The Ultimate Rainmap:
Rainmap Achievements and the Future in Broad-Scale Rain Modelling
Engineering and Physical Sciences Research Council, Oxford, 2007

Can a simple stochastic model generate a plethora of rainfall patterns?

D. Koutsoyiannis, S.M. Papalexiou
Department of Water Resources, National Technical University of Athens
A. Montanari
Department DISTART, Faculty of Engineering, University of Bologna

Part 1: Theoretical framework

How nature works – and how we can model her

Property

- She preserves a few quantities (mass, momentum energy, ...)
- She optimizes a single quantity (Dependent on the specific system - Difficult to find what this quantity is)
- She disallows some states (Dependent on the specific system – Maybe difficult to find)

Mathematical formulation

- One equation per preserved quantity:

  $g_i(\mathbf{s}) = c_i, \quad i = 1, \ldots, k$

  where c_i constants; **s** the size n vector of state variables ($n \geq k$, sometimes $n = \infty$)
- A single "optimation":

  optimize $f(\mathbf{s})$

  [i.e. maximize/minimize $f(\mathbf{s})$] **This is equivalent to many equations** (as many as required to determine **s**) Conversely, many equations can be combined into an "optimation"
- Inequality constraints:

  $h_j(\mathbf{s}) \geq 0, \quad j = 1, \ldots, m$
- In conclusion, we may find how nature works and model her effectively solving the problem:

  optimize $f(\mathbf{s})$
  s.t. $g_i(\mathbf{s}) = c_i, \quad i = 1, \ldots, k$
  $h_j(\mathbf{s}) \geq 0, \quad j = 1, \ldots, m$

Optimizable quantity in simple systems

- Fermat's principle for the light propagation
  - The path taken between two points by a ray of light is the path that can be traversed in the **least time**
  - More correct to substitute "extremal" (or "stationary") for "least" (e.g. in concave mirrors the light path corresponds to maximum time)
- Principle of **least action** (Hamilton's principle – applicable both in classical and in quantum physics)
  - From all possible motions between two points, the true motion has least action
  - More correct to substitute "extremal" (or "stationary") for "least"

Optimizable quantity in complex systems

- Entropy – entropie – Entropie – entropia – entropía – entropi – entrópia – entroopia – entropija – энтропия – ентропія – 熵 – エントロピー – مقياس – אנטרופיה– εντροπία
- The word is ancient Greek (εντροπία, a feminine noun meaning: turning into; turning towards someone's position; turning round and round)
- The scientific term is due to Clausius (1850); the entropy concept was fundamental to formulate the 2nd law of thermodynamics
- Boltzmann (1877) (later complemented by Gibbs, 1948), gave it a statistical mechanical content, showing that entropy of a macroscopical stationary state is proportional to the logarithm of the number *w* of possible microscopical states that correspond to this macroscopical state
- Shannon (1948) generalized the mathematical form of entropy and also explored it further
- At the same time, Kolmogorov (1957) founded the concept on more mathematical grounds on the basis of the measure theory
- Entropy is a measure of uncertainty, or (depending on the discipline) disorder and complexity

The principle of maximum entropy (ME) and the marginal distribution

- The Boltzmann-Gibbs-Shannon entropy for a continuous random variable X with density function $f(x)$ is by definition (e.g. Shannon, 1949; Papoulis, 1991)

$$S = E\left(-\ln f(x)\right) = -\int_{-\infty}^{\infty} f(x)\ln\left(f(x)\right)dx$$

- The principle of ME, as formalized by E.T. Jaynes (1957a, b), states that the (unknown) density function $f(x)$ of a random variable X is the one that maximize the entropy S, subject to any known constrains
- Application of the ME principle using the Boltzmann-Gibbs-Shannon entropy with simple constraints of known mean μ and variance σ^2 results in

$$f(x) = \exp(-\lambda_0 - \lambda_1 x - \lambda_2 x^2) \qquad (1)$$

  where λ_0, λ_1 and λ_2 are parameters depending on the known mean and variance; inspection of (1) shows that it is the normal density function
- In statistical physics, if X denotes the momentum of molecules or atoms in a gas volume, the mean and variance constraints correspond precisely to the principles of preservation of momentum and energy

An entropic approach to rainfall – Step 1

- Let X_i denote the rainfall rate at time i discretized at a fine time scale (tending to zero)
- What we definitely know about X_i is $X_i \geq 0$
- Maximization of entropy with only this condition is not possible
- Now let us assume that rainfall has a specific mean μ
- Maximization of entropy with constraints

  $$X_i \geq 0, \qquad E[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx = \mu$$

  results in the exponential distribution: $f(x) = \exp(-x/\mu)/\mu$
- In addition, let us assume that there is some time dependence of X_i, quantified by $E[X_i, X_{i+1}] = \gamma$; this will introduce an additional constraint for the multivariate distribution

  $$E[X_i\, X_{i+1}] = \int_{-\infty}^{\infty} x_i\, x_{i+1}\, f(x_i, x_{i+1})\, dx_i\, dx_{i+1} = \gamma = \rho\, \sigma^2 + \mu^2$$

  Here ρ is the correlation coefficient ($\rho > 0$) and σ is the standard deviation ($\sigma = \mu$ for the exponential distribution and thus $\gamma = \rho\, \sigma^2 + \mu^2 = (\rho + 1)\, \mu^2 > \mu^2$)
- Entropy maximization in multivariate setting will result (?) in Markovian dependence

An entropic approach to rainfall – Step 2

- The constant mean constraint in rainfall modelling does not result from a natural principle – as for instance in the physics of an ideal gas, where it represents the preservation of momentum
- Although it is reasonable to assume a specific mean rainfall, we can allow this to vary in time
- In this case we can assume that the mean at time i is the realization of a random process M_i which has mean μ and lag 1 autocorrelation $\rho^M > \rho$
- Application of the ME principle will yield that M_i is Markovian with exponential distribution
- Then application of conditional distribution algebra results in

  $$f(x) = 2\, K_0(2\, (x/\mu)^{1/2})/\mu, \; F(x) = 1 - 2\, (x/\mu)^{1/2}\, K_1(2\, (x/\mu)^{1/2})/\mu$$

  where $K_n(x)$ is the modified Bessel function of the second kind
- The moments of this distribution are $E[X^n] = \mu^n\, n!^2$ (note: in exponential distribution $E[X^n] = \mu^n\, n!$) so that

  $$E[\mathrm{X}] = \mu,\ \mathrm{Var}[X] = 3\, \mu^2 \rightarrow C_{\mathrm{V}} = \sigma/\mu = \sqrt{3} > 1$$
- The dependence structure becomes more complex than Markovian (difficult to find an analytical solution)

An entropic approach to rainfall – Step 3

- Proceeding in a similar manner as in step 2, we can now replace the constant mean μ of the process M_i with a varying mean, represented by another stochastic process N_i with mean μ and lag 1 autocorrelation $\rho^N > \rho^M > \rho$
- In this manner we can construct a chain of processes, each member of which represents the mean of the previous process
- By construction, the lag 1 autocorrelations of these processes form a monotonically increasing sequence, i.e. …. $> \rho^N > \rho^M > \rho$
- The scale of change or fluctuation of each process of the chain is a monotonically increasing sequence, i.e. …. $> q^N > q^M > q$, where $q := (-\ln \rho)^{-1}$; the scale of fluctuation represents the time required for the process to decorrelate down to an autocorrelation 1/e
- The (unconditional) mean of all processes is the same, μ
- All moments except the first form an increasing sequence as we proceed through the chain; higher moments increase more
- Analytical handling of the marginal distribution and the dependence structure is very difficult
- However we can easily inspect the idea using Monte Carlo simulation

