inverse of probability that the needle crosses a line).<sup>3</sup> Random sampling is also old. Galton (1890) invented a set of three modified dice to generate samples from a normal distribution. “Student” (pseudonym of W.S. Gosset) in 1908 performed simulation experiments using 3000 cards (in 750 groups of size 4) to find the distribution of the  $t$ -statistic and of the correlation coefficient. Tippett (1927) published a table of random numbers: he took 41 600 digits at random from Census Reports and combined them by fours to give 10 400 numbers. Mahalanobis (1934) published tables of random numbers from normal distribution (208 sets of 50 numbers each). (See more information in Stigler 2002.)

The modern Monte Carlo method was conceived by Stanislaw Ulam in 1946.<sup>4</sup> Soon after the method grew to solve neutron diffusion problems by himself and other great mathematicians and physicists in Los Alamos (John von Neumann, Nicholas Metropolis, Enrico Fermi), and was first implemented on the ENIAC computer (Metropolis, 1989; Eckhardt, 1989). The ‘official’ history of the method began in 1949 with the publication of a paper by Metropolis and Ulam (1949). One prominent characteristic of the Monte Carlo method as a numerical integration method is that it does not suffer from the well-known *curse of dimensionality*, unlike the classical (grid-based) numerical integration method, which does suffer. Thus, it is easily shown (e.g. Metropolis and Ulam, 1949; Niederreiter, 1992) that for a number of dimensions  $d > 4$ , the Monte Carlo integration method (in which the function evaluation points are taken at random) is more accurate (for the same total number of evaluation points) than classical numerical integration, based on a grid representation of the integral space. For large dimensionality, e.g. for  $d > 20$ , the classical method is infeasible while the Monte Carlo method is always feasible.

### 3.4 The onset of stochastic modelling in hydrology

Techniques that could be classified as applications of the Monte Carlo method had appeared in the hydrological literature much earlier than the “official start” of the method in 1949. Hazen

---

<sup>3</sup> LeClerc’s Monte Carlo method to calculate  $\pi$  became popular among scientists and his experiment was later repeated by many. There are many other Monte Carlo algorithms to estimate  $\pi$ . However, these are good only for fun, as much faster and much more accurate deterministic algorithms exist to calculate  $\pi$ . Reitwiesner (1950) calculated by a deterministic algorithm, running on the ENIAC computer, the first 2035 decimal digits of  $\pi$ . Metropolis et al. (1950) examined their randomness, an exercise made thereafter many times showing that the digits of  $\pi$  have no apparent pattern and pass tests for statistical randomness. Dodge (1996) promoted an idea opposite to LeClerc’s: that the digits of  $\pi$  form a “Natural Random Number Generator”. Since January 2019, 31.4 trillion digits of  $\pi$  are known (found by the Chudnovsky algorithm); this information, equivalent to  $\sim 100$  million books of 1000 pages each (note for comparison that the British Library has 25 million books), can serve as a basis for any simulation experiment. However, simple random generators are more economic and convenient.

<sup>4</sup> Notably, Ulam conceived the method while playing solitaires during convalescing from an illness, in his attempt to estimate the probabilities of success of the plays. As Ulam describes the story in some remarks later published by Eckhardt (1989):

*After spending a lot of time to estimate them by pure combinatorial calculations, I wondered whether a more practical method than ‘abstract thinking’ might not be to lay it out say one hundred times and simply observe and count the number of successful plays.*

(1914), made a pioneering study in which he introduced the reservoir storage-yield-reliability relationship, a concept that would remain unexploited in the western hydrological literature yet constituting the scientific basis of modern reservoir design (Klemes 1987). In that study he proposed an empirical simulation technique and formed a synthetic time series by combining historical flow records of different rivers ‘spliced’ sequentially together. Sudler (1927) extended the work of Hazen by resampling from a sequence of historical river flows using cards, which he shuffled to form new sequences of data. Obviously, this method heavily distorts the time dependence of river flows whose importance was not known at that time. For it was Hurst (1951) who understood that importance along with the omnipresence in natural processes of a clustering behaviour of similar events in time, a behaviour that is now called long-term persistence (LTP) or long-range dependence (LRD). It is also known as the Hurst-Kolmogorov behaviour, to give also credit to Kolmogorov (1940) who had proposed a mathematical model representing that behaviour before it was verified in natural processes by Hurst. In his attempt to compare natural and random events, Hurst performed physical experiments to generate random numbers. Specifically he tossed 10 coins (sixpences) simultaneously and repeated this 1025 times (note that 10 binary digits are equivalent to about 3 decimal digits). As he notes, his rate was 100 random numbers per 35 min (while that would be of the order of a microsecond in modern computer environments, even slow ones). He also used another method, shuffling and cutting a pack of 52 cards, in which he improved the rate to 100 random numbers per 20 min.

It would take decades before the hydrological community assimilate Hurst’s discovery of LTP (O’Connell et al. 2016), even though some of the pioneering studies appeared towards the end of the 1960s (Mandelbrot and Wallis, 1969). The initial studies implementing primitive variants of stochastic simulation did not reproduce LTP. Barnes (1954), in designing a reservoir in Australia, used a table of random numbers from normal distribution to generate a 1000-year sequence of synthetic annual data. Thomas and Fiering (1962) generated flows correlated in time, but using only the lag one autocorrelation, obviously neglecting LTP. Beard (1965) and Matalas (1967) generated concurrent flows at several sites. Chow (1969), and Chow and Kareliotis (1970) systematized the use of time series models (in particular—and using their terminology—moving average models, sum of harmonics models and autoregression models) and highlighted their value in the economic planning of water supply and irrigation projects. It is evident from the above pioneering studies, as well as of subsequent myriads of studies, that hydrologists have followed (and today still do) the Time Series School rather than the more rigorous Stochastic School.

## **4 Premises**

Before we proceed to the formulation of the simulation scheme, it is necessary to list the main premises on which the scheme is based. These are consistent with the generic procedure in Koutsoyiannis (2016) and are summarized in the following points:

1. A stochastic model is always required for any stochastic task, such as estimation, testing and synthesis. In stochastics there cannot be model-free, also known as nonparametric, methods. Nonparametric methods are a delusion caused by an adherence to the classical statistical assumption of independence which, when dealing with natural processes is the ultimate—and most inappropriate—parameterization *per se*. As an example, Kendall's rank correlation test (Kendall 1938), or its variant for trend detection (Mann 1945), is an important and widely used statistical test that is regarded as nonparametric. And indeed it is nonparametric if it is applied on samples that are independent by construction, i.e. obtained by individual and independent experiments. But if the data are time series (consecutive observations in time of a single process) then application of the test needs parameterization of the dependence structure. As usually applied, the test includes in its null hypothesis the sub-hypothesis that autocorrelations for all lags are zero. Thus, the rejection of the null hypothesis could be regarded as rejection of independence, even though the popular interpretation favours the existence of a trend. There exist of course consistent parametric variants of the test, which take the stochastic structure into account (e.g. Hamed 2008).
2. As natural time runs continuously, the model needs to be formulated for continuous time to avoid the risk of making artificial constructs. The discrete-time representation, which is certainly necessary in simulation, should be derived from the continuous time one. Second-order stochastic tools, such as autocovariance and power spectrum, are affected by discretization and the effect should always be accounted for. The climacogram<sup>5</sup>, i.e. the variance  $\gamma(k)$  of the process averaged at time scale  $k$ , is an exception, as it is not affected by discretization, and has some additional advantages (Dimitriadis and Koutsoyiannis, 2015) which make it the preferable tool in stochastic modelling.
3. The model needs to be parsimonious in order to be useful (Gauch 2003). Inflationary models, while giving an impression of a good fit, in fact entail (often hidden) high uncertainty. An example of a (parametric) parsimonious model is the Filtered Hurst-Kolmogorov process with a generalized Cauchy-type climacogram (FHK-C; Koutsoyiannis 2016, 2017). This is defined through its climacogram:

$$\gamma(k) = \lambda \left( 1 + \left( \frac{k}{\alpha} \right)^{2M} \right)^{\frac{H-1}{M}} \quad (13)$$

---

<sup>5</sup> The term *climacogram*, from the Greek *κλιμακόγραμμα*, deriving from *κλίμαξ* (climax = scale, as well as ladder; pl. *κλίμακες*) and *γράμμα* (gramma = written, drawn), was coined by Koutsoyiannis (2010) and could be translated in English as *scale(o)gram*, but the latter term is used for another concept. *Climacogram* should not be confused to *climatogram* which has another meaning related to the climatic regime of temperature and precipitation at a site or area. The latter term, deriving from the noun *κλίμα* (clima, originally meaning slope; pl. *κλίματα*), was first used in the Hellenistic period by the astronomer Hipparchus to describe climate (in relationship to the slope of the sun's rays) and is different from the other derivative noun *κλίμαξ*. Interestingly though, both *κλίμαξ* and *κλίμα* are eventually etymologized from the same verb *κλίνειν* (klinein = to slope).

where  $\alpha$  and  $\lambda$  are scale parameters with dimensions of  $[t]$  and  $[x^2]$ , respectively, while  $M$  (fractal parameter) and  $H$  (Hurst parameter) are dimensionless determining the dependence structure at a local (smoothness or fractality) and global (LRD) level, respectively. Both  $M$  and  $H$  take on values in  $(0,1)$ , with  $M < 1/2$  or  $> 1/2$  indicating a rough or a smooth process, respectively, and with  $H < 1/2$  or  $> 1/2$  indicating an antipersistent or a persistent process, respectively. It should be stressed that the quantities  $\alpha$ ,  $\lambda$ ,  $M$  and  $H$  in the above expression (plus the process mean, and perhaps a shape parameter of the marginal distribution and a parameter quantifying irreversibility) are the only model parameters. Any other quantity that would emerge in the generation phase should be theoretically calculated (as internal coefficients of the generation scheme) from these parameters and no longer estimated from data. If the data cannot support the estimation of four parameters of equation (13) for the dependence structure, simplifications could be made. For example, assuming  $M = H = 1/2$  we obtain a two-parameter near-Markov model while with  $M = 1/2$  and  $\alpha \rightarrow 0$  we obtain a two parameter Hurst-Kolmogorov model. Examples of models additional to equation (13) can be found in Koutsoyiannis (2017).

