Stochastic Hydrol. Hydraul. 2 (1988) 245–261

Stochastic Hydrology and Hydraulics

A geostatistical approach for areal rainfall statistics assessment

Th. Lebel

ORSTOM – Laboratoire d'hydrologie, 2051, Avenue du val de Montferrand, B.P. 5045, F-34032 Montpellier Cedex, France

J. P. Laborde

Ecole Nationale Superieure de Geologie de Nancy, Institut National Polytechnique de Lorraine, B.P. 40, F-54051 Vandeœuvre-les-Nancy, France

Abstract: Areal rainfall statistics are more relevant in flood hydrology and water resources management than point rainfall statistics when it comes to help designing dams or hydraulic structures. This paper presents a geostatistically based method to derive the areal statistics from point statistics. Assuming that the distribution models of point rainfall and areal rainfall belong to the same class of models and that the rainfall process is stationary, it is shown how the parameters of the areal distribution model can directly be computed from the parameters of the point distribution models in case of a non stationary process, an approximation is derived that yielded good results when applied to a mountainous region in Southern France. The method also allows the computation of the areal reduction factors in a very general form.

Key words: Geostatistics, areal rainfall distribution, areal reduction factor, Gumbel distribution.

1 Introduction

Several methods of estimating high return periods for floods rely on the statistical analysis of precipitation series.

Some of these methods directly relate the ten or hundred year peak flow to the ten or hundred year rainfall using an empirical relationship.

For its part, the gradex method (Guillot and Duband 1967) is based on the utilisation of the scale parameter of a Gumbel distribution fitted to the rainfall series to extrapolate the observed runoff distribution.

In both cases, the meaningfull rainfall statistics should be computed from areal rainfall series rather than from point series. It is, however, rarely possible to obtain long areal rainfall series for the direct fitting of a probabilistic model. This raises the question of deducing the areal distribution or some areal statistics from the point distribution.

There are mainly two ways to tackle the problem:

1) computing an areal reduction factor in order to compute the x year areal rainfall as the product of the x year point rainfall by the areal reduction factor,
2) estimating the theoretical statistical distribution of the areal rainfall.

The first approach was taken by the U.S. Weather Bureau (1958), Brunet-Moret and Roche (1966) among others, the areal reduction factors being computed in an empirical manner and thus only locally valid. The second was chosen by Rodriguez-Iturbe and Meija (1974), who were, to our knowledge, the first to propose a methodology based on the analysis of the spatial correlation structure of the rainfall process. All the methods mentioned above have in common the assumption of the stationarity in the rainfall process, i.e., the point rainfall distribution is the same at any point in the area considered.

Based on the utilisation of the spatial correlation function widely used in geostatistics, the variogram, this paper shows how to compute the areal rainfall probability density function parameters and the areal reduction factors for any given probability of exceedance. A method is derived for obtaining reliable estimates of those parameters, even when it is not possible to consider the point rainfall process variance as a constant over the area.

2 General framework

2.1 *The point rainfall: definitions and notations*

Considered in a probabilistic context, the rainfall depths, Z, measured at a given location over a given duration of accumulation, are assumed to be the realizations of a $2D$ random process $Z(u;\omega)$, where u is a space coordinate ($u \equiv (x,y)$), and ω is an element of the set of events Ω. The mean and the variance of Z are two functions of the space coordinate, noted $\mu(u)$ and $\sigma^2(u)$. The kth realization of Z, associated with the event ω_k is the function $z^k(u)$. Following a current approach (e.g. Journel and Huijbregts 1978), Z is considered a set of random variables (RV) $Z_i(\omega)$ defined at each gauge location $u_i \equiv (x_i,y_i)$. The kth realization of Z_i, is the value taken by z^k at the point u_i and is noted z_i^k, which holds in fact for $z_i(\omega_k)$.

