IFAC
INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL

1ST WORKSHOP ON CONTROL METHODS FOR WATER RESOURCE SYSTEMS

19 - 20 SEPTEMBER 2019 DELFT, NETHERLANDS

# Stochastic simulation of time irreversible processes for use in hydrosystem control problems

Demetris Koutsoyiannis

Department of Water Resources and Environmental Engineering
School of Civil Engineering, National Technical University of Athens, Greece
(dk@itia.ntua.gr, http://www.itia.ntua.gr/dk/)

Visiting professor, Dipartimento di Ingegneria Civile, Edile ed Ambientale,
Sapienza - Università degli studi di Roma, Roma, Italia

**Presentation available online: http://www.itia.ntua.gr/1985/**

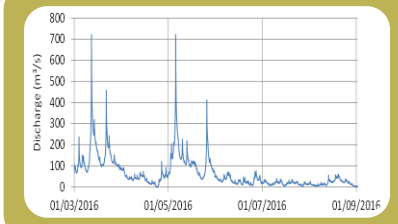
# Part A
# Introduction to stochastic simulation

# Schools of thought converging to stochastic simulation

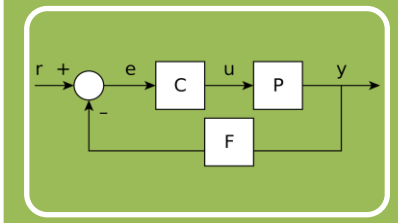

Monte Carlo method

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T g\left(\underline{x}(t)\right)dt = \mathrm{E}[g(\underline{x}(t))]$$

Theory of stochastic processes




Time series analysis




Systems analysis and control

# Prehistory of Monte Carlo simulation

- 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 π (if line distance = needle length, π = 2 / Probability of crossing a line).
- Galton (1890) invented a set of 3 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).]

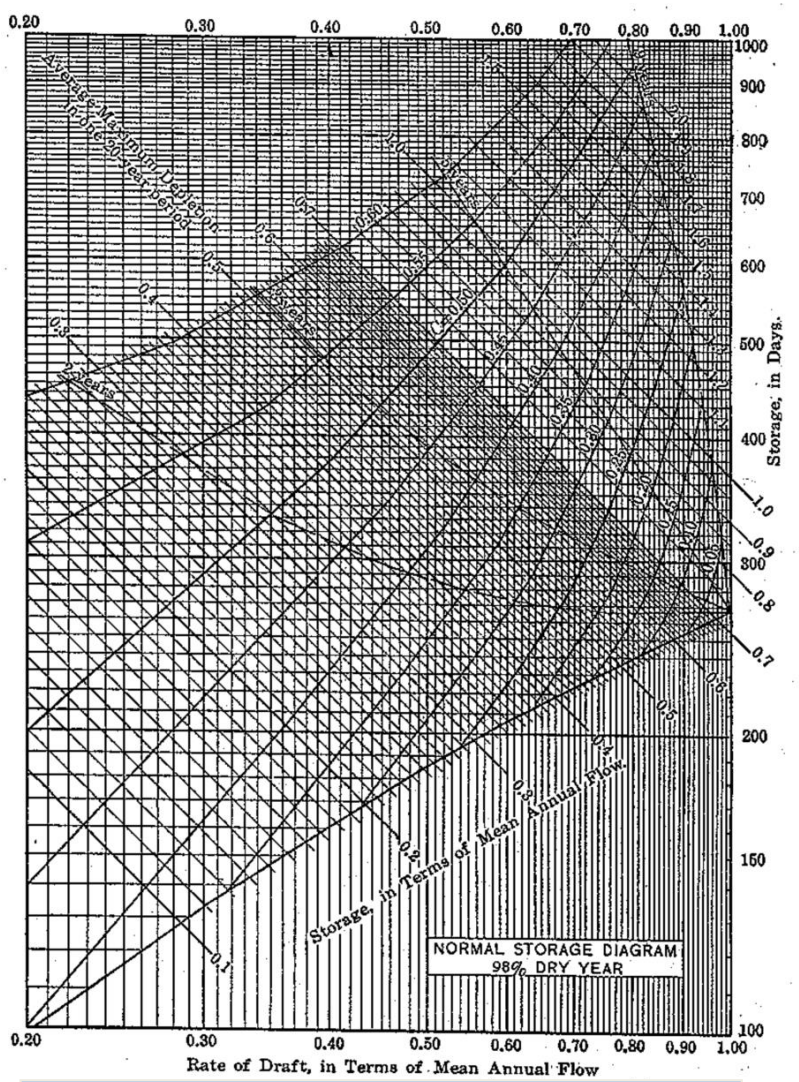
SANKHYĀ : THE INDIAN JOURNAL OF STATISTICS

PLATE 26.

| 201 | 202 | 203 | 204 | 205 | 206 | 207 | 208 |
|---|---|---|---|---|---|---|---|
| −0·467 | +1·189 | +1·008 | +1·888 | +0·190 | +1·671 | +0·606 | −0·902 |
| −0·749 | +0·146 | +2·117 | +1·228 | +0·005 | +1·152 | −1·857 | −0·118 |
| +2·880 | −1·275 | +0·904 | +1·014 | −0·434 | −0·098 | −0·120 | −1·564 |
| −1·950 | +0·706 | +0·899 | +0·878 | −1·067 | −0·510 | +1·081 | −0·225 |
| −1·185 | −0·015 | −0·846 | +1·850 | −1·699 | −1·112 | −1·134 | −0·017 |

# Prehistory of Monte Carlo simulation in hydrosystems

- Hazen (1914): 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
- Hurst (1951): Performed physical experiments to generate random numbers:
  - Tossing 10 coins (sixpences) 1025 times at a rate of 100 random numbers per 35 min; note that 10 binary digits are equivalent to about 3 decimal digits.
  - Shuffling and cutting a pack of 52 cards at a rate of 100 random numbers per 20 min.




Hazen's graph of the reservoir storage-yield-reliability relationship

# Stanislaw Ulam, the solitaire and the conception of the Monte Carlo method

## STAN ULAM, JOHN VON NEUMANN, *and the* MONTE CARLO METHOD

*by Roger Eckhardt*

The Monte Carlo method is a statistical sampling technique that over the years has been applied successfully to a vast number of scientific problems. Although the computer codes that implement Monte Carlo have grown ever more sophisticated, the essence of the method is captured in some unpublished remarks Stan made in 1983 about solitaire.

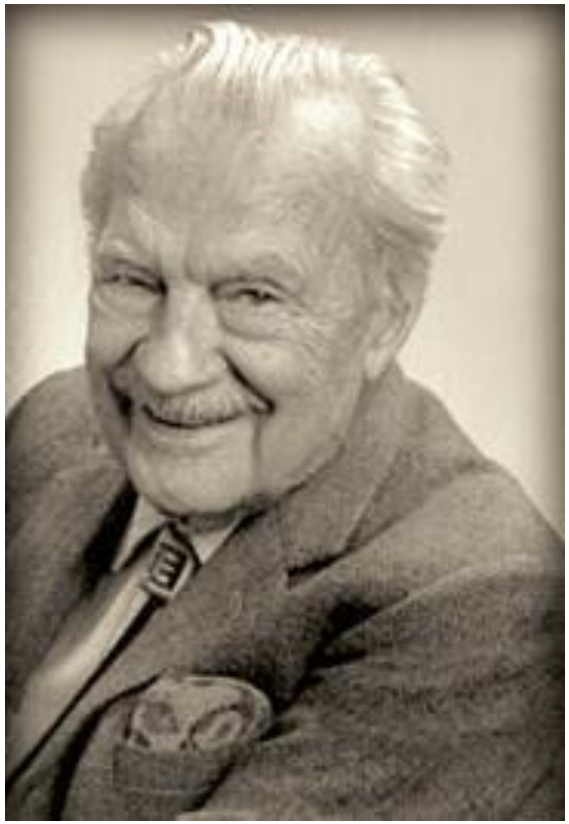
"The first thoughts and attempts I made to practice [the Monte Carlo method] were suggested by a question which occurred to me in 1946 as I was convalescing from an illness and playing solitaires. The question was what are the chances that a Canfield solitaire laid out with 52 cards will come out successfully? After spending a lot of time trying 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. This was already possible to envisage with the beginning of the new era of fast computers, and I immediately thought of problems of neutron diffusion and other questions of mathematical physics, and more generally how to change processes described by certain differential equations into an equivalent form interpretable as a succession of random operations. Later... [in 1946, I] described the idea to John von Neumann and we began to plan actual calculations."

Von Neumann was intrigued. Statistical sampling was already well known in mathematics, but he was taken by the idea of doing such sampling using the newly developed electronic computing techniques. The approach seemed especially suitable for exploring the behavior of neutron chain reactions in fission devices. In particular, neutron multiplication rates could be estimated and used to predict the explosive behavior of the various fission weapons then being designed.

In March of 1947, he wrote to Robert Richtmyer, at that time the Theoretical Division Leader at Los Alamos (Fig. 1). He had concluded that "the statistical approach is very well suited to a digital treatment," and he outlined in some detail how this method could be used to solve neutron diffusion and multiplication problems in fission devices for the case "of 'inert' criticality" (that is, approximated as momentarily static config-

**Stanislaw Ulam** (13 April 1909 – 13 May 1984): Polish-American mathematician; since 1943 he worked in Los Alamos National Laboratory (Manhattan Project under leadership of Robert Oppenheimer)

# Nicholas Metropolis and the "birth certificate" of the Monte Carlo method



**Nicholas Metropolis** (11 June 1915 – 17 October 1999): Greek-American physicist; since April 1943 he worked in the Manhattan Project in Los Alamos

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

*Number 247* SEPTEMBER 1949 *Volume 44*

## THE MONTE CARLO METHOD

NICHOLAS METROPOLIS AND S. ULAM
*Los Alamos Laboratory*

We shall present here the motivation and a general description of a method dealing with a class of problems in mathematical physics. The method is, essentially, a statistical approach to the study of differential equations, or more generally, of integro-differential equations that occur in various branches of the natural sciences.

ALREADY in the nineteenth century a sharp distinction began to appear between two different mathematical methods of treating physical phenomena. Problems involving only a few particles were studied in classical mechanics, through the study of systems of ordinary differential equations. For the description of systems with very many particles, an entirely different technique was used, namely, the method of statistical mechanics. In this latter approach, one does not concentrate on the individual particles but studies the properties of *sets of particles*. In pure mathematics an intensive study of the properties of

# Integration: Classical numerical method

- In the numerical integration of a function $f$ of a scalar variable $u$, a definite integral is approximated by the relationship (known as the trapezoidal rule):

$$\int_0^1 f(u)\mathrm{d}u \approx \sum_{n=0}^{m} w_n f\left(\frac{n}{m}\right) \tag{1}$$

  where $m$ is a positive integer and $w_n$ denotes a weight, equal to $1/2m$ for the endpoints $n = 0$ and $n = m$, and equal to $1/m$ for all intermediate $n$.

- Likewise, in the numerical integration of a function of a vector variable of size $s$ in the space $I^s := [0,1]^s$, the relationship becomes:

$$\int_{I^s} f(\boldsymbol{u})\mathrm{d}\boldsymbol{u} \approx \sum_{n_1=0}^{m} \dots \sum_{n_s=0}^{m} w_{n_1} \dots w_{n_s} f\left(\frac{n_1}{m}, \dots, \frac{n_s}{m}\right) \tag{2}$$

- The computational nodes form a rectangular grid with equidistance $1/m$.
- Their number is $N = (m + 1)^s$ and the computational error is $O(m^{-2}) = O(N^{-2/s})$.
- Consequently, for a specified acceptable error, $N$ increases exponentially with $s$ (**curse of dimensionality**).

# Integration: The Monte Carlo method

- In the Monte Carlo integration, the $N$ points for the evaluation of $f(u)$ are taken at random (rather than at the nodes of a grid) and the weight is $1/N$, so that (Niederreiter, 1992):

$$\int_{I^s} f(\boldsymbol{u})\mathrm{d}\boldsymbol{u} \approx \frac{1}{N}\sum_{n=1}^{N} f(\boldsymbol{x}_n) \tag{3}$$

  where $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N$ are independent random points over the space $I^s$.

