ESTIMATION, REGIONALIZATION, AND SPATIAL AVERAGING OF EXTREME RAINFALL AT SMALL TIME-STEPS

by Ch. OBLED, T. LEBEL * and M. SLIMANI **

Institut de Mecanique de Grenoble, Groupe d' Hydrologie
B.P. 6838402 Saint Martin d'Heres Cedex FRANCE

A very practical problem in the design of hydraulic structures is that of extreme value flood analysis. Commonly based on long discharges series, the problem must be approached differently when such data is lacking. In such a case, a possible solution consists in using information derived from a more currently available variable, i. e. rainfall, together with other information, either hydrological or topomorphological.One such example is the Gradex method (Guillot and Duband 1967, Lebel and Guillot 1983), which extrapolates the discharge probability function according to that of an appropriate representative rainfall for the watershed considered. Other widely used methods of estimating given return period floods rely on rainfall statistics such as the 5- or 10-year daily rainfall (e.g. see the NERC Flood studies Report -1975-, or the French Ministry of Agriculture SOCOSE-CRUPEDIX method-1980-).

An argument for selecting rainfall rather than other flood generating factors is that it may be less sensitive to local effects such as land or vegetation cover, and thus better suited to regionalization.

These reasons have led to rising interest in the Climatology of Extreme Rainfall, at first mainly on a daily basis. However, concerns with smaller rural watersheds as well as needs in urban hydrology soon raised interest in shorter time-steps too, with the drawback of much less dense networks of recording raingages. This leads to the problem of regionalizing sparse information, as well as that of the representativeness of point values for spatially extensive watersheds.

* Now with ORSTOM - Montpellier - France
** Now with Institut National Agronomique Tunisien - Tunis - Tunisia

Thus, the following paper tries to address 3 sequentially encountered problems :

1 - estimation of the extreme value distribution of point rainfall for short time-steps.
2 - regionalization of the point statistics.
3 - statistics of spatially averaged rainfall.

An extensive case study was performed on a region subject to violent and sudden rainfall of Mediterranean origin in Central and South East of France (see Obled and Creutin -1982-, or Obled -1983-, for brief meteorological description of these regions).

The region sketched in Figure 1-a is equipped with a network of 47 long-standing recording raingages, with most record lengths in the range 13-22 years, a few exceeding 25 years and one reaching 62 years.

I - EXTREME VALUE DISTRIBUTION FOR POINT VALUES.

I - 1 Avalaible data and choice of a probability law.

Only half the stations were analyzed for all available clock hour time-steps (e.g. 8h 00 - 9h 00), while for the others, only the monthly maximum were selected for time steps of :

1 h, 2 h, 4 h, 6 h, 12 h, 24 h

A drawback of the procedure, coped with through a Weiss correction factor, is that the time period may only begin or end on a rounded clock hour.

The probability model widely accepted for extreme rainfall is the Gumbel model (Extreme Value Type 1 : E.V.1). This may be backed either by extensive sampling (Hershfield and Kohler 1960), or through analytical considerations (Lebel and Guillot 1983). In our case, the choice was mostly empirical.

The initial distribution of monthly maxima (3 per year during the rainy autumn season) displays significant departures from the Gumbel model through a curvature or broken line effect at the bottom tail of the distribution. However, the goodness of fit steadily improves when the maxima are taken over increasingly long periods (e.g. 1 per year or 1 per 2 years - see Figure 2).

I - 2 Estimation procedures .

The Gumbel model for a random variable X is :

$$F(x) = Pr(X < x) = \exp (- \exp (- \frac{x - b}{a}))$$

where the parameters a and b are related to the 1st and 2nd order moments μ and σ^2 of the population through :