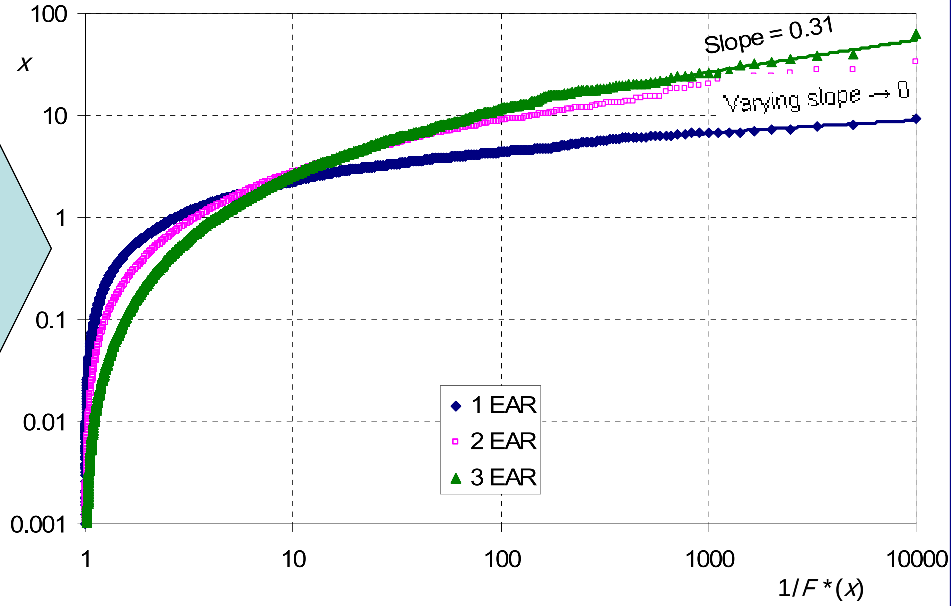
A demonstration using a chain with three processes

- Simulation of a Markovian process with exponential distribution is easy and precise – a special case of the Gamma autoregressive model (GAR; Lawrance and Lewis, 1981; Lawrance, 1982; Fernadez and Salas, 1991) for shape parameter equal to 1 (will be denoted EAR)
- Simulations with a length 10 000 were performed for the following cases (for comparison)

	Case	1 EAR	2 EAR		3 EAR		
	Process	L	M	L	N	M	L
Processes in chain	Mean	1	1	-	1	-	-
	Lag 1 autocorrelation	0.47	0.9	0.25	0.99	0.85	0.2
	Scale of fluctuation	1.33	9.5	0.72	99.5	6.2	0.62
Final process	Mean	1	1		1		
	Standard deviation	1	1.73		2.75		
	Lag 1 autocorrelation	0.47	0.47		0.47		

Simulation results – distribution function

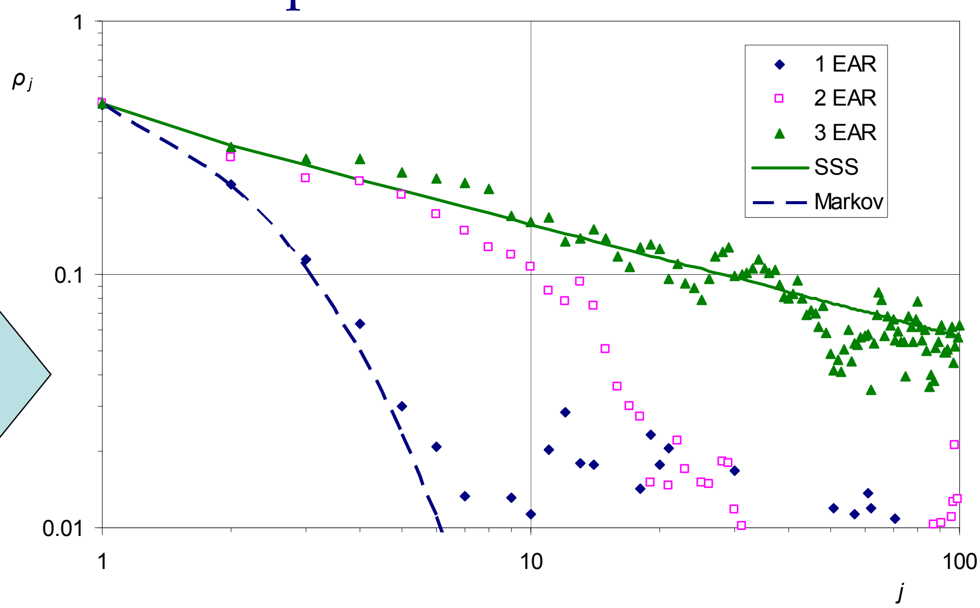
Logarithmic plot of "rainfall intensity" (x) vs. empirically estimated exceedence probability ($F^*(x) := 1 - F(x)$, where $F(x)$ is the distribution function)



As the number of processes in the chain increases, the distribution tail moves toward higher "rainfall intensity" values and its shape changes from exponential type to power type

Simulation results – dependence structure

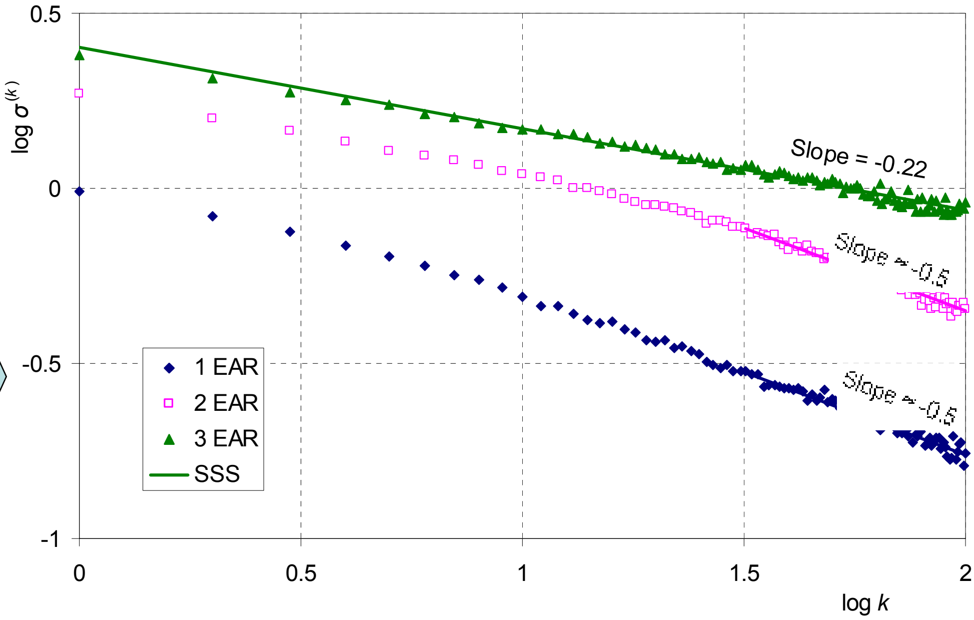
Logarithmic plot of autocorrelation coefficient ρ_j vs. lag j