4. Parameter estimation needs to consider statistical bias, which is present in all estimators (Koutsoyiannis 2016), including very simple and common ones, such as variance and correlation (a known exception is the mean whose standard estimator is not affected by bias in the presence of dependence; higher order noncentral moments are also unbiased in theory, but suffer from unknowability; Koutsoyiannis 2019a).
5. Awareness of stochastics (the mathematics of stochastic variables and processes), theoretical consistency and logical rigour are always necessary to avoid misleading or erroneous calculations, results and interpretations.

## 5 Simulation method

Let  $\underline{x}(t)$  be a stationary stochastic process representing the instantaneous quantity of a certain hydrological or other physical process in continuous time  $t$ . For simplicity and without loss of generality we assume that it has zero mean and (instantaneous) variance:

$$\gamma_0 := \text{var}[\underline{x}(t)] \quad (14)$$

which is arguably (Koutsoyiannis, 2017) assumed to be finite. The discrete-time representation of the process (equation (5)) can also be written as:

$$\underline{x}_\tau := \frac{1}{D} \int_{(\tau-1)D}^{\tau D} \underline{x}(u) du = \frac{\underline{X}(\tau D) - \underline{X}((\tau-1)D)}{D} \quad (15)$$

where

$$\underline{X}(t) := \int_0^t \underline{x}(\xi) d\xi \quad (16)$$

is the cumulative process, which if  $\underline{x}(t)$  aims to represent a natural processes, should necessarily be nonstationary. Its variance at time  $t$ , known as cumulative climacogram, is:

$$\Gamma(t) := \text{var}[\underline{X}(t)] = t^2\gamma(t) \quad (17)$$

where  $\gamma(\cdot)$ , is the variance of the averaged process. This latter, as a function of time scale  $k$ , is called climacogram:

$$\gamma(k) := \text{var}\left[\frac{\underline{X}(k)}{k}\right] \quad (18)$$

The autocovariance function  $c(h)$  of the continuous-time process  $\underline{x}(t)$  for time lag  $h$  is (Koutsoyiannis 2016):

$$c(h) := \text{cov}[\underline{x}(t), \underline{x}(t+h)] = \frac{1}{2} \frac{d^2\Gamma(h)}{dh^2} \quad (19)$$

and the power spectrum  $s(w)$  for frequency  $w$  is the Fourier transform of the autocovariance function, i.e.:

$$s(w) := 4 \int_0^{\infty} c(h) \cos(2\pi wh) dh, \quad c(h) = \int_0^{\infty} s(w) \cos(2\pi wh) dw \quad (20)$$

The discrete time autocovariance for integer time lag  $\eta$  is (Koutsoyiannis 2016):

$$c_{\eta} = \frac{1}{D^2} \left( \frac{\Gamma(|\eta+1|D) + \Gamma(|\eta-1|D)}{2} - \Gamma(|\eta|D) \right) \quad (21)$$

and is related to the discrete-time power spectrum by

$$s_d(\omega) = 2c_0 + 4 \sum_{\eta=1}^{\infty} c_{\eta} \cos(2\pi\eta\omega), \quad c_{\eta} = \int_0^{1/2} s_d(\omega) \cos(2\pi\omega\eta) d\omega \quad (22)$$

where  $\omega := wD$  is dimensionless frequency. The power spectrum has some analogies with another stochastic tool, the so-called climacospectrum, which is directly given in terms of the climacogram. Specifically, it is proportional to the difference of the variances of the averaged process at time scales  $k$  and  $2k$ :

$$\zeta(k) := \frac{k(\gamma(k) - \gamma(2k))}{\ln 2} \quad (23)$$

This tool is useful in the model fitting phase (see section 7).

To simulate  $\underline{x}_{\tau}$  we use the generalized moving average scheme (Koutsoyiannis 2000):

$$\underline{x}_{\tau} = \sum_{j=-J}^J a_j \underline{v}_{\tau-j} \quad (24)$$

where  $a_j$  are weights to be calculated from the autocovariance function,  $\underline{v}_j$  is white noise averaged in discrete-time (and not necessarily Gaussian), also known as *innovation process*, and  $J$  is theoretically infinite, so that in all theoretical calculations we will assume  $J = \infty$ , while in the generation  $J$  is a large integer chosen so that the resulting truncation error be negligible. Here

we stress that the above scheme is just the contrary to that suggested by Whittle (1952) in the quotation given in section 3.2. Specifically, (a) we use a purely moving average scheme instead of the autoregressive scheme favoured by Whittle and (b) we do not relate our scheme with observations, as the observations have been already used in the model fitting phase, which is totally isolated from the generation scheme. Since we do not use the observations, the moving average scheme has many advantages in calculations.

Writing equation (24) for  $\underline{x}_{\tau+\eta}$ , multiplying it with (24) and taking expected values we find the convolution expression for  $J = \infty$ :

$$c_\eta = \sum_{l=-\infty}^{\infty} a_l a_{\eta+l} \quad (25)$$

We need to find the sequence of  $a_\eta, \eta = \dots, -1, 0, 1, \dots$ , so that (25) holds true. A known solution (Koutsoyiannis 2000) is the symmetric moving average (SMA) scheme in which  $a_{-\eta} = a_\eta$ . Nonetheless, the problem has infinitely many non-symmetric solutions, which are asymmetric moving average (AMA) schemes. A more common one is the ordinary backward asymmetric moving average (OBAMA) scheme in which  $a_\eta = 0$  for any  $\eta < 0$ ; this latter is typically formulated in a different manner and denoted as simply moving average—MA, but as we study a richer family of schemes, the distinct name OBAMA is necessary.

Here, the following generic solution of the AMA scheme, giving the coefficients  $a_\eta$ , is proposed:

$$a_\eta = \int_{-1/2}^{1/2} e^{2\pi i(\theta(\omega) - \eta\omega)} A^R(\omega) d\omega \quad (26)$$

where  $\theta(\omega)$  is any (arbitrary) odd real function (meaning  $\theta(-\omega) = -\theta(\omega)$ ) and

$$A^R(\omega) := \sqrt{2s_d(\omega)} \quad (27)$$

We prove in Appendix A that the sequence of  $a_\eta$ :

- (1) consists of real numbers, despite the expression in equation (26) involving complex numbers;
- (2) satisfies precisely equation (25); and
- (3) is easy and fast to calculate using the fast Fourier transform (FFT).

We can now convert this theoretical result into a numerical algorithm. This consists of the following steps:

- (a) From the continuous-time stochastic model, expressed through its climacogram  $\gamma(k)$ , we calculate its autocovariance function in discrete time (assuming time step  $D$ ):

$$c_\eta = \frac{(\eta + 1)^2 \gamma(|\eta + 1|D) + (\eta - 1)^2 \gamma(|\eta - 1|D)}{2} - \eta^2 \gamma(|\eta|D) \quad (28)$$

(This step is obviously omitted if the model is already expressed in discrete time through its autocovariance function.)

- (b) We choose an appropriate number of coefficients  $J$  that is a power of 2 and perform inverse FFT (using common software) to calculate the discrete-time power spectrum and the frequency function  $A^R(\omega)$  for an array of  $\omega_j = j w_1, j = 0, 1, \dots, J, w_1 := 1/JD$ :

$$s_d(\omega_j) = 2c_0 + 4 \sum_{\eta=1}^J c_\eta \cos(2\pi\eta\omega_j), \quad A^R(\omega_j) = \sqrt{2s_d(\omega_j)} \quad (29)$$

- (c) We choose  $\theta(\omega)$  (see below) and we form the arrays (vectors)  $\mathbf{A}^R$  and  $\mathbf{A}^I$ , both of size  $2J$  indexed as  $0, \dots, 2J - 1$ , with the superscripts R and I standing for the real and imaginary part of a vector of complex numbers, respectively:

$$[\mathbf{A}^R]_j = \begin{cases} A^R(\omega_j) \cos(2\pi\theta(\omega_j))/2, & j = 0, \dots, J \\ [\mathbf{A}^R]_{2J-j}, & j = J + 1, \dots, 2J - 1 \end{cases} \quad (30)$$

$$[\mathbf{A}^I]_j = \begin{cases} -A^R(\omega_j) \sin(2\pi\theta(\omega_j))/2, & j = 0, \dots, J - 1 \\ 0 & j = J \\ -[\mathbf{A}^I]_{2J-j}. & j = J + 1, \dots, 2J - 1 \end{cases} \quad (31)$$

- (d) We perform FFT on vector  $\mathbf{A}^R + i \mathbf{A}^I$  (using common software), and get the real part of the result for  $j = 0, \dots, J$ , which is precisely the sequence of  $a_\eta$ .