$$a = 0.78\,\sigma \quad \text{and} \quad b = \mu - 0.577\,a$$

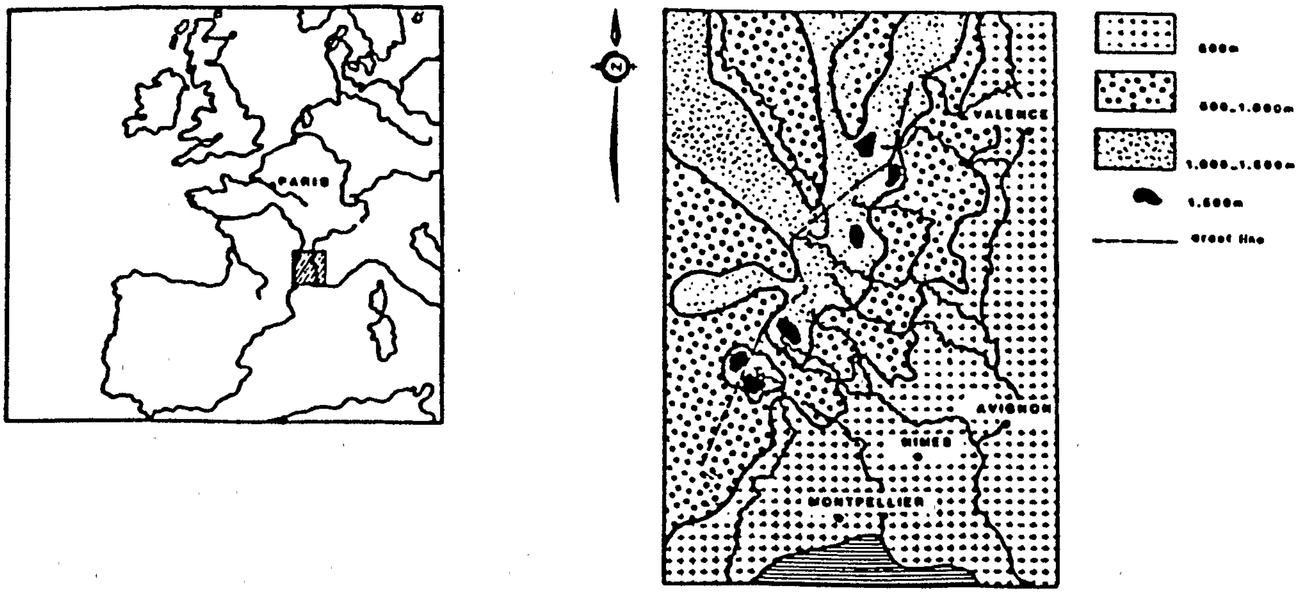
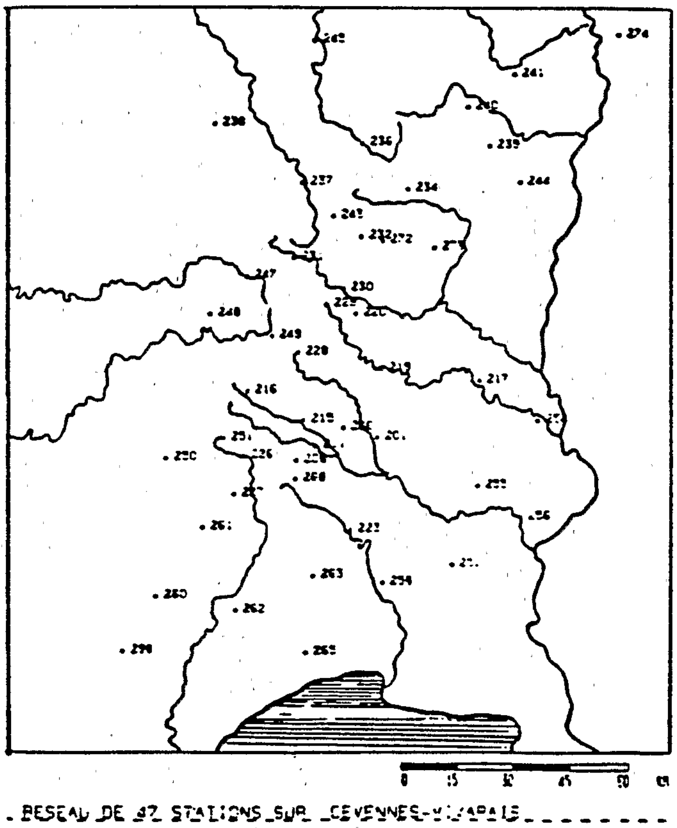