As the number of processes in the chain increases, the shape of the autocorrelation function changes from Markovian (exponential decay – short range dependence) to power type (long range dependence)

The latter type is a characteristic of the Hurst phenomenon, which can be represented by a simple scaling stochastic process (SSS process)

Simulation results – variation of the aggregated process



Logarithmic plot of standard deviation $\sigma^{(k)}$ of the process aggregated at scale k, vs. scale k

The slope of the logarithmic plot (as $k \to \infty$) is $H - 1$ where H is the Hurst exponent

The slope in the "1 EAR" case is –0.5, i.e. $H = 0.5$, meaning no Hurst behaviour

The slope in "3 EAR" is –0.22, i.e. $H = 0.78$, suggesting a Hurst behaviour

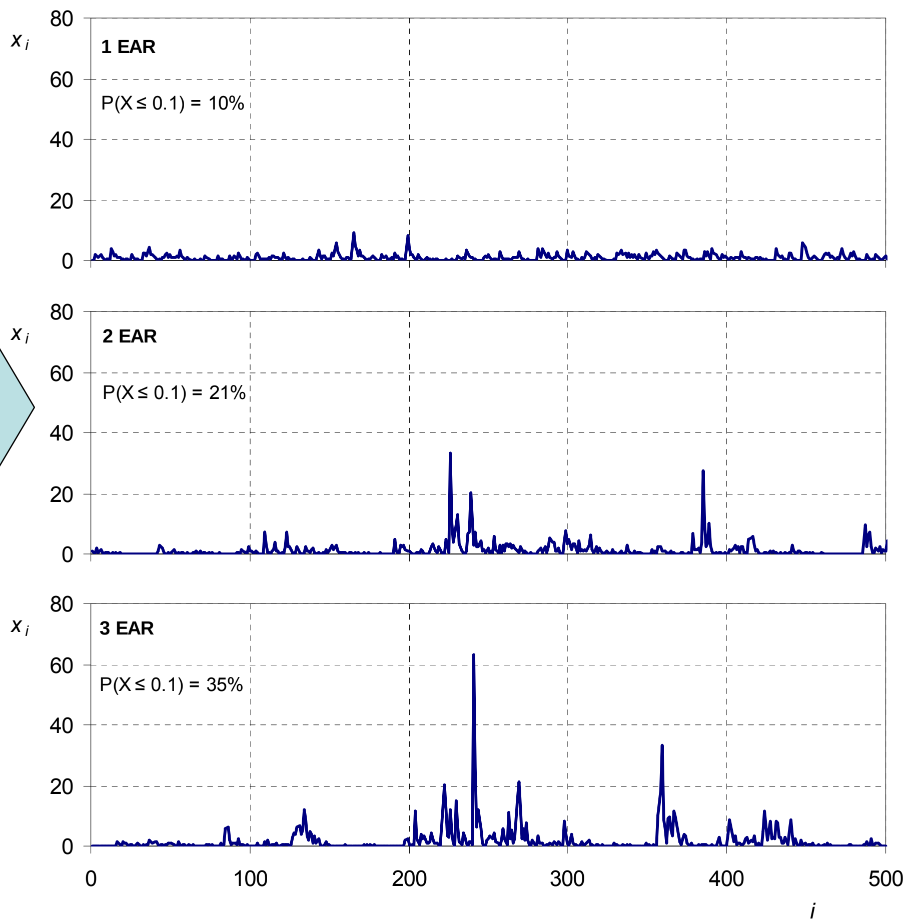
Simulation results – general behaviour

Plots of parts of the generated time series (selected so as to include the maximum over 10 000 generated values)

As the number of processes in the chain increases the general shape changes:

From monotony to rich patterns

From steadiness to intermittency



Paradoxes of the classical entropic framework and required adaptations

- The classical entropy maximization theory cannot produce a stochastic process that is a realistic representation of rainfall in a single optimization
- Despite the fact that the compound processes resulting from 2, 3, … chain processes have greater variance (and uncertainty) than a single EAR process, the classical definition of entropy assigns to the former a lower value of entropy
- Despite the fact that the long-range dependence implies higher uncertainty than short-range dependence, the classical definition of entropy assigns to the former a lower value of entropy
- The first paradox can be remedied by generalizing the entropy definition
- The second paradox can be remedied using a multi-scale setting of the entropy maximization

The Tsallis entropy and long distribution tails

- A generalization of the Boltzmann-Gibbs-Shannon entropy has been proposed by Tsallis (1998, 2004)

$$S_q = \frac{1 - \int_{-\infty}^{\infty} \left(f(x)\right)^q dx}{q - 1}$$

with $q = 1$ corresponding to the Boltzmann-Gibbs-Shannon entropy
- Maximization of Tsallis entropy with known μ and σ^2 yields

$$f(x) = [1 + \xi\,(\lambda_0 + \lambda_1 x + \lambda_2 x^2)]^{-1-1/\xi}, \quad x \geq 0 \qquad (2)$$

where λ_0, λ_1, λ_2 and ξ are parameters; it can be shown that (2) is mathematically equivalent to the so-called Tsallis distribution (Tsallis et al., 1995; Prato and Tsallis, 1999)
- Clearly, this has an over-exponential (power-type) distribution tail
- An alternative approach is to assume a normalizing nonlinear transformation $z = g(x)$ (based on (2)) and then apply the classical definition of entropy on z

The principle of maximum entropy on a multivariate setting, and the long autocorrelation tails

- Maximum entropy + Dominance of a single time scale → Time independence
- Maximum entropy + Time dependence + Dominance of a single time scale → Markovian (short-range) time dependence
- Maximum entropy + Time dependence + Equal importance of time scales → Time scaling (long-range dependence / Hurst phenomenon)
- As there is no reason that nature would choose a specific time scale for entropy maximization, long autocorrelation tails are reasonable
- The omnipresence of long autocorrelation tails in numerous long hydrologic time series, validates the applicability of the ME principle
- For details see Koutsoyiannis (2005b)

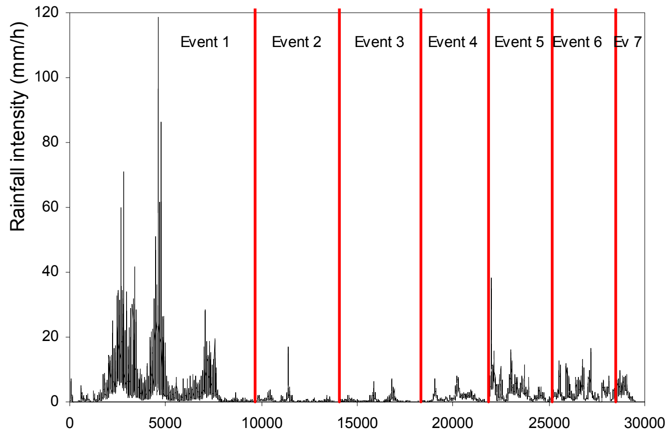
Part 2: Testing of the theoretical framework based on a high temporal resolution rainfall data set

Premise

- An high resolution data set has been obtained by the Hydrometeorology Laboratory at the University of Iowa using devices that are capable of high sampling rates, once every 5 or 10 seconds (Georgakakos et al., 1994)
- The data set offers a basis for fundamental investigations that could provide insights for the characterization and mathematical modeling of the rainfall process
- A first target of this study is to investigate whether the data verify or falsify the applicability of the maximum entropy hypothesis, i.e., whether or not
  - the coefficient of variation (σ/μ) at fine scale is greater than 1
  - the marginal distribution has long (power type) tail
  - the autocorrelation function has a long (power type) tail
- A second target is to investigate whether or not all events, despite large differences among them, could be regarded as outcomes (sample functions) of a single stochastic process
- The overall question is: Could a single stochastic process based on maximum entropy considerations produce a plethora of different types of events statistically resembling the actual events?

