



**HAL**  
open science

## Stochastic model for generating hourly hyetographs

Flavie Cernesson, Jacques Lavabre, Jean Marie Masson

► **To cite this version:**

Flavie Cernesson, Jacques Lavabre, Jean Marie Masson. Stochastic model for generating hourly hyetographs. Atmospheric Research, 1996, 42, pp.149-161. hal-01575729

**HAL Id: hal-01575729**

**<https://hal.science/hal-01575729>**

Submitted on 20 Nov 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Stochastic model for generating hourly hyetographs

Flavie Cemesson \ Jacques Lavabre \ Jean-Marie Masson <sup>b</sup>

<sup>a</sup> CEMAGREF, Le Tholonet, B.P. 31, 13612 Aix-en-Provence, Cedex 1, France

<sup>b</sup> LHM-USTL, Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier, Cedex 5, France

## 1. Introduction

Rainfall data is one of the foundations of hydrology. For some purposes, however, the necessary data is not always available in sufficiently short time steps over a sufficiently long period. With the growth of computer power over the last 30 yr, more or less complex tools for simulating plausible rainfall time series at a given point have appeared. All these models are based on the principle that rainfall can be considered as a random intermittent process described by the laws of probability.

Hyetograph simulation models can have a variety of different structures, depending on the choice of variables used for describing the rainfall episode and the intermittent nature of rainfall. The models developed by Rodriguez-Iturbe et al. (1987), Entekhabi et al. (1989), Istok and Boersma (1989) and Cowperrwait (1991) are based on the Neyman-Scott (Neyman and Scott, 1958) aggregation process. The starting assumption is that storms can be considered as aggregating according to a group hierarchy described by a Poisson's law. While the aggregation process is a useful tool for describing the

frequency of occurrence of a rainfall episode (Waymire and Gupta, 1981), it involves many parameters (although it must be said that these models operate on a continuous basis). Croley et al. (1978), Tourasse (1981) and Acreman (1990) use other definitions of the storm in developing their models. All the variables describing the structure of the rainfall are independent. These models can simulate hourly hyetographs. Models based on disaggregated daily rainfall (Hershenson and Woolhiser, 1987; Econopouly et al., 1990) require fewer parameters and can simulate rainfall episodes. The dimensionless approach (Huff, 1967; Bonta and Rao, 1989) has culminated in a model (Garcia Gazman and Aranda-Oliver, 1993) capable of generating storms. For the moment, the use of fractals (Bocquillon and Moussa, 1992; Hubert, 1992) is a way of analysing rainfall episodes yielding a rainfall typology. This list does not pretend to be exhaustive, it shows the range of models currently under development.

Hourly hyetograph generation models are written to fill the gap in the rainfall time series, mostly obtained on a daily basis. Hourly time series have obvious benefits in rainfall runoff modelling in short time steps, of interest for example in flood control, resource management, monitoring of erosion-critical areas, river control design, etc. Lardet (1992) bases a flood forecasting system on an hourly rainfall generation model. Leviandier (1992) uses simple rainfall shapes derived from daily rainfall data to determine the maximum runoffs needed for water engineering design. It is with this objective in mind that we are developing our stochastic hourly hyetograph generation model. Used in combination with a rainfall runoff model, it should provide the designer with a range of flood scenarios as input to his performance studies.

We will describe observations made while building the model. It was based on data

from a rainfall recorder on the Mediterranean seaboard of France. The input data, methodological aspects in defining the rainfall episode, and the structure of the model is first described. The second part deals with a few applications, more particularly estimating rare frequency rainfalls. In addition to being one step in assessing model performance, the study of maximum rainfalls derived from simulated hyetographs is also the first application of the model.

## **2. Data and methods**

### *2.1. Rainfall records*

The Mediterranean climate displays extreme aerial and temporal variability, and various sets of records from the French Mediterranean seaboard have been analysed to arrive at a single architecture for the model. Series from Néoulous (Pyrénées Orientales), Marignane (Bouches du Rhône) and Croix d'Anselme (Var) rainfall recorders are referred to in this article. The features of these stations are listed in Table 1. All the recorders are Précis Mécanique tipping bucket rain gauges. The Néoulous apparatus is heated. Data is recorded by strip chart and analysed by the "variable-time-step" method.

