European Geosciences Union General Assembly 2007, Vienna, Austria, 15 - 20 April 2007

Session NP3.04: Geophysical extremes: Scaling aspects and modern statistical approaches

Scaling properties of fine resolution point rainfall and inferences for its stochastic modelling

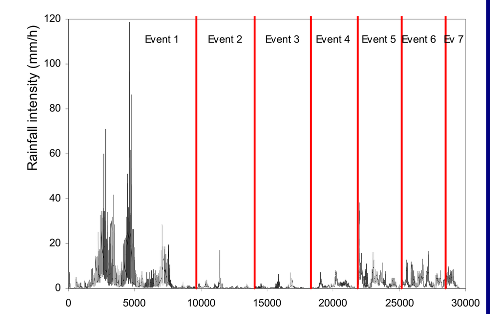
S.M. Papalexiou, D. Koutsoyiannis, Department of Water Resources, National Technical University of Athens;
A. Montanari, Department DISTART, Faculty of Engineering, University of Bologna

1. Abstract

The well-known data set of the University of Iowa comprising fine temporal resolution measurements of seven storm events is analysed. Scaling behaviours are observed both in state and in time. Utilizing these behaviours, it is concluded that a single and rather simple stochastic model can represent all rainfall events and all rich patterns appearing in each of the separate events making them look very different one another. From a practical view point, such a model is characterized by distribution tails de-creasing slowly (in an asymptotic power-type law) with rainfall intensity, as well as by high autocorrelation at fine time scales, decreasing slowly (again in an asymptotic power-type law) with lag. Such a distributional form can produce enormously high rainfall intensities at times and such an autocorrelation form can produce hugely different patterns among different events. Both these behaviours are just opposite to the more familiar processes resembling Gaussian white noise, which would produce very "stable" events with infrequent high intensities. In this respect, both high distribution tails and high autocorrelation tails can be viewed as properties enhancing randomness and uncertainty, or entropy.

2. The original data

This high resolution record consists of seven storms that were measured 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 statistics of each event (see table below), reveals great differences among them.



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
Standard 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	0.79	0.89	0.94	0.89	0.87	0.97	0.89

3. Motivation

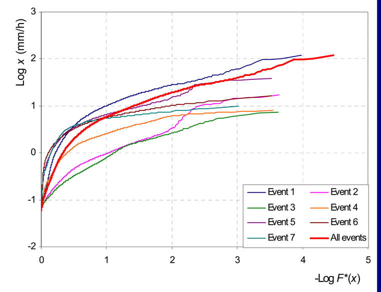
- This unique data set allows inspection of the rainfall process at very fine time scales and was the subject of several advanced and extensive analyses (e.g. Carsteanu and Foufoula-Georgiou, 1996; Kumar and Foufoula-Georgiou, 1997) . Apart from such sophisticated analyses, this data set offers a basis for simpler yet more fundamental investigations that could provide insights for the characterization and mathematical modeling of the rainfall process.
- A major target of this study is to investigate whether all events, despite the large differences, could be regarded as the outcomes (sample functions) of a single stochastic process.
- If a single model could then adequately produce all different type of events, exhibiting such great differences among them, the question that naturally arise is how such a model would look like.
- Of great importance is to answer whether the tails of the marginal distribution function and of the autocorrelation function of such a model would be long (power type), or short (exponential type). In both cases, long tails imply high uncertainty and comply with the maximum entropy principle applied with certain constraints (Koutsoyiannis, 2005a, b)

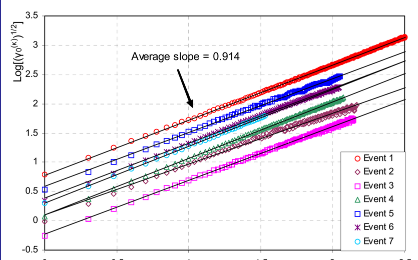
4. Stochastic approach versus deterministic approach