We note that by choosing  $J$  as a power of 2, the vectors  $\mathbf{A}^R$  and  $\mathbf{A}^I$  will have size  $2J$  which is also a power of 2, thus achieving maximum speed in the FFT calculations. More details are contained in a supplementary file which includes numerical examples along with the simple code needed to do these calculations on a spreadsheet.

It may be useful to note the following additional points about the method:

- Equation (26) gives not a single solution, but a variety of infinitely many ones, all of which preserve exactly the second-order characteristics of the process.
- A particular solution is characterized by the chosen function  $\theta(\omega)$ .
- Even assuming  $\theta(\omega) = \theta_0 \text{sign } \omega$  with constant  $\theta_0$ , again there are infinitely many solutions.
- The availability of infinitely many solutions enables preservation of additional statistics, e.g. those related to time asymmetry.
- In addition, we always have several options related to the distribution of the white noise  $\underline{v}_T$ , which in general is not Gaussian, thus enabling preservation of moments of any order. A general scheme for preserving moments of any order has been outlined in Koutsoyiannis (2019a), while successful applications for moments up to fourth order with SMA have been presented in Dimitriadis and Koutsoyiannis (2018). It should be stressed that in typical sample sizes, high order moments should be evaluated theoretically through the distributional parameters rather than estimated from the data, as their sample estimates are unreliable (Lombardo et al. 2014).

The special case  $\theta(\omega) = 0$  corresponds to the symmetric scheme (SMA), in which:

$$A^S(\omega) \equiv A^R(\omega) = \sqrt{2s_d(\omega)}, \quad a_\eta^S = \int_0^{1/2} \sqrt{2s_d(\omega)} \cos(2\pi\eta\omega) d\omega = a_{-,\eta}^S \quad (32)$$

where we have used the superscript S to denote the symmetric case. Another interesting special case is  $\theta(\omega) = 1/4 \text{ sign } \omega$  (or  $2\pi\theta(\omega) = \pi/2 \text{ sign } \omega$ ): This corresponds to an antisymmetric AMA scheme (ANTAMA) with

$$A^A(\omega) = A^R(\omega)\delta(\omega) + iA^R(\omega), \quad a_\eta^A = \delta_0 + \frac{1}{2} \int_0^{1/2} \sqrt{2s_d(\omega)} \sin(2\pi\eta\omega) d\omega \quad (33)$$

where the superscript A denotes the antisymmetric case,  $\delta(\omega)$  is the Dirac delta function, and

$$\delta_0 := \frac{\sqrt{2s_d(0)}}{2(2J+1)} \approx \sqrt{\frac{\gamma_J}{2J+1}} \quad (34)$$

with  $\delta_0$  approaching zero as  $J$  becomes large. Any other case of constant  $\theta_0$  (where  $\theta(\omega) = \theta_0 \text{ sign } \omega$ ) can be expressed in terms of the above two limiting cases through:

$$a_\eta = \delta_0 + (a_\eta^S - \delta_0) \cos(2\pi\theta_0) + (a_\eta^A - \delta_0) \sin(2\pi\theta_0) \quad (35)$$

(the proof is omitted). In particular, the case  $\theta_0 = 1/8$  (or  $2\pi\theta_0 = \pi/4$ ) yields the interesting result:

$$a_\eta = \frac{\sqrt{2}}{2} (a_\eta^S + a_\eta^A) - (\sqrt{2} - 1)\delta_0 \quad (36)$$

If for some  $\theta_0$  and for  $\eta < 0$  it happens that  $a_\eta \approx 0$ , then it can be verified that:

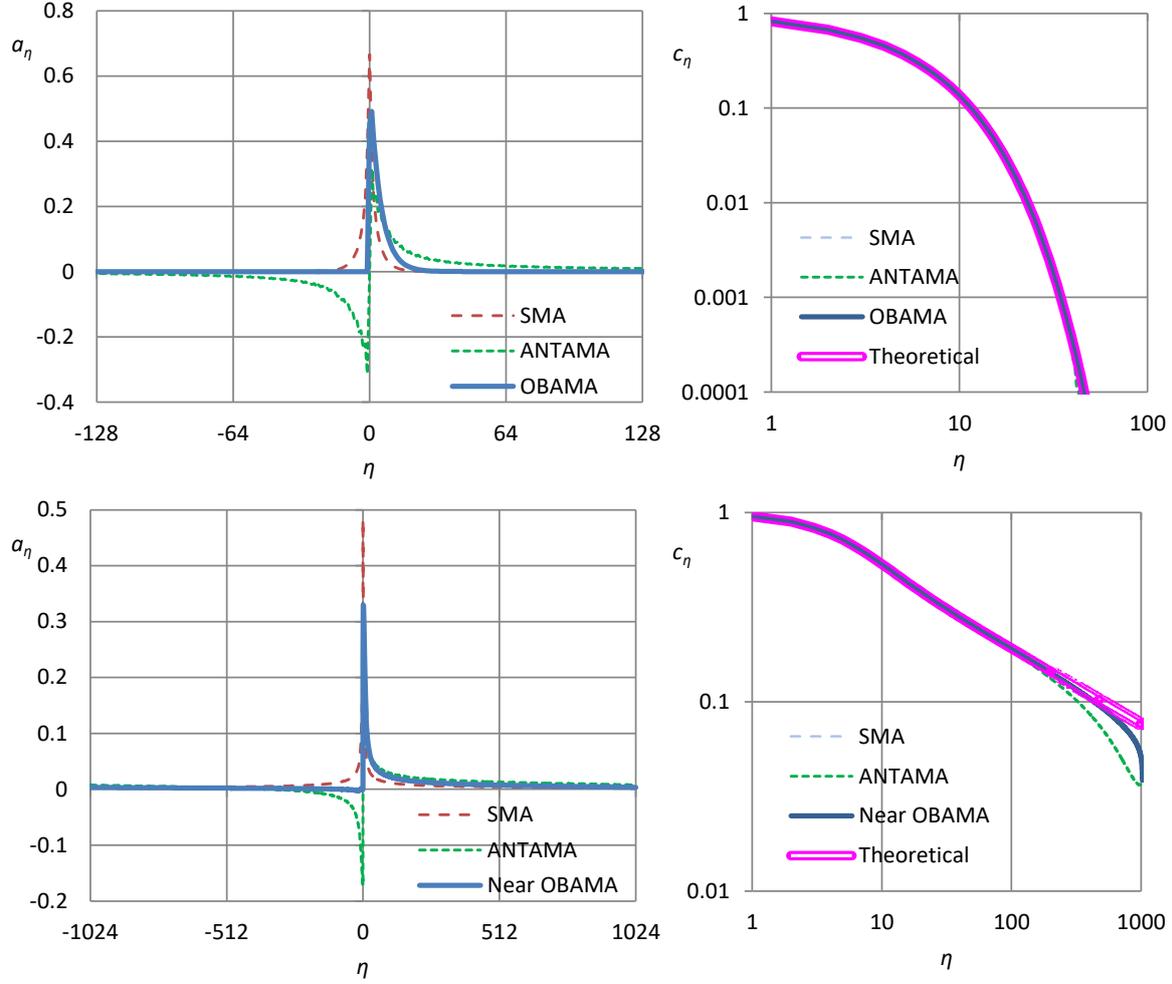
$$a_\eta \approx \begin{cases} 0, & \eta < 0 \\ a_\eta^S \cos(2\pi\theta_0) + (1 - \cos(2\pi\theta_0) - \sin(2\pi\theta_0))\delta_0, & \eta = 0 \\ \sqrt{2}a_\eta^S + (2 - \sqrt{2})\delta_0, & \eta > 0 \end{cases} \quad (37)$$

Such a sequence with almost zero coefficients for negative  $\eta$ , will be close to the OBAMA scheme. It is interesting to notice that in this approximate solution only  $a_0$  depends on the constant  $\theta_0$ , while for  $\eta > 0$  the coefficients are approximately equal to those in the SMA, multiplied by  $\sqrt{2}$ .

However, if full accuracy is needed, we must have in mind that a constant  $\theta_0$  does not give a precise OBAMA. A generic analytical solution that would give a function  $\theta(\omega)$  for OBAMA is not simple (this problem is known as factoring of the power spectrum; see Papoulis 1991, p. 402). However, solutions for simple special cases are not too difficult to find (e.g. for rational spectra; Papoulis 1991, p. 402-404). In particular, for a Markov process in continuous time (which in discrete time behaves as an ARMA(1,1) process; see equation (39)), as well as for any ARMA(1,1) process (including its special cases AR(1) and MA(1)), the following solution is exact (the proof is omitted):

$$\theta(\omega) = \frac{1}{2\pi} \cos^{-1} \left( \frac{\frac{1}{\sqrt{s_d(\omega)}}}{\left( \frac{1}{\sqrt{s_d(0)}} + \frac{1}{\sqrt{s_d(1/2)}} \right)} + \frac{\sqrt{s_d(\omega)}}{\sqrt{s_d(0)} + \sqrt{s_d(1/2)}} \right) \quad (38)$$

For stochastic structures with LRD, the OBAMA scheme is difficult to achieve with full precision, but an approximation can be easily found as described in Appendix B.



**Figure 2** Illustration of the symmetric (SMA), antisymmetric (ANTAMA) and ordinary-backward (OBAMA) cases of the generic AMA model for (upper) a Markov and (lower) a FHK-C process. The parameter values are  $\alpha = 10$ ,  $\lambda = 1$  (in both processes),  $H = 0.8$ ,  $M = 0.7$ , and the number of weights is 2049 ( $J = 1024 = 2^{10}$ ). The panels of the left column show the coefficients  $a$  and those of the right column the autocovariance function.