The original data

Plot of the high resolution data set (consisting of seven storms occurred in Iowa in 1990-91

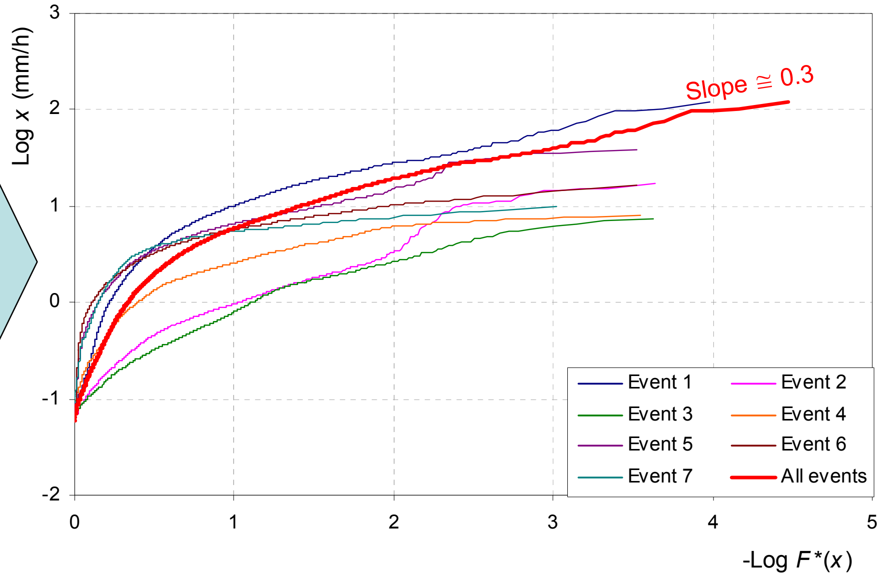


Statistics of the seven storms and the compound record of all storms

Event #	1	2	3	4	5	6	7	All
Sample size	9697	4379	4211	3539	3345	3331	1034	29536
Average (mm/h)	3.89	0.50	0.38	1.14	3.03	2.74	2.70	2.29
St. deviation (mm/h)	6.16	0.97	0.55	1.19	3.39	2.20	2.00	4.11
Coefficient of variation	1.58	1.95	1.45	1.04	1.12	0.81	0.74	1.79
Skewness	4.84	9.23	5.01	2.07	3.95	1.47	0.52	6.54
Kurtosis	47.12	110.24	37.38	5.52	27.34	2.91	-0.59	91.00
Hurst Exponent	0.94	0.76	0.92	0.95	0.90	0.87	0.97	0.94

Scaling in state

Logarithmic plots of rainfall intensity (x) vs. empirically estimated (by the Weibull formula) exceedence probability ($F^*(x) := 1 - F(x)$, where $F(x)$ is the distribution function) for the seven events

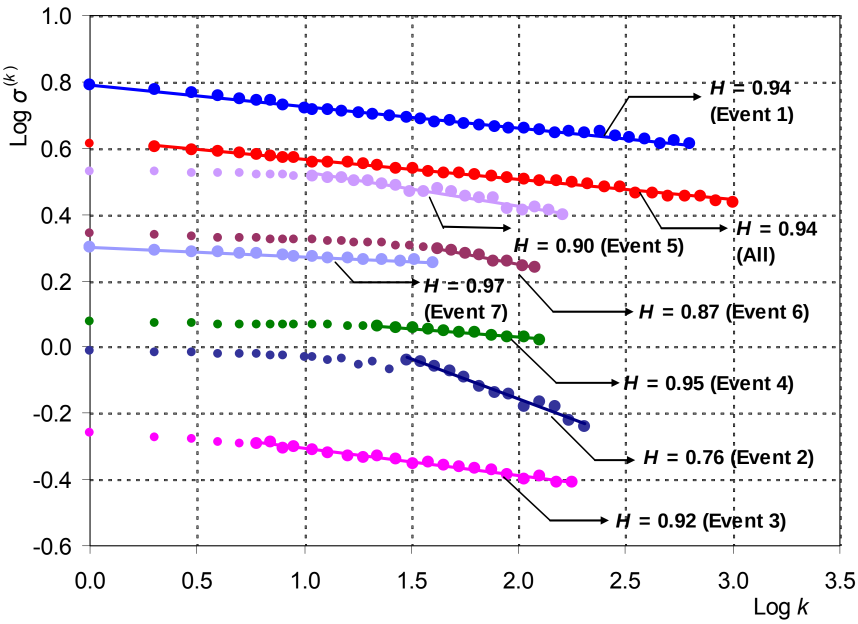


The probability plot of the compound record of all events seems to justify a long distribution tail, which in a double logarithmic plot is depicted as a constant nonzero slope of the empirical distribution (or an asymptotic relationship of the form $x \sim [1/F^*(x)]^\kappa$ for large x).

In five of the seven events (1 to 5 – those with the largest durations) the variation σ/μ is higher than 1, which suggests a power type tail even for a single event

Scaling in time

Logarithmic plots of standard deviation $\sigma^{(\kappa)}$ of rainfall intensity vs. time scale k for the seven events and the compound record



Each of the events separately indicates a Hurst behaviour (a straight line arrangement of points corresponding to different time scales) with Hurst exponent ranging form 0.76 to 0.97

The Hurst behaviour is very clear in the compound record of all events, with Hurst exponent 0.94

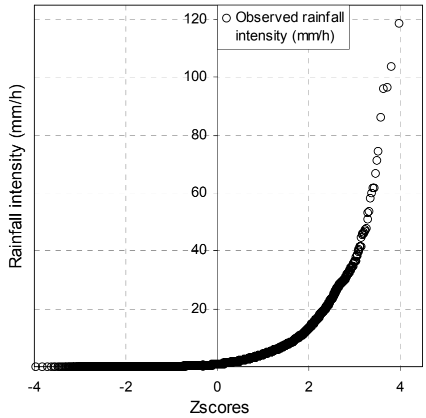
A normalizing transformation

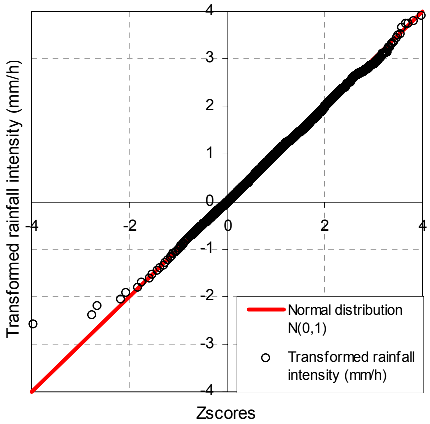
- As the normal distribution is very convenient in building a stochastic model, a normalizing transformation $Z = g(X)$ has been applied to the variable X (rainfall intensity), instead of using the non-normal distribution (2).
- The transformation

$$z = \left(\alpha x^{-\nu} + \beta\right)\left(\psi + \sqrt{\left(1 + \frac{1}{\kappa}\right)\ln\left(\kappa\left(x - \psi\right)^2 + 1\right)}\right)$$

effectively transforms the observed data to normal, also being consistent with the Tsallis distribution (implies a power type tail for X)