- For an arbitrary integration space $B$ the relationship becomes:

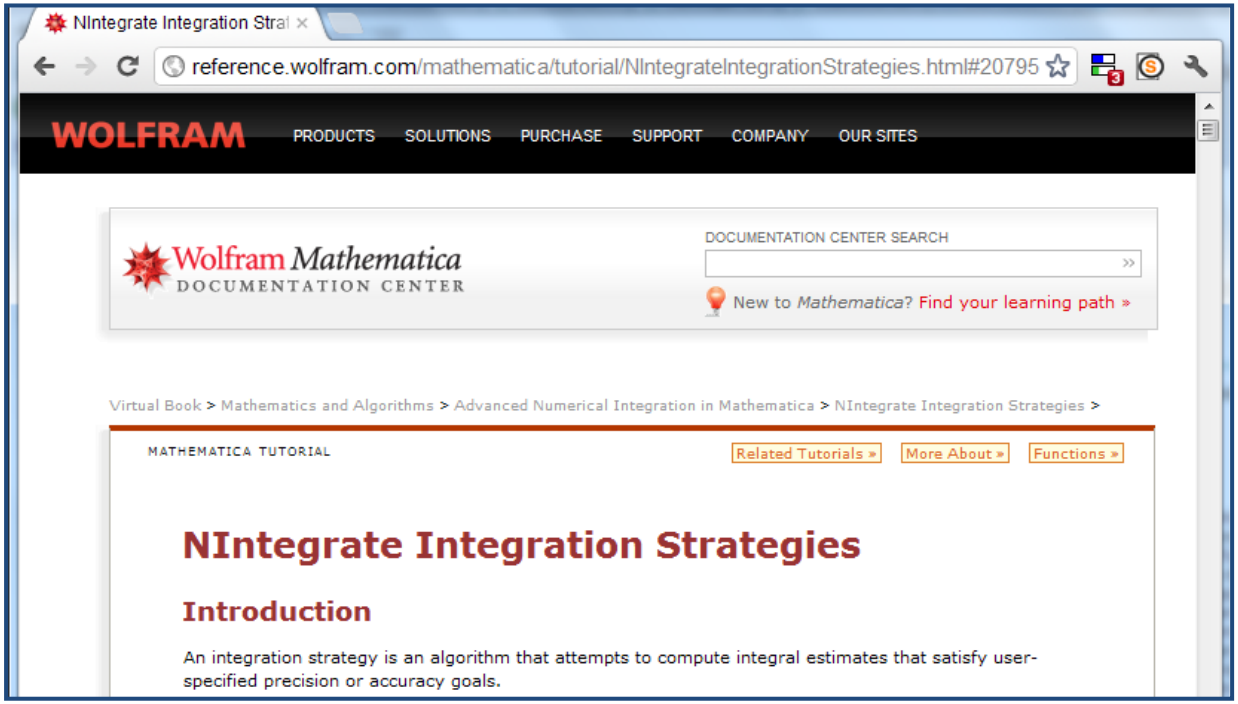
$$\int_{B} f(\boldsymbol{u})\mathrm{d}\boldsymbol{u} \approx \frac{1}{N}\sum_{n=1}^{N} f(\boldsymbol{x}_n)U_B(\boldsymbol{x}_n)\,, \qquad U_B(\boldsymbol{x}_n) \coloneqq \begin{cases} 1, & \boldsymbol{x}_n \in B \\ 0 & \boldsymbol{x}_n \notin B \end{cases} \tag{4}$$

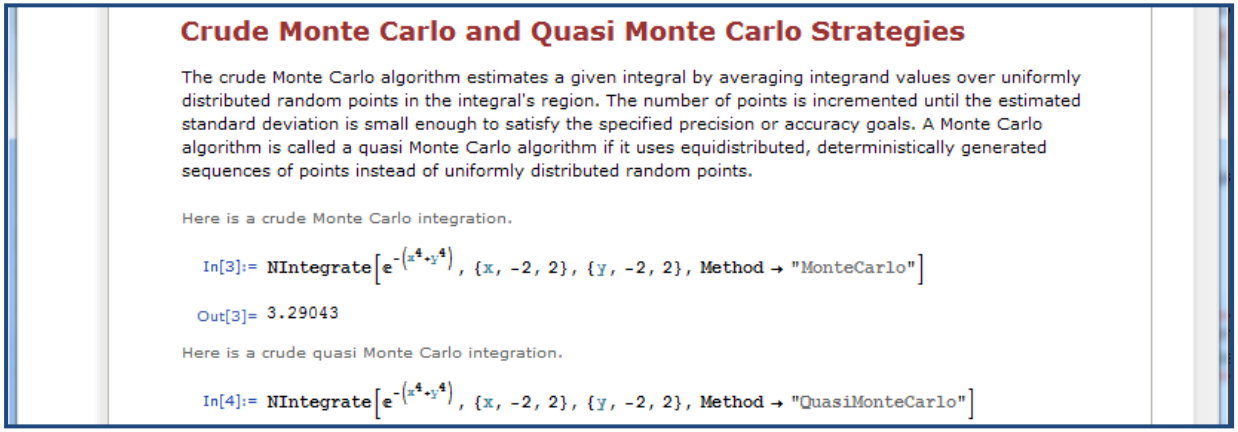
  According to a classical statistical law, the computational error is $O(N^{-1/2})$.

- **Observation**: The error **does not depend on the dimensionality** $s$.
- **Conclusion**: Comparing the errors of the classical and Monte Carlo methods, we readily obtain that the latter is preferable when the dimensionality $> 4$ .
- **Remark**: For large dimensionality $s$, e.g. > 20, the classical method is infeasible while the Monte Carlo is always feasible.

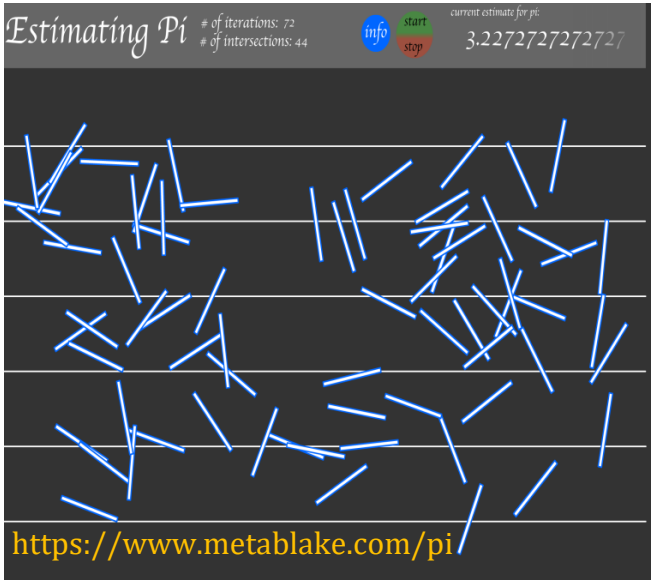
# The Monte Carlo method is part of routine numerical modelling

- The screen on the right shows how the *Mathematica* software implements various versions of the Monte Carlo method for numerical integration.
- This is not just an additional option within a repertoire of available options; for high-dimensional spaces it is the only possibility.



# Archimedes' constant π in Monte Carlo simulation




- LeClerc's Monte Carlo method to calculate π became popular among scientists and his experiment was later repeated by many, most famous of whom became the Italian Lazzarini for reporting in 1902 an agreement up to sixth digit with just 3408 trials.
- There are many other Monte Carlo algorithms to estimate π (one will appear in Koutsoyiannis, 2020). However, these are good only for fun, as much faster and much more accurate deterministic algorithms exist to calculate π.
- Reitwiesner (1950) calculated by a deterministic algorithm, running on the ENIAC computer, the first 2035 decimal digits of π. Metropolis et al. (1950) examined their randomness, an exercise made thereafter many times showing that the digits of π have no apparent pattern and pass tests for statistical randomness.
- Dodge (1996) promoted an idea opposite to LeClerc's: that the digits of π form a "Natural Random Number Generator".
- Since January 2019, 31.4 trillion digits of π are known (found by the Chudnovsky algorithm*; equivalent to ~100 million books, 1000 pages each—note, the British Library has 25 million books) and thus can serve as a basis for any simulation experiment. However, simple random generators are more economic and convenient.

* https://cloud.google.com/blog/products/compute/calculating-31-4-trillion-digits-of-archimedes-constant-on-google-cloud

# Memorable moments in the history of stochastics

**Ludwig Boltzmann**

(1844 –1906, Universities of Graz and Vienna, Austria, and Munich, Germany)

**1877** Explanation of the concept of **entropy** in probability theoretic context.

**1884/85** Introduction of the notion of **ergodic*** systems which however he called "isodic"

* The term is etymologized from Greek words but which ones exactly is uncertain (options: (a) έργον + οδός; (b) έργον + είδος; (c) εργώδης; see Mathieu, 1988).

**George D. Birkhoff**

(1884 – 1944; Princeton, Harvard, USA)

**1931** Discovery of the **ergodic (Birkhoff–Khinchin) theorem**

**Aleksandr Khinchin**

(1894 – 1959; Moscow State University, Russia)

**1933** Purely measure-theoretic proof of the **ergodic (Birkhoff–Khinchin) theorem**

**1934** Definition of **stationary stochastic processes** and probabilistic setting of the **Wiener-Khinchin theorem** relating autocovariance and power spectrum

**Andrey N. Kolmogorov**

(1903 – 1987; Moscow State University, Russia)

**1931** Introduction of the terms **process** to describe change of a certain system and **stationary** to describe a probability density function that is unchanged in time

**1933** Definition of the concepts of **probability** & **random variable**

**1937-1938** Probabilistic exposition of the **ergodic (Birkhoff–Khinchin) theorem** and **stationarity**

**1947** Definition of **wide sense stationarity**

# The foundation of stochastics by the Moscow School of Mathematics

- Kolmogorov (1931) introduced the term *stochastic process*, identifying "process" with "change".* He also used the term *stationary* to describe a probability density function that is unchanged in time.
- Kolmogorov (1933) introduced the definition of *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).
- Khinchin (1934) gave a more formal definition of a *stochastic process* and *stationarity*.
- Kolmogorov (1938) gave a concise presentation of the concepts:

  [...] a stationary **stochastic process** [...] **is a set of random variables** $x_t$ depending on the parameter $t$, $-\infty < t < +\infty$, such that the distributions of the systems $(x_{t_1}, x_{t_2}, \dots, x_{t_n})$ and $(x_{t_1+\tau}, x_{t_2+\tau}, \dots, x_{t_n+\tau})$ coincide for any $n, t_1, t_2, \dots, t_n$ and $\tau$.

* Kolmogorov cited Bachelier (1900; French mathematician) as having already used stochastic processes but Bachelier did not use that name. 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, 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 equivalent term "random") must have been related to political and ideological reasons (incongruence with the *dialectic materialism*): models beyond strict deterministic were considered with a priori suspicion (Mazliak, 2018).

# The onset of Time Series Analysis by economists

- Perhaps the first definition of a time series was given 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.*

- Ten years earlier, W.M. Persons (1919) (American economist), in studying the problem "*When to buy or sell*", he introduced analysis of time series, which he called *statistical series*, and asserted that they:

  *result from the combination of four elements: secular* ***trend***, ***seasonal*** *variation,* ***cyclical fluctuation****, and a* ***residual*** *factor.* [Trap 1]

- He also proposed methods for "*Eliminating secular trends*" and "*Eliminating seasonal variation*". [Trap 2]
- Slutsky (1927) (Ukrainian/Russian/Soviet mathematical statistician and economist) demonstrated that what was regarded as **cyclical component is nothing but a statistical artefact** with no essential meaning (see e.g. Kyun and Kim 2006; Barnett, 2006).
- 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.

# The Uppsala school and the stochastic foundation of time series analysis

- Wold (1938) (Norwegian-born econometrician and statistician with career in Sweden) proved that a time series (more precisely, a stochastic process) 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) (New-Zealand-born mathematician and statistician) 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($p$), MA($q$), ARMA($p$,$q$), ARIMA($p$,$d$,$q$); they became popular with these names and have been also 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). These are made by replacing the integer parameter $d$ in ARIMA($p$,$d$,$q$) with a real one (fractional differencing) and are usually termed ARFIMA($p$,$d$,$q$).

# The onset of time series modelling in hydrology

- 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.
- 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 seems that hydrologists have followed (and today still do) the "Time Series School" rather than the more rigorous "Stochastic School".

# Problems with the "Time Series School"

- The term *time series* is ambiguous, sometimes denoting a a series of observations as in the original definition of Bailey (1929) or a realization of a stochastic process, and other times denoting the stochastic process *per se*.
  [Note: Here the term is used with the first meaning, a series of numbers, not of random variables.]
- Time series analysis separated from the stochastic theory may be meaningless. Popular computer programs have made calculations easy and fast, but numerical results may mean nothing if we are ignorant of stochastics.
- With the exception of AR(1) and ARMA(1,1), time series models are too artificial because, being complicated discrete-time models, they do not correspond to a continuous time process, while natural processes typically evolve in continuous time.
- Their identification, typically based on the estimation of the autocorrelation function from data, usually neglects estimation bias and uncertainly, which in stochastic processes (as opposed to purely random processes) are often tremendous (Lombardo et al., 2014).
- Today they are unnecessary as synthetic series from a process with any arbitrary autocorrelation structure can be easily generated otherwise (see below).

# Problems with the "Time Series School" (2)

- From their onset, time series models (excepting AR(1) and ARMA(1,1)), 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:

  *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*. [Trap 3]

  *…*

  *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*. [Trap 4]
- This is similar to fitting a distribution function using observations to estimate many moments thereof, without using a parametric expression for the distribution (cf. Lombardo et al., 2014, "**Just two moments**".)
- Today we know how to (a) use a parsimonious model structure and separate it from parameters and (b) calculate model coefficients theoretically (not estimate from data).

# Problems with the "Time Series School" (3)

- The decomposition of a time series to components, trends, seasonal variation and residuals, is popular even today.
- 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**.
  - It is hardly conceivable how **time per se** could be regarded as an **explanatory variable** in a complex process and what is the **logical basis** in expressing the statistics of a process as a deterministic function of time.
  - Accumulation of data series with long time spans has shown 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**.
- In addition, "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).