The stations are about 300 km apart at very different altitudes. Rainfall in the Mediterranean area is strongly influenced by orographic effects and this selection of

Table 1

Characteristics of the stations. The position is given in Lambert III (station 56) or in extensive Lambert II (stations 01 and 18)

Station	Code	Owner	Position	Duration
Néoulous	01	Direction Départementale de l'Agriculture — 66	$x=649.4km$ $y=120km$ $z=1100m$	1965–1990 26 yr
Marignane	18	Météo France	$X=834.0 km$ $y=130.5 km$ $z=4m$	1955–1979 25 yr
La Croix d'Anselme	56	CEMAGREF	$X=926.22 km$ $y=112.4 km$ $z=355 m$	1966–1988 23 yr

three stations enables us to describe three variants of the French Mediterranean climate. Marignane, practically at sea level, has the least rainfall, the annual mean being around 500 mm. The other two stations at Néoulous and Croix d'Anselme have similar mean annual rainfalls of around 1000 mm.

With respect to the aim of the method, we have decided to select only rainfall episodes producing significative runoffs. Rainfall episodes are selected when they contain at least one daily rainfall in excess of 20 mm. The great variability of rainfall in the course of the year imposes a seasonal breakdown. Two seasons, winter from December to May, and summer from June to November are detected (Cemesson et al., 1995). There are roughly as many intense rainfall episodes in summer as in winter at Néoulous (105 episodes in summer, 94 in winter) and at Marignagne (76 episodes in summer, 84 in winter) whereas Croix d'Anselme has a rainier winter than summer (99 episodes in summer, 159 in winter). The selected episodes are quite numerous to make the statistical analysis correctly.

## 2.2. Mode!

It was decided to describe rainfall episodes by the Croley et al. (1978) methods as used by Tourasse (1981), with modifications, although the general principle remains the same.

A rainfall episode is an intermittent event consisting of a series of dry periods and storms. A storm has as many peaks as there are relative maxima. Groups of peaks are sometimes observed (i.e. no dry periods between peaks), as well as isolated peaks. Where there are two peaks, one immediately after the other, we defined a relative minimum which is conventionally assigned to the first peak.

A storm is completely described by its number of peaks, the shape of each peak, and the amount of rainfall it yields. There are seven variables describing the discontinuous nature of the process and the structure of the peak. First, a rainfall episode contains NG storms. Each storm contains NA peaks.

A peak is characterised by:

- its duration DA (in h);
- its mean intensity HMA (in units of 0.1 mm/h);

- the relative position RPX of the crest of the peak curve, i.e. the ratio between the time to maximum intensity and the total duration of the peak;
- maximum intensity ratio RX between maximum and mean intensity.

A storm-peak typology has been defined for a better description of rainfall intensity and duration. So, we distinguish between the principal peak delivering the greatest depth of rainfall in the episode and the other peaks. Each episode consists of a single principal peak plus, where applicable, secondary peaks.

The seventh variable is the non-zero dry period between successive storms DIA (in h). There are two subsidiary variables, number of episodes per year and per season NE, and time of onset of storm, TSE.

Statistical analysis of the distributions of the variables describing the rainfall episodes requires selecting an appropriate probability function, shown in Table 2 (Cemesson et al., 1995). Calibrating the model parameters involves determining them by the method of moments or by numerical means. Hourly hyetographs are generated on the basis of a random draw of each variable from its probability distribution and breaking down the variables into hourly rainfalls, according to the distribution rules.

The model is essentially based on a proper reconstruction of the mean intensity and duration of the storm peak, because these are the two variables "delivering" the water. For optimum description of the distributions of these variables, we must use composite distributions, having three or four parameters. Laws describing DIA, NG and NA are less important but the episode structure is based on them. Because of the strong