- Rainfall has been traditionally regarded as a random process with several peculiarities, mostly related to intermittency and non Gaussian behaviour. However, many have been not satisfied with the idea of a pure probabilistic or stochastic description of rainfall and favoured a deterministic modeling option (e.g. Eagleson, 1970, p. 184).
- A deterministic perception of the rainfall process may seem in accord to the high temporal dependence (autocorrelation) of the rainfall process at small lag times. However, if one focuses on the change of rainfall intensity in time, the justification of a deterministic view weakens and simultaneously, that of a stochastic approach is strengthened.
- The great variability in statistics within individual rainfall events of the Iowa data set fortifies further a stochastic perception.
- More recently, developments of nonlinear dynamical systems and chaos allowed many to apply algorithms from these disciplines in rainfall and claim for having discovered low dimensional deterministic dynamics in rainfall (see reviews in Sivakumar, 2000, 2004). However, such results have been disputed by others (e.g. Schertzer et al., 2002; Koutsoyiannis, 2006).

5. Scaling in state and time

Logarithmic plots of rainfall intensity quantile (x) vs. empirically estimated (by the Weibull formula) of the exceedence probability ($F^*(x) := 1 – F(x)$). $F(x)$ is the distribution function) for the seven events. In five of the seven events (1 to 5) the variation σ/μ is higher than 1, which implies a power type tail. The upper and lower envelope of the probability plots of the events (mostly represented by events 1 and 3, respectively) seem to justify a long distribution tail, which in a double logarithmic plot is depicted as a constant nonzero slope κ of the empirical distribution (or an asymptotic relationship of the form $x \sim [1/F^*(x)]^\kappa$ for large x). Both envelops indicate a κ of around 0.30.





Logarithmic plots of standard deviation $(\gamma_0^{(\kappa)})^{1/2}$ of rainfall intensity vs. time scale k. The scaling behaviour in time, or the Hurst behaviour, is manifested as a straight line arrangement of points corresponding to different time scales. The slopes of individual events ranges form 0.79 to 0.97, while the average slope of all events is as high as 0.914, a value that verifies the time scaling behaviour.

6. 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)

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

- The principle of ME, due to E.T. Jaynes (1957a, b), states that a random variable X with unknown density function $f(x)$, has a density function $f(x)$ so as to 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 and the non-negativity constraint (mandatory in most hydrometeorological variables including rainfall) results in

$$f(x) = \exp(-\lambda_0 - \lambda_1 x - \lambda_2 x^2), \quad x \geq 0 \qquad (1)$$

where λ_0, λ_1 and λ_2 are parameters depending on the known mean and variance. Inspection of (1) shows that it is none other than the truncated normal density function.

7. The Tsallis entropy and the marginal distribution

- A generalization of the Boltzmann-Gibbs-Shannon entropy, effectively used in numerous scientific disciplines and also valuable in hydrology has been proposed by Tsallis (1998, 2004):

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

with $q = 1$ corresponding to the Boltzmann-Gibbs-Shannon entropy.
- Koutsoyiannis (2005a) reports that the truncated normal distribution fails to describe cases in which the variation $\sigma/\mu > 1$. To find a ME solution for such cases one should abandon standard entropy and use Tsallis entropy. Maximization of Tsallis entropy with known μ and σ^2 yields an over-exponential (power-type) distribution,

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

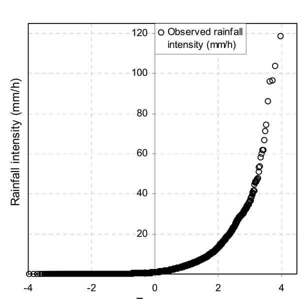
where λ_0, λ_1, λ_2 and ξ are parameters. It can be shown that (2) is mathematically equivalent to the so called Tsallis distribution (Tsallis, 1995; Prato and Tsallis, 1999).
- The fact that high variation σ/μ is common in hydrological variables at fine time scales is a strong indication of the applicability of Tsallis ME principle in hydrology. The most essential difference of (2) with respect to (1) is the implied long (over-exponential) tail of distribution.