# Notes on optimization and control of hydrosystems

- "*It is amazing how many are unaware that* ***the primary reason for feedback in control is uncertainty***" (Bennett, 1996).
- The best way to describe uncertainty is provided by stochastics.
- **Linear programming** methods are extremely effective but they presuppose linear equations which are hardly the case in hydrosystems. The stochastic variants of linear programming are rather naïve and not quite useful (except for simple sub-problems).
- **Dynamic programming** methods are good for problems that **can be divided into sequential stages** but this is a special case not often met in hydrosystems. The stochastic extensions of these methods presuppose drastic assumptions that oversimplify system dynamics and may make the results irrelevant to real system.
- Optimization and control problems in hydrosystems are characterized by:
  - **uncertainty** in terms of inputs, with peculiar stochastic behaviour;
  - **nonlinearity** with respect to dynamics, operation constraints and objectives;
  - **nonconvexity** in terms of objective functions and constraints, so that numerous local optima appear very often;
  - **hysteresis and storage effects** which demand to express the performance measure (objective function) in a global manner (not on an instantaneous basis);
  - **many variables** yet demanding a **parsimonious** representation;
  - **multiobjective** setting (involving several performance criteria).
- **Question: Can we construct a generic methodology for such problems?**

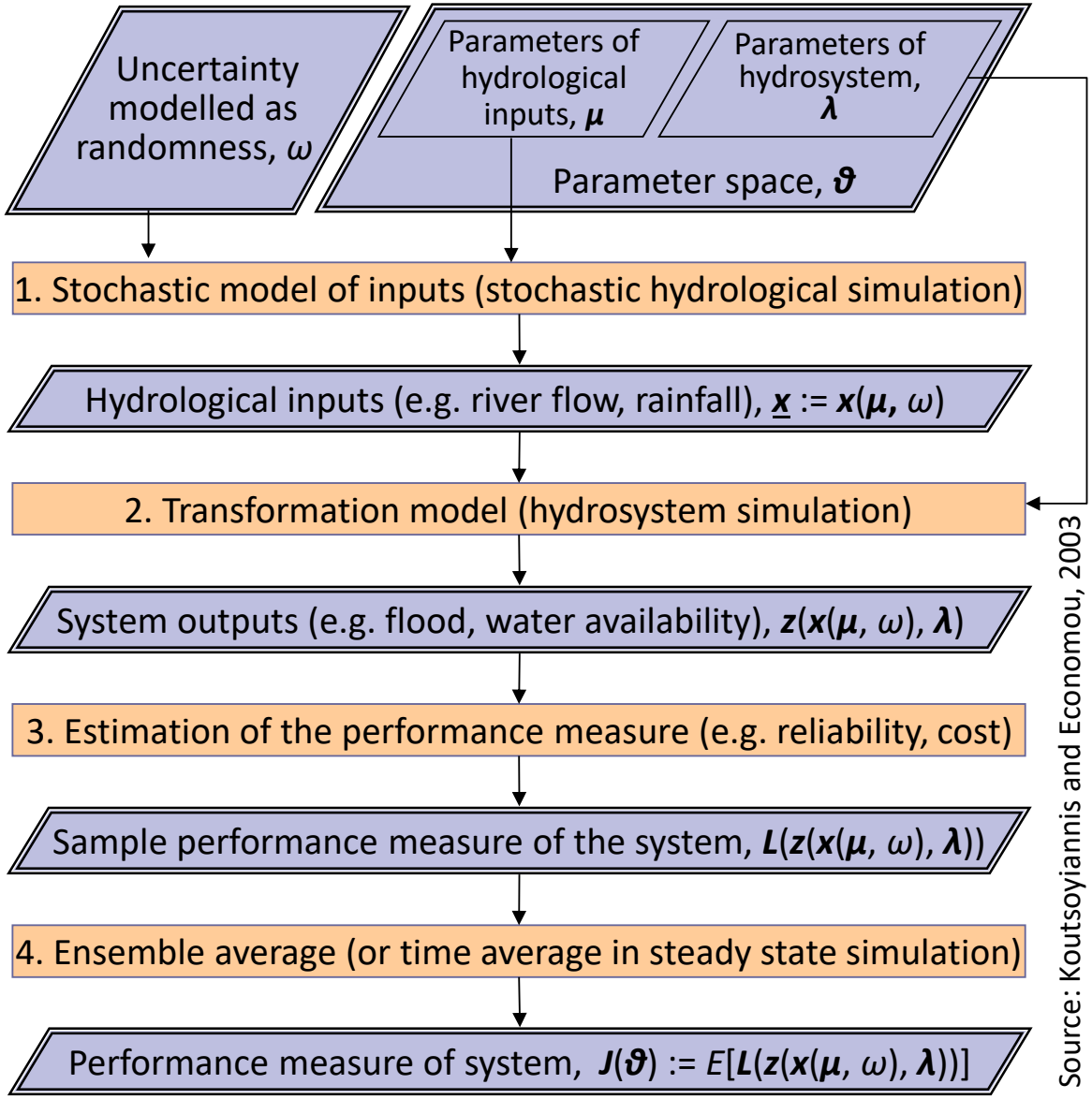
# A general methodological scheme for hydrosystems

Mathematically, hydrosystem engineering and management problems include two sub-problems:

An **integration** problem to find a performance measure of the hydrosystem, $\boldsymbol{J}(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \mathrm{E}[\boldsymbol{L}(\boldsymbol{z}(\boldsymbol{x}(\boldsymbol{\mu}, \omega), \boldsymbol{\lambda}))]$ [Note: expectation means integration.]

A constrained **optimization** problem, in which we seek the hydrosystem operation parameters $\boldsymbol{\lambda}$ that optimize the performance $\boldsymbol{J}(\boldsymbol{\mu}, \boldsymbol{\lambda})$.

For both sub-problems stochastic simulation methods offer a feasible and consistent solution.



# Parameterization – simulation – optimization in hydrosystem control problems

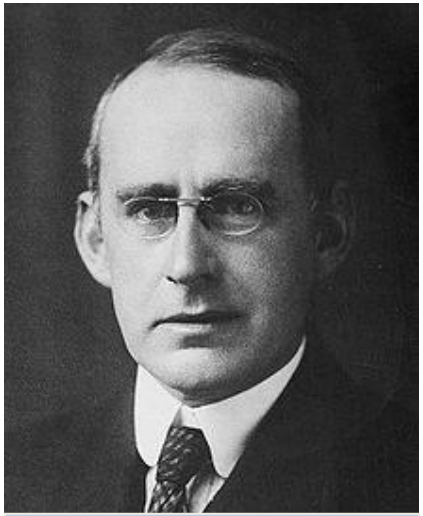
- Parameterization constitutes a parsimonious (low-dimensional) representation of the system control; for specific examples see:
  - Koutsoyiannis and Economou (2003) for toy examples supporting the theory;
  - Nalbantis and Koutsoyiannis (1997) for an algebraic parameterization;
  - Giuliani et al., (2014a,b) for a parameterization based on universal approximators such as Artificial Neural Networks and Radial Basis Functions.
- The representation of the system operation per se can be as complex as required, thus not making reductions in the faithful representation of its dynamics.
- The objective functions, as well as possible constraints and penalty functions, are evaluated faithfully by performing detailed Monte Carlo simulation, considering the entire system structure with all its components. Therefore, any type of uncertainty and risk is naturally incorporated into the simulation. Then optimal control is attained by optimizing the objective function by a nonlinear optimization procedure, which is made possible because the low dimensionality of the search space.
- Till now this methodology has been applied on large timescales. If it is to be applied for small ones, what are the **main differences in terms of the required simulation**?
  - **Reply**: Characteristic behaviours of hydrological inputs include: **intermittency**, high **autocorrelation** as a result of **both long-range dependence** and **roughness** (fractality), high **skewness**, and **time irreversibility**.

# Part B
# Time's arrow and time irreversibility

# Time line of time's arrow

- Time irreversibility in physics goes back to 1807, when Fourier established the equation of heat conduction (Fourier 1822), which are not symmetric with respect to past and future directions of time, unlike those of Newtonian mechanics.
- Fourier was followed by Thomson and Maxwell, while Clausius (1850, 1854, 1865) laid the foundation for the second law of thermodynamics, introduced the concept (and the term) of *entropy* and connected time irreversibility with the second law.
- Boltzmann (1877) explained the concept of entropy in probability theoretic context, and penetrated into the concept of irreversibility, on which he had a productive debate with Planck (Hollinger and Zenzen 1985).
- The term "time's arrow" was introduced by Eddington (1928):

  *I shall use the phrase 'time's arrow' to express this one-way property of time which has no analogue in space. It is a singularly interesting property from a philosophical standpoint. We must note that: (1) It is vividly recognized by consciousness. (2) It is equally insisted on by our reasoning faculty, which tells us that a reversal of the arrow would render the external world nonsensical. (3) It makes no appearance in physical science except in the study of organization of a number of individuals. Here the arrow indicates the direction of progressive increase of the random element.*



**Sir Arthur Stanley Eddington** (1882 –1944) English astronomer, physicist, and

# Time's arrow in stochastics

- A simple definition of **time reversibility** is provided in the framework of stochastics:

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

  $$F(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) := P\{\underline{x}(t_1) \le x_1, \underline{x}(t_2) \le x_2, \dots, \underline{x}(t_n) \le x_n\} \quad (5)$$

  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 $n$, $t_1, t_2, \dots, t_{n-1}, t_n$,

  $$F(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = F(x_1, x_2, \dots, x_n; -t_1, -t_2, \dots, -t_n) \quad (6)$$

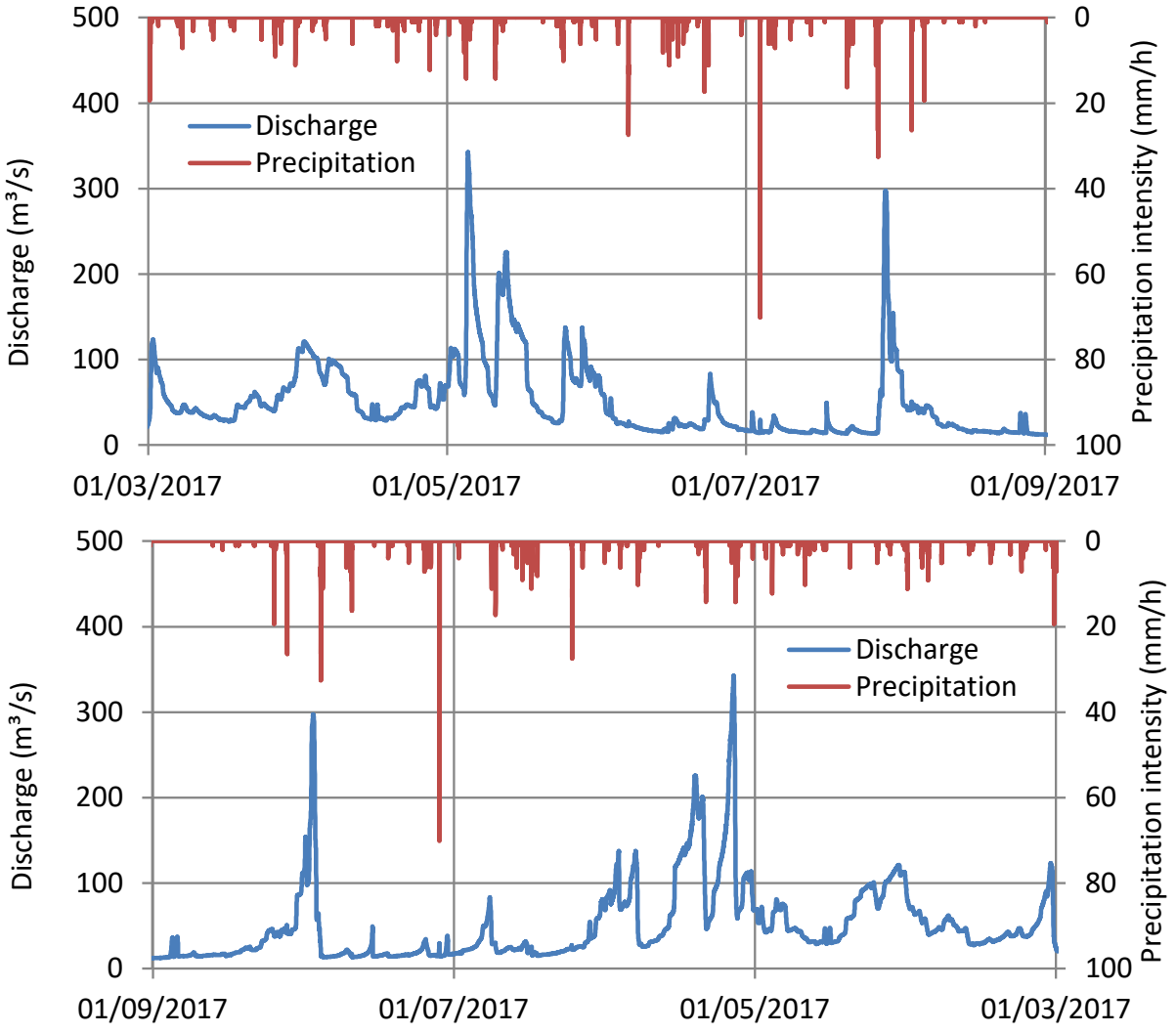
  If times $t_i$ are equidistant, i.e. $t_i - t_{i-1} = D$, this can be also written as:

  $$F(x_1, x_2, \dots, x_{n-1}, x_n; t_1, t_2, \dots, t_{n-1}, t_n) = F(x_1, x_2, \dots, x_{n-1}, x_n; t_n, t_{n-1}, \dots, t_2, t_1) \quad (7)$$

- A process that is not time-reversible is called **time-asymmetric, time-irreversible** or **time-directional**.
- A **time reversible process is also stationary** (Lawrance, 1991).
- If a 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.
- This relationship of Gaussianity and reversibility holds for **scalar** (univariate) stochastic processes only. 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).

# Is time irreversibility visible?

- Is it visible in two time series plotted together?
  - What is the relationship with causality?
- Is it visible in a single time series?
  - What is the relationship with skewness?
- If we increase the time scale of observation, will irreversibility be visible in a single time series?
  - What is the relationship with the central limit theorem?




Precipitation and discharge for a six-month period for the USGS site North Branch Potomac River Near Cumberland, MD, USA

# Stochastic quantification of time irreversibility

The time asymmetry of a process is studied through the differenced process, i.e.:

$$\underline{\tilde{x}}_{\tau}^{(\kappa)} := \frac{\underline{x}_{\tau\kappa} - \underline{x}_{(\tau-1)\kappa}}{\kappa} \qquad (8)$$

in discrete time $\tau$ and discrete time scale $\kappa$.