When analysing a point rainfall series, we focus on the distribution of a given RV Z_i, independently of the spatial variability of Z in the $2D$ geographic space $\{u(x,y)\colon u \in \mathbf{R}^2\}$. This distribution is characterized by its probability density function (pdf) f_i. Its first two moments are:

$$E\{Z_i\} = \int_\Omega z_i(\omega) f_i(z_i(\omega))\, d\omega$$

$$Var\{Z_i\} = E\{[Z_i]^2\} - [E\{Z_i\}]^2$$

$$= \int_\Omega [z_i(\omega)]^2 f_i(z_i(\omega))\, d\omega - [E\{Z_i\}]^2$$

Clearly, the expectation operator has a climatological meaning, i.e., it applies to the whole set of possible values $\{z\}$ associated with the events $\{\omega\colon\omega \in \Omega\}$. $E\{Z_i\}$ and $Var\{Z_i\}$ are in fact the values μ_i and σ_i^2 taken by the functions $\mu(u)$ and $\sigma^2(u)$ respectively, at the point u_i. These values are not exactly known, but estimated as:

$$\hat{\mu}_i = m_i = \frac{1}{K}\sum_{k=1}^{K} z_i^k \tag{1}$$

and

$$\hat{\sigma}_i^2 = s_i^2 = \frac{1}{K-1}\sum_{k=1}^{K}\left[z_i^k - m_i\right]^2, \tag{2}$$

where K is the number of events observed at gauge i.

2.2 *The spatial variability of rainfall*

Rainfall probably is the variable in hydrometeorology that varies the most rapidly in space. This has long been a main concern to hydrologists interested in water resources management and several methods are commonly used to account for it. Among them an increasingly favored one is the whole set of techniques known as 'geostatistics'. Because geostatistics was developed as a new field of statistics with the aim of solving mine recovery problems, the emphasis was put on the single realization context ($K = 1$), and the spatial variability of Z is mainly considered as being the variability observed between the values of this single realization. Due to the impossibility of computing m_i and s_i^2 the variations of $\mu(u)$ and $\sigma^2(u)$ are not readily available. The basis of the geostatistical approach is to get rid of these rather than to use them as part of our knowledge of the process. While this technique may sometimes prove useful in rainfall studies, it is by far preferable to use as much information as possible as long as this information originates from the same statistical population. This may be achieved by studying the rainfall fields in a climatological context. A comprehensive development of an interpolation method in such a context was undertaken by Gandin (1965), based on the spatial covariance function:

$$C(u,u') = E\left\{[Z(u) - \mu(u)]\cdot[Z(u') - \mu(u')]\right\}, \quad (u,u') \in \mathbf{R}^2$$

As long as it is thought that good models of the functions $\mu(u)$, $\sigma^2(u)$ and $C(u,u')$ can be developed, the correlation $\rho(u,u')$ between any two points (whether instrumented or not) is computed, which in turn permits the calculation of some rainfall statistics. This is however rather infrequent, due to the poor density of long rainfall series, especially at small time steps (less than 24 hours). The estimates of the values of the functions, μ, σ^2 and C, are thus obtained in a very limited number of points, which prevents the derivation of their analytical representation, unless some further assumptions are made. Since first applied by Delhomme and Delfiner (1973), the geostatistical framework has been widely used in areal rainfall analysis. Assuming the stationarity and the isotropy of the differences $[Z(u+h) - Z(u)]$, the covariance function is replaced by the variogram:

$$\gamma = \frac{1}{2}E\left\{[Z(u+h) - Z(u)]^2\right\}, \quad (u,h) \in \mathbf{R}^2 \tag{3}$$

In order for to estimate γ, the function $\mu(u)$ is assumed to be constant:

$$\mu(u) = m_z, \quad u \in \mathbf{R}^2, \tag{4}$$

but the inference of γ does not require prior knowledge of μ_z.

In case of a finite variance (i.e., the process Z itself is stationary), we have the following relations:

$$\gamma = \sigma_Z^2 - C(h) \tag{5a}$$

and

$$\rho(h) = \frac{C(h)}{\sigma_Z^2} = 1 - \frac{\gamma}{\sigma_Z^2}, \tag{5b}$$

where σ_Z^2 is the variance of the process ($\sigma_Z^2 = C(o)$), and $C(h)$ and $\rho(h)$ are the covariance and correlation functions, respectively (depending on the distance between u and u' only since the process was assumed isotropic). Calculations based on the variogram may thus be carried out using the correlogram and vice-versa. On the other hand, when the variance is a priori not finite, these calculations are possible only when using the variogram. A further complication occurs if the process is assumed to be non stationary. This requires the inference of either a model of the drift and the underlying variogram, or of a so-called Generalized Covariance (GC) (Matheron 1972).