Figure 1 : Study Area

1 - a Location and relief
1 - b Drainage and Recording Raingages Networks



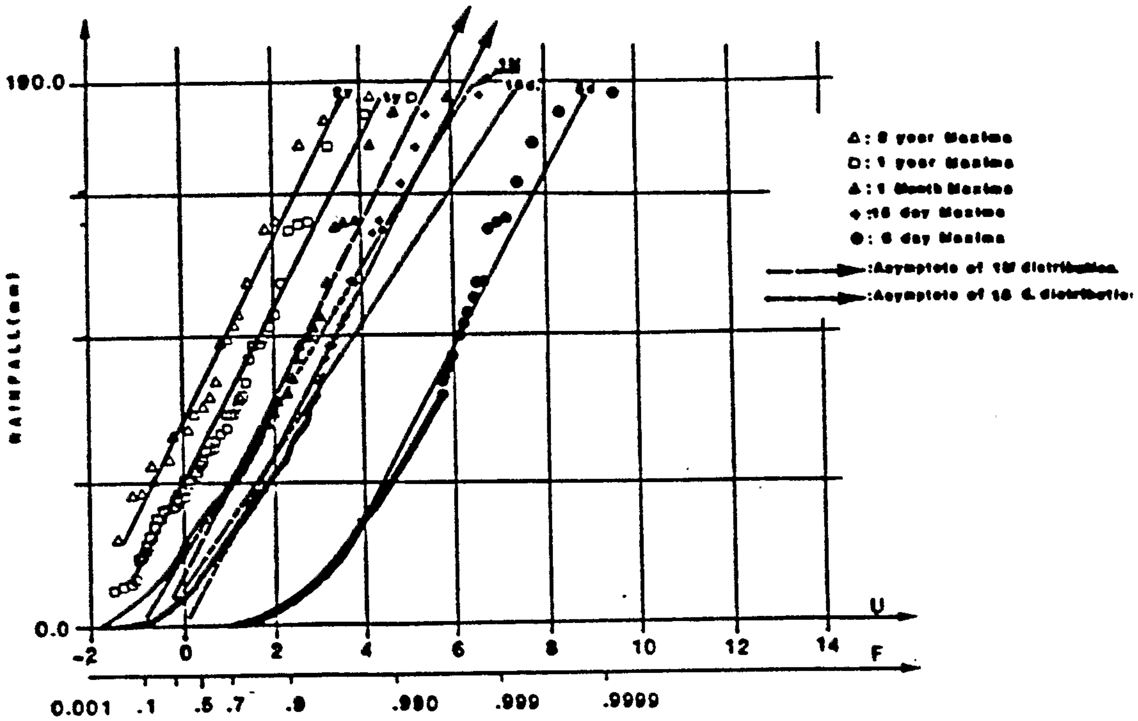


Figure 2 : Empirical Distribution of Fixed Duration Maxima
for six-hour rainfall

The Gumbel scale parameter a is sometimes called the" Gradex" in the french literature (for gradient of the exponential).

The problem of correctly estimating a and b from a finite sample was approached in several ways :

a) through synthetically generating series of presumed extreme rainfalls following exactly a Gumbel distribution.

b) through generating series of continuous hourly rainfalls Y , or 2 h - rainfall etc..., according to the time- step considered. An appropriate model to do this is a mixture of 2 populations with a sum of exponentials as a probability function:

$$F(y) = Pr(Y< y) = 1 - \gamma \exp(- y/a) - \beta \exp (- y/c)$$

Here, the distribution of X, the maxima over fixed length periods is only assymptotically gumbelian.

c) through splitting the 62 year long series of hourly rainfall at Montpellier Bel Air station into subsamples of 20 years each.

For every dataset obtained from a), b) and c), several estimating procedures were applied, namely:

1) Moment Method (MO) , using sample mean and standard deviation, with a correction factor for small sample sizes (Gumbel 1954)

2) Maximum Likelihood Method (ML), as described by Clarke

(1973), and using a small sample size correction proposed by Fiorentino and Gabriele (1984).

3) Mean Square Fit (MS) of the upper part of the empirical distribution on Gumbel paper in case of pronounced curvature or break in the lower point swarm.

In fact, only the MO and MS methods were found acceptable in the experiments a), b), or c) above. The ML method, even with the proposed correction factors, always underestimates the Gumbel parameters, which is unacceptable for safety objectives.

Finally, the fixed origin effects, mainly sensitive for 1 h and 2 h rainfalls, were corrected, according to Weiss, by a coefficient of 1.14 and 1.07 respectively.

For each time-step Δt, 3 parameters were selected for a regional study and thus computed at each station :

Gumbel scale parameter	$a(\Delta t)$
10 year rainfall	$P_{10}(\Delta t)$
100 year rainfall	$P_{100}(\Delta t)$

along with their sampling estimation variances , as summarized by Laborde (1984) :

$$\sigma_a = 1.05\ {}^{a}/\sqrt{n} \qquad \sigma_{P_{10}} = 3.82\ {}^{a}/\sqrt{n} \qquad \sigma_{P_{100}} = 6.15\ {}^{a}/\sqrt{n}$$

Although the choice of the Gumbel distribution seems hardly questionnable in our case, the authors agree that the estimation techniques may still be refined to obtain more robust estimates of quantiles , and even more importantly , better estimates of sampling variations (see e. g. Greenwood et al. - 1979- or Greis and Wood -1981-)

II - REGIONALIZATION OF POINT STATISTICS.

Here too, the approach was 3-fold.

II - 1 - Direct mapping of point values, through either spline surface fitting or kriging (Figure 3).

The first technique may be considered purely analytical and provides as smooth a surface as possible. The quantiles derived in section I show good organization for each time-step. However, this pattern is not unique and changes gradually from one step to annother (e. g. P_{10}(1 h) as compared to P_{10}(2 h) and so on...).