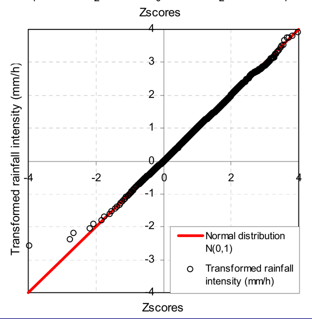
8. The transformation

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

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

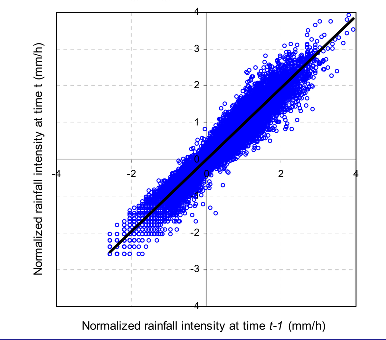
effectively transforms the observed data to normal, as the figures on the right attest, while it is in consistency to the Tsallis distribution. Namely, it yields a hyper-exponential tail.
- The parameters of the transformation were estimated by minimizing the square error (SE) of the model and empirical distribution function.
- The inverse transformation $X = \phi^{-1}(Z)$, essential for de-normalizing synthetic normal series, has been approached numerically, due to the lack of analytical solution.

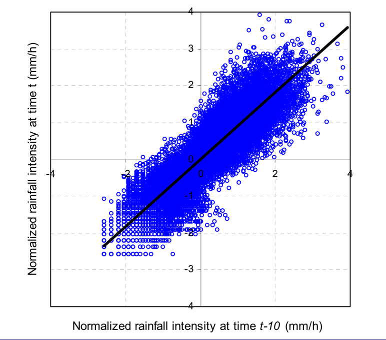




9. The principle of maximum entropy and the linearity of multivariate distribution

The maximum entropy principle, according to which the uncertainty (standing as an interpretation of entropy) is as high as possible, implies linear relationships in consecutive items of a stochastic process. More specifically, provided that a specific transformation of a process has normal marginal distribution, application of the maximum entropy principle results that the multivariate distribution of any number of variables of this transformed process will be multivariate normal (Papoulis, 1991). Besides, it is well known that multivariate normal distribution entails linear relationships among variables. The following figures of the normalized rainfall intensity, affirm this reasoning.



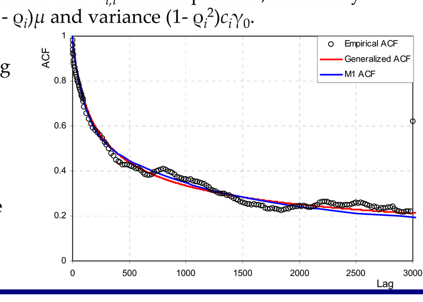


10. The principle of maximum entropy and the long autocorrelation tails

- 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)
- The long autocorrelation tails behavior is a result of the principle of maximum entropy
- The omnipresence of long autocorrelation tails in numerous long hydrologic time series, validates the applicability of the ME principle
- For details see Koutsoyiannis (2005b)

11. The fitted models: Long range dependence model

- The basic characteristic of a long range dependence model is that its autocovariance is a power function of lag. In this study, the empirical evidence favoured the adoption of a generalized autocovariance structure (GAS) $\gamma_j = \gamma_0\,(1 + \kappa\beta j)^{-1/\beta}$ where κ and β are constants with $\beta > 0$ (see Koutsoyiannis, 2000). Clearly GAS is a power function of lag.
- A simple approximation of GAS can be attained by the sum of four independent AR(1) processes (for details of an approach with three AR(1) see Koutsoyiannis, 2002). The autocovariance function (ACF) of the approximate model (referred to as M1) for lag j is $\gamma_{M1,j} = \sum_{i=1}^{4} c_i \rho_i^j$ where ρ_i is the lag one autocorrelation coefficient of the ith AR(1) process and c_i constants satisfying $\sum_{i=1}^{4} c_i = 1$
- The stochastic process W_t that represents the model is $W_t = \sum_{i=1}^{4} X_{i,t}$ where $X_{i,t} = \varrho_i X_{i,t-1} + V_{i,t}$ is the ith AR(1) processes and $V_{i,t}$ are independent, identically distributed, random variables with mean $(1-\varrho_i)\mu$ and variance $(1-\varrho_i^2)c_i\gamma_0$.
- Firstly, the GAS was fitted to the empirical ACF of the transformed data, by minimizing the square error (SE).
- The transformed data ~N(0,1), so $\mu = 0$ and $\gamma_0 = 1$.
- The parameters of the model M1, c_i and ϱ_i, were evaluated by minimizing the SE of the $\gamma_{M1,j}$ and the fitted GAS. The figure on the right attests the satisfactory approximation of the empirical ACF.