Acceptance or rejection of the assumption of rainfall stationarity depends heavily on the rainfall type of interest, the area over which the rainfall is studied, and the available records. But, from the estimation point of view, the most important factor conditioning our analysis, is the density of the measurement network with respect to the range of correlation of the data. As a matter of fact it will determine the limit distance of utilisation of the structural function (whether covariance, variogram or GC). We are then more interested in the plausibility of a local stationarity than in the overall stationarity of the process. Therefore, even though it is obvious that rainfall is not a stationary process at large scales, it is worthwhile to notice that several authors have concluded that the combination of a constant drift and a variogram was an appropriate representation of the rainfall structural function for the purpose of local interpolation. (e.g., Chua and Bras 1980, Creutin and Obled 1982, Lebel and Bastin 1985, Desbordes 1987). This conclusion is also supported by Journel (1986) who incidentally remarks that in general "Universal kriging, or for that matter IRF k, is needed only in cases of extrapolation when the point being estimated is beyond the range of any datum".

2.3 Kriging in a climatological context

Of course this raises the question of the utility of the kriging approach as compared to Gandin's. As stated earlier, the main impediment to the use of Gandin's method is that the inference of the covariance function requires the knowledge of the mean and the variance at each point of interpolation. Thus, the fields of the observed means and variances has to be interpolated, resorting to a method allowing the interpolation in a single realization context (such as kriging for instance). The main advantage of geostatistics as compared to covariance based methods is to provide a framework which is valid in many different situations, even in the simplest case of a formal equivalence with Gandin's climatological approach. In this way, Delhomme and Delfiner (1973), estimating the areal rainfall over the Ouadi Kadjemeur, found it hazardous to infer the variogram storm by storm with a limited number of raingauges. They introduced the concept of average variogram, which was further developed by Bastin et al. (1984), who separated the variogram model into a time-varying but space-invariant factor and a time invariant variogram. More recently, a mean square interpolation error method was derived by Lebel and Bastin (1985) to indentify a scaled climatological variogram. This scaled variogram g is the structural function of a scaled process W defined as:

$$Z(\omega,u) = a(\omega)W(u), \tag{6}$$

where $A(\omega)$ is a scaling RV and W is assumed to be ergodic. The experimental variogram is built by mixing several realizations $w^k, k = 1,K$, that is by computing $(w_i^k - w_j^k)^2$ for every pair of stations (i,j) and every realization index k. The value a^k taken by a for a given realization ω_k is computed as the experimental standard deviation s^k of the observed values z_i^k:

$$s^k = \left\{ \frac{1}{N} \sum_{i=1}^{N} [z_i^k - m_k]^2 \right\}^{1/2},$$

where N is the number of measurement sites and

$$m_k = \frac{1}{N} \sum_{i=1}^{N} z_i^k,$$

is the arithmetic mean of the observations. As a matter of fact, for strong rainfall events spreading over the whole region of study, s^k is a good measure of the rainfall event magnitude, and it may be assumed that the scaling removes the variability in magnitude from one event to another. Furthermore if W is stationary and has finite variance, g is bounded to one, and is very simply related to the correlogram by the following expression:

$$g(h) = 1 - \rho(h). \tag{7}$$

3 Mean areal rainfall distribution

3.1 Basic assumptions

The main areal rainfall,

$$Z_A = \frac{1}{a}\int_A Z(u)\,du,$$

over a given area A, with the total surface of a, is a random variable, and we are interested in inferring its cumulative distribution function (CDF), F_A, or at least some of its statistics. This can be done in a way similar to what is currently done for any RV Z_i, provided a long areal rainfall series is available. However, the scarcity of long term raingauge series (especially for recording raingauge needed at small time steps) makes it often impossible to derive such a series. A direct assessment of F_A, using several long point series, is nevertheless possible on a few assumptions:

A1) The point and areal CDFs belong to the same M set of probabilistic models;

A2) This set of models is characterized by two parameters (Θ_1,Θ_2), a given pair of values of (Θ_1,Θ_2) correspo nding to a particular model within this set M;

A3) The fitting of a model to a given series requires the inference of the mean and the variance of the parent distribution only.

These assumptions seem reasonable as long as: 1) the area under consideration is not too large (under 1000 km^2), and 2) the rainfall data originate from homogeneous meteorological events.