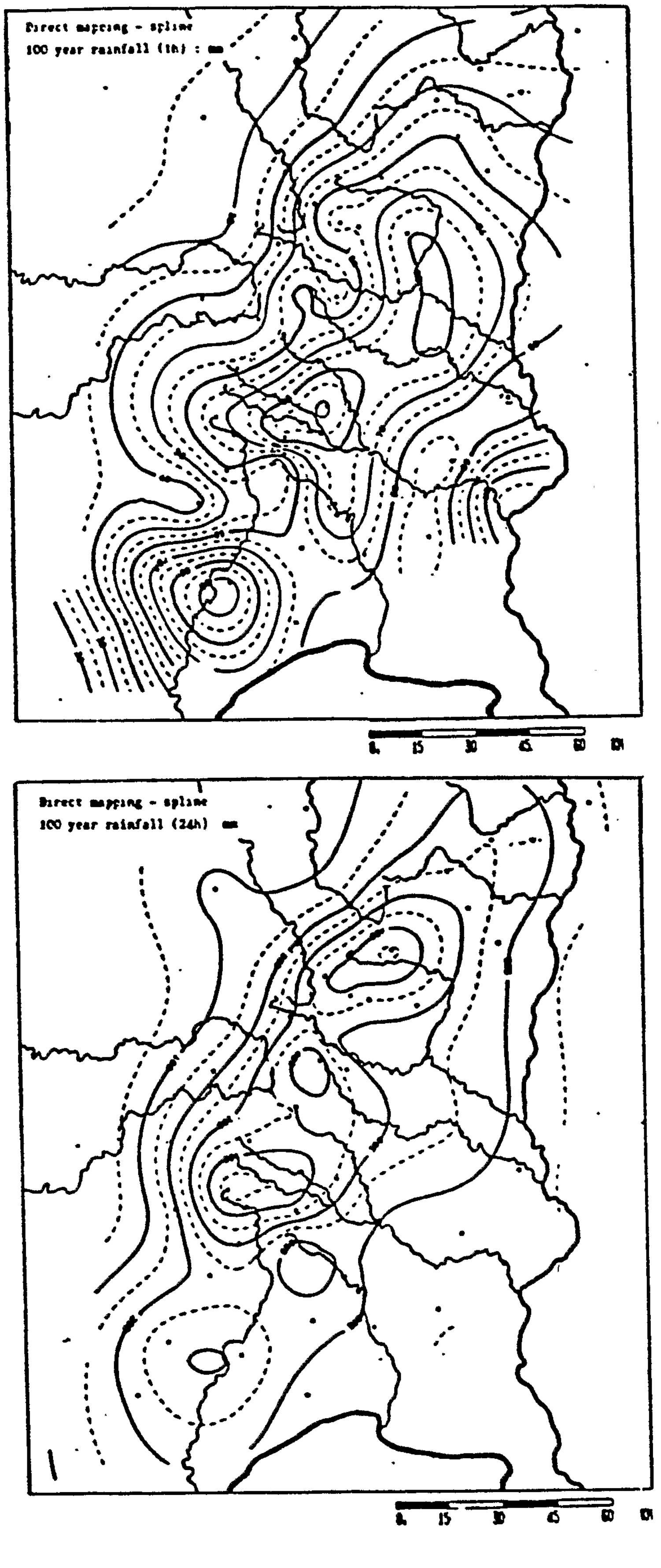


Figure 3 : Direct Mapping by Spline Surface Fitting of the 100-year rainfall for time-steps of 1 h and 24 h.

Kriging is a geostatistical technique that binds mapping to an underlying spatial random process, with its structure summarized in a variogram. (see Figure 4). In our case, the data were reasonably structured , with zero correlation length (or" range" in geostatistical terminology) increasing regularly from :

50 km for P_{10}(1h) to 80 km for P_{10}(24 h)

and 40 km for P_{100}(1h) to 60 km for P_{100}(24h)

Although further data would be needed for a firm conclusion, the decrease in range with larger return periods seems physically acceptable, suggesting that strongest events are more spatially concentrated.

However, kriging tends to use outlying information conservatively and may thus underestimates extreme values, since it tends to fall back on the marginal field average in poorly sampled regions. Conversely, the purpose of such studies, namely design safety, leads to a preference for bias towards overestimation, so spline surface fitting was prefered in this 1st step.

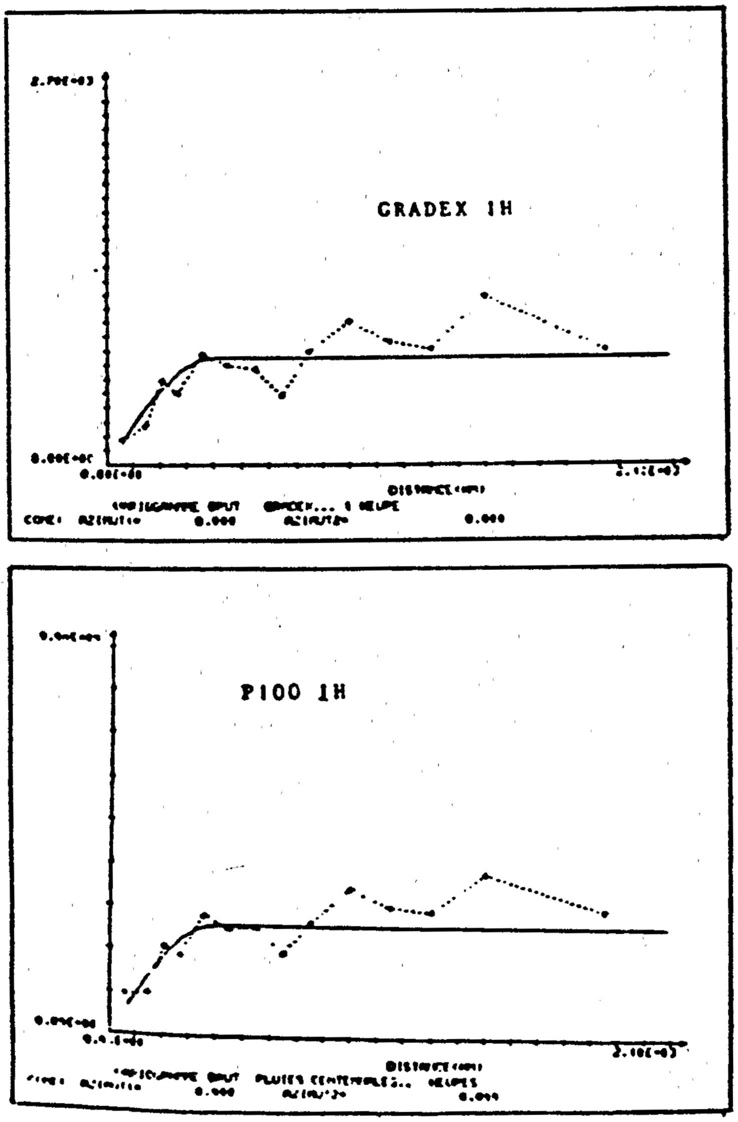


Figure 4:

Variograms of some 1 h-rainfall characteristics (note an approximate "range" of 40 km)

II - 2 Relationship with relief and morphological factors.