The method is illustrated in Figure 2 using two example processes. The first is the Markov process, whose basic properties are:

$$c(h) = \lambda e^{-|h|/\alpha}, \quad \gamma(k) = \frac{2\lambda}{k/\alpha} \left( 1 - \frac{1 - e^{-k/\alpha}}{k/\alpha} \right), \quad c_\eta = \frac{\lambda(1 - e^{-D/\alpha})^2}{(D/\alpha)^2} e^{-(\eta-1)D/\alpha}, \quad \eta \geq 1 \quad (39)$$

The second is the FHK model defined in equation (13), which gives its climacogram, whilst all its other properties are evaluated through equations (21)-(22). Figure 2 shows three special cases,

SMA (equation (32)), ANTAMA (equation (33)) and OBAMA for the two processes. For the OBAMA case and the Markov process the exact equation (38) is used while for the FHK-C process the coefficients  $a_\eta$  are determined from the approximate equation (B2) in Appendix B (with optimized coefficients of the equation). Some slight (rather invisible) deviations from zero are present in the left-bottom panel, which at a later step will be set to zero and the small resulting effect will be further handled as a truncation error (in the manner described by Koutsoyiannis 2016) to obtain an exact OBAMA scheme.

All in all, this simple method of the AMA scheme renders ARMA models unnecessary, particularly because of the generic, analytical and fast solution it offers. Here it is important to stress that, while optimization of coefficients involved in the function  $\theta(\omega)$  could sometimes be necessary as discussed above and specified in Appendix B, it is not necessary in general. Any odd real function  $\theta(\omega)$ , chosen arbitrarily, will give  $a_\eta$  that will satisfy equation (25) (apart from a truncation error) and thus can directly be used in generation. Even if the sequence of  $\theta(\omega_j)$  is constructed at random (e.g. as a sequence of random numbers in the interval  $[0,1/4]$ ), again equation (25) will be satisfied and the resulting  $a_\eta$  can directly be used in generation. (This case is also implemented on the spreadsheet provided as supplementary information).

## 6 Additional methodological issues

### 6.1 The AMA scheme in forecast mode

It is generally believed that AR models are better suited to forecast as they depend on past variables rather than on multiple innovation variables. According to this logic, among the MA models, only backward schemes would be suitable for forecast, while the generic AMA scheme (including the special case of SMA) would not, because it involves convolutions for both positive and negative lags  $\eta$ . Such beliefs are not quite correct though. Arguably, the forecast problem could be regarded as totally independent from the stochastic model per se, while forecasts can be obtained from any stochastic model, not only from a backward one. Had this not been the case, we would be in a deadlock in providing forecasts even with AR models. To see this, we may consider the case of an AR( $p$ ) scheme used to approximate a process with LRD. Because of LRD, the value of  $p$  will be large, typically larger than the number ( $n$ ) of known past observations ( $p > n$ ). Therefore, the mathematical expression of the AR( $p$ ) as a weighted sum of past values could not be applied for forecast because some of the required past values would not be known.

Koutsoyiannis (2000) formulated a simple and general solution of the forecast problem, which can be applied for any type of linear model and hence for AMA. This calculates the prediction  $\hat{x}_\tau$  for any future time  $\tau$  combining, on the one hand, any simulation  $\underline{x}_\tau$  independent of the observations, and, on the other hand, the observations  $\hat{\underline{z}}$ , through the equations:

$$\hat{x}_\tau = \underline{x}_\tau + \mathbf{c}_\tau^T \mathbf{C}^{-1} (\hat{\underline{z}} - \underline{z}), \quad \text{var}[\hat{x}_\tau | \hat{\underline{z}}] = \gamma_1 - \mathbf{c}_\tau^T \mathbf{C}^{-1} \mathbf{c}_\tau \quad (40)$$

where  $\hat{\underline{z}} := [\hat{x}_0, \hat{x}_{-1}, \dots, \hat{x}_{-n}]^T$  is a vector containing the observations at the present ( $\hat{x}_0$ ) and past times ( $\hat{x}_{-1}, \dots, \hat{x}_{-n}$ ), and  $\underline{z} := [x_0, x_{-1}, \dots, x_{-n}]^T$  is a vector containing the independently

simulated items of the stochastic process at the present and past, whilst  $\mathbf{c}_\tau := \text{cov}[\underline{x}_\tau, \underline{z}]$  and  $\mathbf{C} := \text{cov}[\underline{z}, \underline{z}]$  are a vector and a matrix of covariances, respectively, with items determined from the chosen stochastic model (again, not estimated from the data).

## 6.2 Time asymmetry

Time irreversibility can very easily be handled within the AMA framework. Assuming that the white noise  $\underline{v}_\tau$  (in discrete time  $\tau$ ) has variance 1 and coefficient of skewness  $C_S^{(v)}$ , we will have:

$$\text{var}[\underline{x}_\tau] = \sum_{j=-J}^J a_j^2, \quad \mu_3[\underline{x}_\tau] = \sum_{j=-J}^J a_j^3 C_S^{(v)}, \quad C_S := \frac{\mu_3[\underline{x}_\tau]}{(\text{var}[\underline{x}_\tau])^{3/2}} = \frac{\sum_{j=-J}^J a_j^3}{\left(\sum_{j=-J}^J a_j^2\right)^{3/2}} C_S^{(v)} \quad (41)$$

where  $\mu_3[\underline{x}_\tau]$  and  $C_S$  are the third moment and the coefficient of skewness of process  $\underline{x}_\tau$ , respectively.

Time asymmetry is quantified through the skewness of the differenced process  $\tilde{\underline{x}}_\tau$ , according to equation (4), which by virtue of (24) is written as:

$$\tilde{\underline{x}}_\tau := \underline{x}_\tau - \underline{x}_{\tau-1} = \sum_{j=-J}^J (a_j - a_{j-1}) \underline{v}_{\tau-j} \quad (42)$$

with  $a_{-J-1} = 0$ . Thus its skewness will be:

$$\tilde{C}_S := \frac{\mu_3[\tilde{\underline{x}}_\tau]}{(\text{var}[\tilde{\underline{x}}_\tau])^{3/2}} = \frac{\sum_{j=-J}^J (a_j - a_{j-1})^3}{\left(\sum_{j=-J}^J (a_j - a_{j-1})^2\right)^{3/2}} C_S^{(v)} \quad (43)$$

The ratio:

$$\frac{\tilde{C}_S}{C_S} = \frac{\sum_{j=-J}^J (a_j - a_{j-1})^3}{\left(\sum_{j=-J}^J (a_j - a_{j-1})^2\right)^{3/2}} \frac{\left(\sum_{j=-J}^J a_j^2\right)^{3/2}}{\sum_{j=-J}^J a_j^3} \quad (44)$$

is independent of  $C_S^{(v)}$  and primarily depends on  $\theta(\omega)$ , which determines the sequence of  $a_\eta$ . The case  $\theta(\omega) = 0$ , i.e. the SMA, results in complete time symmetry. However, a constant  $\theta_0 \neq 0$  (appropriately chosen) can make the ratio  $\tilde{C}_S/C_S$  as high as we wish, thus enabling preservation of time asymmetry.

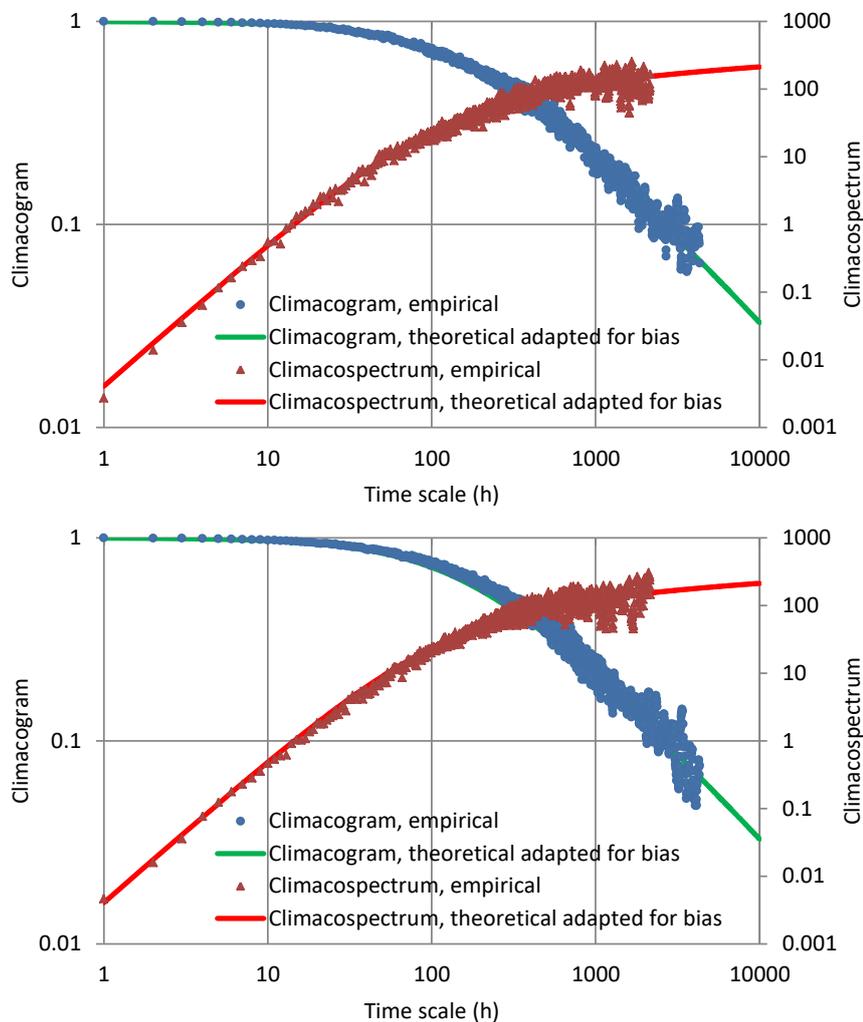
The above results make it clear that without skewness in the original process  $\underline{x}_\tau$  (e.g. in the case of Gaussian processes), there cannot be time asymmetry, a result that is consistent with Weiss's (1975).