3.2 Computation of the distribution parameters when Z is stationary

If Z is a stationary process, with mean μ_Z and variance σ_Z^2, the variogram γ is related to the covariance by expression (5a). The mean and the variance of the areal rainfall Z_A is then given by (Matheron 1972):

$$\mu_A = \mu_Z \tag{8a}$$

$$\sigma_A^2 = \sigma_Z^2 - \frac{1}{a^2}\iint_A \gamma(u,u')\,du\,du'. \tag{8b}$$

Assumptions A1, A2, and A3 are then sufficient to estimate the areal rainfall distribution parameters from Eqs. (8a) and (8b), by means of a numerical integration of the variogram γ over the surface A. This integration is routinely performed when computing the variance of the estimation error of $\hat{Z}_A$ by kriging.

In the particular case of the Gumbel extreme value type 1 (EV1) distribution, we have:

$$\Theta_1(u) = 0.78\sigma(u) \tag{9a}$$

and

$$\Theta_2(u) = \mu(u) - 0.577\Theta_1(u), \tag{9b}$$

which translates into:

$$\Theta_1(u) = 0.78\sigma_Z \tag{9c}$$

and

$$\Theta_2(u) = \mu_Z - 0.577\Theta_1(u), \tag{9d}$$

when $\mu(u) = \mu_Z$ and $\sigma(u) = \sigma_Z$.

Since the variogram is actually an estimated model of the spatial structure of the process Z, combining (8) and (9a) provides an estimate of the scale parameter Θ_{1A} of the areal rainfall distribution:

$$\hat{\Theta}_{1A} = 0.78\left(\hat{\sigma}_Z^2 - \frac{1}{a^2}\iint_A \gamma(u,u')\, du du'\right)^{1/2}$$

or

$$\hat{\Theta}_{1A} = \left(\hat{\Theta}_{1Z} - .608/A^2\iint_A \gamma(u,u')\, du du'\right)^{1/2}, \tag{10a}$$

where Θ_{1Z} is the scale parameter of the point rainfall distribution.

The location parameter Θ_{2A} is estimated as:

$$\hat{\Theta}_{2A} = \hat{\mu}_A - 0.577\hat{\Theta}_{1A} \tag{10b}$$

3.3 The non-stationary case

In regions where the spatial variability of rainfall is large with respect to the distance of interpolation, it would be unrealistic to assume that both μ_Z and σ_Z^2 are constant. An example is given in Fig. 1, where a map of the scale parameter $\hat{\Theta}_1$ is drawn for a region of rather rugged topography. If one focuses on the watershed located in the southern part of the region, it is impossible to assume that $\Theta_1(u)$ (and hence $\sigma(u)$ according to relation (9a)) is constant. Equation (9a) does not hold and thus any direct computation of σ_A^2 using Eq. (8b) is precluded. A natural step would be to turn to a GC in order to account for this non-stationarity, but this move raises three objections. (i) As stated in Section 2 ordinary kriging often proves as efficient as universal kriging in terms of interpolation accuracy, even though a non constant drift is present at large scales. (ii) The inference of a GC allows the computation of the variance of the linear combination $\sum\lambda_i Z_i$ that filters out the drift (it does not however allow the computation of the variance σ_A^2). (iii) GC's account for the spatial variation of μ_Z but not for that of σ_Z^2. Consequently, rather than resorting to a sophisticated model of the covariance function, it is more suitable to find a way to both take into account the variability of σ_Z^2 in space and to obtain a good, however rough, estimate of the correlation between any two points of the area A. Furthermore, since in real life applications Z is replaced by an estimate, a practical step consists of using the linear combination of the observed RVs:

$$\hat{Z}_A = \sum_{i=1}^{N}\lambda_i Z_i, \tag{11}$$

the variance of which is used as an approximation of σ_A^2 as,

$$Var\{\hat{Z}_A\} = \sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j Cov(Z_i,Z_j). \tag{12}$$

The weighting coefficients λ_i are computed by solving a linear system, such as kriging. Denote g_{ij} the value of g for two RVs Z_i and Z_j, and ρ_{ij} the correlation between these two RVs. Then Eq. (7) leads to:

$$\rho_{ij} = 1 - g_{ij}$$

or

$$Cov(Z_i, Z_j) = \rho_{ij} \cdot \sigma_i \cdot \sigma_j,$$

where σ_i^2 and σ_j^2 are respectively the variances of Z_i and Z_j. Making use of this relation in Eq. (12), yields:

$$\hat{\sigma}_A^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j \rho_{ij} \sigma_i \sigma_j$$

or

$$\hat{\sigma}_A^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j (1 - g_{ij}) \sigma_i \sigma_j.$$