In Figure 3, the relief sketched in figure 1 seems to play a major role in the well organized pattern of isolines. Previous studies by Laborde (1982,1984) in North East of France have shown that a significant part of the variation in extreme rainfall parameters may be found in topographic descriptors.

A large number of such variables, summarizing slope, altitude, exposure, embanking, etc ..., were thus computed for each of the 47 stations.

They were then regressed on each parameter (a, P_{10}, P_{100}) for each time-step. After screening, no more than 4 variables are needed (among 100 proposed), and exposure, distances, both to the sea and to the ridge of the mountain range, were almost always selected. The regression equation, calibrated over the sample of 47 stations x_i, may then be applied at any point x_0 were the morphometric variables are available, thus providing a regression estimate Z_R, where Z holds for a or P10 etc...

Here too, a gradually changing pattern emerges. Explained variance (R^2) ranges from 51 % to 68 % for time-steps of 1 h to 24 h, suggesting a stronger relation with the regional environment for totals over longer durations.

The relief of the whole region was then digitized over 500 m space-steps, and the regression equations obtained over the 47 stations were then extended to each gridpoint where the morphometric explanative variables could easily be computed.

This provided a very detailed map, generated by the relief alone, but that cannot be used as such because of the large part of unexplained variance.

II - 3 Cokriging.

Kriging in section II-1 can be considered as a regression of each unknown point on its measured neighbours only, i.e. the variable $Z(x_0)$ at point x_0 is regressed on the same variables $Z(x_i)$ at the $i = 1, \ldots n$ observed stations . Here " observed" stands for " estimated from a sample of extreme rainfalls". Inversely, section II- 2 considered regressions over external relief variables only, but which were computed at the exact point x_0.

Cokriging (see for exemple Vauclin et al. 1984) is an attempt to merge the 2 steps described above in a kind of spatial ARMA model where the variable at the unmeasured point x_0 is estimated as :

$$Z(x_0) = \sum_{i=1}^{n} \alpha_i Z(x_i) + \sum_{j=1}^{k} \beta_j Y_j + \epsilon(x_0)$$

where Y_j could be the k original relief variables used in step II-2.

In fact, in order to reduce the number of degrees of freedom, we directly used in the Σ_1 the relief estimated value $Z_R(x_0)$, obtained from the regression of step II-2 , as the only external variable Y. However, this variable may be taken both at x_0 and at a few nearby surrounding mesh points. Figure 5 shows the estimating scheme.

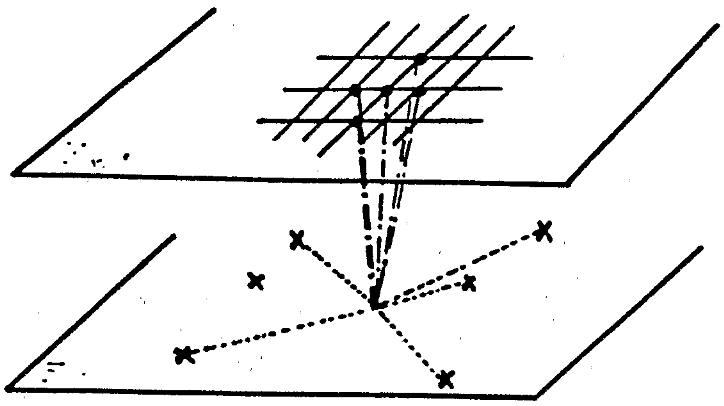


Figure 5 : Estimation Scheme for Cokriging , mixing "observed" values Z(x) with "external" information $Z_R(x)$

As in the kriging method, cokriging provides a map of the theoretical estimation variance, i.e. of

$$\sigma_\epsilon(x_0)$$

that may be used as a confidence interval index of the map obtained for $Z(x_0)$.

Without entering into the problem of estimating the required cross-structure functions, the conclusions that can be drawn from figure 6 are :

- introduction of the relief does not significantly change the pattern obtained in section II-1 for dense regions.
- the detail introduced by rapidly changing local relief may be somewhat unrealistic. Once combined with the map of estimation variance, it may be partly considered as noise and only large features need be considered.
- nevertheless, in poorly sampled zones, as well as at the edges of the map, the use of relief partially palliates the lack of nearby observed stations and significantly reduces border effects, as seen in figure 3.