## 7 Application

Here we repeat the application in Koutsoyiannis (2019b) for the streamflow time series but using the new analytical framework. The data are from the database of the U.S. Geological Survey ([https://nwis.waterdata.usgs.gov/md/nwis/uv/?site\\_no=01603000&agency\\_cd=USGS](https://nwis.waterdata.usgs.gov/md/nwis/uv/?site_no=01603000&agency_cd=USGS), retrieved: 2018-09-16) for the site USGS 01603000 North Branch Potomac River Near Cumberland, MD (39°37'18.5"N, 78°46'24.3"W, catchment area 2271 km<sup>2</sup>). They cover the

period 2013-10-01 to 2018-08-31 for a time step of 15 min, which was aggregated to hourly, while discharges were converted from cubic feet per second to  $\text{m}^3/\text{s}$ . Missing values (4% for a total of 172 416 values) were left unfilled. Seasonality is not negligible and was treated by standardising the values of each month by an appropriate coefficient (see Koutsoyiannis 2019b). We note here that this standardization is just a rough modelling approximation, adequate for our aim to investigate time irreversibility, while an approach for operational use would require the construction of a cyclostationary model.

The process  $x_\tau$  is modelled as a FHK-C process; the fitting is impressively good and the parameters are  $M = 0.56$  (indicating a slightly smooth process),  $H = 0.6$  (a persistent process),  $\alpha = 160$  h and  $\lambda = \gamma(0) = 1.01$  (for the standardized process). The fitting is shown in Figure 3 in terms of the climacogram, which is good for visualizing the behaviour at large time scales, and the climacospectrum, which provides a better visualization for small time scales.

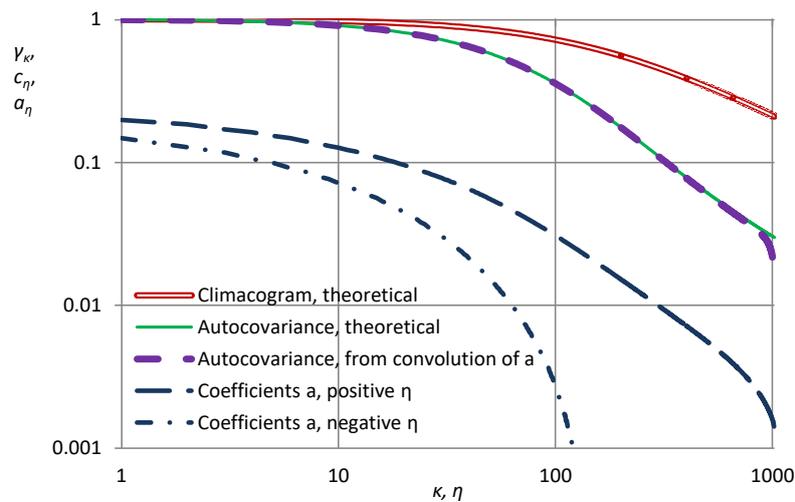


**Figure 3** Comparison with the FHK-C model of the climacogram and climacospectrum (upper) of the original series, after standardization for dealing with seasonality; (lower) of the generated series.

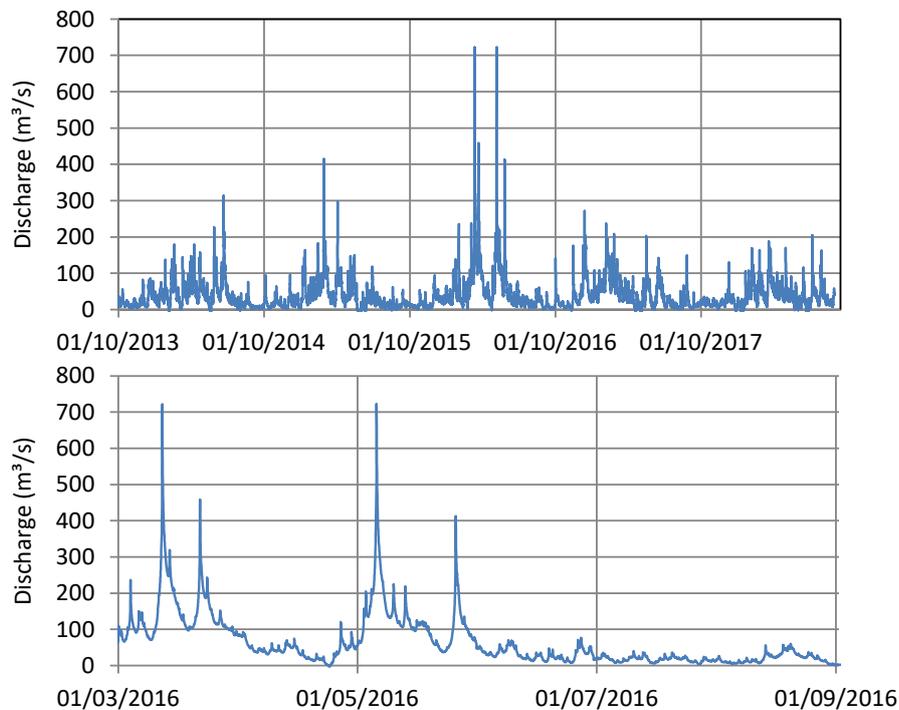
Additional parameters, quantifying the state and time asymmetry, are the skewness coefficients of the original and differenced process. The analysis in Koutsoyiannis (2019b) suggested prominent time irreversibility at time scales from 1 h to about 4 d. This is quantified

by a skewness coefficient of the differenced process, which is much higher than the skewness of the original process  $\underline{x}_\tau$  (10.99 for  $\tilde{x}_\tau$  and 2.98 for  $\underline{x}_\tau$  at the hourly scale, a ratio  $\tilde{C}_S/C_S = 3.69$ ).

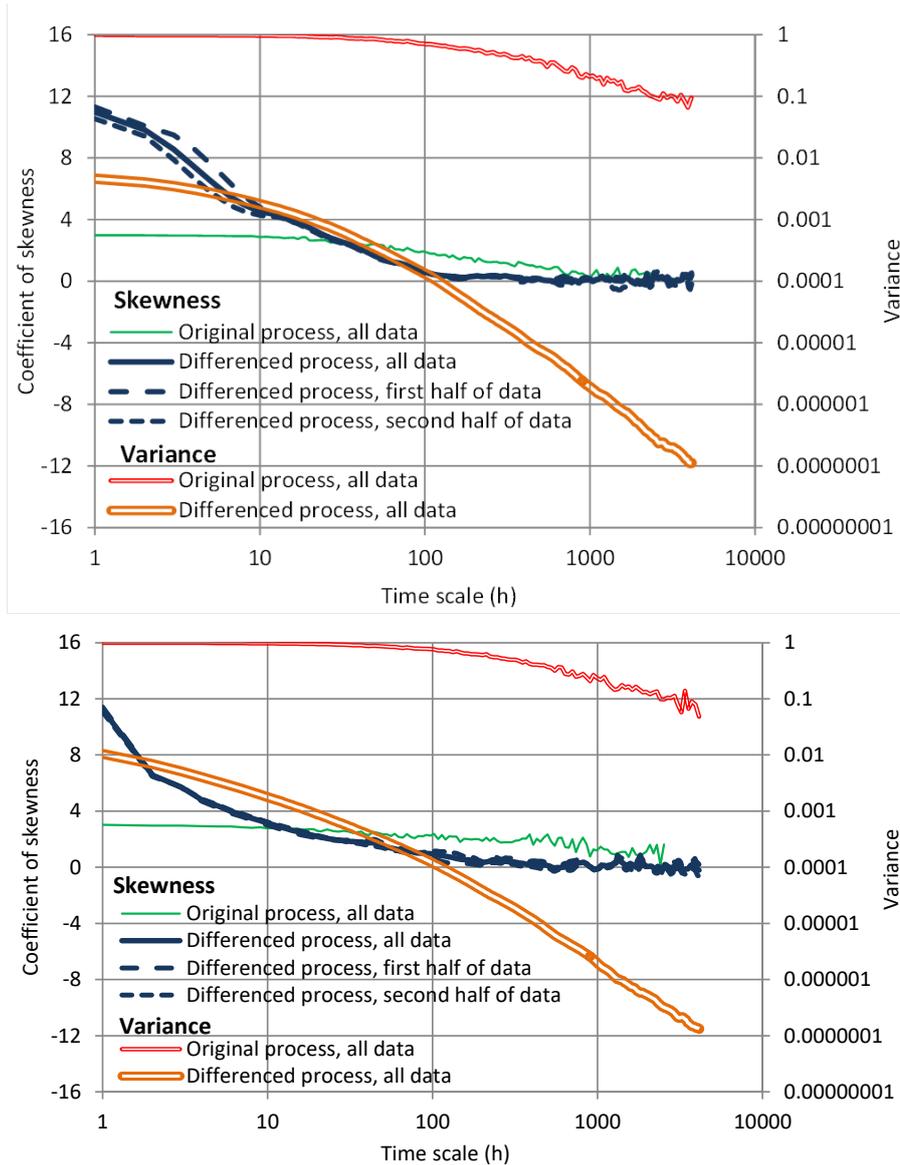
For the application we use the proposed method with constant  $\theta_0$  and with  $J = 1024$ . We note that the SMA case results in zero skewness of the differenced process (time symmetry) while in the ANTAMA case the skewness tends to infinity, which indicates that the method does not have any upper limit of time irreversibility that it can handle. Choosing  $\theta_0 = 0.0638$ , we make the ratio of the skewness of  $\tilde{x}_\tau$  and  $\underline{x}_\tau$  equal to 3.69, as required. The coefficients  $a_\eta$  determined by direct application of the method are shown in Figure 4, along with the climacogram and the autocorrelation function of the original process.