Table 2  
Theoretical functions describing each variable distribution

Variable	Functions	
NE	Poisson law	$P(x \leq k) = \sum_{i=0}^k \exp(-p) p^i / i!$
TSE	Geometric fonction TSE+!	$E(x) = p / \text{Var}(x) = p$ $p(x \leq k) = 1 - (1-p)^k$ $E(x) = 1/p$ $\text{Var}(x) = (1-p)/p^2$
NG	Geometric fonction	cf(TSE)
NA	Geometric fonction	cf(TSE)
DIA	DIA < 13 h Poisson law 13 ≤ DIA ≤ 24 h, uniform law	cf(NE)
RX	Exponential law	$F(x) = 1 - \exp[-(x - p_2/p_1)]$ $E(x) = p_1 - p_2$ , $s = p_1$
RPX	Normal law	( $m = p_1$ , $s = p_2$ )
HMA	Summer	si $x \leq x_0$ , $F(x) = 1 - \exp[-((x - x_0)/a_1 + y_0)]$
	Two combinations of two exponential functions separating at the pivot point	si $x > x_0$ , $F(x) = 1 + \exp[-((x - x_0)/a_2 + y_0)]$
	Winter	
	2 Weibull fonctions	$F(x) = 1 - \exp[-((x - 1)/a)^b]$
DA	Poisson law	cf(NE)
	Uniform law	



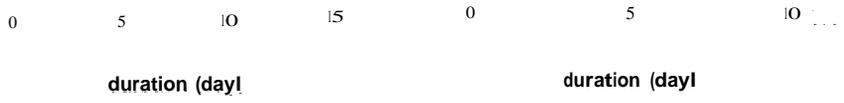


Fig. 1. Station 56 — comparison of the distribution for simulated values (continuous line) and observed values (dash). Studied variables are the rainfall depth (mm) and the total duration (h) for summer and winter.

The distributions of variables "rainfall depth" and "total duration of episodes", calculated from the observed and simulated series, were compared (Fig. 1), and the model correctly reproduced the mean and standard deviation of each distribution.

The model-generated hourly hyetographs are very similar to observed hyetographs. Fig. 2 compares a simulated and observed hyetograph of equivalent total rainfall, total duration and maximum hourly rainfall. The regular shape of the simulated curve arises from the triangular or trapezoidal distributions used.

### 3. Maximum rainfall study

There are two reasons for studying maximum rainfalls for given durations. One is to find the model efficiency criterion. Maximum rainfall frequency analysis is also an imperfect but immediate use for simulated hyetographs.

#### 3.1. Maximum rainfalls as efficiency criteria

Performance assessment and validation of a stochastic-type model is based on statistical analysis of variables not used explicitly in generating hyetographs. Results can be refined by comparing the simulated and observed 1-, 2-, 4-, 6-, 12- and 24-h maximum rainfall distributions. These variables, calculated for durations shorter than the

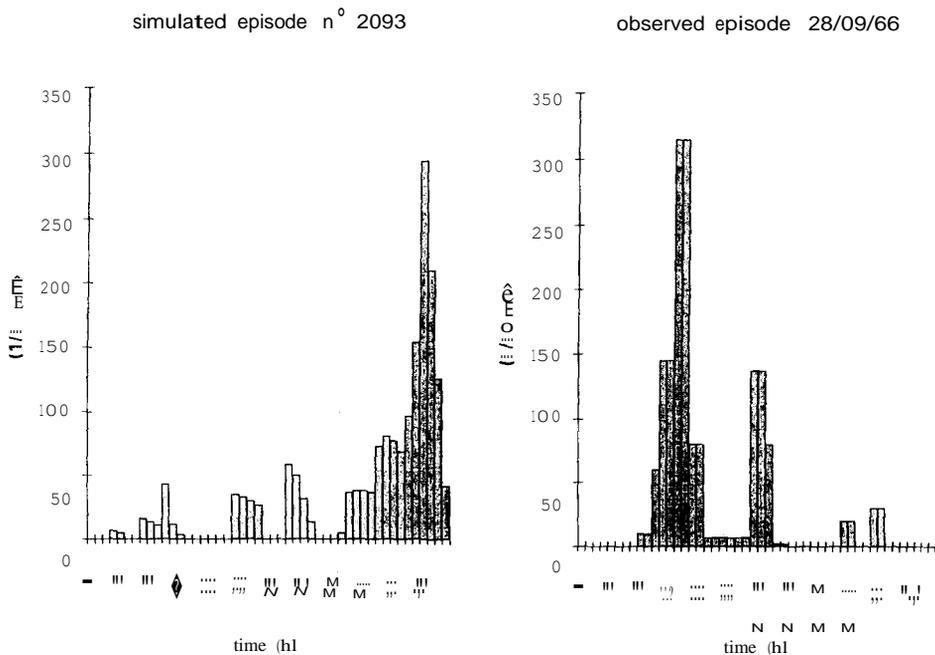


Fig. 2. Example of observed episode and simulated episode: — observed episode: total amount = 166 mm, total duration = 48 h, hourly maximal rainfall = 31.5 mm, — simulated episode: total amount = 177 mm, total duration = 48 h, hourly maximal rainfall = 29.4 mm.