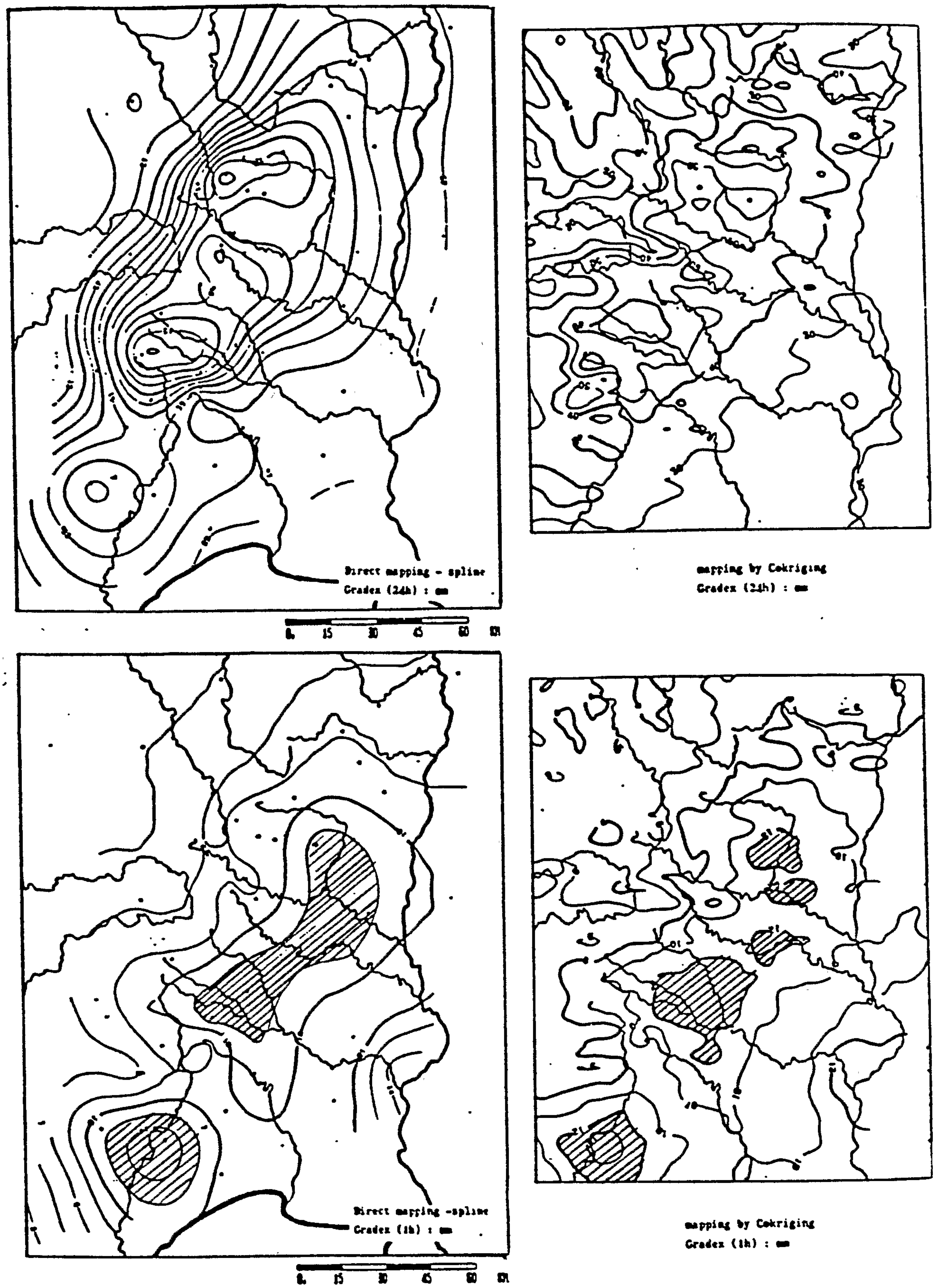


Figure 6: Comparison of direct Spline Mapping with Cokriging using Relief Estimation (Note that the relief has been digitized only over a part of the whole domain. The 1h-12 mm contour has been hachured to allow easier comparison)

A final but very interesting conclusion comes from the comparison of different time-steps : the zones of maxima for a given time-step smoothly shift from the ridge zone (maximum elevation) for long durations (24H, 12 h), towards the middle and the bottom half of the sloping relief (on the Mediterranean side) for short durations (2 h, 1 h). This backs up the very few results concerning the climatology of short duration rainfalls in mountains regions (Fletcher et al. 1981).

III - SPATIALLY AVERAGED RAINFALL .

In fact, the actual design problem is often set up in terms of a flood volume, itself related to a rainfall volume collected by the watershed. In other words, a given rainfall event may strongly vary in space, although picked out from a spatially homogeneous probability distribution, but only the basin average is of interest for the designer.

Furthermore, the extreme value distribution may also be heterogeneous, i.e. its parameters may vary significantly in space, as seen in section II for our region. This thus causes problems for the elongated watersheds of the Cevennes region, which spread through isolines 80 to 140 mm/h for the 2 hour - 100 year rainfall for exemple.

So, there is a strong need to obtain some kind of integrated representative value at the scale of a given watershed, as well as the distribution of its extremes. The appropriate time-step here is usually accepted to be the concentration time of the watershed, that may be roughly derived and fixes the time-step.