**Figure 4** Climacogram, autocorrelation and coefficients  $a_\eta$  determined by the proposed method.



**Figure 5** (Upper) Discharge time series generated (time references are arbitrary) and (lower) close up for a six-month period containing the highest generated discharges; plots refer to the “naturalized” (back-transformed) series.



**Figure 6** Climacograms and skewness coefficients of the original and differenced time series for: (upper) the real-world data and (lower) the generated data; plots refer to the standardized series.

For the generation of  $v_i$  we use lognormal white noise with skewness determined so that the skewness of the original process be 2.98 as required. A time series with length equal to that of the real world data was derived for the process  $x_\tau$  using equation (24), and was then “naturalized” by applying the inverse seasonal transformation. Plots for the generated time series are shown in Figure 5, where in the close up (lower panel) the irreversibility is evident, with steeper rising branches and milder falling ones. Comparisons of the generated time series with the model, as well as with the real-world data, in terms of climacogram and climacospectrum, are shown in Figure 3. Comparisons of the characteristics related to state and time asymmetry are shown in Figure 6. It is observed that the generated time series is consistent with the model and the original data in terms of all important statistics, marginal and joint, as well as related to time irreversibility. One slight discrepancy is that the curve related to time asymmetry in Figure 6, while it captures the skewness of the real-world differenced data for the

hourly scale, has a steeper slope for intermediate time scales; however, controlling this slope was not in the scope of this study. All in all, the method seems successful in simulation of streamflow time series even at fine time scales.

## 8 Conclusions

The main contribution of this article is the introduction of an equation that determines analytically the weights  $a_\eta$  of an asymmetric moving average (AMA) scheme, which can be used in any problem of stochastic simulation of time irreversible processes. This equation is rewritten here as:

$$a_\eta = \int_{-1/2}^{1/2} \sqrt{2} e^{2\pi i(\theta(\omega) - \eta\omega)} \sqrt{s_d(\omega)} d\omega \quad (45)$$

where  $s_d(\omega)$  is the power spectrum of the discrete-time representation of the process of interest and  $\theta(\omega)$  is any arbitrary odd real function. The same equation can also be used for time reversible processes, even though in this case it suffices to use the simpler symmetric (SMA) scheme, which can be obtained as a special case of the AMA for  $\theta(\omega) = 0$ . In this case the equation simplifies to:

$$a_\eta^S = \int_0^{1/2} \sqrt{2} \cos(2\pi\eta\omega) \sqrt{s_d(\omega)} d\omega \quad (46)$$

A simple expression can also be obtained for any case where  $\theta(\omega)$  is a constant  $\theta_0$  multiplied by the sign of  $\omega$ . Assuming that an adequately large number of weights is used (so that the quantity  $\delta_0$  be negligible) this expression becomes:

$$a_\eta = a_\eta^S \cos(2\pi\theta_0) + a_\eta^A \sin(2\pi\theta_0) \quad (47)$$

where

$$a_\eta^A = \frac{1}{2} \int_0^{1/2} \sqrt{2} \sin(2\pi\eta\omega) \sqrt{s_d(\omega)} d\omega \quad (48)$$

is the antisymmetric solution (ANTAMA) of the equation.

The numerical evaluation of these weights is very easy and fast, as it only needs application of FFT, while the generation of the processes of interest is even easier; both are illustrated in a spreadsheet provided as supplementary information. The applicability of the method has no restrictions and no iterations are needed. Once the process dependence is known through any second-order tool (climacogram, autocovariance, power spectrum, etc.) and the properties related to time reversibility or irreversibility are specified, the method can be readily applied for stochastic generation.

The extended historical review presented in this article reveals that the proposed method is against the current of simulation approaches that were developed since the 1950s. Specifically

the approach proposed (a) uses a purely moving average scheme, without any autoregressive term, (b) involves innovation terms and weights that extend to both directions of time, and (c) does not link the generation scheme with the observations—rather, the observations are used in the model fitting phase, which is independent of the simulation. Because of the moving average scheme we are able to easily handle and preserve moments higher than second. Because of the involvement of terms in both directions of time in an asymmetric mode, we can easily reproduce time irreversibility in a controlled manner. And because of the separation of model fitting from the calculation of the weights of the linear model, we are able to handle any type of dependence in time for whatever range (short or long) and fractality (smooth or rough). For the same reason, we are able to make the model as parsimonious as we wish (and as the data allow) without being trapped to involve many parameters in order to preserve, for example, autocovariances of large lags.

The entire framework is so simple and powerful that one can hardly believe that it is presented for first time. Perhaps it is a reinvention of something already known. Whether it has been invented or not in the past, fact is that it is not well known (and certainly not known to the author), otherwise the ARMA models and their variants would not be so popular. Hopefully this article could contribute to give ARMA-type models a position in the history of stochastics.

## Appendix A: Proof of the basic properties of the AMA solution

To show point (1) below equation (27), we observe that the sequence of  $a_\eta$  is the finite Fourier transform of:

$$A(\omega) := e^{2\pi i \theta(\omega)} A^R(\omega) = (\cos(2\pi\theta(\omega)) + i \sin(2\pi\theta(\omega))) A^R(\omega) \quad (A1)$$

We recall that the power spectrum  $s(\omega)$  is an even function, and so are  $s_d(\omega)$  and  $A^R(\omega)$ . Besides  $\cos(2\pi\theta(\omega))$  is also even, while  $\sin(2\pi\theta(\omega))$  is odd, given that  $\theta(\omega)$  is odd. Consequently,  $A(\omega)$  has even real part and odd imaginary part. Hence, its Fourier transform is real (Bracewell, 2000, p. 13), which proves that  $a_\eta$  is real. It can also be written as a real expression:

$$a_\eta = 2 \int_0^{1/2} \cos(2\pi(\theta(\omega) - \eta\omega)) A^R(\omega) d\omega \quad (A2)$$

To show point (2), we observe that  $A(\omega)$  is the inverse finite Fourier transform of  $a_\eta$ , i.e.,

$$A(\omega) = 2 \sum_{l=-\infty}^{\infty} a_l e^{2\pi i l \omega}$$

Besides, the inverse finite Fourier transform of the quantity  $c'_\eta := \sum_{l=-\infty}^{\infty} a_l a_{\eta+l}$  is

$$\begin{aligned}
B(\omega) &= 2 \sum_{\eta=-\infty}^{\infty} c'_\eta e^{2\pi i \eta \omega} = 2 \sum_{\eta=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} a_l a_{\eta+l} e^{2\pi i \eta \omega} = 2 \sum_{l=-\infty}^{\infty} a_l \sum_{\eta=-\infty}^{\infty} a_{\eta+l} e^{2\pi i \eta \omega} \\
&= 2 \sum_{l=-\infty}^{\infty} a_l \sum_{j=-\infty}^{\infty} a_j e^{2\pi i (j-l)\omega} = 2 \sum_{l=-\infty}^{\infty} a_l e^{-2\pi i l \omega} \sum_{j=-\infty}^{\infty} a_j e^{2\pi i j \omega} = \frac{A(\omega)^* A(\omega)}{2}
\end{aligned} \tag{A3}$$

where  $A(\omega)^*$  is the conjugate of  $A(\omega)$ , equal to  $e^{-2\pi i \theta(\omega)} A^R(\omega)$ . Consequently,

$$B(\omega) = \frac{A(\omega)^* A(\omega)}{2} = \frac{e^{-2\pi i \theta(\omega)} A^R(\omega) e^{2\pi i \theta(\omega)} A^R(\omega)}{2} = \frac{(A^R(\omega))^2}{2} = s_d(\omega) \tag{A4}$$

where we have utilized the definition of  $A^R(\omega)$  in (27). Hence the Fourier transform of  $c'_\eta$  equals the power spectrum, which means that  $c'_\eta$  is identical to the autocovariance function  $c_\eta$ .