This has mean zero and variance

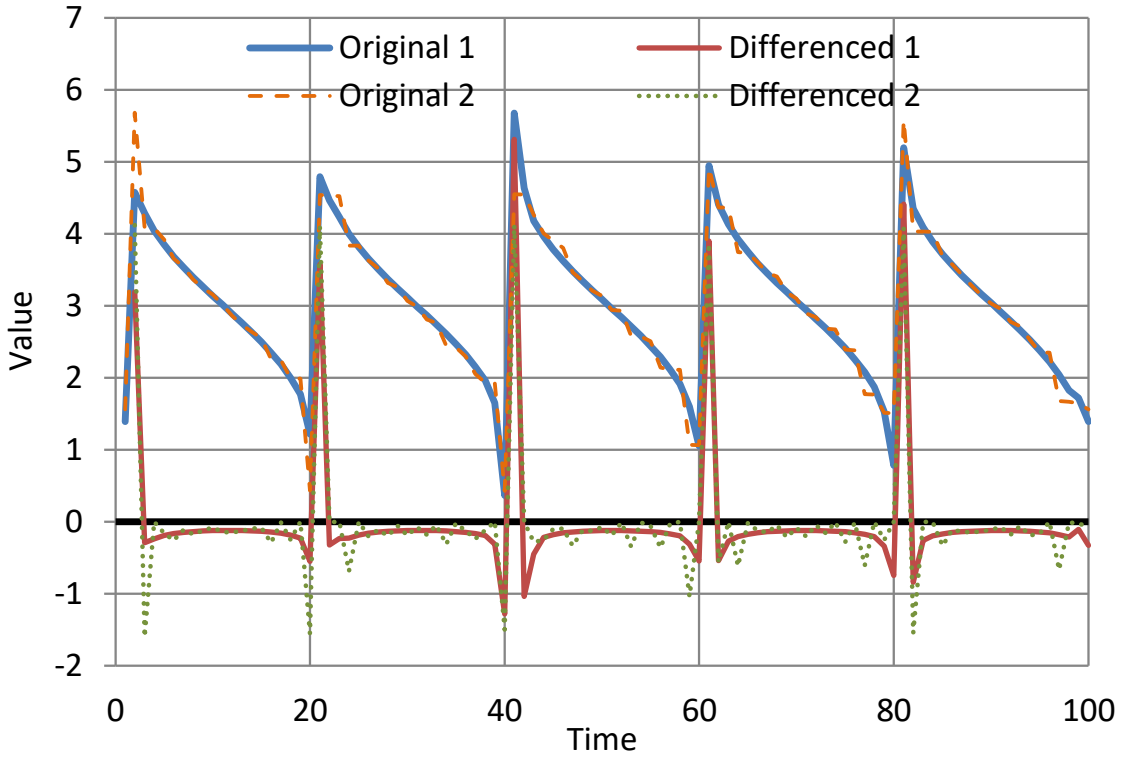
$$\tilde{\gamma}(\kappa) := \mathrm{var}\left[\underline{\tilde{x}}_{\tau}^{(\kappa)}\right] = \frac{2v_{\kappa}}{\kappa^2} \qquad (9)$$

with $v_{\kappa} := c_0 - c_{\kappa}$ denoting the structure function of the original process $\underline{x}_{\tau}$ where $c_{\kappa}$ is the autocovariance. Then the time asymmetry is quantified by the skewness coefficient:

$$\tilde{C}_{\mathrm{s}}(\kappa) := \frac{\tilde{\mu}_3(\kappa)}{\left(\tilde{\gamma}(\kappa)\right)^{3/2}} \qquad (10)$$

or by the probability:

$$P_{-}^{(\kappa)} := P\left\{\underline{\tilde{x}}_{\tau}^{(\kappa)} \leq 0\right\} = \tilde{F}(0) \qquad (11)$$

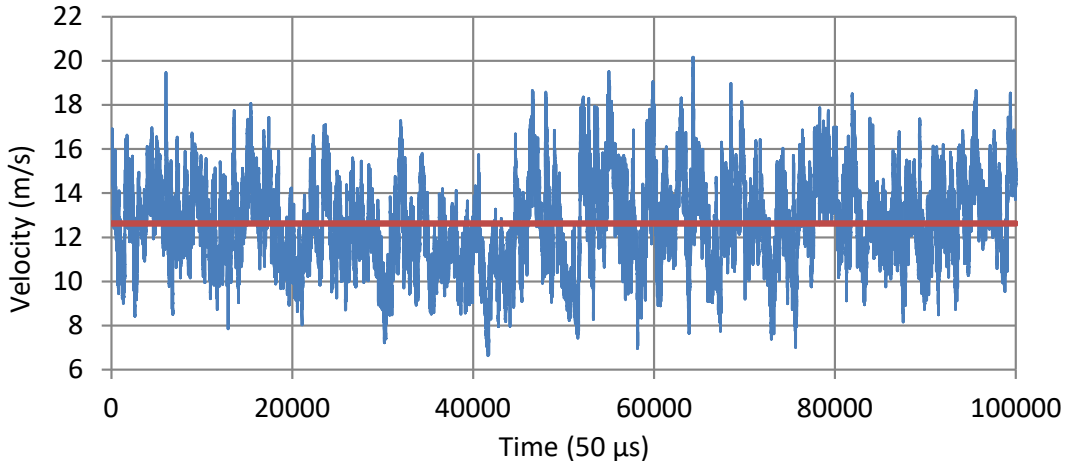


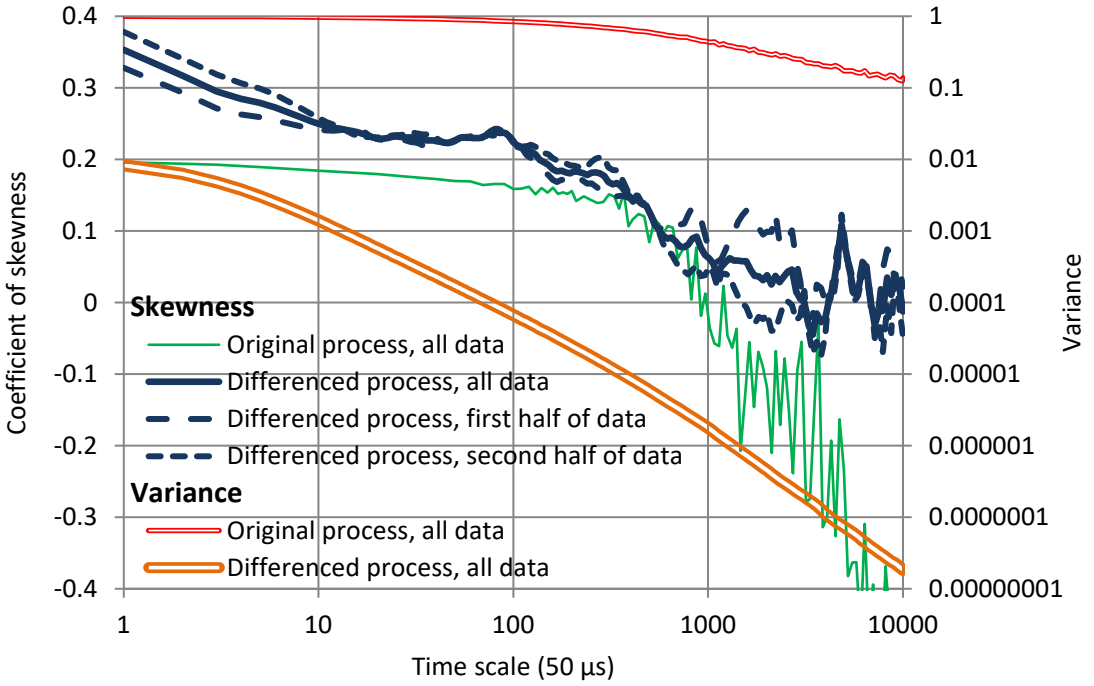

Two synthetic time series generated by maximizing time irreversibility properties of a process restricted to be *marginally* Gaussian ($N(3, 1)$) with lag-one autocorrelation 0.5. Solution 1 maximizes the skewness of the differenced process. Solution 2 maximizes the frequency that the differenced process 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.

# Exploration: turbulence

A long series of nearly isotropic grid turbulence provides a view of the structure of a process at the finest time scales. We use part of grid data of turbulence from the Corrsin Wind Tunnel at a high Reynolds number (Kang et al. 2003), namely the series of velocity along the flow direction at the first of the probes, which we averaged for time scale of $D$ = 50 μs, thus forming a time series with length 600 000.

The skewness of the original process becomes zero at a scale of about 0.05 s, while the differenced process still has positive values (and thus time irreversibility) at time scales up to about 0.1 s.

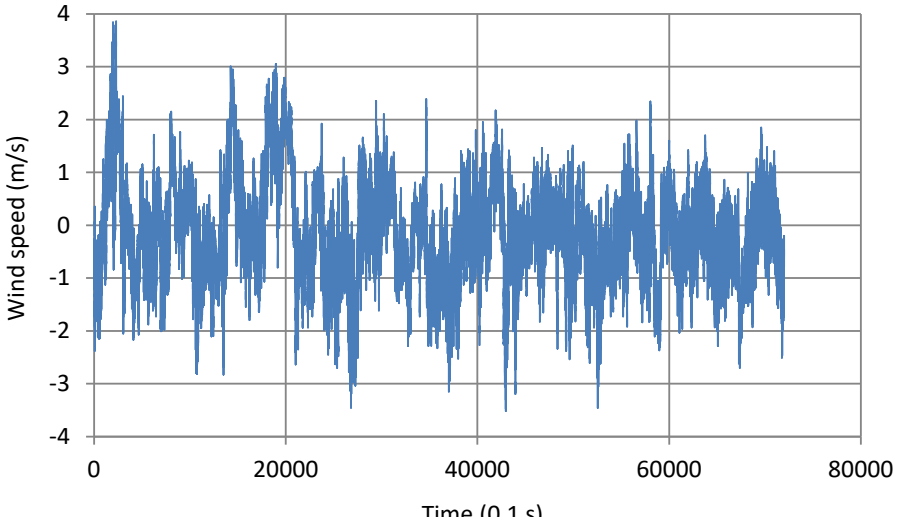


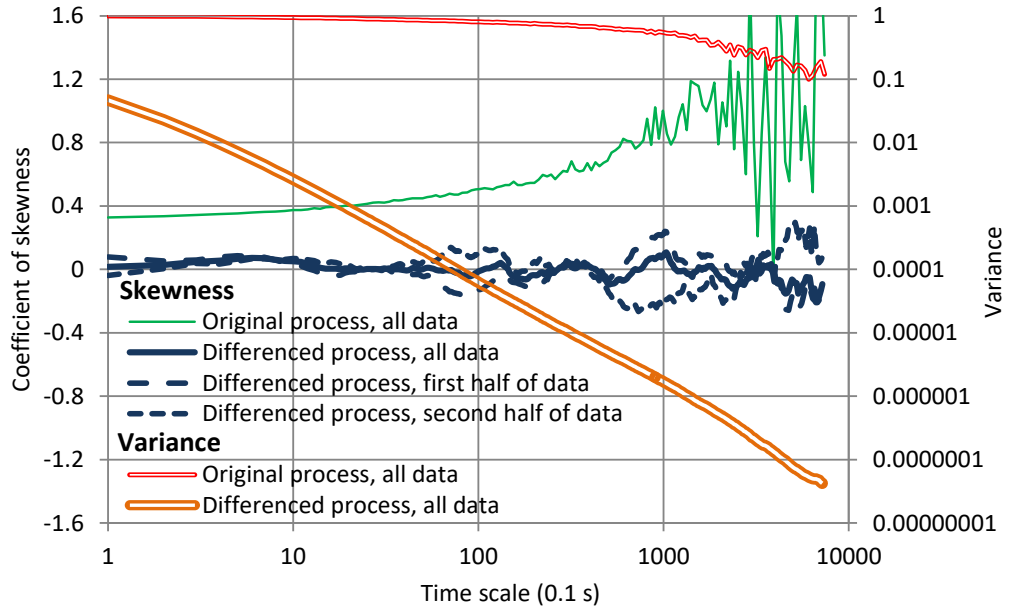



# Exploration: wind speed

Here we study wind data at the much larger, yet very fine, time scale of 0.1 s. We use part of the data recorded at a 10 Hz resolution for a period of one month by a sonic anemometer on a meteorological tower located at Beaumont KS (Doran, 2011). Namely we use 71 998 data values for the period starting at 1999-10-11 20:30 UTC (perhaps 6 h earlier in local time) and ending 2:30 h later.

Neither remarkable skewness nor time directionality are observed, a finding consistent with the previous example of turbulent velocity, as now the minimum time scale is 0.1 s.



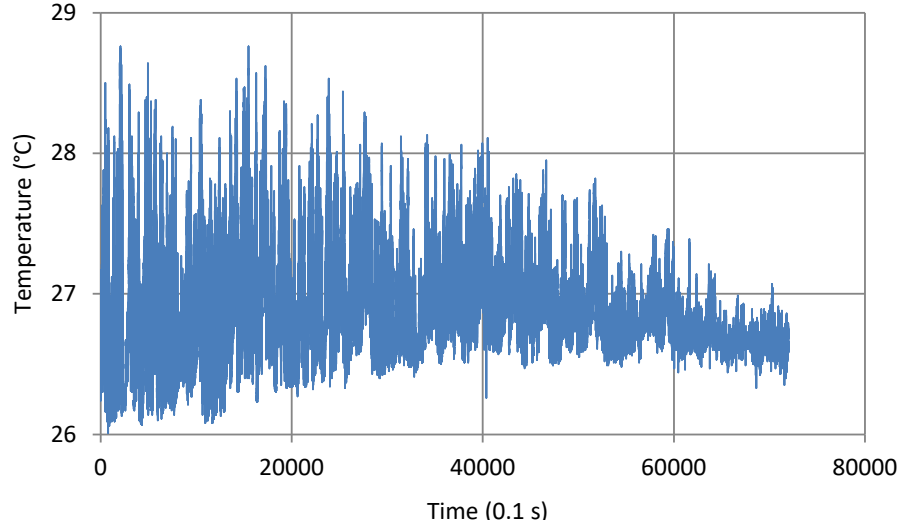



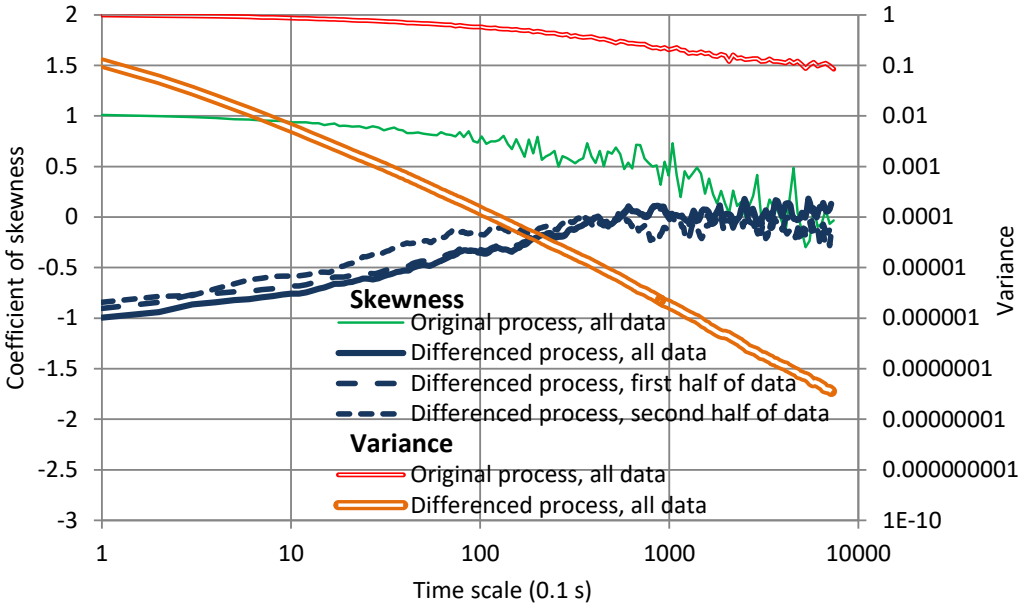
# Exploration: temperature

Here we study the temperature measurements at the same site as in the wind speed and the same period, from the same data set as above.