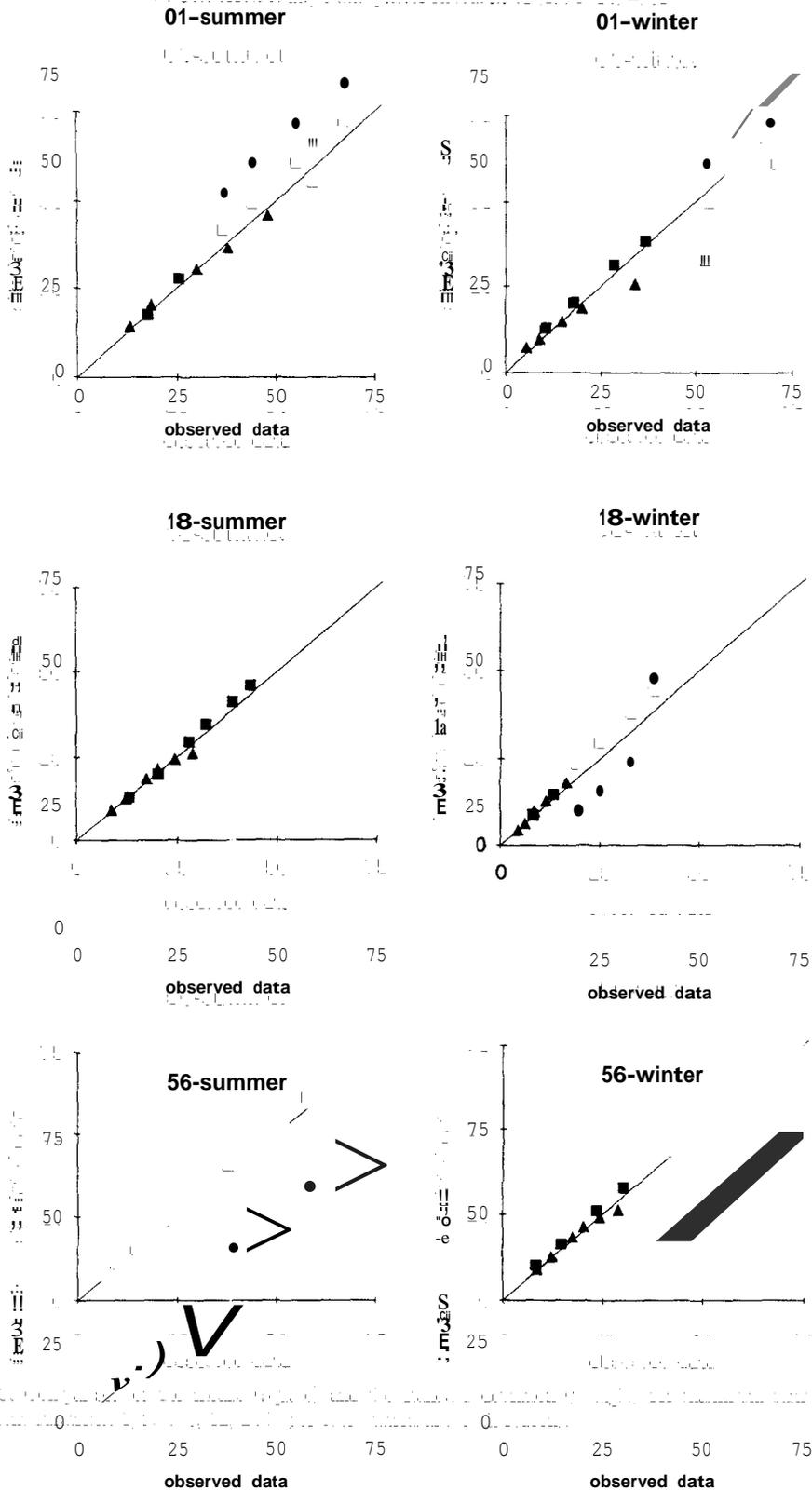


Fig. 3. Comparison of the means (square) and the standard deviation (triangle) for maximum rainfall of different durations: 1, 2, 4, 6, 12, 24 h (for each station and each season).