Now, coming to point (3), the calculation of the sequence of coefficients  $a_\eta$ , and hence the generation of the required time series, we first note that, unlike in the above theoretical derivations, summation is in practice made for a finite number of terms, i.e., for  $\eta = -J, \dots, 0, \dots, J$ , where  $J$  is finite. The coefficients  $a_\eta$  will decrease with increasing  $\eta$  and will be negligible beyond an appropriately chosen, adequately large,  $J$ . Possible truncation errors can be handled as described by Koutsoyiannis (2016—even though that solution was for the SMA scheme). For the numerical part, since we know the frequency functions  $A^R(\omega)$  (from the power spectrum) and  $\theta(\omega)$  (appropriately or even arbitrarily or randomly chosen), for an array of frequencies  $\omega_j$ , we are able to write equations (30)-(31) and complete the algorithm described in steps (a)-(d) in section 5.

## Appendix B: The OBAMA scheme and other cases requiring optimisation

If the process exhibits LRD and for some reason an OBAMA scheme is desired, one option would be to generalize equation (38) to the form:

$$\theta(\omega) = \theta_0 + \frac{1}{2\pi} \cos^{-1} \left( \left( \frac{1}{C_1 \sqrt{s_d(\omega)}} + \frac{\sqrt{s_d(\omega)}}{C_2} \right)^{C_0} \right), \quad \omega > 0 \tag{B1}$$

where  $\theta_0, C_0, C_1, C_2$ , are coefficients to be optimized numerically, so that the deviation from zero of all  $\eta < 0$  (quantified e.g. by the mean square error) be minimal. Continuity at 0 of  $\theta(\omega)$  is not necessary, but if required it can be achieved setting  $\theta(0) = 0$ .

A second option which was found quite satisfactory in numerical explorations, is to define  $\theta(\omega)$  as the smooth minimum of two rational functions of frequency,  $\theta_i(\omega), i = 1, 2$ , with first degree polynomials, i.e.:

$$\theta(\omega) = \frac{1}{\zeta} \ln(e^{\zeta \theta_1(\omega)} + e^{\zeta \theta_2(1/2-\omega)}), \quad \theta_i(\omega) := \frac{C_{1,i} \omega}{C_{2,i} + \omega} + C_{0,i} \tag{B2}$$

where  $C_{ji}; i = 1, 2; j = 0, 1, 2$  are coefficients with values to be optimized and  $\zeta$  a big negative number, say,  $\zeta = -10$ . Optimizing each  $\theta(\omega_j)$ , without using parametric relationships, is

another option but in the LRD case the frequencies of interest  $\omega_j$  will be numerous and, therefore, an expression such as (B1) or (B2), whose coefficients are optimized, may be an advantage.

We can use the same expressions for other tasks. For example, if for some reason we need an ordinary-backward scheme of ARMA type, again the framework provided can directly convert itself into an ARMA scheme. Thus, to convert AMA to an ARMA(1,  $q$ ) model, for which  $a_\eta = 0$  for  $\eta < 0$  and  $a_\eta = \beta a_{\eta-1} = \beta^{\eta-q} a_q$  for  $\eta > q$  and for some appropriate  $\beta$ , we can form the objective function (representing the cumulative departure of the achieved sequence of coefficients from those of an ARMA(1,  $q$ ) model):

$$e(a_0, \dots, a_q; \beta) := \sum_{\eta=-J}^0 a_\eta^2 + \sum_{\eta=q}^J (a_\eta - \beta^{\eta-q} a_q)^2 \quad (\text{B3})$$

minimize it finding the vector of coefficients  $(a_0, \dots, a_q; \beta)$ , and reformulate the generating equation (24) as:

$$\underline{x}_\tau = \beta \underline{x}_{\tau-1} + \sum_{\eta=0}^q a_\eta \underline{v}_{\tau-\eta} \quad (\text{B4})$$

so that a limited number of previous noise variables  $\underline{v}_{\tau-\eta}$  and the latest process variable  $\underline{x}_{\tau-1}$  are only needed. This case is also implemented in the spreadsheet provided as supplementary information. The same spreadsheet file contains also other cases, such as the conversion into an AR( $q$ ) model or to the sum of three AR(1) models, where the algorithms used are hopefully self-explanatory.

**Acknowledgments.** I thank the eponymous reviewers William Farmer and Alberto Montanari for their extraordinarily constructive comments which encouraged me to broaden, completely restructure and, eventually, substantially improve the initial manuscript. During the preparation of the second version of the manuscript, I had the pleasure to present its content in a seminar in the Sapienza University of Rome as well as in the First Workshop on Control Methods for Water Resource Systems in Delft. I wish to express my gratitude to Francesco Napolitano for inviting and organizing my visit in Rome during September 2019, and Alla Kolechkina and Ronald van Nooyen for the invitation in the Workshop in Delft.

**Funding information.** No funding was provided for this research.

**Conflicts of interest.** No conflict of interest exists.

## References

- Bailey, A.L., 1929. A summary of advanced statistical methods. US Statistical Bureau, <https://babel.hathitrust.org/cgi/pt?id=mdp.39015067991326&view=1up&seq=7>.
- Barnes, F.B., 1954. Storage required for a city water supply. *J. Inst. Eng., Australia*, 26, 198-203.
- Barnett, V., 2006. Chancing an interpretation: Slutsky's random cycles revisited. *Euro. J. History of Economic Thought*, 13(3), 411-432, doi: 10.1080/09672560600875596.

- Beard, L.R., 1965. Use of interrelated records to simulate streamflow. *Proc. ASCE, J. Hydraul. Div.*, 91(HY5), 13-22.
- Birkhoff, G. D., 1931. Proof of the ergodic theorem. *Proc. Nat. Acad. Sci.*, 17, 656–660.
- Boltzmann, L., 1877. Über die Beziehung zwischen dem zweiten Hauptsatz der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respektive den Sätzen über das Wärmegleichgewicht. *Wiener Ber.*, 76, 373–435 (in German).
- Boltzmann, L., 1884/85. Über die Eigenschaften monocyclischer und anderer damit verwandter Systeme. *Crelles J.*, 98, 68–94.
- Bortkiewicz, L., 1917. Die Iterationen — Ein Beitrag zur Wahrscheinlichkeitstheorie. Springer, Berlin.
- Box, G.E., and Jenkins, G.M., 1970. *Time Series Models for Forecasting and Control*. Holden Day, San Francisco, USA.
- Bracewell, R.N., 2000. *The Fourier Transform and Its Applications*. 3rd Ed., McGraw-Hill, Boston, Mass., USA.
- Chow, V.T., 1969. Stochastic analysis of hydrologic systems. Final Report, University of Illinois. Urbana, Illinois, [https://www.ideals.illinois.edu/bitstream/handle/2142/90345/Chow\\_1969.pdf?sequence=2](https://www.ideals.illinois.edu/bitstream/handle/2142/90345/Chow_1969.pdf?sequence=2).
- Chow, V.T., and Karelitis, S.J., 1970. Analysis of stochastic hydrologic systems. *Water Resources Research*, 6(6), 1569-1582.
- Dimitriadis, P., and Koutsoyiannis, D., 2015. Climacogram versus autocovariance and power spectrum in stochastic modelling for Markovian and Hurst–Kolmogorov processes, *Stochastic Environmental Research & Risk Assessment*, 29 (6), 1649–1669, doi:10.1007/s00477-015-1023-7.
- Dimitriadis, P., and Koutsoyiannis, D., 2018. Stochastic synthesis approximating any process dependence and distribution, *Stochastic Environmental Research & Risk Assessment*, 32 (6), 1493–1515, doi:10.1007/s00477-018-1540-2.
- Dodge, Y., 1996. A natural random number generator. *International Statistical Review*, 64(3), 329-344.
- Eckhardt, R., 1989. Stan Ulam, John von Neumann and the Monte Carlo method. *From Cardinals to Chaos*, ed. By N. G. Cooper, Cambridge University, NY.
- Galton, F., 1890. Dice for statistical experiments. *Nature*, 42, 13–14.
- Gauch, H.G., Jr., 2003. *Scientific Method in Practice*. Cambridge University Press, Cambridge, UK.
- Hamed, K.H., 2008. Trend detection in hydrologic data: the Mann–Kendall trend test under the scaling hypothesis. *Journal of Hydrology*, 349(3-4), pp.350-363.
- Hazen, A., 1914. Storage to be provided in impounding reservoirs for municipal water supply. *Trans. Amer. Soc. Civil Eng.*, 77, 1539-1640.
- Hosking, J. R.M., 1981. Fractional differencing, *Biometrika*, 68(1), 165–176, doi: 10.1093/biomet/68.1.165.
- Hurst, H.E., 1951. Long term storage capacities of reservoirs. *Trans. Am. Soc. Civil Eng.*, 116, 776–808.
- Kendall, M.G., 1938. A new measure of rank correlation. *Biometrika*, 30(1/2), 81-93.
- Khintchine, A., 1933. Zu Birkhoffs Lösung des Ergodenproblems. *Math. Ann.*, 107, 485-488.
- Khintchine, A., 1934. Korrelationstheorie der stationären stochastischen Prozesse. *Mathematische Annalen*, 109 (1), 604–615.
- Klemes, V., 1987. One hundred years of applied storage reservoir theory. *Water Resources Management*, 1(3), 159-175.
- Kolmogorov, A.N., 1931. Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung, *Mathematische Annalen*, 104, 415-458. (English edition: Kolmogorov, A.N., 1992. On analytical methods in probability theory. *Selected Works of A. N. Kolmogorov - Volume 2, Probability Theory and Mathematical Statistics*, A. N. Shiriyayev, ed., Kluwer, Dordrecht, The Netherlands, pp. 62-108).