The plot of the time series shows an interesting time directionality (or dependence) with decreasing variability with the increase of time. However, we may conjecture that this behaviour is local, related to the diurnal temperature cycle.

Even in this part of the series both state asymmetry and time asymmetry cease at time scales of 100 s or more. It is interesting, though, that for time scales smaller than that the skewness of the original process is positive, while that of the differenced is negative.

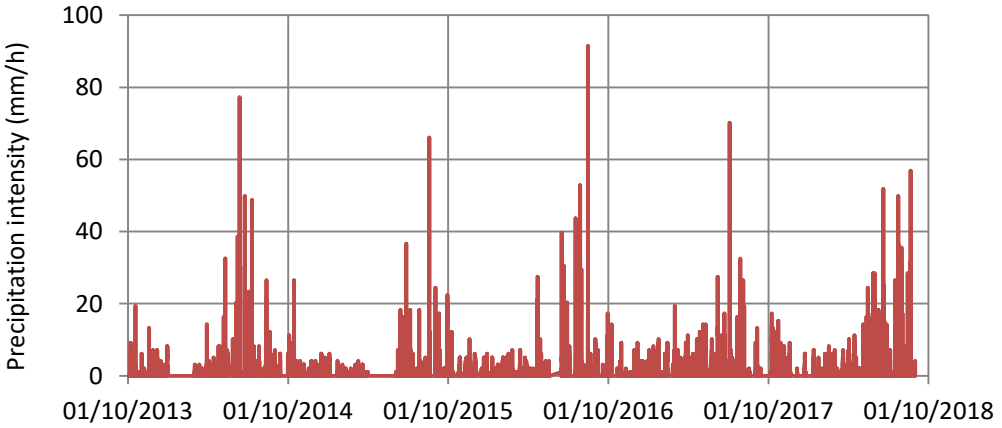


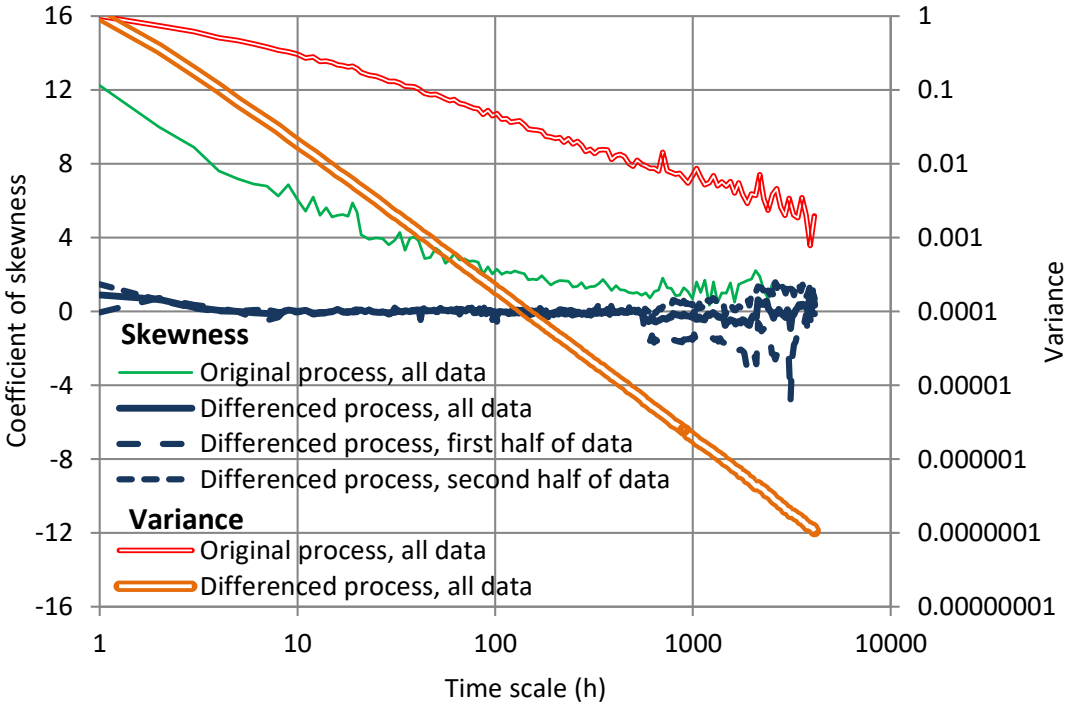



# Exploration: precipitation

The precipitation time series already seen in p. 25 is further studied here at the hourly scale. Diurnal cycle is found to be marked and is treated by standardising by hourly coefficients (Koutsoyiannis 2019a).

The un-differenced transformed process has a very high coefficient of skewness (~12), but in the differenced one the skewness and thus the irreversibility is rather negligible.



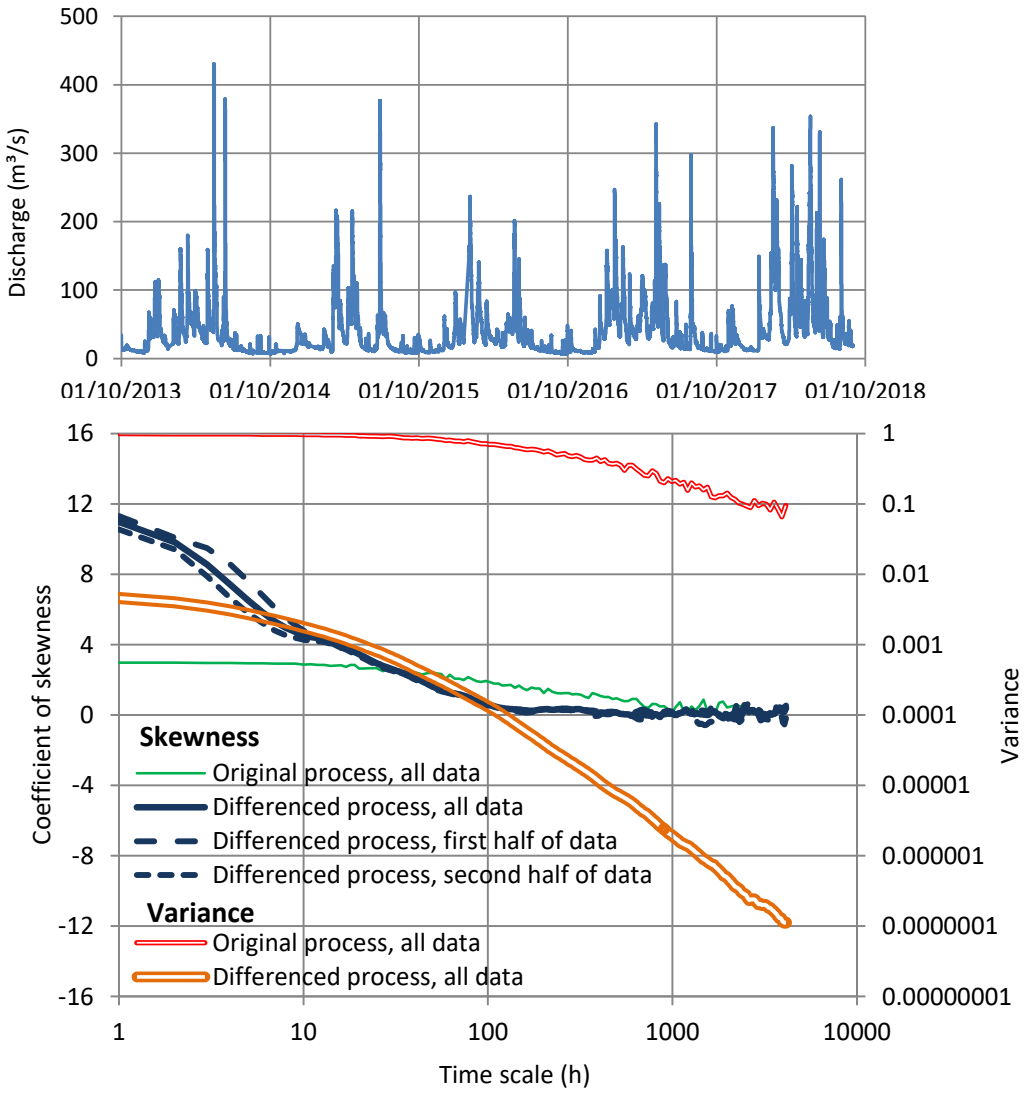



# Exploration: streamflow

The streamflow time series already seen in p. 25 is further studied here at the hourly scale. The annual cycle is found to be marked and is treated by standardising by monthly coefficients (Koutsoyiannis 2019a).

The skewness coefficients of the original and differenced processes after the transformation are 2.98 and 10.99, respectively, for the hourly scale (not very different from those before the transformation).

The latter value indicates strong irreversibility and its attenuation is rather slow, requiring about 100 h or 4 days to cease. This means that irreversibility is relevant for flood simulations on operational time scales, and can only be neglected on the monthly time scale and beyond.



# Part C
# A generic stochastic generator for time irreversible (and reversible) processes

# Fundamental assumptions connecting stochastics with data

- Central to the notion of a stochastic process are the concepts of **stationarity** and **nonstationarity**, two widely misunderstood and broadly misused concepts (Montanari and Koutsoyiannis, 2014; Koutsoyiannis and Montanari, 2015). Their **definitions apply only to stochastic processes** (e.g., time series cannot be stationary, nor nonstationary).
- Stationarity is also related to **ergodicity**, which in turn is a prerequisite to make inference from data, that is, induction.
- If the system that is modelled in a stochastic framework has **deterministic dynamics** then **a stationary system is also ergodic and vice versa**, and **a nonstationary system is also non-ergodic and vice versa**. (Mackey 1992, p. 52)
- 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 device a stochastic process that is ergodic**.

**Reminder of definitions**

Following Kolmogorov (1931, 1938) and Khinchin (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 statistics for any $t$ and $t'$.

Based on the **ergodic theorem** (Birkhoff, 1931; Khinchin, 1933; see also Mackey, 1992, 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** $\mathrm{E}[g(\underline{x}(t))]$, i.e., $\lim_{T\to\infty} \frac{1}{T}\int_0^T g\left(\underline{x}(t)\right) dt = \mathrm{E}[g(\underline{x}(t))]$.

# Some principles and interpretations about natural processes that are modelled as stochastic processes

- There are two distinguishable characteristics visible at different time scales.
- **Local behaviour**: As time scale tends to zero there may be **smoothness** (or **roughness**).
- **Global behaviour:** As time scale tends to infinity there may be **persistence** (or **antipersistence**) related to several mechanisms of change acting on different time scales.
- **Roughness** and **persistence** cause or amplify **positive autocorrelation**. Smoothness and antipersistence cause or amplify negative autocorrelation. Note though that **as time scale tends to zero**, the **autocorrelation should always be positive**, regardless of smoothness or antipersistence.
- Autocorrelation is **not memory**; it is not a cause but a consequence of change mechanisms.
- It is more intuitive to use variance than autocorrelation to describe change. The variance of the averaged process as a function of time scale of averaging, is termed **climacogram**:

$$\gamma(k) \coloneqq \mathrm{var}\left[\frac{\underline{X}(k)}{k}\right],\ \underline{X}(t) \coloneqq \int_0^t \underline{x}(\xi)\mathrm{d}\xi \tag{12}$$

- The following expression of a climacogram defines a model appropriate for many applications: the **filtered Hurst-Kolmogorov** process with **Cauchy**-type (FHK-C) climacogram:

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

where $\alpha$ and $\lambda$ are scale parameters with dimensions of $[t]$ and $[x^2]$, and $M$, $H$ are dimensionless parameters with values in (0,1). $M$ (for Mandelbrot) characterizes the smoothness (fractality) of the process and $H$ (for Hurst) characterizes the global behaviour or persistence.

# Climacogram and entropy production

- The conditional variance (conditional climacogram) once past and present are known is

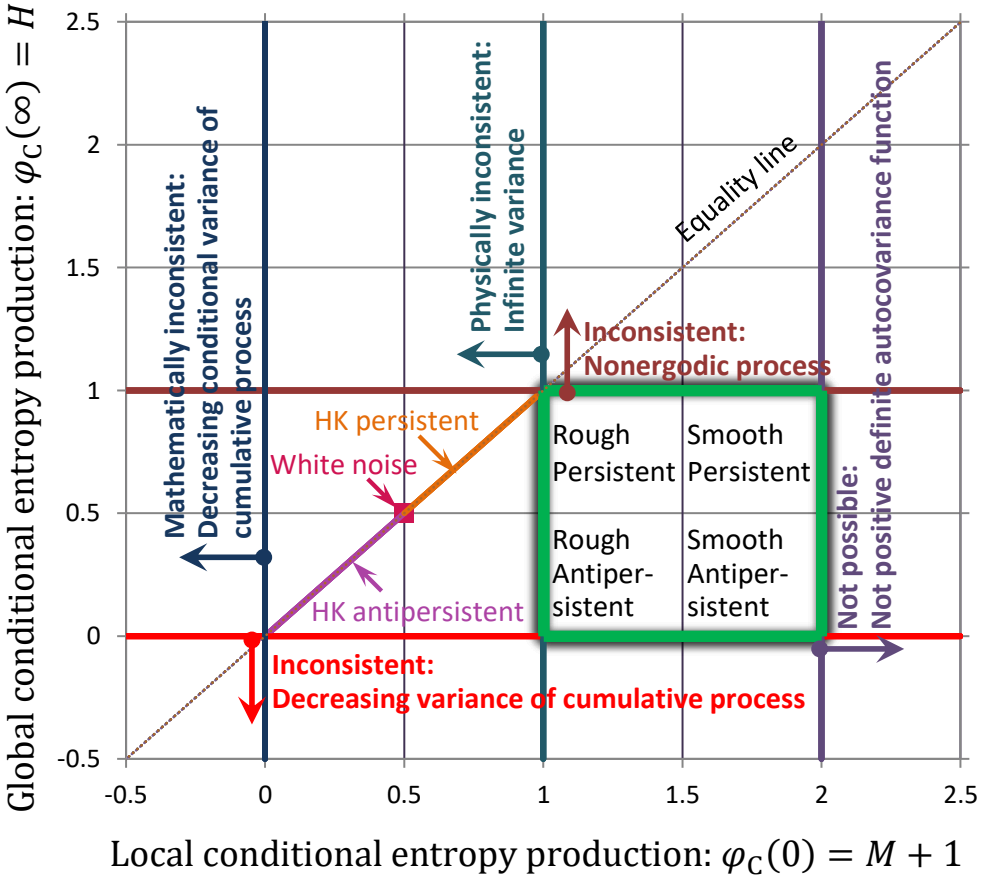
$$\gamma_C(k) \approx \varepsilon\big(\gamma(k) - \gamma(2k)\big) \qquad (14)$$

where $\varepsilon$ is a constant determined from asymptotic properties of the process (Koutsoyiannis, 2017).