- The parameters of the transformation were estimated by minimizing the square error (SE) of the model and the empirical distribution function
- The inverse transformation $X = g^{-1}(Z)$, necessary for de-normalizing the synthetic normal series, has been handled numerically, due to lack of analytical solution



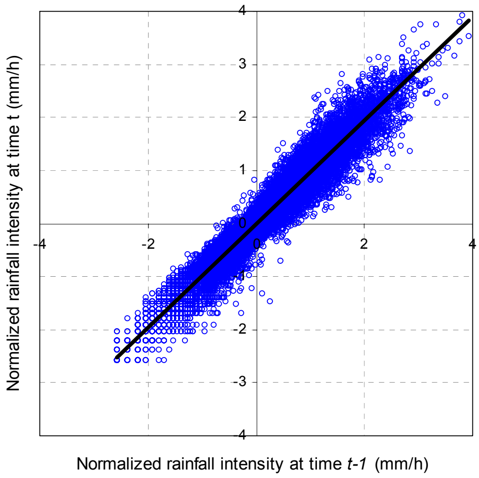


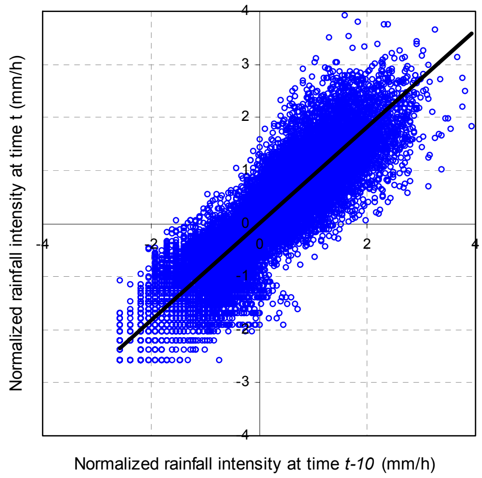
The principle of maximum entropy and the linearity in multivariate distribution

The maximum entropy principle, implies linear relationships in consecutive items of a Gaussian stochastic process

Specifically, provided that a specific transformation of a process has normal marginal distribution, application of the maximum entropy principle produces multivariate normal distribution for any number of variables (Papoulis, 1991)

Multivariate normal distribution entails linear relationships among variables; the following figures verify the theory





Part 3: Models and simulation results

Model 1 – with long-range dependence (M1)

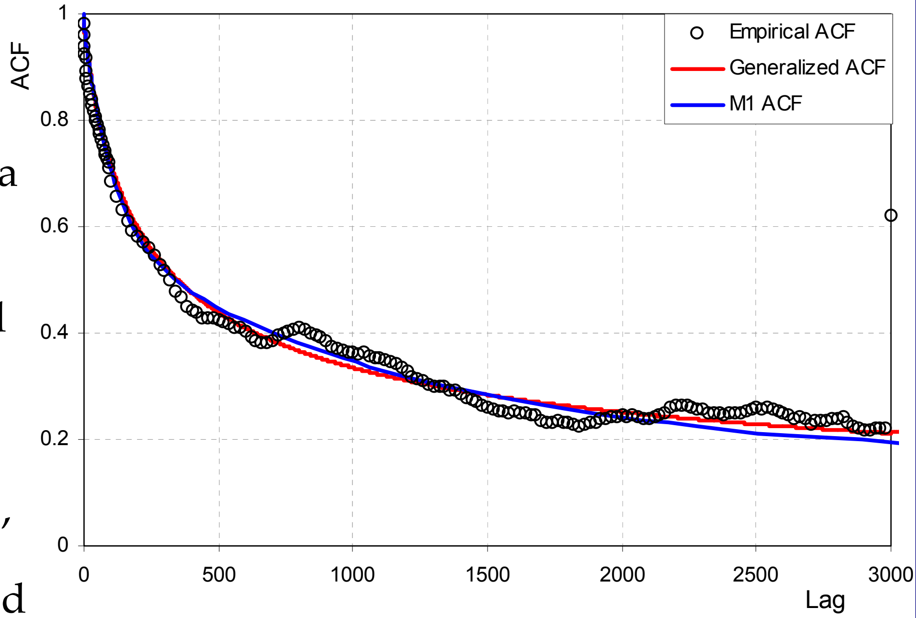
- A generalized autocovariance structure (GAS) is adopted: $\gamma_j = \gamma_0 \, (1 + \alpha\beta j)^{-1/\beta}$, where γ_j, is the lag j autocovariance and α and β are constants (Koutsoyiannis, 2000); clearly, the case $\beta > 0$ corresponds to power type tail
- Approximation of a process with GAS can be obtained using a condition-wise chain of Markovian processes as in Part 1
- In a Gaussian setting, the condition-wise chain can be replaced by a sum of independent AR(1) processes (see details in Koutsoyiannis, 2002)
- Three to four such chained processes are enough to obtain good approximation of GAS for lags 1000-10 000
- With four AR(1) processes, the approximate model M1 and its autocovariance function (ACF) for lag j are

  $$W_i = \sum_{l=1}^{4} Y_{l,i} \qquad \gamma_{\mathrm{M1},j} = \sum_{l=1}^{4} c_l \rho_l^j$$

  where ρ_i is the lag one autocorrelation coefficient of the *i*th AR(1) process and c_i are constants (summing up to 1)
- Each of the four AR(1) processes is $Y_{i,t} = \rho_i Y_{i,t-1} + V_{i,t}$ where $V_{i,t}$ are independent, identically distributed, random variables with mean $(1 - \rho_i)\mu$ and variance $(1 - \rho_i^2)c_i\gamma_0$.

Fitting of model 1

- GAS was fitted to the empirical ACF of the transformed data by minimizing the square error
- The transformed data ~N(0,1), so $\mu = 0$ and $\gamma_0 = 1$
- The parameters of the model M1, (c_i and ρ_i, totally 7) were evaluated by minimizing the square error of $\gamma_{M1,j}$ and the fitted GAS



- The figure suggests satisfactory approximation of the empirical ACF for lags up to 3 000