12. The fitted models: Short range dependence model

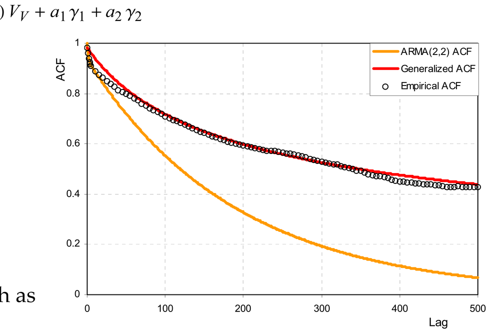
- From a theoretical point of view autoregressive-moving-average models ARMA(p,q), are short range dependence models. For lag > q, the autocovariance (autocorrelation) function reduces to that of an autoregressive model AR(p). An ARMA(2,2) takes the form

$$X_t = b_2\,V_{t-2} + b_1\,V_{t-1} + V_t + a_2\,X_{t-2} + a_1\,X_{t-1}$$

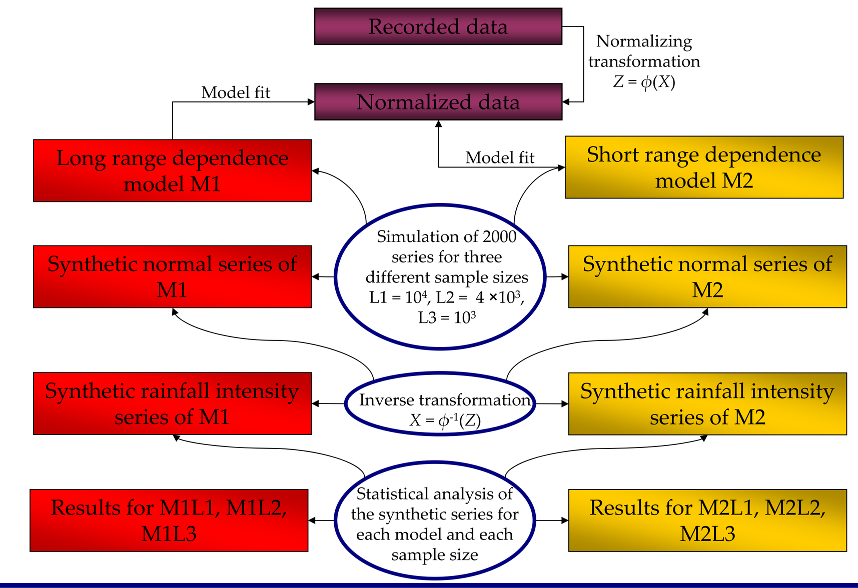
where α_i, b_i, are the model coefficients and V_t are independent, identically distributed random variables with mean μ_V and variance V_V. The unknown model parameter were evaluated by solving the non-linear system below, provided that the mean μ_X and variance V_X of the process is 0 and 1 respectively (normalized data).