- The entropy production (vis. production of uncertainty) in logarithmic time, conditional on known past and present, is expressed as a function of the log-log derivative (LLD) of the conditional variance, $\gamma_C^{\#}(t)$:

$$\varphi_C(t) = 1 + \tfrac{1}{2}\,\gamma_C^{\#}(t) \qquad (15)$$

- Smooth and persistent processes yield the highest entropy production.



# Other second-order stochastic tools and their relationships to climacogram

- *Autocovariance function*, $c(h)$ for time lag $h$, defined as

$$c(h) := \operatorname{cov}[x(t), x(t+h)] = \frac{1}{2}\frac{\mathrm{d}^2\left(h^2\gamma(h)\right)}{\mathrm{d}h^2} \tag{16}$$

- *Power spectrum* (also known as *spectral density*), $s(w)$ for frequency $w$, defined as the Fourier transform of the autocovariance function, i.e.,

$$\begin{gathered} s(w) := 4\int_0^\infty c(h)\cos(2\pi w h)\,\mathrm{d}h \\ c(h) = \int_0^\infty s(w)\cos(2\pi w h)\,\mathrm{d}w, \quad \gamma(k) = \int_0^\infty s(w)\ \operatorname{sinc}^2(\pi w k)\,\mathrm{d}w \end{gathered} \tag{17}$$

- *Structure function* (also known as *semivariogram* or *variogram*),

$$v(h) := \frac{1}{2}\operatorname{var}\left[\underline{x}(t) - \underline{x}(t+h)\right] = \gamma_0 - c(h) \tag{18}$$

- *Climacospectrum*, $\zeta(k)$, for time scale $k$:

$$\zeta(k) := \frac{k\left(\gamma(k) - \gamma(2k)\right)}{\ln 2} = \frac{k\,\gamma_{\mathrm{C}}(k)}{\varepsilon \ln 2} \tag{19}$$

This resembles the power spectrum and combines the asymptotic behaviours of the climacogram and the structure function.

# The generic generator

- Any stationary stochastic process $\underline{x}_\tau$ can be generated by the moving average scheme (Koutsoyiannis 2000):

$$\underline{x}_\tau = \sum_{j=-J}^{J} a_j \underline{v}_{\tau-j} \tag{20}$$

where $a_j$ are weights to be calculated from the autocovariance function, $\underline{v}_j$ is white noise averaged in discrete-time and $J$ is a large integer (theoretically, $J = \infty$).

- The autocovariance is given by the convolution expression:

$$c_\eta = \sum_{j=-J}^{J} a_j a_{\eta+j} \tag{21}$$

- Given the stochastic model, the weights $a_\eta$ can be calculated from its second-order characteristics by the following explicit relationship (Koutsoyiannis, 2019b):

$$a_\eta = \int_{-1/2}^{1/2} \sqrt{2} \mathrm{e}^{2\pi \mathrm{i}(\theta(\omega)-\eta\omega)} \sqrt{s_\mathrm{d}(\omega)}\, \mathrm{d}\omega \tag{22}$$

where $\omega$ denotes frequency, $s_\mathrm{d}(\omega)$ is the power spectrum of the discrete-time representation of the process (see below) and $\theta(\omega)$ is any arbitrary odd real function.

☺ Notice the appearance of $2, 1/2, \sqrt{2}$, e, π, i (imaginary unit) in equation (22).

# The generic generator (2)

- The equations:

$$\underline{x}_\tau = \sum_{j=-J}^{J} a_j \underline{v}_{\tau-j}, \quad a_\eta = \int_{-1/2}^{1/2} \sqrt{2}\mathrm{e}^{2\pi\mathrm{i}(\theta(\omega)-\eta\omega)}\sqrt{s_\mathrm{d}(\omega)}\,\mathrm{d}\omega \qquad (20)\ \&\ (22)$$

  define the **asymmetric moving average (AMA)** scheme, which can be used in any problem of stochastic simulation of time irreversible and reversible processes.
- The sequence of $a_\eta$ given by equation (22):
  - consists of real numbers, despite the expression involving complex numbers;
  - reproduces precisely the required autocovariance function (equation (21)); and
  - is easy and fast to calculate using the fast Fourier transform (FFT).
- Equation (22) 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 $\theta(\omega) = \theta_0 =$ constant, again there are infinitely many solutions.
  - This 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 $\underline{v}_\tau$ (which in general **is not Gaussian**), thus enabling preservation of moments of any order (Koutsoyiannis, 2018; Dimitriadis and Koutsoyiannis, 2018).

# Computational algorithm (not involving data at this phase)

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

$$c_j = \frac{(j+1)^2\gamma(|j+1|D)+(j-1)^2\gamma\big((|j-1|D\big)}{2} - j^2\gamma(|j|D) \tag{23}$$

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

2. We choose an appropriate number of coefficients $J$ that is a power of 2 and use the FFT to calculate the discrete-time power spectrum and the frequency function $A^{\mathrm{R}}(\omega)$ for an array of $\omega_j = j\,w_0, j = 0,1,\dots,J, w_0 \coloneqq 1/JD$:

$$s_{\mathrm{d}}(\omega_j) = 2c_0 + 4\sum_{\eta=1}^{J} c_\eta \cos(2\pi\eta\omega_j), \quad A^{\mathrm{R}}(\omega_j) = \sqrt{2s_{\mathrm{d}}(\omega_j)} \tag{24}$$

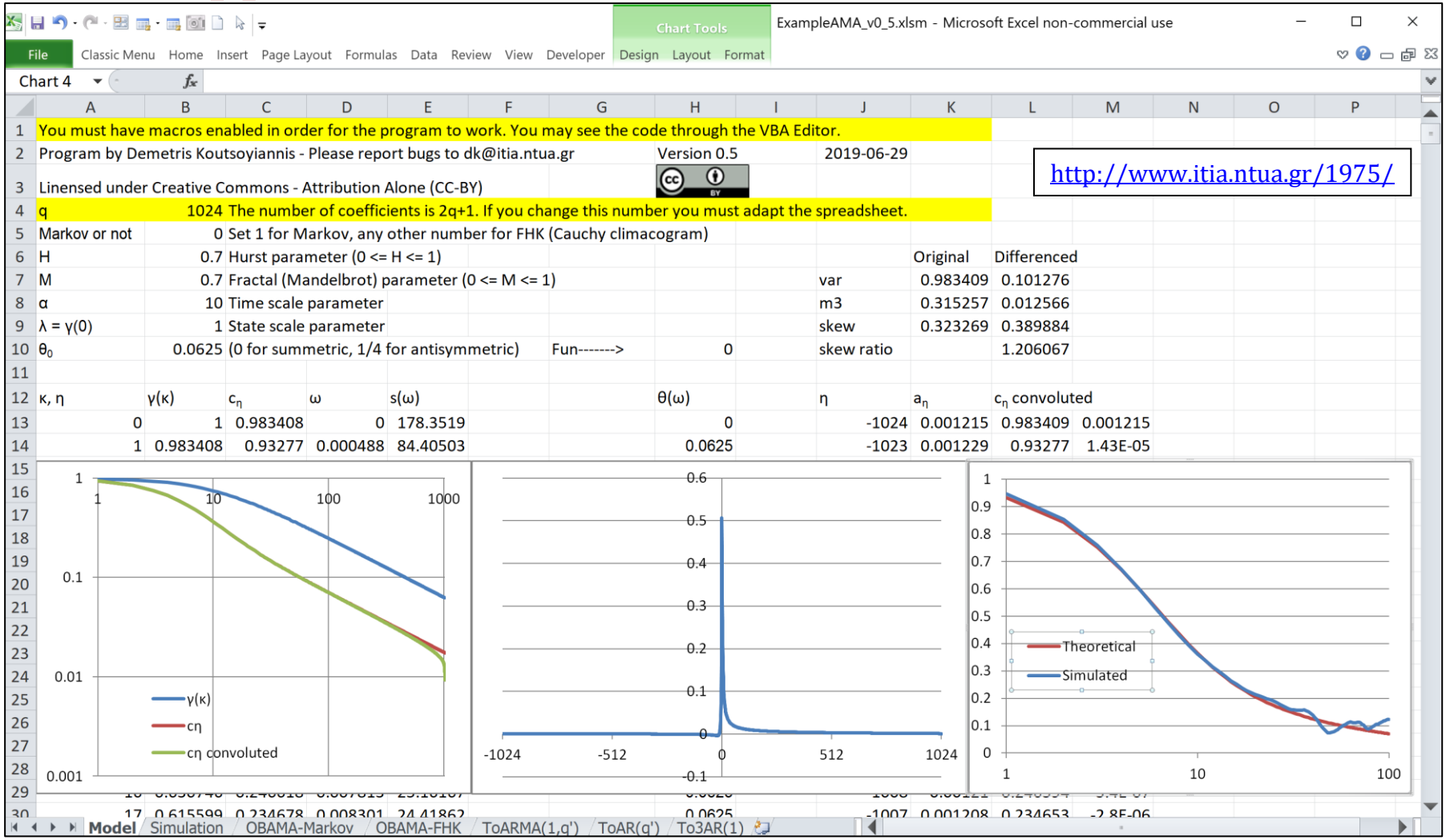
3. We choose $\theta(\omega)$ and form the arrays (vectors) $\boldsymbol{A}^{\mathrm{R}}$ and $\boldsymbol{A}^{\mathrm{I}}$, both of size $2J$ indexed as $0,\dots,2J-1$, with the superscripts R and I standing for a real and an imaginary vector, respectively:

$$[\boldsymbol{A}^{\mathrm{R}}]_j = \begin{cases} \dfrac{A^{\mathrm{R}}(\omega_j)\cos\big(2\pi\theta(\omega_j)\big)}{4J}, & j = 0,\dots,J \\ [\boldsymbol{A}^{\mathrm{R}}]_{2J-j}, & j = J+1,\dots,2J-1 \end{cases} \tag{25}$$

$$[\boldsymbol{A}^{\mathrm{I}}]_j = \begin{cases} \dfrac{A^{\mathrm{R}}(\omega_j)\sin\big(2\pi\theta(\omega_j)\big)}{4J}, & j = 0,\dots,J-1 \\ 0 & j = J \\ -[\boldsymbol{A}^{\mathrm{I}}]_{2J-j}. & j = J+1,\dots,2J-1 \end{cases} \tag{26}$$

4. We perform (inverse) FFT on vectors $\boldsymbol{A}^{\mathrm{R}}$ and $\boldsymbol{A}^{\mathrm{I}}$, and get the real part of the result for $j = 0,\dots,q$, which is precisely the sequence of $a_\eta$.

# Ease of application




The method is easily implemented in any computational environment; this example is for a common spreadsheet—available online, http://www.itia.ntua.gr/1975/.

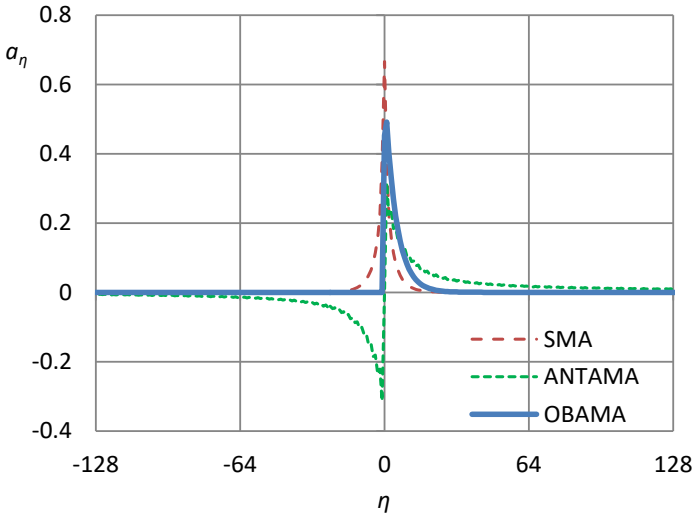
# From genericity to special cases

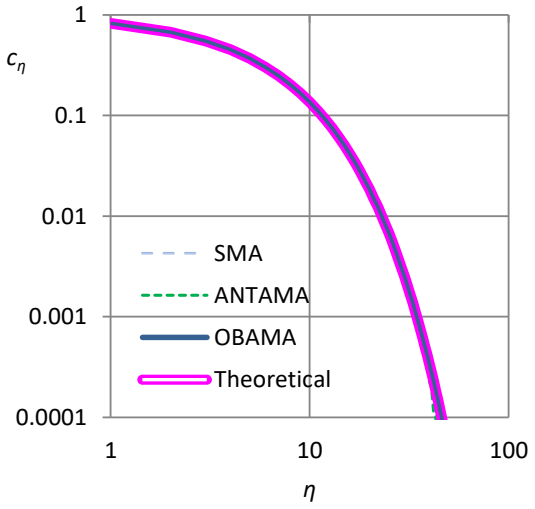
**Generic case**: Asymmetric moving average scheme (**AMA**).