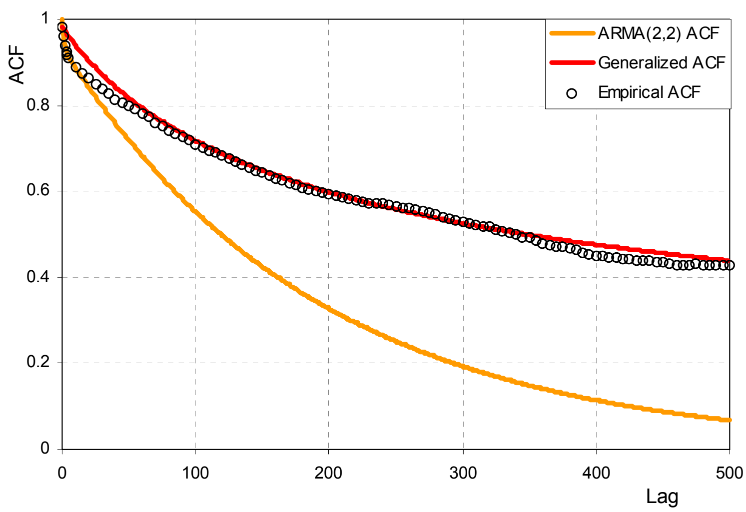
Model 2 – with short-range dependence

- For comparison, an ARMA(2,2) model was also examined, namely

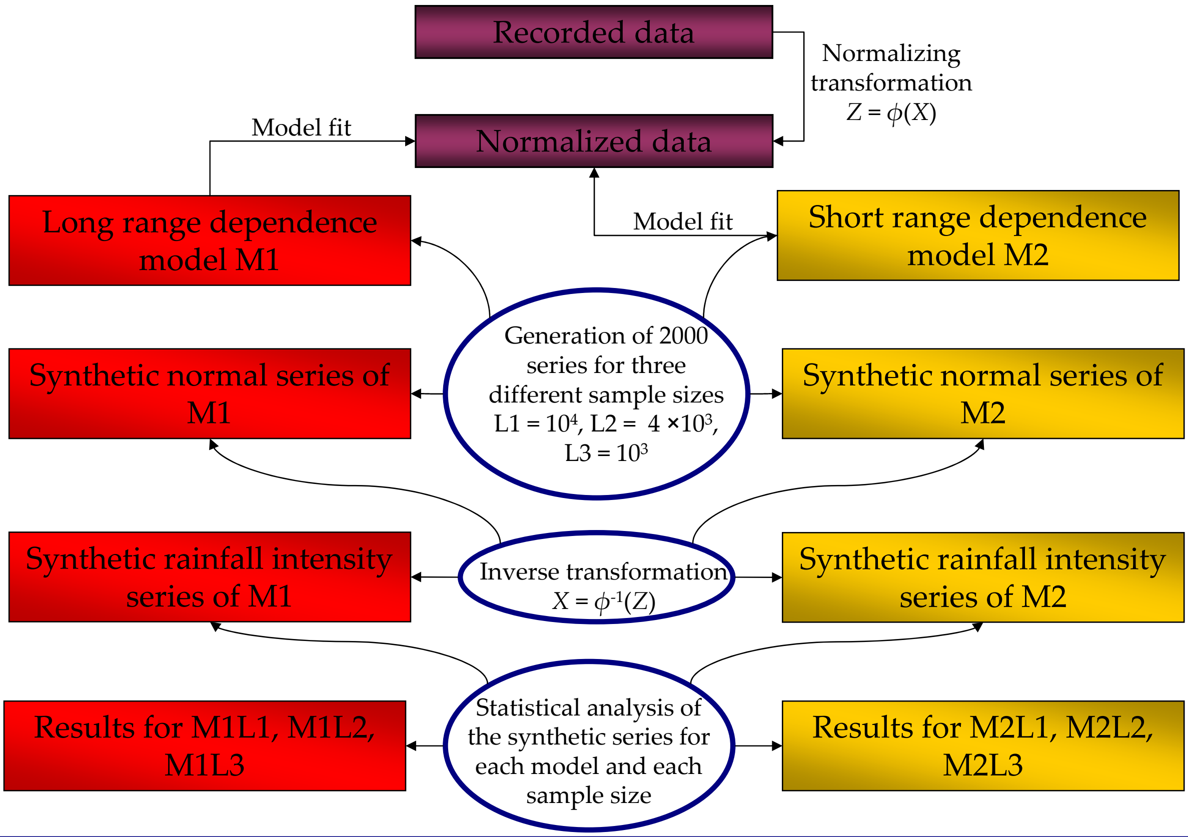
$$Z_i = a_2\, Z_{i-2}\, a_1\, X_{i-2} + b_2\, V_{i-1} + b_1\, V_i$$

where a_i, b_i, are the model coefficients and V_i are independent, identically distributed random variables with mean μ_V and standard deviation σ_V

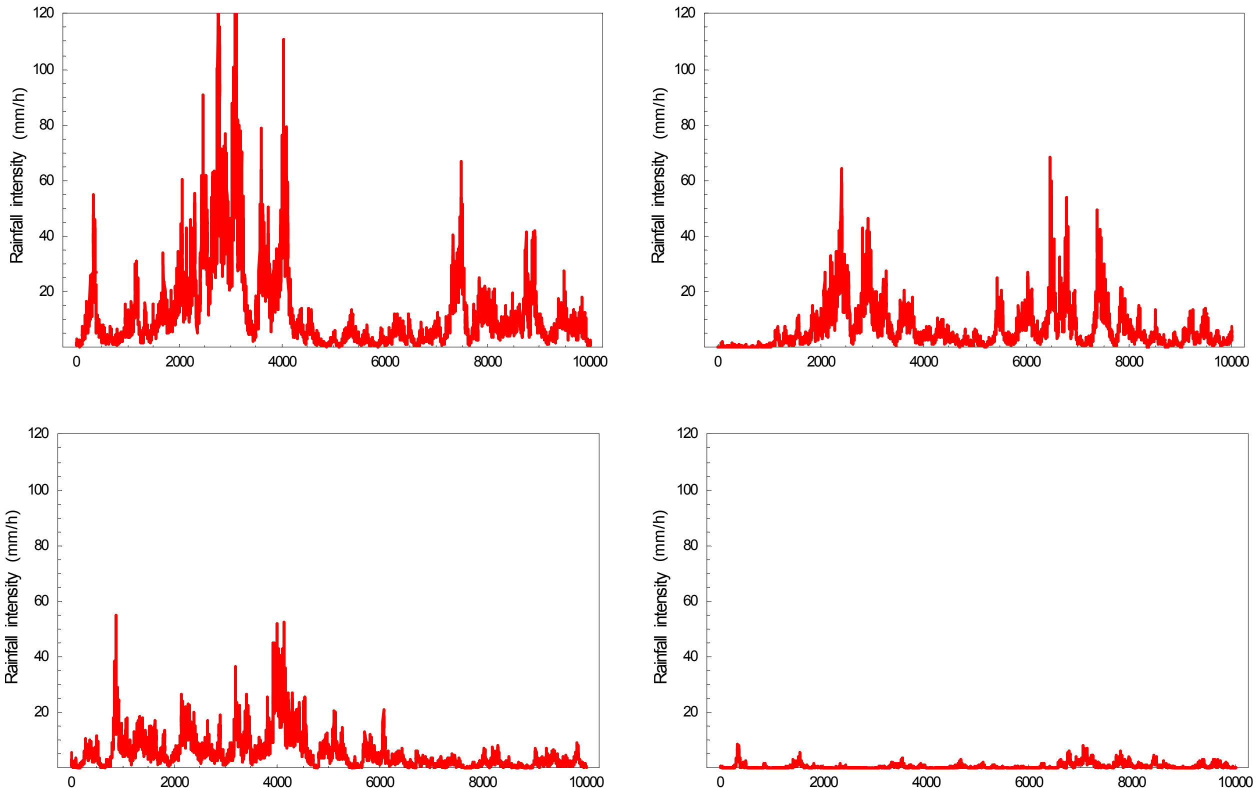
- The model was fitted so as to reproduce the variance and the first four autocovariances of the transformed process
- Despite the fact that ARMA(2,2) is a short-range dependence model, the figure shows that it is can have positive autocorrelation values for lags as high as 500