$$\gamma_0 = (b_1\,(a_1 + b_1) + b_2\,(a_1^2 + b_1\,a_1 + a_2 + b_2) + 1)\,V_V + a_1\,\gamma_1 + a_2\,\gamma_2$$
$$\gamma_1 = (b_1 + (a_1 + b_1)\,b_2)\,V_V + a_1\,\gamma_0 + a_2\,\gamma_1$$
$$\gamma_2 = b_2\,V_V + a_2\,\gamma_0 + a_1\,\gamma_1$$
$$\gamma_3 = a_2\,\gamma_1 + a_1\,\gamma_2$$
$$\gamma_4 = a_2\,\gamma_2 + a_1\,\gamma_3$$
$$\mu_V = \frac{(-a_1 - a_2 + 1)\,\mu_X}{b_1 + b_2 + 1}$$

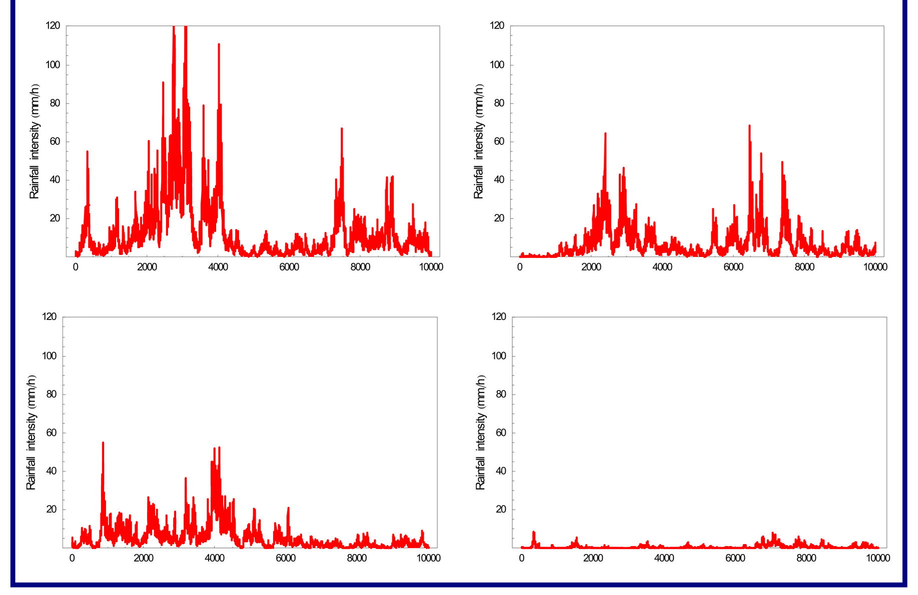
- Despite the fact that ARMA(2,2) is a short range dependence model, the figure on the right attests, that it is capable of keeping positive autocorrelation values for lag as high as 500.