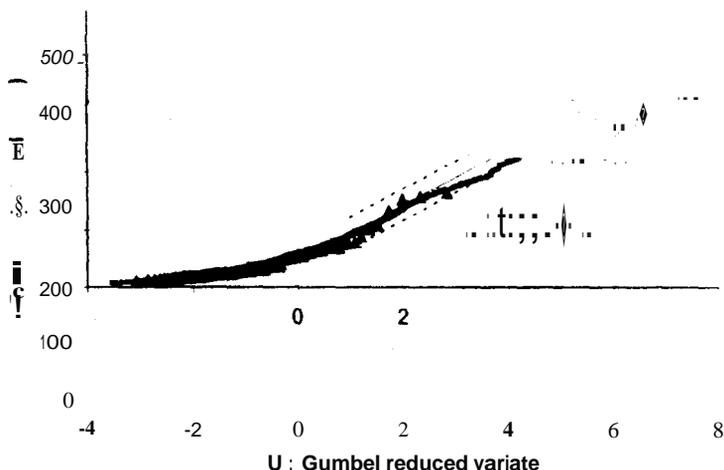


Fig. 4. Station 01 — summer. Simulated and observed distributions of 6-h maximum rainfall. Triangles and dashes represent respectively observed and simulated data.

total mean duration (approximately 2 days regardless of recording station and season) throw light on the internal structure of the episodes and validate the model's capacity for reproducing realistic hyetographs.

The model was calibrated for each station and each season. Test sets are 1000-yr simulations for all three stations, Néoulous (referenced 01), Marignane (referenced 18) and Croix d'Anselme (referenced 56) and for both seasons. Mean and standard deviation are calculated for each of the distributions of maximum rainfall of different durations: 1, 2, 4, 6, 12 and 24 h (Fig. 3).

The means of the simulated series are generally of the same order of magnitude as the means of the observed series. This is also true for standard deviations. Looking in more detail, the summer simulations are satisfactory for the three stations although the standard deviation is systematically overestimated for Croix d'Anselme. Differences are greater in winter; at Néoulous, maximum 12- and 24-h rainfall is distinctly underestimated. Means and standard deviations are overestimated at Marignane for all durations.

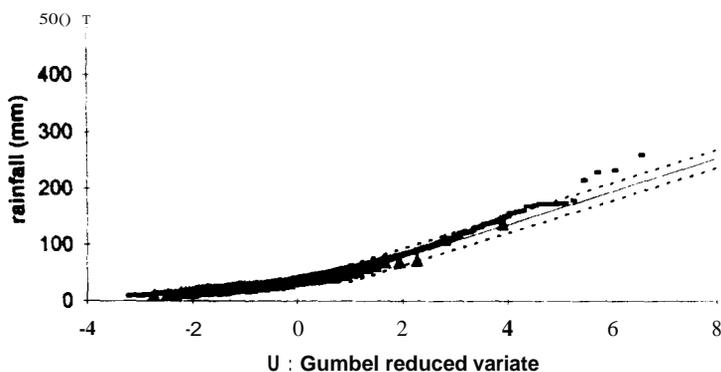


Fig. 5. Station 18 — summer. Simulated and observed distributions of 6-h maximum rainfall. Triangles and dashes represent respectively observed and simulated data.

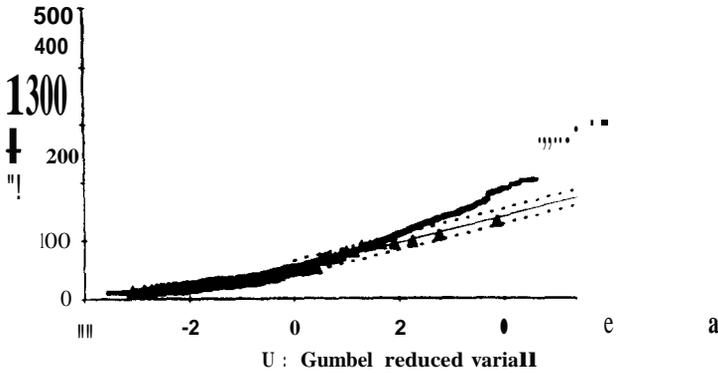


Fig. 6. Station 56 — summer. Simulated and observed distributions of 6-h maximum rainfall. Triangles and dashes represent respectively observed and simulated data.

At Croix d'Anselme, means and standard deviations are correctly reproduced except for the 1-h event.