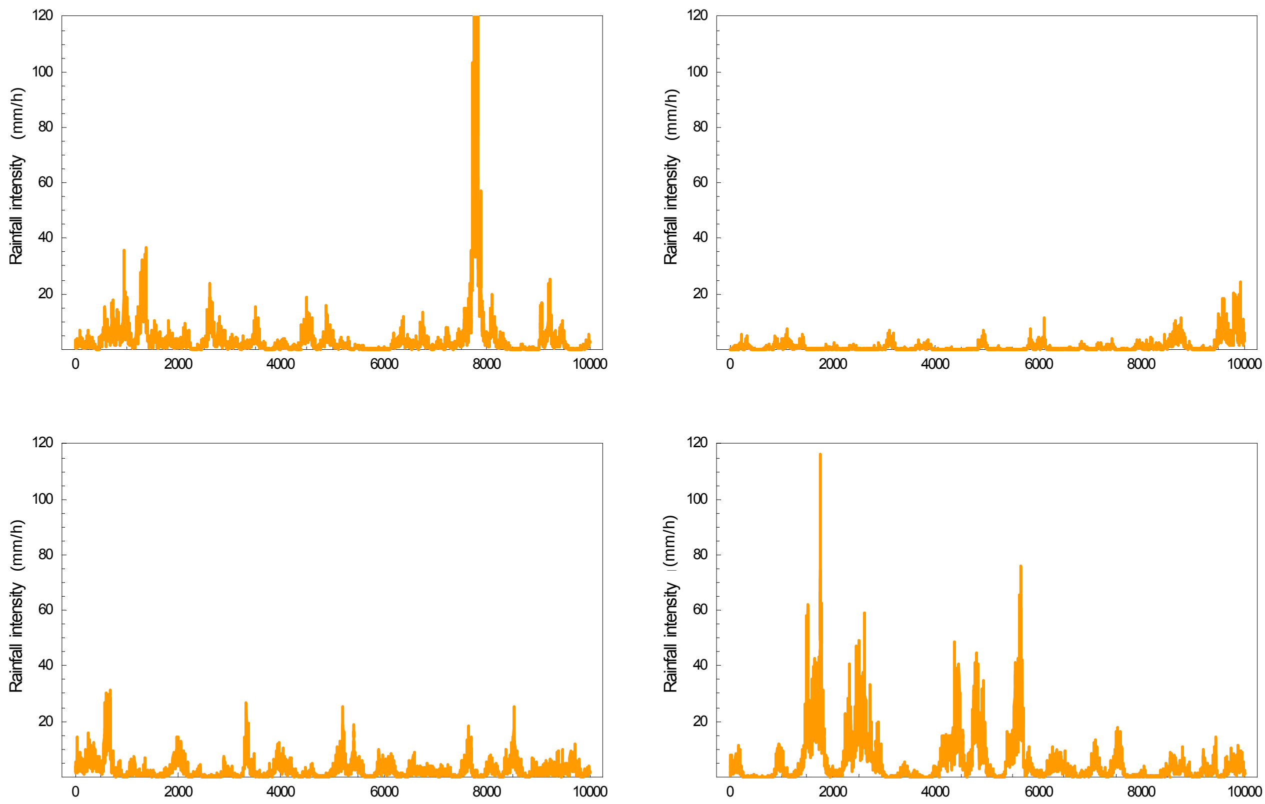
Simulation procedure



Synthetic series of the long range dependence model

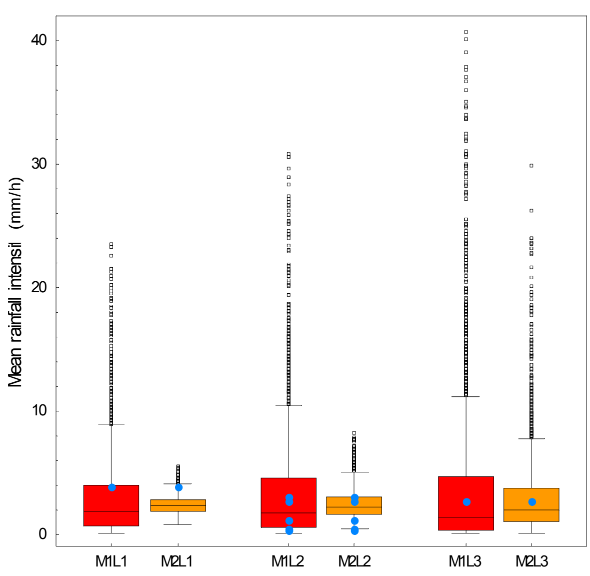


Synthetic series of the ARMA(2,2) model

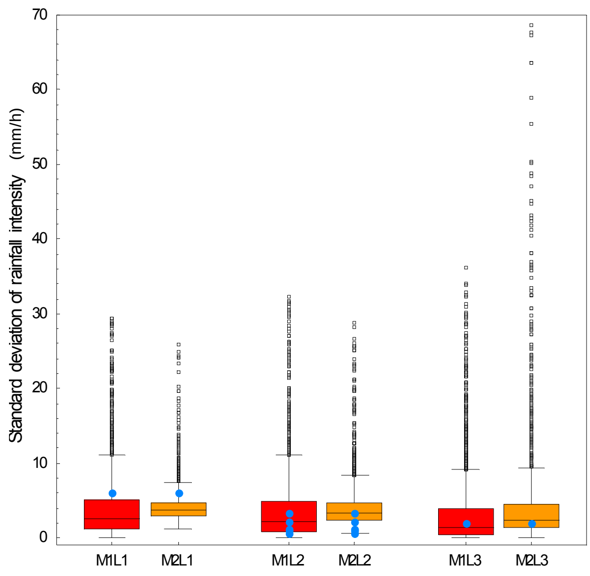


Simulation results: Mean and standard deviation

- Box plots of the estimated mean rainfall intensity of the synthetic series produced by models M1 and M2 for sample sizes L1, L2 and L3; blue dots represent the observed means

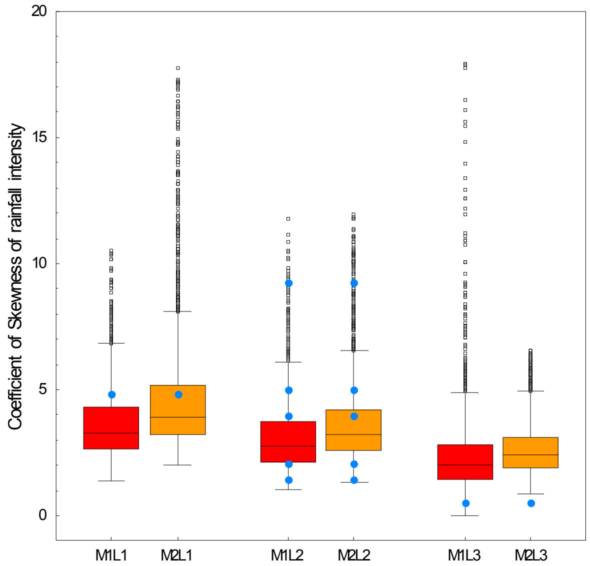


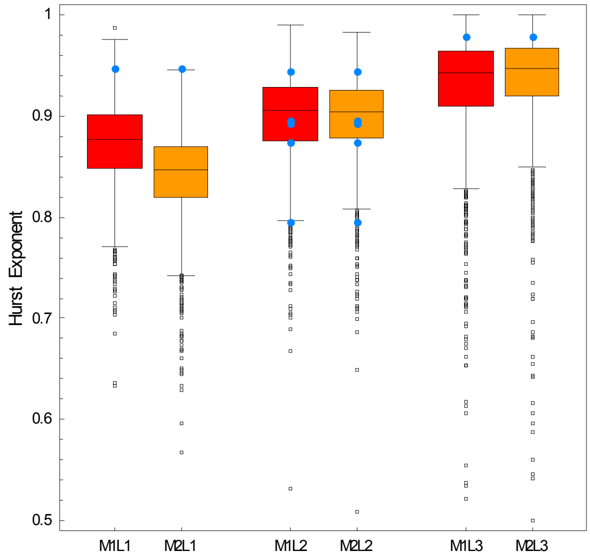
- Box plots of the estimated standard deviation of the rainfall intensity of the synthetic series produced by models M1 and M2 for sample sizes L1, L2 and L3; blue dots represent the observed standard deviations