- Kolmogorov, A. N., 1933. *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Ergebnisse der Mathematik und Ihrer Grenzgebiete), Berlin. (2nd English Edition: *Foundations of the Theory of Probability*, 84 pp. Chelsea Publishing Company, New York, 1956).
- Kolmogorov, A. N., 1937. Ein vereinfachter Beweis der Birkhoff-Khintchineschen Ergodensatzes. *Mat. Sb.*, 2, 367-368.
- Kolmogorov, A.N., 1938. A simplified proof of the Birkhoff-Khinchin ergodic theorem. *Uspekhi Mat. Nauk*, 5, 52-56. (English edition: Kolmogorov, A.N., 1991, *Selected Works of A. N. Kolmogorov - Volume 1, Mathematics and Mechanics*, Tikhomirov, V. M. ed., Kluwer, Dordrecht, The Netherlands, pp. 271-276).
- Kolmogorov, A.N., 1940. Wiener'sche Spiralen und einige andere interessante Kurven im Hilbertschen Raum. *Dokl. Akad. Nauk SSSR*, 26, 115-118. (English edition: Kolmogorov, A.N., 1991. Wiener spirals and some other interesting curves in a Hilbert space. *Selected Works of A. N. Kolmogorov - Volume 1, Mathematics and Mechanics*, Tikhomirov, V. M. ed., Kluwer, Dordrecht, The Netherlands, pp. 303-307).
- Kolmogorov, A.N., 1947. Statistical theory of oscillations with continuous spectrum. *Collected papers on the 30th anniversary of the Great October Socialist Revolution*, Vol. 1, Akad. Nauk SSSR, Moscow-Leningrad, pp. 242-252. (English edition: Kolmogorov, A.N., 1992. *Selected Works of A. N. Kolmogorov - Volume 2, Probability Theory and Mathematical Statistics*, A. N. Shirayev ed., Kluwer, Dordrecht, The Netherlands, pp. 321-330).
- Koutsoyiannis, D., 2000. A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series. *Water Resources Research*, 36 (6), 1519-1533, doi: 10.1029/2000WR900044.
- Koutsoyiannis, D., 2010. A random walk on water. *Hydrology and Earth System Sciences*, 14, 585-601, doi: 10.5194/hess-14-585-2010.
- Koutsoyiannis, D., 2016. Generic and parsimonious stochastic modelling for hydrology and beyond. *Hydrological Sciences Journal*, 61 (2), 225-244, doi: 10.1080/02626667.2015.1016950.
- Koutsoyiannis, D., 2017. Entropy production in stochastics. *Entropy*, 19 (11), 581, doi: 10.3390/e19110581.
- Koutsoyiannis, D., 2019a. Knowable moments for high-order stochastic characterization and modelling of hydrological processes, *Hydrological Sciences Journal*, 64 (1), 19-33, doi: 10.1080/02626667.2018.1556794.
- Koutsoyiannis, D., 2019b. Time's arrow in stochastic characterization and simulation of atmospheric and hydrological processes, *Hydrological Sciences Journal*, 64 (9), 1013-1037, doi: 10.1080/02626667.2019.1600700.
- Koutsoyiannis, D., and Montanari, A., 2015. Negligent killing of scientific concepts: the stationarity case. *Hydrological Sciences Journal*, 60 (7-8), 1174-1183.
- Kyun, K. and Kim, K., 2006. *Equilibrium business cycle theory in historical perspective*. Cambridge University Press.
- Lawrance, A.J., 1991. Directionality and reversibility in time series. *International Statistical Review*, 59(1) 67-79.
- Lombardo, F., Volpi, E., Koutsoyiannis, D., and Papalexiou, S.M., 2014. Just two moments! A cautionary note against use of high-order moments in multifractal models in hydrology, *Hydrology and Earth System Sciences*, 18, 243-255, doi: 10.5194/hess-18-243-2014.
- Mackey, M.C., 2003. *Time's Arrow: The Origins of Thermodynamic Behavior*, Dover, Mineola, NY, USA, 175 pp.
- Mahalanobis, P.C., 1934. Tables of random samples from a normal population. *Sankya*, 1, 289-328.
- Mandelbrot, B.B. and Wallis, J.R., 1969. Computer Experiments with Fractional Gaussian Noises: Part 1, Averages and Variances. *Water Resources Research*, 5, 228-241.
- Mann, H.B., 1945. Nonparametric tests against trend. *Econometrica*, 13 (3), 245-259.
- Matalas, N.C., 1967. Mathematical assessment of synthetic hydrology, *Water Resour. Res.*, 3(4), 937-945.

- Mathieu, M., 1988. On the origin of the notion 'Ergodic Theory', *Expositiones Mathematicae*, 6, 373-377.
- Mazliak, L., 2018. The beginnings of the Soviet encyclopedia. The utopia and misery of mathematics in the political turmoil of the 1920s. *Centaurus*, 60(1-2), 25-51.
- Metropolis, N., 1989. The beginning of the Monte Carlo method. *From Cardinals to Chaos*, ed. by N. G. Cooper, Cambridge University, NY.
- Metropolis, N., Reitwiesner, G., and von Neumann, J., 1950. Statistical treatment of values of first 2,000 decimal digits of  $e$  and of  $\pi$  calculated on the ENIAC. *Mathematical Tables and Other Aids to Computation*, 4, 109-112.
- Metropolis, N. and Ulam, S.M., 1949. The Monte Carlo method. *J. Amer. Statist. Assoc.*, 44, 335-341.
- Montanari, A., and Koutsoyiannis, D., 2014. Modeling and mitigating natural hazards: Stationarity is immortal!, *Water Resources Research*, 50(12), 9748-9756, doi:10.1002/2014WR016092.
- Niederreiter, H., 1992. *Random Number Generation and Quasi-Monte Carlo Methods*, Society for Industrial and Applied Mathematics, Philadelphia.
- O'Connell, P.E., Koutsoyiannis, D., Lins, H.F., Markonis, Y., Montanari, A., and Cohn, T.A., 2016. The scientific legacy of Harold Edwin Hurst (1880 - 1978). *Hydrological Sciences Journal*, 61 (9), 1571-1590, doi: 10.1080/02626667.2015.1125998.
- Papoulis, A., 1991. *Probability, Random Variables and Stochastic Processes*, 3rd edn., McGraw-Hill, New York.
- Persons, W.M. 1919. *Measuring and Forecasting General Business Conditions*. Harvard University & American Institute of Finance, <https://babel.hathitrust.org/cgi/pt?id=hvd.32044018749697>
- Reitwiesner, G., 1950. An ENIAC Determination of  $\pi$  and  $e$  to more than 2000 Decimal Places. *Mathematical Tables and Other Aids to Computation*, 4, 11-15.
- Slutsky, E., 1925, Über stochastische Asymptoten und Grenzwerte, *Metron*, 5(3), 3-89.
- Slutsky, E., 1927. Slozhenie sluchainykh prichin, kak istochnik tsiklicheskich protsessov. *Voprosy kon'yunktury*, 3, 34 - 64. 1927 [English edition: Slutsky, E., 1937. The summation of random causes as the source of cyclic processes. *Econometrica: Journal of the Econometric Society*, 105-146].
- Slutsky, E., 1928a. Sur un critérium de la convergence stochastique des ensembles des valeurs éventuelles). *Comptes rendus Acad. Sci.*, 187, 370.
- Slutsky, E., 1928b. Sur les fonctions éventuelles continues, intégrables et dérivables dans le sens stochastiques. *Comptes rendus des séances de l'Académie des Sciences*, 187, 878.
- Slutsky, E., 1929. Quelques propositions sur les limites stochastiques éventuelles. *Comptes rendus des séances de l'Académie des Sciences*, 189, 384.
- Stigler, S.M., 2002. *Statistics on the Table: The History of Statistical Concepts and Methods*. Harvard University Press.
- Sudler, C., 1927. Storage required for the regulation of streamflow. *Trans. Am. Soc. Civ. Eng.*, 91, 622-660.
- Thomas, H.A., and Fiering, M.B. 1962. Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. *Design of Water Resource Systems*, by A. Maass, M.M. Hufschmidt, R. Dorfman, H.A. Thomas, Jr., S.A. Marglin and G.M. Fair, Harvard University Press, Cambridge, Mass.
- Tippett, L.H.C., 1927. *Random Sampling Numbers*. Tracts for Computers No. XV, Camb. Univ. Press.
- Tong, H. and Zhang, Z., 2005. On time-reversibility of multivariate linear processes. *Statistica Sinica*, 15(2), 495-504.
- Walker, G., 1931. On periodicity in series of related terms, *Proceedings of the Royal Society of London, Ser. A*, 131, 518-532.
- Weiss, G., 1975. Time-reversibility of linear stochastic processes. *Journal of Applied Probability*, 12(4), 831-836.
- Whittle, P., 1951. *Hypothesis Testing in Times Series Analysis*. PhD thesis, Almqvist & Wiksells, Uppsala.

- Whittle, P., 1952. Tests of fit in time series. *Biometrika*, 39(3/4), 309-318.
- Whittle, P., 1953. The analysis of multiple stationary time series. *Journal of the Royal Statistical Society B*, 15(1), 125-139.
- Wold, H.O., 1938. *A Study in the Analysis of Stationary Time-Series*. PhD thesis, Almqvist and Wicksell, Uppsala.
- Wold, H.O., 1948. On prediction in stationary time series. *The Annals of Mathematical Statistics*, 19(4), 558-567.
- Yule, G.U. 1927. On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers, *Philosophical Transactions of the Royal Society of London A*, 226, 267-298.