3.4 Application to the Gumbel distribution

The Gumbel distribution is widely used in extreme rainfall analysis. It constitutes a good example in detail of the application of the above expressions to a particular model. Substituting into Eq. (9a), it is possible to compute $\hat{\Theta}_{1A}$:

$$\hat{\Theta}_{1A} = \left\{ \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j (1 - g_{ij}) \hat{\Theta}_i \hat{\Theta}_j \right\}^{1/2}. \qquad (13)$$

From Eq. (11) it is straightforward to estimate μ_A as:

$$\hat{\mu}_A = \sum_{i=1}^{N} \lambda_i \mu_i,$$

and Eq. (9b) takes the form of,

$$\Theta_{2A} = \sum_{i=1}^{N} \lambda_i \mu_i - 0.577 \Theta_{1A}. \qquad (14)$$

The two parameters of the Gumbel distribution of the areal rainfall are thus identified. The methodology is valid for any distribution which is a function of the first two moments of the random variable only.

3.5 Example

Following a comprehensive study of the rainfall series at various time steps (1 to 24 hours) over the Cevennes region (south eastern France), Slimani and Lebel (1986) found that the EV1 distribution fits the experimental distribution of the point rainfall monthly maxima well. The study was limited to the fall season (September-October-November) to guarantee climatological homogeneity of the data, this season being the period of major rainfall events. This study led to the contour mapping of the scale parameter of the EV1 distribution used as an index of the extreme rainfall risk on which the gradex method relies (Obled et al. 1986). In so far as they give a quick indication on the value of Θ_1 at any given point, such maps are very usedful (see another example in Laborde (1984)). They do not however provide a direct and reliable estimate of Θ_{1A} when the surface of A is large with respect to the spatial variability of Θ_1. The availability of a method allowing for the direct computation of a reliable estimate Θ_{1A} based on some point values Θ_{1i}, is thus a complement of great practical interest. In this respect and in order to assess the performance of the method proposed above, it was decided to compare three estimates of the scale parameter Θ_{1A} of the Gumbel distribution of the areal rainfall monthly maxima over a few watersheds, the surface of which ranged from 150 to 550 km^2:

$(\Theta_{1A})^{\dagger}$ is obtained by fitting the EV1 distribution to the experimental distribution of the estimator $\hat{Z}_A$ of the areal rainfall. $\hat{Z}_A$ is computed using the climatological kriging method as described in Lebel and Bastin (1985).

$(\Theta_{1A})^*$ is the mean of the point rainfall distribution parameters Θ_{1i}, $i = 1, n$

$(\Theta_{1A})^g$ is computed using Eq. (13) and the climatological variogram used in previous studies by Lebel and Bastin (1985). This variogram is a spherical model, the range of which is approximately given by the following relationship:

$R = 25t^{0.3}$,

where R is the range in km and t is the duration of rainfall accumulation in hours.

In order for $(\Theta_{1A})^{\dagger}$ to be a good estimate of Θ_{1A}, the areal rainfall Z_A has to be computed with as dense a network as possible. To fulfill this requirement, data from the period 1971-1980 were chosen. This period is shorter than the one on which the contour mapping of the Fig. 1 is based, but a greater number of recording raingages is then available over the area of interest (18 against 13 in 1966 for instance), which means a gage area (Ag) of about 100 km^2 (Fig. 2).

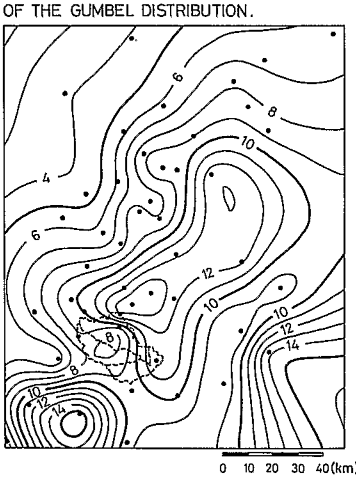


Figure 1

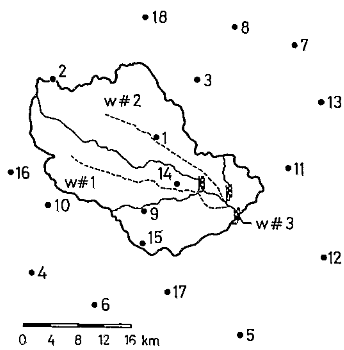


Figure 2

Figure 1. Contour mapping of the scale parameter of a Gumbel distribution fitted to the observed point rainfall monthly maxima (1966-1980) of the fall season (Cevennes region of France)