Simulation results: Coefficient of skewness and Hurst exponents

- Box plots of the estimated coefficient of skewness of rainfall intensity of the synthetic series produced by models M1 and M2 for sample sizes L1, L2 and L3; blue dots represent the observed ones
- Box plots of the estimated Hurst exponents of rainfall intensity of the synthetic series produced by models M1 and M2 for sample sizes L1, L2 and L3; blue dots represent the observed ones

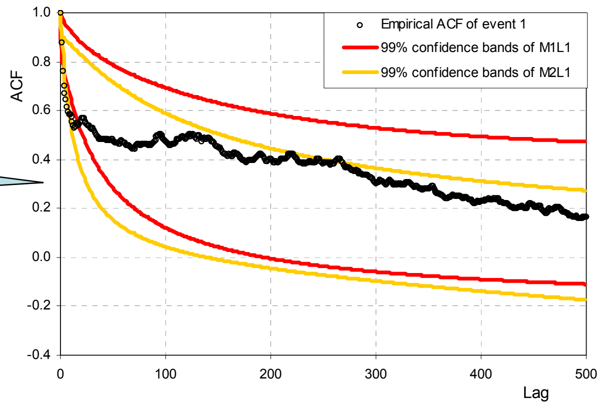




Simulation results: Confidence bands of autocorrelation

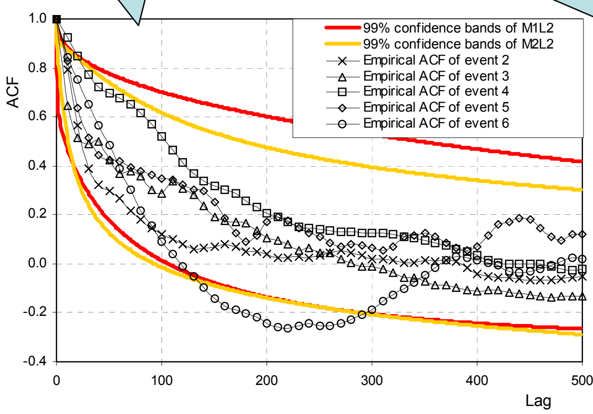
- Empirical autocorrelation function of rainfall events and 99% Monte Carlo confidence bands of ACF for the two models

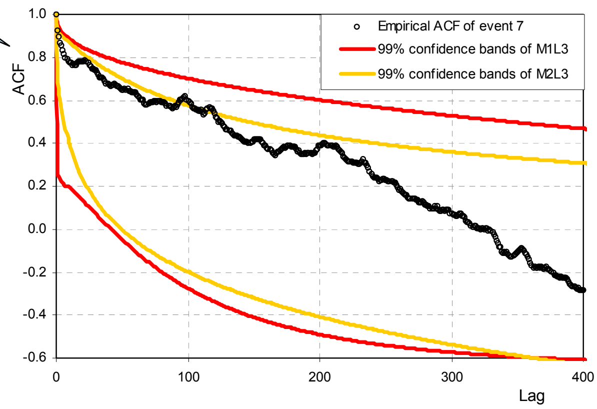
Length L1 = 10^4
event 1



Length L2 = 4×10^3
events 2-6

Length L3 = 10^3
event 7





Conclusions and discussion

- The principle of maximum entropy provides a sound theoretical basis for studying the rainfall process
- A single and rather simple stochastic model based on this principle can represent all rainfall events and all rich patterns appearing in each one making them look very different one another
- From a practical view point, such a model is characterized by high autocorrelation at fine scales, slowly decreasing with lag, as well as by distribution tails slowly decreasing with rainfall intensity
- Both these long tails entail high uncertainty (high entropy)
- Whether the tails are power type is difficult to conclude because both the power-law functions are by definition asymptotic properties; thus merely empirical studies (necessarily implying finite sample sizes) are not enough to verify this behaviour
- It is important that the empirical evidence presented in the current study does not falsify the hypothesis that both tails are long; this hypothesis is strengthened by the principle of maximum entropy and the fact that the M1 (long-range) model had performance superior to the M2 (short-range) model
- In fact the uncertainty/entropy should be higher than demonstrated here because only seven events, despite their high resolution, cannot be representative of the entire rainfall process

References

- Fernadez B. and J.D. Salas (1990), Gamma-autoregressive models for stream flow simulation, *J. Hydraul. Eng.*, 116(11), 1403-1414.
- Georgakakos, K. P., A. A. Carsteanu, P. L. Sturdevant, and J. A. Cramer, (1994), Observation and analysis of Midwestern rain rates, *J. Appl. Meteorol.*, 33, 1433-1444.
- Jaynes, E. T., (1957a), Information theory and statistical mechanics, I, Phys. Rev., 106, 620-630.
- Jaynes, E. T., (1957b), Information and statistical mechanics, II, Phys. Rev., 108, 171-190.
- Koutsoyiannis, D., (2000), A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series, *Water Resources Research*, 36(6), 1519-1533.
- Koutsoyiannis, D., (2002), The Hurst phenomenon and fractional Gaussian noise made easy, *Hydrological Sciences Journal*, 47(4), 573-595.
- Koutsoyiannis, D. (2005a), Uncertainty, entropy, scaling and hydrological stochastics, 1, Marginal distributional properties of hydrological processes and state scaling, *Hydrological Sciences Journal*, 50(3), 381-404.
- Koutsoyiannis, D. (2005b), Uncertainty, entropy, scaling and hydrological stochastics, 2, Time dependence of hydrological processes and time scaling, *Hydrological Sciences Journal*, 50(3), 405-426.
- Lawrance A.J. (1982), The innovation distribution of a gamma distributed autoregressive process *Scandinavian J. Statistics*, 9(4), 234-236.
- Lawrance A.J. and P.A.W. Lewis (1981), A new autoregressive time series model in exponential variables [NEAR(1)], *Adv. Appl. Prob.*, 13(4), 826-845.
- Papoulis, A. (1991), *Probability, Random Variables, and Stochastic Processes* (third edn.), McGraw-Hill, New York.
- Prato, D., and C. Tsallis (1999), Nonextensive foundation of Lévy distributions, *Phys. Rev. E*, 60(2), 2398-2401.
- Shannon, C. E. and Weaver, W. (1949), The Mathematical Theory of Communication. The University of Illinois Press, Urbana, Illinois.
- Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics, *J. Stat. Phys.*, 52, 479-487.
- Tsallis, C. (2004). Nonextensive statistical mechanics: construction and physical interpretation, in *Nonextensive Entropy, Interdisciplinary Applications* (edited by M. Gell-Mann & C. Tsallis), Oxford University Press, New York, NY.
- Tsallis, C., S. V. F. Levy, A. M. C. Souza, and R. Maynard (1995), Statistical-mechanical foundation of the ubiquity of Lévy distributions in nature", *Phys. Rev. Lett.*, 75, 3589–3593.