**Special cases**

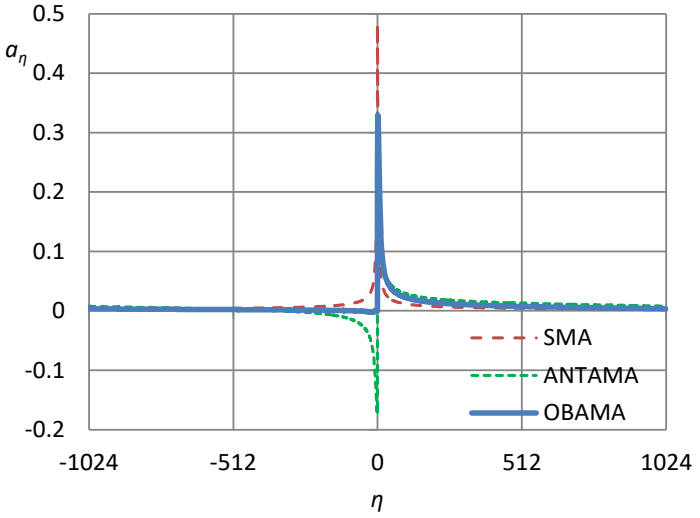
- $\theta(\omega) = 0 \Rightarrow$: Symmetric moving average (**SMA**), $a_{-\eta} = a_{\eta}$
- $\theta(\omega) = 1/4$ : Antisymmetric moving average (**ANTAMA**), $a_{-\eta} + a_{\eta} = \frac{\sqrt{2s_d(0)}}{2J+1} \approx 2\sqrt{\frac{\gamma_J}{(2J+1)}} \approx 0$ (for large $J$)
- Ordinary backward AMA (**OBAMA**): $a_{\eta} = 0$ for all $\eta < 0$ (but it is not always easy to find which $\theta(\omega)$ produces this case).

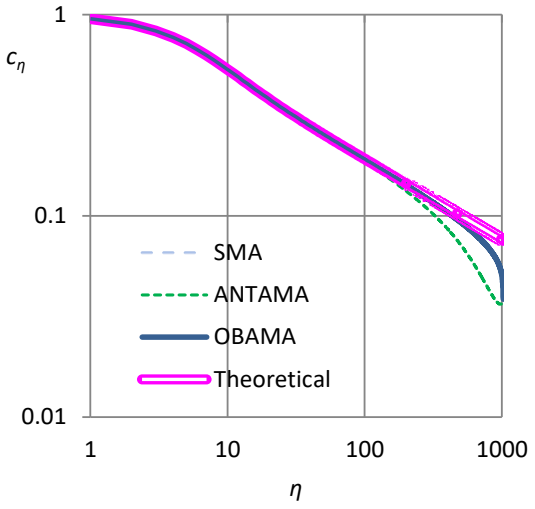
Markov, $\alpha = 20$, $\lambda = 1$, $J = 1024$




Filtered Hurst-Kolmogorov, $H = 0.8$, $M = 0.7$, $\alpha = 20$, $\lambda = 1$, $J = 1024$



# Time asymmetry

- Assuming that $\underline{v}_\tau$ has variance 1 and skewness coefficient $C_{S_v}$, we will have:

$$\mathrm{var}[\underline{x}_\tau] = \sum_{j=-J}^{J} a_j^2, \quad \mu_3[\underline{x}_\tau] = \sum_{j=-J}^{J} a_j^3 \, C_{S_v}, \quad C_S := \frac{\mu_3[\underline{x}_\tau]}{(\mathrm{var}[\underline{x}_\tau])^{3/2}} = \frac{\sum_{j=-J}^{J} a_j^3}{\left(\sum_{j=-J}^{J} a_j^2\right)^{3/2}} C_{S_v} \quad (27)$$

- For the differenced process we have:

$$\tilde{\underline{x}}_\tau := \underline{x}_\tau - \underline{x}_{\tau-1} = \sum_{j=-J}^{J} (a_j - a_{j-1}) \underline{v}_{\tau-j} \quad (28)$$

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

$$\tilde{C}_S := \frac{\mu_3[\tilde{\underline{x}}_\tau]}{(\mathrm{var}[\tilde{\underline{x}}_\tau])^{3/2}} = \frac{\sum_{j=-J}^{J} (a_j - a_{j-1})^3}{\left(\sum_{j=-J}^{J} (a_j - a_{j-1})^2\right)^{3/2}} C_{S_v} \quad (29)$$

- The ratio:

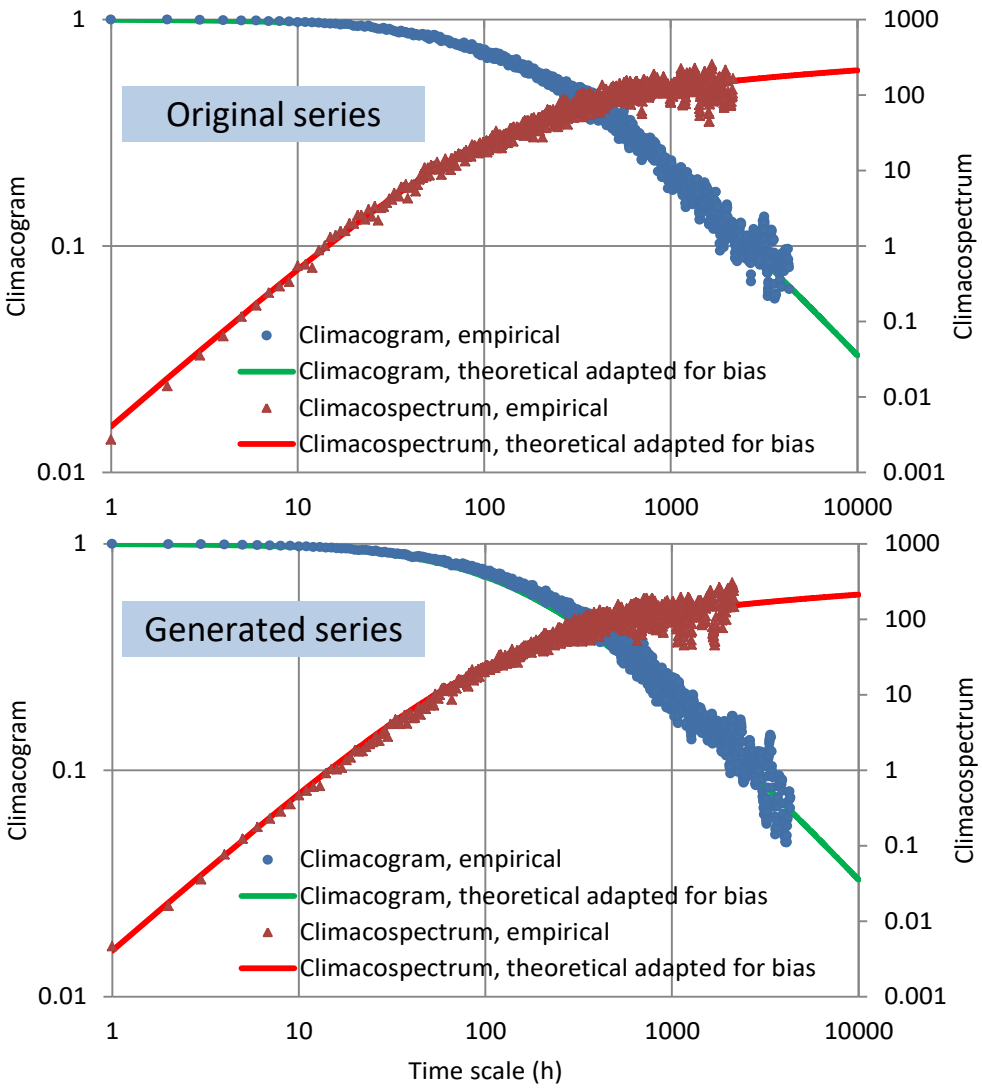
$$\frac{\tilde{C}_S}{C_S} = \frac{\sum_{j=-J}^{J} (a_j - a_{j-1})^3}{\left(\sum_{j=-J}^{J} (a_j - a_{j-1})^2\right)^{3/2}} \frac{\left(\sum_{j=-J}^{J} a_j^2\right)^{3/2}}{\sum_{j=-J}^{J} a_j^3} \quad (30)$$

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

- It should be made clear that **without skewness in the original process $\underline{x}_\tau$ (e.g. in the case of Gaussian processes), there cannot be time asymmetry**.

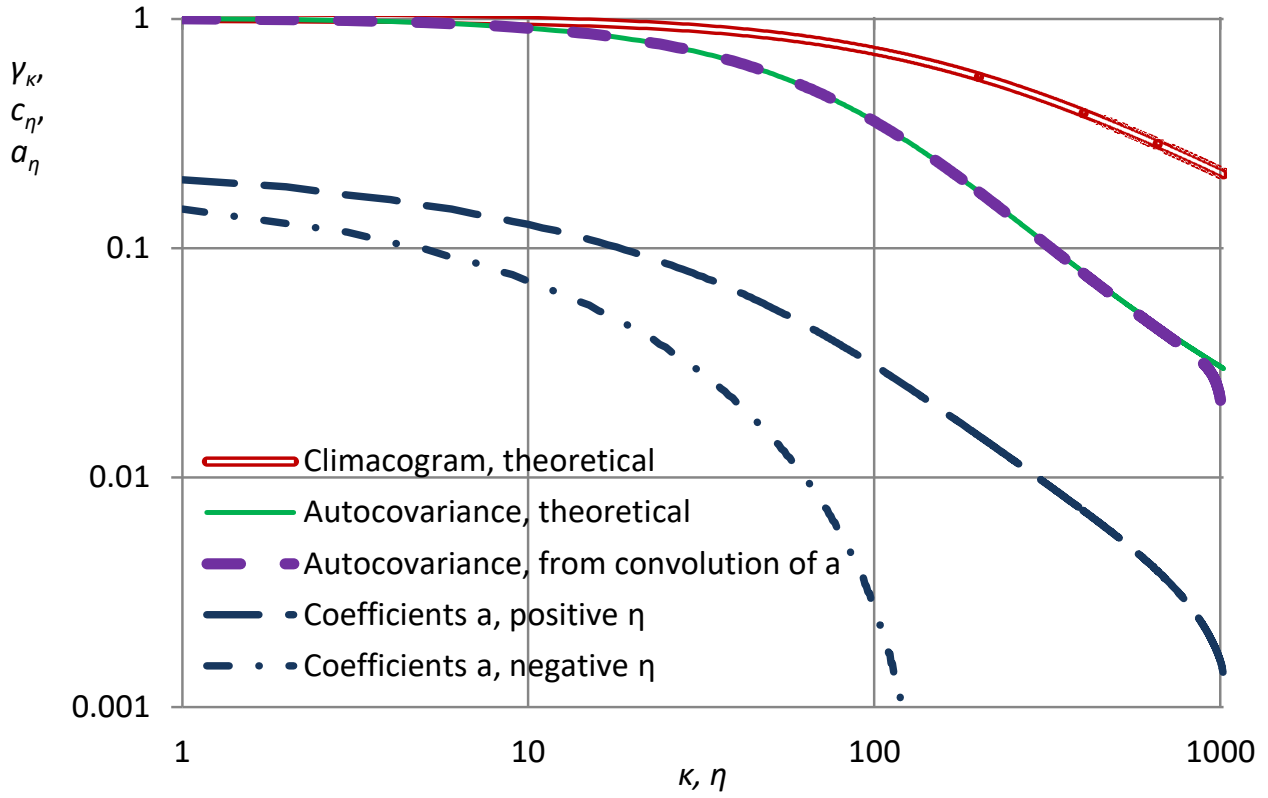
# Case study – model fitting (data are used at this phase)

- We use hourly streamflow data from the site USGS 01603000 North Branch Potomac River Near Cumberland, 2013-2018. We treat seasonality by standardising by monthly coefficients (Koutsoyiannis 2019a).
- We model $\underline{x}_\tau$ as a FHK-C process; the fitting is impressively good with parameters $M = 0.56$ (slight smoothness), $H = 0.6$ (persistence), $\alpha = 160$ h, $\lambda = \gamma(0) = 1.01$ (for the standardized process). The fitting shown in the figure was done in terms of the climacogram and the climacospectrum.
- Additional parameters, quantifying the time and state asymmetry, are the skewness coefficients of $\underline{\tilde{x}}_\tau$ and $\underline{x}_\tau$ whose estimates are 10.99 and 2.98, respectively—a ratio $\tilde{C}_S/C_S = 3.69$.