Hopefully, the designer should then be able to easily derive an estimate of the rainfall volume for any given return period without going back to the original data, for instance using the maps derived in section II.

Such a possibility was tested using the following 3 step approach:

III- 1 Studies of long series of spatially averaged rainfall.

For a few heavily instrumented watersheds (50 to 500 km ^{2}), series of basin average rainfall have been computed, using a kriging estimation process, at time-steps of 1 to 24 h (Lebel and Creutin 1983) over a period of about 12 years.

First, the Gumbel distribution was tested for this new variable and then, the Moment Method was used to estimate the appropriate parameters. It appears that this spatially averaged rainfall over small watersheds shows even more gumbelian behavior than point rainfall, especially for small time-steps.

These series were thus used as a kind of empirical, or sample-based, reference for more analytical approaches.

III - 2 Average statistics derived from point statistics.

In fact, the sample statistics required, at least for the Moment Method, were estimates of the mean and variance of the spatially averaged extreme rainfall :

$$P_M = \frac{1}{B}\int_B P(x)\,dx \quad \text{over a basin area B.}$$

These may in turn be estimated from neighbouring point values as weighted estimators, since an event value may be obtained as :

$$P_M^* = \sum_{i=1}^{n} \lambda_i P(x_i)$$

then the population mean may be estimated as :

$$\mu_M = E\langle P_M\rangle = E\langle P_M^*\rangle = \sum \lambda_i \mu_i$$

and

$$\text{var}\langle P_M^*\rangle = \sum_i \sum_j \lambda_i \lambda_j \,\text{cov}\langle P(x_i), P(x_j)\rangle$$

$$= \sum_i \sum_j \lambda_i \lambda_j (1-\gamma_{ij})\,\sigma_i \sigma_j$$

with γij the scaled structure function (Gandin 1965) or variogram between observed points or stations x_i and xj.

However, this variance var$\langle P_M^*\rangle$ underestimates the true variance since the estimate P_M^* has a lower variance than the true integrated value P_M.

The difference is the estimation variance, which may itself be evaluated as :

$$\sigma_E^2(P_M^*) = \sum_i \lambda_i \int_B \gamma(x_i,x)\,dx - \sum_i \sum_j \lambda_i\lambda_j \gamma_{ij} - \int_B \gamma(x,x')\,dx\,dx'$$

However, as a first approximation, σ_E^2 may be omitted. In this case, the basin average Gumbel parameter a_B may be estimated as a_B^* through the MO method taking:

$$a_B^* = 0.78\left[\sum_i \sum_j \lambda_i \lambda_j (1-\gamma_{ij})\,\sigma_i \sigma_j\right]^{1/2}$$

instead of

$$a_B = 0.78\ \text{var}(P_M)^{1/2}$$

This may even be written :

$$a_B^* = \left[\Sigma_i \ \Sigma_j \ \lambda_i \lambda_j (1 - \gamma_{ij}) \, a_i a_j\right]^{1/2}$$

The λ_i's are the weights used in estimating the average rainfall and are given by the well known kriging system :

$$\begin{bmatrix} & & & 1 \\ & & & \cdots\cdots \\ & \gamma_{ij} & & 1 \\ & & & \cdots\cdots \\ 1\ldots\ldots & 1 & \ldots\ldots 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \\ \lambda_i \\ \\ \mu \end{bmatrix} = \begin{bmatrix} \frac{1}{B}\int_B \gamma(x,x_i)\,dx \\ \\ 1 \end{bmatrix}$$

And thus, an estimate of the Gumbel parameter a_B may be derived directly from the parameter point values a_i's.

The same holds for estimating μ_M, i.e.

$$E \langle P_M \rangle = E \langle \frac{1}{B} \int_B P(x)\,dx \rangle$$

Other studies show that γ_{ij} in fact depends only on the distance d_{ij} between x_i and x_j, and on the time-step duration Δt (Lebel and Creutin 1983).

Thus, the Gumbel distribution for basin averaged rainfall may be inferred from point statistics (or map), current rainfall geostatistical structure function (or an appropriate model $\gamma(d_{ij},\Delta t)$), through the simple calculation of the kriging weights λ_i's.

III - 3 Simpler estimates.

However, this simple computation may itself be considered tedious, and a cruder estimate may be obtained from a simple averaging:

$$a_B^{**} = \Sigma_i \ \Sigma_j \ a_i \, a_j$$

with i,j screening the whole meshed domain B considered.

In fact, this estimate is not as crude as would first appear , in that it compensates the underestimation of var(P_M^*) viz. var(P_M) by considering that the λ_i's are all equal, i.e. that x_i, x_j are fully correlated at 1.0 .

Practically, such an averaging can easily be performed over the maps obtained in section II.

These 3 steps were compared over several watersheds with reasonnably good results.

Example: parameters a and P_{10} for

	$\Delta t = 1$ h		and $\Delta t = 24$ h	
areal rainfall computations	8	35mm	53	220 mm
"kriging" of point parameters	7	41	43	195 m
crude averaging	12	52	61	245mm

for the Gardon d'Anduze watershed (550 km^2).

areal rainfall computations	9	37mm	58	240 mm
"kriging" of point parameters	9	43mm	49	220mm
crude averaging	12	51mm	65	240 mm

for the Gardon de Mialet (only 265 km^2 but higher on the mountain slope).