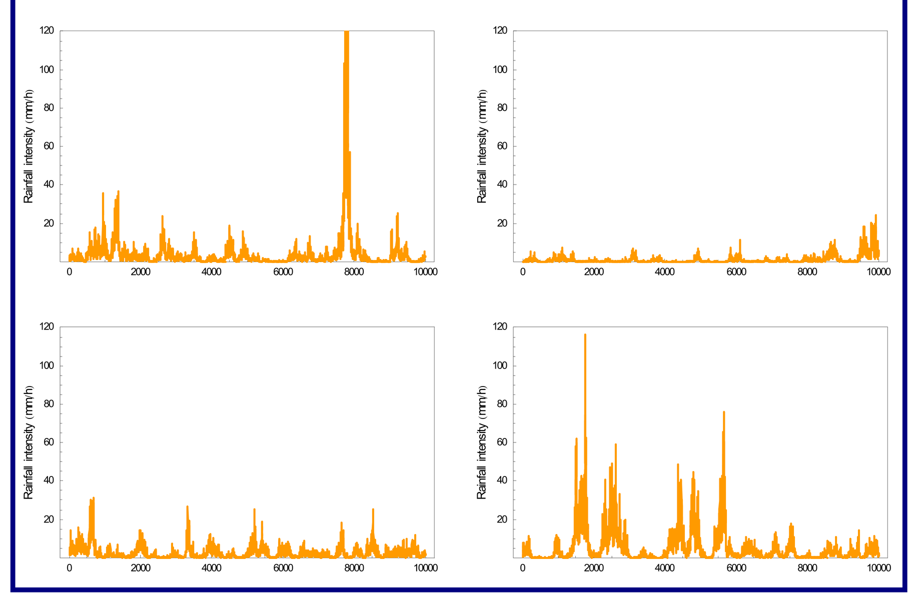
13. Simulation procedure



14. Synthetic series of the long range dependence model

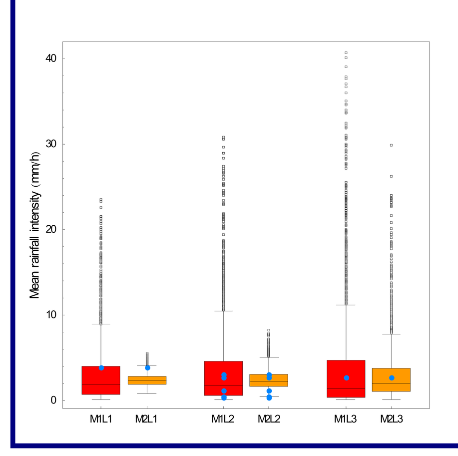


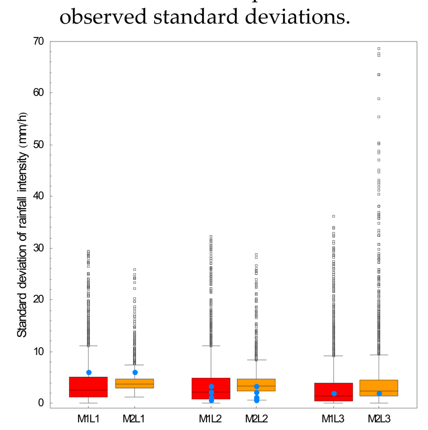
15. Synthetic series of the ARMA(2,2) model



16. Simulation results: Mean and Standard Deviation

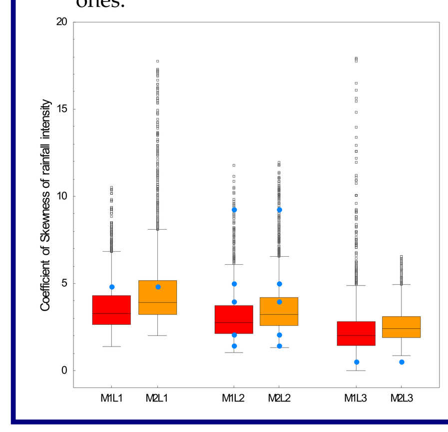
- Box plots of the estimated mean rainfall intensity of the synthetic series produced by models M1 and M2 for sample sizes L1, L2 and L3. Blue dots represent the observed means.
- Box plots of the estimated standard deviation of the rainfall intensity of the synthetic series produced by models M1 and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

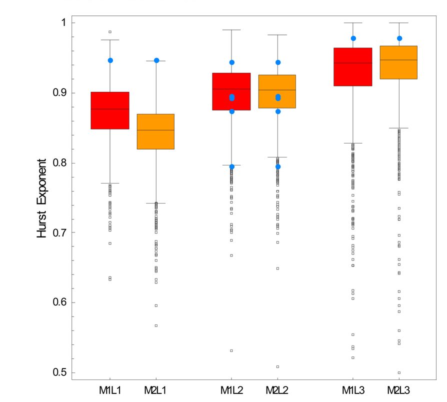




17. Simulation results: Coefficient of skewness and hurst exponents

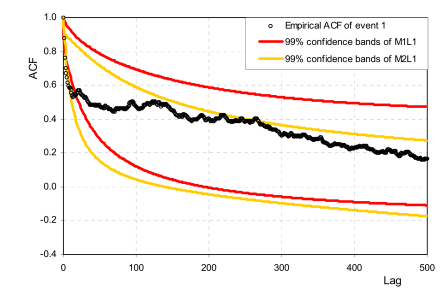
- Box plots of the estimated coefficient of skewness of rainfall intensity of the synthetic series produced by models M1 and M2 and for sample sizes L1, L2 and L3. Blue dots represent the observed ones.
- Box plots of the estimated Hurst exponents of rainfall intensity of the synthetic series produced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