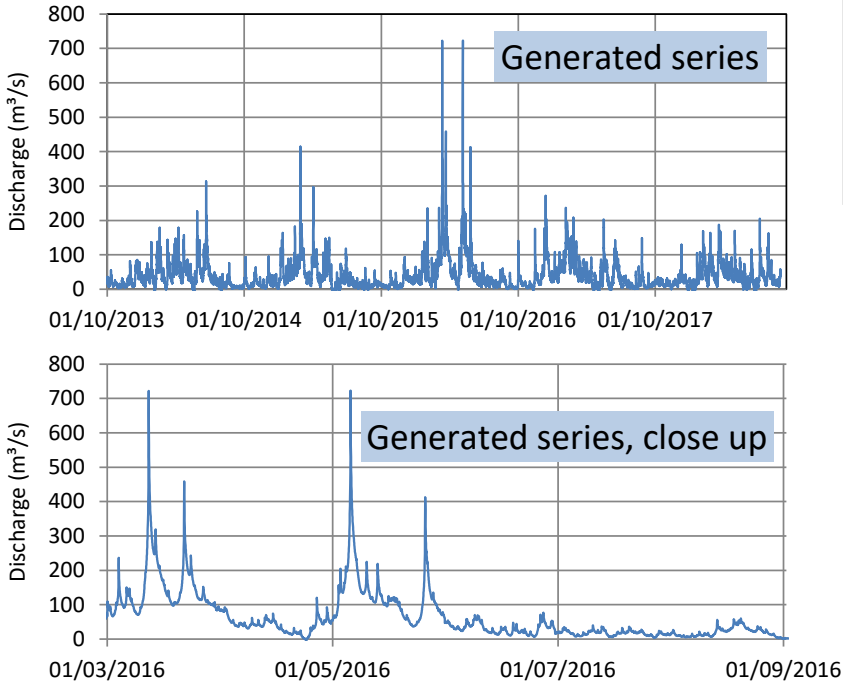
# Case study – generation scheme (data are NOT used at this phase)

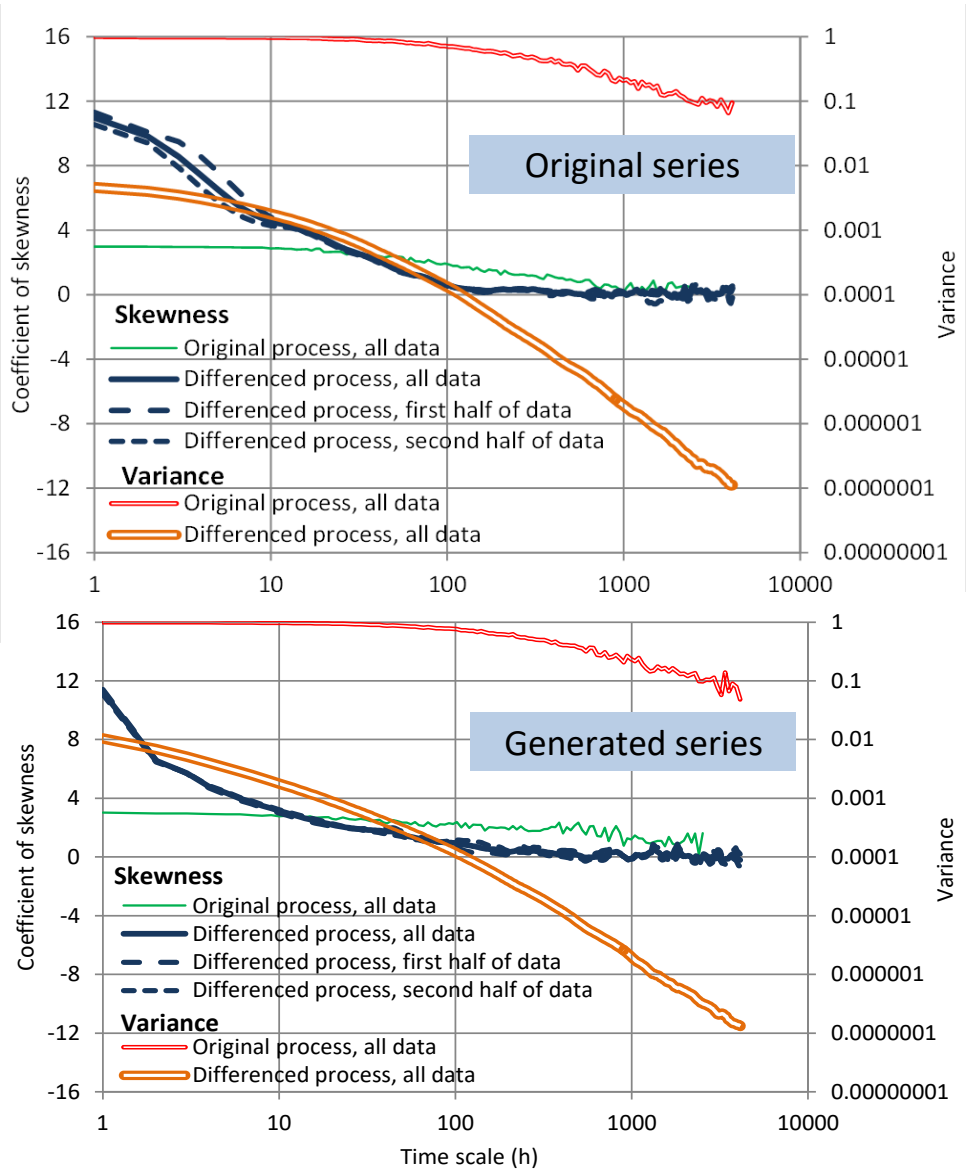
- For the application we use the proposed method with constant $\theta_0$ and with $q$ = 1024.
- The SMA case would result in zero skewness of the differenced process (time symmetry).
- In the ANTAMA case the skewness tends to infinity, which indicates that the method does not have any upper limit of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



# Case study – results

Both visual inspection and comparison of quantified indices suggest good reproduction of the statistical behaviour, including time irreversibility.



# AMA in forecast mode

- The general belief is that:
  - Autoregressive (AR) models are better suited to forecast as they depend on past variables rather than on noise variables.
  - Among the moving average (MA) models, only the backward (i.e., OBAMA) schemes can be suitable for forecast, while the generic AMA scheme (including the special case of SMA) is not, because it involves convolutions for both positive and negative lags $\eta$.

  This is not correct though.
- The forecast problem should be regarded as totally independent from the stochastic model per se; as an example we may consider the case of an AR($p$) scheme used to approximate a process with long-range dependence. 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 cannot be applied for forecast.
- There is a general solution of the forecast problem (Koutsoyiannis, 2000), which can be applied for any type of linear model and hence for AMA. This calculates the prediction $\underline{\hat{x}}_\tau$ for any future time $\tau$ based, on the one hand, on any simulation $\underline{x}_\tau$ independent of the observations, and, on the other hand, on the observations $\underline{\hat{\boldsymbol{z}}}$, by:

$$\underline{\hat{x}}_\tau = \underline{x}_\tau + \boldsymbol{c}_\tau^{\mathrm{T}} \boldsymbol{C}^{-1} \left( \underline{\hat{\boldsymbol{z}}} - \underline{\boldsymbol{z}} \right), \qquad \mathrm{var}\left[ \underline{\hat{x}}_\tau | \underline{\hat{\boldsymbol{z}}} \right] = \gamma_1 - \boldsymbol{c}_\tau^{\mathrm{T}} \boldsymbol{C}^{-1} \boldsymbol{c}_\tau \tag{31}$$

where $\underline{\boldsymbol{z}} \coloneqq \left[ \underline{x}_0, \underline{x}_{-1}, \dots, \underline{x}_{-n} \right]$, $\underline{\hat{\boldsymbol{z}}} \coloneqq \left[ \underline{\hat{x}}_0, \underline{\hat{x}}_{-1}, \dots, \underline{\hat{x}}_{-n} \right]$, $\boldsymbol{c}_\tau \coloneqq \mathrm{cov}\left[ \underline{x}_\tau, \underline{\boldsymbol{z}} \right]$ and $\boldsymbol{C} \coloneqq \mathrm{cov}\left[ \underline{\boldsymbol{z}}, \underline{\boldsymbol{z}} \right]$.

# Concluding remarks

- Atmospheric processes, including turbulence, do not exhibit time irreversibility except for the finest time scales. The same applies to the rainfall process at scales above hourly.
- Time asymmetry is prominent in the streamflow process at all scales below several days and it is important to preserve in stochastic simulations.
- Classical time series models cannot preserve time asymmetry.
- This offers another excellent opportunity to abandon classical time series models and thus escape from traps related to them.
- The framework developed to tackle time asymmetry can be used in any simulation problem, and offers several advantages even for time symmetric processes.
- The processes that exhibit time irreversibility cannot be Gaussian and linear at the same time. However, the framework developed can handle non-Gaussian process in an analytical manner through high-order moments (determined theoretically from the distribution, not estimated from the data).
- The framework developed can also be used for making stochastic forecasts.

# Main points of the proposed framework and its main differences from popular methods

- 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.
- The model needs to be parsimonious in order to be useful. Inflationary models, while giving an impression of a good fit, in fact entail (often hidden) high uncertainty.
- 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 (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 is an exception, as it is not affected, and has some additional advantages (Dimitriadis and Koutsoyiannis, 2015) which make it the preferable tool in stochastic modelling.
- Parameter estimation needs to consider statistical bias, which is present in all estimators except in that of the mean.
- 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.

# 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.
- Bachelier, M.L., 1900. Théorie de la spéculation, A*nn. Ecole Norm. Super.*, 17, 21-86.
- Beard, L.R., 1965. Use of interrelated records to simulate streamflow. *Proc. ASCE, J. Hydraul. Div.*, 91(HY5), 13-22.
- Bennett, S., 1996. A brief history of automatic control. *IEEE Control Systems Magazine*, 16(3), 17-25.
- 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 Hauptsatze 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 monocyklischer 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.
- 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.
- Chow, V.T., and Kareliotis, S.J., 1970. Analysis of stochastic hydrologic systems. *Water Resources Research*, 6(6), 1569-1582.
- Clausius, R., 1850. Über die bewegende Kraft der Wärme und die Gesetze, welche sich daraus für die Wärmelehre selbst ableiten lassen. *Annalen der Physik*, 79, 500–524, doi: 10.1002/andp.18501550403 (in German). English translation: Clausius, R. 1851. On the moving force of heat, and the laws regarding the nature of heat itself which are deducible therefrom. *London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 4, 2 (VIII), 102–119.
- Clausius, R., 1854. Über eine veränderte Form des zweiten Hauptsatzes der mechanischen Wärmetheorie. *Annalen der Physik*, 481–506, doi: 10.1002/andp.18541691202 (in German). English translation: Clausius, R.,1856. On a modified form of the second fundamental theorem in the mechanical theory of heat. *London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 4, 2, 86.
- Clausius, R., 1865. Über verschiedene für die Anwendung bequeme Formen der Hauptgleichungen der mechanischen Wärmetheorie. *Annalen der Physik und Chemie*, 125 (7), 353–400. doi: 10.1002/andp.18652010702 (in German).
- 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.
- Doran, C., 2011. Anemometer - Sonic at ABLE Beaumont Site Data. Version 1.0. UCAR/NCAR – Earth Observing Laboratory, http://data.eol.ucar.edu/dataset/45.910. Accessed 2018-09-17.

# References (2)


- Eddington, A., 1928. *The Nature of the Physical World*. Cambridge University Press, Cambridge, UK.
- 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.
- Fourier, J., 1822. Théorie Analytique de la Chaleur. Firmin Didot Père et Fils, Paris.
- Galton, F., 1890. Dice for statistical experiments. *Nature*, 42, 13–14.
- Giuliani, M., Herman, J.D., Castelletti, A., and Reed, P., 2014a. Many-objective reservoir policy identification and refinement to reduce policy inertia and myopia in water management. *Water Resources Research*, 50(4), 3355-3377.
- Giuliani, M., Mason, E., Castelletti, A., Pianosi, F. and Soncini-Sessa, R., 2014b. Universal approximators for direct policy search in multi-purpose water reservoir management: A comparative analysis. *IFAC Proceedings Volumes*, 47(3), 6234-6239.
- Hazen, A., 1914. Storage to be provided in impounding reservoirs for municipal water supply. *Trans. Amer. Soc. Civil Eng.*, 77, 1539-1640.
- Hollinger, H.B., and Zenzen, M.J., 1985. *The Nature of Irreversibility: A Study of its Dynamics and Physical Origins*. D. Reidel Publishing Company, Dordrecht, Holland.
- 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 Engrs.*, 116, 776–808.
- Kang, H.S., Chester, S., and Meneveau, C., 2003. Decaying turbulence in an active-grid-generated flow and comparisons with large-eddy simulation. *J. Fluid Mech.*, 480, 129–160.
- 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.
- Kolmogorov, A.N., 1931. Uber die analytischen Methoden in der Wahrscheinlichkcitsrechnung, *Mathematische Annalen*, 104, 415-458. (English translation: On analytical methods in probability theory, In: Kolmogorov, A.N., 1992. *Selected Works of A. N. Kolmogorov - Volume 2, Probability Theory and Mathematical Statistics*, A. N. Shiryayev, 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. Wienersche Spiralen und einige andere interessante Kurven in Hilbertschen, *Raum. Dokl. Akad. Nauk* URSS, 26, 115–118 (In German).
- 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. Shiryayev ed., Kluwer, Dordrecht, The Netherlands, pp. 321-330).

# References (3)


- 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., 2017. Entropy production in stochastics. *Entropy*, 19 (11), 581, doi: 10.3390/e19110581.
- Koutsoyiannis, D., 2019a. Time's arrow in stochastic characterization and simulation of atmospheric and hydrological processes. *Hydrological Sciences Journal*, doi:10.1080/02626667.2019.1600700.
- Koutsoyiannis, D., 2019b. Simple stochastic simulation of time irreversible and reversible processes. *Hydrological Sciences Journal* (in review) (http://www.itia.ntua.gr/en/docinfo/1975/.
- D. Koutsoyiannis, 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, 2019c.
- Koutsoyiannis, D., 2020. *Stochastics of Hydroclimatic Extremes – A Cool Look at Risk* (in preparation).
- Koutsoyiannis, D. and Economou, A., 2003. Evaluation of the parameterization-simulation-optimization approach for the control of reservoir systems. *Water Resources Research*, 39(6), doi:10.1029/2003WR002148.
- 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.
- 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 π 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.
- Nalbantis, I. and Koutsoyiannis, D., 1997. A parametric rule for planning and management of multiple-reservoir systems. *Water Resources Research*, 33(9), pp.2165-2177.
- Niederreiter, H., 1992. *Random Number Generation and Quasi-Monte Carlo Methods*, Society for Industrial and Applied Mathematics, Philadelphia.

# References (4)


- 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 π and e to more than 2000 Decimal Places. *Mathematical Tables and Other Aids to Computation*, 4, 11-15.
- Slutzky, E.,1925, Über stochastische Asymptoten und Grenzwerte, Metron, 5(3), 3-89.
- Slutsky, E., 1927. Slozhenie sluchainykh prichin, kak istochnik tsiklicheskikh protsessov. *Voprosy kon''yunktury*, 3, 34 – 64. 1927 [English edition: Slutzky, 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 ùn critérium de la convergence stochastique des ensembles des valeurs éventuelles). *Comptes rendus Acad. Sci.*, 187, 370.
- Slutzky, 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.
- Slutzky, 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. Sot. 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 GM. Fair, Harvard University Press, Cambridge, Mass..
- Tippett, L.H.C., 1927. *Random SamplingNumbers*. 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, Uppsala: Almqvist & Wiksells.
- 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., 1938. *A Study in the Analysis of Stationary Time-Series*. PhD thesis, Uppsala, Almquist and Wicksell.
- 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.