Figure 2. Recording raingage network over the Gardon watershed

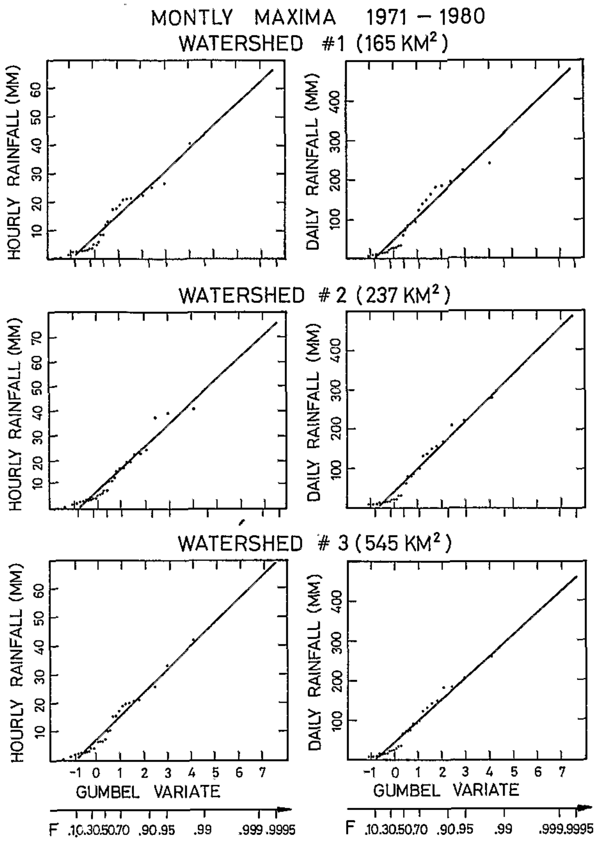


Figure 3. Fitting of the Gumbel distribution to monthly areal rainfall maxima. The areal rainfall is estimated by kriging of the rainfall measured at the 18 raingages shown in Fig. 2

Prior work by Lebel and Bastin (1985) showed that with such a density in this region, the theoretical variance of estimation error $\sigma^2(Z_A - \hat{Z}_A)$ varies from 0.001 to 0.04 depending on the watershed area. Considering these low values of the scaled variance of estimation error, it may be acceptable to approximate the distribution of Z_A by that of $\hat{Z}_A$, keeping in mind that the variance of $\hat{Z}_A$ slightly underestimates that of Z_A. The hourly values of $\hat{Z}_A$, were computed for three watersheds and for all the significant rainfall events of the fall season over the period 1971-1980. These values were then aggregated to make up samples of 2, 4, 6, 12, and 24 hour rainfalls. It was then possible to plot the experimental distributions of the 30 monthly maxima for the various time steps which are fitted well by the EV1 distribution for all three watersheds (Fig. 3). It is especially noteworthy that the goodness of fit is as good or better than that of most fitting to point rainfall series. It may consequently be concluded that in this region the EV1 distribution is a depending model of the monthly maxima areal rainfall distribution for watersheds of which the area is less than 1000 km^2. Thus $(\Theta_{1A})^{\dagger}$ is considered the reference scale parameter to which are compared the two other estimates $(\Theta_{1A})^{g}$ and $(\Theta_{1A})^{*}$.

In the first step $(\Theta_{1A})^g$ and $(\Theta_{1A})^*$ are compared with the network of 18 stations used to compute $\hat{Z}_A$ and are compared to $(\Theta_{1A})^\dagger$. The values of $(\Theta_{1A})^g$ are close to the reference values, while $(\Theta_{1A})^*$ overestimates them (Table 1). This is a direct consequence of the implicit assumption of a perfect correlation between all the stations underlying the computation of $(\Theta_{1A})^*$.