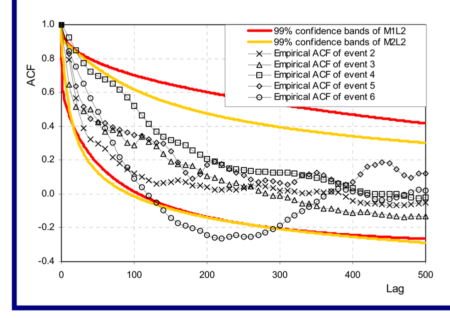


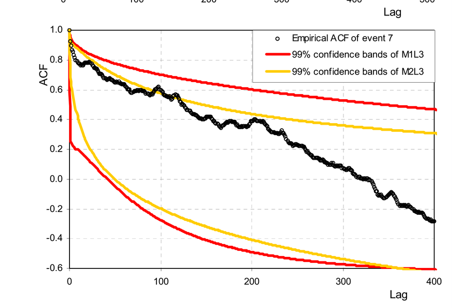


18. Simulation results: Confidence bands of ACF (99%)

- Empirical ACF of rainfall events and 99% Monte Carlo confidence bands of ACF for the two models. The observed events have been classified into three categories: events with sample size approximately equal to L1, L2 and L3 respectively. Sample size length clearly affects the range of the ACF confidence bands.







19. Conclusions

- A single and rather simple stochastic model can represent all rainfall events and all rich patterns appearing in each of the separate events 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.
- The application of the principle of maximum entropy (also using the Tsallis entropy) establishes a solid theoretical basis for power type tails both in marginal distribution and autocorrelation function.
- Whether the tails of both the marginal distribution and autocorrelation functions are power type is difficult to conclude because both these power-law functions are by definition asymptotic properties. In this respect, it seems impossible to verify such asymptotic laws by empirical studies, which necessarily imply finite sample sizes. It is important that the empirical evidence presented in the current study does not falsify the hypotheses that both tails are long while this hypotheses is strengthened by the principle of maximum entropy.
- Both the ARMA(2,2) and the long range dependence model M1 can produce a great variety of rainfall events. However, the comparison between them clearly reveals the superiority of the second one as it is capable of producing a larger variety of rainfall patterns. Moreover, ranges of statistics are wider so as to include all observed events.

20. References

- Carsteanu, A., and E. Foufoula-Georgiou, (1996), Assessing dependence among weights in a multiplicative cascade model of temporal rainfall, *Journal of Geophysical Research-Atmospheres,* 101(D21), 26363-26370.
- Eagleson, P.S., (1970), Dynamic Hydrology, McGraw Hill, New York
- 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.
- Koutsoyiannis, D. (2006), On the quest for chaotic attractors in hydrological processes, *Hydrological Sciences Journal*.
- Kumar, P., and E. Foufoula-Georgiou, (1997), Wavelet analysis for geophysical applications, *Reviews of Geophysics*, 35(4), 385-412.
- 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.
- Schertzer, D., Tchguirinskaia, I., Lovejoy, S., Hubert, P., Bendjoudi, H., & Larchvêque, M. (2002), Which chaos in the rainfall–runoff process?. *Hydrol. Sci. J.* 47(1), 139-148.
- Shannon, C. E. and Weaver, W. (1949), The Mathematical Theory of Communication. The University of Illinois Press, Urbana, Illinois.
- Sivakumar, B. (2000), Chaos theory in hydrology: important issues and interpretations, *J. Hydrol.*, 227, 1-20.
- Sivakumar, B. (2004), Chaos theory in geophysics: past, present and future, *Chaos, Solitons and Fractals*, 19, 441–462.
- 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.