### 3.2. Maximum rainfalls as first applications of the model

The rainfall model is intended to be coupled with a rainfall runoff model for flood estimating purposes (Cemesson and Lavabre, 1994), i.e. estimating the design flood used as input for the design of dams and other river works. The method is a statistical approach which must accommodate the data available at or near the development site.

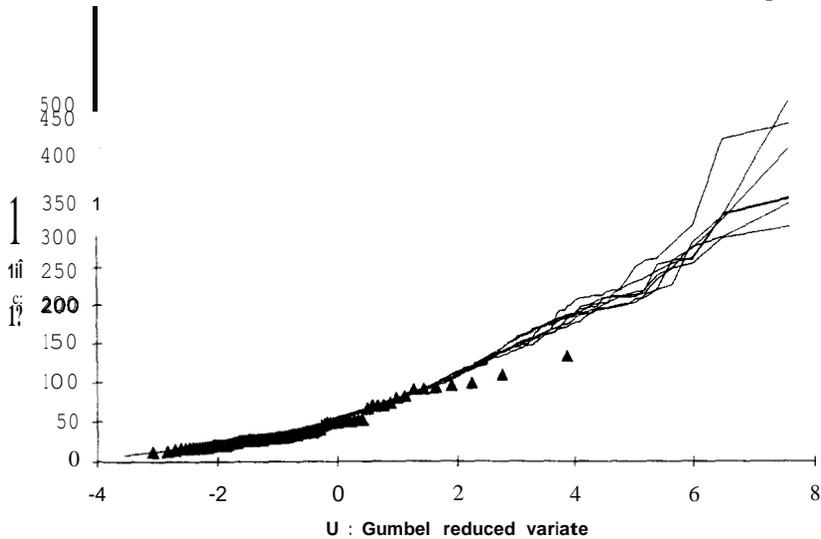


Fig. 7. Station 56 — summer distributions of 6-h maximum rainfall. The seven simulated series are represented by a continuous line, each break indicates a simulated value. Observed series is represented by triangles.

Some rainfall-based design flood estimation methods have the advantage of using data measured at shorter time steps than streamflow data. Design flood estimation methods are based on fitting a theoretical distribution to maximum rainfall series for a given duration. The risk tolerated (expressed as return period  $T$ ) is related to the frequency of non-exceedance of some maximum rainfall by the equation

$$T = 1 / (1 - F)$$

in which  $F$  is the frequency of non-exceedance and  $T$  is the return period in years.

It will be easily understood that if the analytical expression for  $F$  as a function of rainfall  $P$  is known, we can find the rainfall  $P$  for a given return period.

In France, maximum rainfall distributions are usually described by a Gumbel distribution:

$$F(P) = \exp - [\exp - (P/P_0)]$$

Taking  $u = -\ln(-\ln F(P))$ , we can linearise the expression to:

$$P = a'' u + P_0$$

Parameter  $a$  is called Gradex i.e. gradient of extreme values (Duband and Guillot, 1968; Duband et al., 1988) and  $P_0$  is the position parameter.

Fig. 4, Figs. 5 and 6 show observed and simulated 6-h maximum rainfall distributions at three stations. The empirical frequency is calculated with Hazen's formula:

$$F(P) = \frac{r - 0.5}{n}$$

in which  $r$  is the rank of rainfall  $P$  ranked in increasing order, and  $n$  is the total sample size.

Figs. 4–6 also show the Gumbel distribution fitted to the observed maximum rainfall sample. The observed and simulated series for the three stations give just about the same values of 6-h maximum rainfall for Gumbel standardised variables  $u$  less than about 2 (which represents common frequencies of occurrence with a return period less than 10 yr). The clusters of points overlap on the graphs. Beyond positive  $u$  values, the simulated and observed plots bend slightly. The simulated curves tend to give higher maximum rainfall values for a given value of  $u$  greater than 2.

In the rare frequency area, with high  $u$  values, the three series behave very differently. The simulated series for station 01 (Néoulous) keep their linear shape. It is the same for station 18 (Marignane) if the five highest simulated rainfall values are ignored. The record for simulated rainfall was 460 mm in 6 h. Although high, it is a plausible value considering the extreme episodes actually occurring in the region.