As expected, the kriging approach slighty underestimates, while full averaging overestimates. These 2 values, nevertheless, provide a reasonable range for the designer.

Note that these approaches arec only valid for watershed extension of 1 or 2 correlation lengths or "ranges" at the most. Outside this limit, the Gumbel hypothesis may no longer be acceptable.

CONCLUSIONS

Although the different steps have not yet been developed to the same extent, the overall approach is highly integrated and aims at merging as much information as possible in the form of easily handled maps, together with a simple technique for estimating basin average statistics. Real world applicaions to small size watersheds (<500 km^2) in the investigated region are underway.

ACKNOWLEDGEMENTS:

We are deeply indebted to the different services, EDF-DTG, SRAE's, DDE 's etc... which kindly provided the data . D. DUBAND, P.GUILLOT and J.P. LABORDE are greatly thanked for suggestions and helpful discussions. This study was partly funded by the Direction Departementale du Gard .

REFERENCES.

BERAN M. 1981 ' The "Gradex " method, letter to the Editor ' IAHS Bulletin , Vol. 26, 2, p. 223-224

CLARKE R.T. 1973 ' Mathematical models in Hydrology' Irrig. and Drain. Pap. n°19 . F.A.O. publ. United Nations . Rome.

COOPER C.F. 1967 ' Rainfall intensity and elevation in Southwestern Idaho' Water. Ress. Res. Vol 3, 1, p 131-137

CROFT A.R. and MARSTON R.B. 1950 ' Summer rainfall characteristics in Northern Utah' Trans. Amer. Geophys. Union , Vol 31, 1, p. 83-95

DUBAND D. 1967 ' Composition des lois de probabilite des pluies journalieres' Internat. Hydrol. Symp. Colorado State Univ., Proc. Vol 1, p . 516-526

FIORENTINO M. and GABRIELE S. 1984 'A correction for the bias of Maximum Likelihood estimators of Gumbel parameters ' J. of Hydrol., Vol. 39, 1, p.39-49

FLETCHER H.E. et Al. 1981 'Precipitation characteristics of summer storms at Straight Canyon Barometer watershed, Utah' Research Pap. INT-274, U.S. Dept of Agriculture, Nat. Forest Service - May 1981 - 11 p.

GREENWOOD J.A., LANDWEHR J.M., MATALAS M.C. and WALLIS J.R. 1984 ' Probability weighted moment: definition and relation to parameters of several distribution expressed in inverse form' Water . Ress. Res., Vol 15, 16, p. 1049-1054

GREIS N.P. and WOOD E. 1981 'Regional flood frequency estimation and network design' Water Ress. Res. , Vol. 17, 4, p. 1167-1177

GUILLOT P. and DUBAND D. 1967 ' The Gradex method to compute extreme flood probability from the observation of rainfall' Internat. Hydrol . Symp., Colorado State Univ., Fort Collins, Proc. Vol 1, p. 506-515

GUMBEL E.J. 1954 ' Statistical theory of extreme values and some practical applications ' Nat. Bureau of Standards, Appl. Math. Series 3-3

HERSHFIELD D. and KOHLER M. 1960 'An empirical appraisal of the Gumbel extreme value procedure' J. of Geophys. Res., Vol 65, 6, p. 1737-1746

LABORDE J.P. 1982 ' Cartographie automatique des caracteristiques pluviometriques: prise en compte des relations morphometrie-pluviometrie ' La Houille Blanche N°4, p. 331-338

LABORDE J. P. 1984 'Analyse des donnees et cartographie automatique en hydrologie - Elements d'hydrologie lorraine' These de doctorat d'Etat. I.N.P. de Lorraine. Nancy -France

LEBEL T. and CREUTIN J.D. 1983 'Areal rainfall estimation, forescast, and simulation over small watersheds subject to major flashfloods' 5th Conf. on Hydromet., Amer. Met. Soc., Tulsa, OK., Proc. p.291-298

LEBEL T. and GUILLOT P. 1983 '985 Statistical analysis of point and areal hourly rainfall: application to the design flood by the Gradex method' 5th Conf. on Hydromet., Amer. Met. Soc. Tulsa, OK. Proc. p. 17-25

NATURAL ENVIRONMENT RESEARCH COUNCIL 1975 ' Floods Studies Report' 5 vol. NERC London U.K.

OBERLIN G. 1980 'Une methode sommaire d'estimation des crues sur de petits bassins versants de metropole: la methode SOCOSE' La Houille Blanche N° 4/5, p. 271-273

OBLED CH. 1983 'Hydrometeorology in France : New results and developments ' 5th Conf. on Hydromet., Amer. Met. Soc. Tulsa, OK., Proc., p 1-8

VAUCLIN M. et al. 1983 ' The use of Cokriging with limited field soil observations' J. of Soil Science Soc. of Amer., Vol 127, 2, p 175-184

Third International Conference on Statistical Climatology

June 23-27, 1986,

Vienna, Austria