The differences in extreme values can be explained by the random generation of variables. Fig. 7 shows each of the distributions for the seven 1000-yr simulations and the distribution of maximum rainfall recorded at station 56 (Croix Anselme). While the curves remain close together up to  $u$  values of around 6, the three highest generated values sometimes vary very considerably. It is worth noting the highly random nature of this model, which simulates 1000-yr samples that are all different. This feature can be used in terms of confidence interval in estimating extreme rainfall values.

Table 3

6-h maximum rainfall (mm) for 10-yr, 50-yr, 100-yr, 1000-yr return periods. The codes (a), (b), (c), indicate, respectively, the minimum, median and maximum values, calculated from 10 simulations of 1000 years

Return period (yr)	Néoulous		Marignane		La Croix d'Anselme	
	observed	simulated	observed	simulated	observed	simulated
10	159	155	86	92	104	125
50	236	218	133	150	140	187
100	269	239 (a)	154	164 (a)	155	195 (a)
		258 (b)		173 (b)		208 (b)
		286 (c)		202 (c)		218 (c)
1000	379	337 (a)	221	232 (a)	207	273 (a)
		388 (b)		288 (b)		316 (b)
		487 (c)		346 (c)		418 (c)

A Gumbel distribution is fitted to each observed series by linear regression. This results in the following equations:

$$\text{Station 01: } P = 47.9u + 49$$

$$\text{Station 18: } P = 29.4u + 18.5$$

$$\text{Station 56: } P = 22.3u + 52.7$$

in which  $P$  is the 6-h maximum rainfall (mm) and  $u$  is the Gumbel standardised variable.

With these equations, we can extrapolate maximum rainfall values for long return periods. The same method can be used for each simulated series but the generated values can also be used directly. It was decided to present the values for long return periods (100 and 1000 yr) as the minimum, maximum and median values from ten 1000-yr simulations. This presentation reflects the variability of the series simulated for long return periods. Maximum rainfalls with return periods of 10, 50, 100 and 1000 yr are listed in Table 3.

Quantiles estimated from the simulated series are generally higher than those calculated from observed data. The degree of overestimation is very obvious for the Croix d'Anselme station and slightly less for Marignane. For Néoulous, the simulations give the same results as the observations.

#### 4. Conclusion

The model has been tested against three rainfall stations on the Mediterranean seaboard: Croix d'Anselme, Marignane and Néoulous, in differing topographic and climatological environments.

The model properly reproduces the means and standard deviations of 1-, 2-, ..., 24-h maximum rainfalls. Maximum rainfalls are not used in the model, which ensures good time restitution of hourly rainfalls. In its present architecture, the model appears to be transposable to other rainfall stations on the Mediterranean seaboard. Study of maximum rainfalls is one of the applications of the model. Regarding extreme values of maximum

rainfall, the model yields more conservative results than the conventional extrapolation method except at Néoulous. Does this discrepancy between Néoulous and the other two stations reveal a sampling problem (subject to data shortages) or specific localised differences in climate?

Estimating rare episodes from the observed series is somewhat delicate since it is dependent on the sample selected, because of the standard deviation changing, even slightly. The model can generate any desired number of episodes for a given number of years, the only limit being computer time for generating rainfall. Directly adjusting the simulated sample for fit avoids extrapolation problems. We still have to determine the duration to obtain some stability for a given return period. A 1000-yr series appears suitable for estimating the 100-yr rainfall.

Estimates of 6-h rainfall of long return periods at Croix d'Anselme are higher than those obtained by simple extrapolation of the probability function assumed to represent the observed sample. This may be due to using a strictly exponential function and different results might have been obtained with more strongly skewed distributions. Apart from this, there is room for discussion as to the applicability of this type of model for investigating extreme rainfalls. Intuitively, the model has the advantage of working with all variables describing the internal variability of rainfall episodes and may give a better statistical description of the process. In addition to the choice of several variables, the model's intrinsic probability functions refer to large samples, which guarantees some robustness in the choice of functions and their suitability.

Despite the successive random draws which might lead one to expect a tendency towards mean values, the model remains extremely variable and does not stabilise over long successive simulations. This property can in fact be used to determine a confidence interval on quantiles of different frequencies.

## Acknowledgements

We thank Gisèle Andre, Eric Ribero and Marie-France Vuillet for their contributions to develop the model and Claude Michel for his advice.

## References

- Acreman, M.C., 1990. A simple stochastic model of hourly rainfall for Farnborough. *Hydrol. Sci., J. Sci. Hydrol.*, 35(2): 119–148.
- Bocquillon, C. and Moussa, R., 1992. Caractérisation fractale d'une série chronologique d'intensité de pluie. *Rencontres hydrologiques Franco-Romaines, Paris 2–5 septembre 1991, UNESCO Paris*, pp. 363–370.
- Bonta, J.V. and Rao, A.R., 1989. Regionalization of storm hyetographs. *Water Resour. Bull.*, 25(1): 211–217.
- Cernesson, F. and Lavabre, J., 1994. Modèle Shypre: Simulation de crues grâce à la modélisation pluie-débit. *Congrès de la SHF, 23èmes journées de l'Hydraulique "Crues et inondations"*, Nîmes, 14–16 September 1994, pp. 229–236.
- Cernesson, F., Lavabre, J. and Masson, J.M., 1995. Modèle de génération stochastique de hyétoigrammes horaires. I. Construction du Modèle, II. Applications — CEMAGREF, Aix-en-Provence, 35 pp.
- Cowpertwait, P.S.P., 1991. Further developments of the Neyman Scott clustered point process for modeling rainfall. *Water Resour. Res.*, 27(7): 1431–1438.

- Croley, T. et al., 1978. Ralston Creek hourly precipitation model. *Water Resour. Res.*, 14(3): 485–490.
- Duband, D. and Guillot, P., 1968. La méthode du Gradex pour le calcul de la probabilité des crues à partir des pluies. S.H.F., Xèmes journées de l'Hydraulique, Paris, question 1, rapport 7, 7 pp.
- Duband, D., Michel, C., Garros, H. and Astier, J., 1988. Evaluation des crues extrêmes et de la crue de projet par la méthode du Gradex. Commission Internationale des Grands Barrages. XYième congrès des Grands Barrages, San Francisco, Q63–R60: 1009–1047.
- Econopouly, T.W., Davis, D.R. and Woolhiser, D.A., 1990. Parameter transferability for a daily rainfall disaggregation model. *J. Hydrol.*, 118: 209–228.
- Entekhabi, D., Rodriguez-Iturbe, I. and Eagleson, P., 1989. Probabilistic representation of the temporal rainfall process by a modified Neyman–Scott rectangular pulses model: parameter estimation and validation. *Water Resour. Res.*, 25(2): 295–302.
- Garcia Gazman, A. and Aranda-Oliver, F., 1993. A stochastic model of dimensionless hyetograph. *Water Resour. Res.*, 29(7): 2363–2370.
- Hershenson, J. and Woolhiser, D.A., 1987. Disaggregation of daily rainfall. *J. Hydrol.*, 95: 299–322.
- Hubert, P., 1992. Analyse multifractale des champs temporels d'intensité des précipitations. Rencontres hydrologiques Franco-Romaines, Paris, 2–5 septembre 1991, UNESCO, Paris, pp. 379–386.
- Huff, F.A., 1967. Time distribution of rainfall in heavy storms. *Water Resour. Res.*, 3(4): 1007–1019.
- Istok, J.D. and Boersma, L., 1989. A stochastic cluster model for hourly precipitation data. *J. Hydrol.* 106: 257–285.
- Lardet, P., 1992. Prédiction des crues: contribution à l'utilisation opérationnelle des modèles pluie-débit. Thèse de doctorat de l'Institut National Polytechnique de Grenoble, 205 pp.+ annexes.
- Leviandier, T., 1992. Méthodologie d'estimation des crues de projet en Algérie. CEMAGREF Antony. OMM, 60 pp.
- Neyman, J. and Scott, E., 1958. Statistical approach to problems of cosmology. *J.R. Stat. Soc. Ser. b (methodological)*, 20(1): 1–43.
- Rodriguez-Iturbe, I., Cox, D.R., Isham, F.R.S. and V., 1987. Some models for rainfall based on stochastic point processes. *R. Soc. London, A410*: 269–288.
- Tourasse, P., 1981. Analyses spatiales et temporelles des précipitations et utilisation opérationnelle dans un système de prévision de crues. Application aux régions cévenoles. Thèse IMG, Université scientifique et médicale, et Institut National Polytechnique de Grenoble, Grenoble, 190 pp.
- Waymire, E. and Gupta, V.K., 1981. The mathematical structure of rainfall representations. *Water Resour. Res.*, 17(5): 1261–1294.