In the second step two networks of lower density were used to test the robustness of $(\Theta_{1A})^g$ and $(\Theta_{1A})^*$, while $(\Theta_{1A})^\dagger$ remained the reference value. The first network is made up of 13 stations (#1 to 13 in Fig. 2) operated at least since 1965, and the second one of 8 stations (#1 to 8 in Fig. 2) operated since 1963. For the purpose of comparison with $(\Theta_{1A})^\dagger$ the values of the point estimates $\hat{\Theta}_1$ computed from the period 1971-1980 were kept, even though better estimates were available using the full record length. In fact, as can be seen from Table 2, there is a slight difference only between the point estimates computed from the period 1963-1980 and those computed from the period 1971-1980. It is clear from Figs. 4a and 4b that the lower the network density, the worse the performance of $(\Theta_{1A})^*$, while $(\Theta_{1A})^g$ remains a close estimate of the reference value $(\Theta_{1A})^\dagger$. These results are related to the decrease of the correlation between the point values used in the estimation process when the network density is reduced. Since the computation of $(\Theta_{1A})^*$ implicitely assumes a correlation of one between these point values, the estimate performs relatively well when the correlations are high but rapidly worse as the correlation coefficients decrease. On the other hand, the steadily good behaviour of $(\Theta_{1A})^g$ tends to support the idea that the climatological variogram used to estimate the correlation coefficients as a function of the distance between the point values is a robust tool. In other words the error introduced by modeling of structural function remains relatively small as compared to other errors involved in the estimation process.

Table 1. Comparison of three estimates of the scale parameter of the Gumbel distribution (network of 18 stations, values in mm)

	W#1(165 km^2)						W#3(545 km^2)					
	1	2	4	6	12	24	1	2	4	6	12	24
			(hours)						(hours)			
$(\Theta_{1A})^\dagger$	7.7	14.1	20.7	25.2	34.1	56.2	8.1	14.3	21.8	26.5	34.6	53.3
$(\Theta_{1A})^g$	8.1	13.9	20.2	25.3	34.8	55.8	8.6	8.0	14.5	21.9	26.5	35.2
$(\Theta_{1A})^*$	10.2	17.8	25.3	30.0	41.1	60.4	10.1	18.7	26.4	32.3	40.6	57.0

$(\Theta_{1A})^\dagger$: Scale parameter of the Gumbel distribution fitted to the monthly maxima of the areal rainfall estimates.
$(\Theta_{1A})^g$: Direct computation of Θ_{1A} using Eq. (13).
$(\Theta_{1A})^*$: Arithmetic mean of the point rainfall distribution Gumbel scale parameters. Note: Data for *W*2 are omitted

Table 2. Influence of the sampling period on the estimation of the scale parameter Θ_1

RAINGAGE and Period		Duration (hours)					
		1	2	4	6	12	24
#10	1971-1980	9.4	14.1	20.2	27.2	41.1	64.8
	1961-1980	9.9	14.5	21.7	28.2	44.0	63.7
#2	1971-1980	9.1	12.3	20.4	28.1	42.0	51.8
	1963-1980	8.6	11.7	19.2	25.4	38.5	52.2

Θ_1: value of the scale parameter (mm) of the Gumbel distribution fitted to the point rainfall monthly maxima (September, October, November).

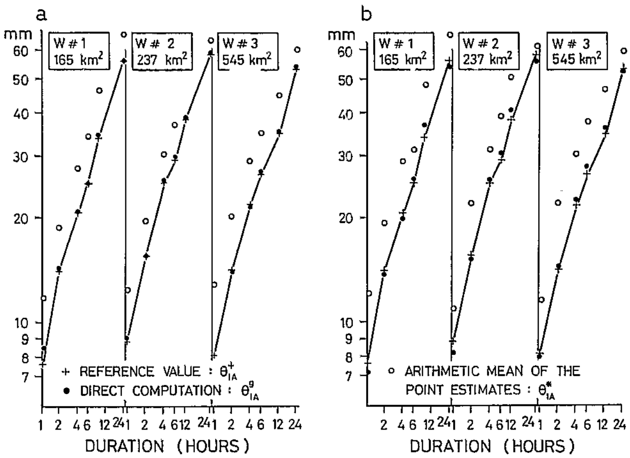


Figure 4. Comparison of the three estimates of the Gumbel distribution scale parameter. 4a: 13 station network (#1 to 13 in Fig. 2). 4b: 8 station network (#1 to 8 in Fig. 2)

4 Areal reduction factors

4.1 General formulation

The concept of an areal reduction factor applies to areas over which the point statistics are assumed stationary. This means that the parameters Θ_1 and Θ_2 of the point rainfall distribution are constants Θ_{1Z} and Θ_{2Z}. On the other hand the rainfall CDF can be expressed as a function of a reduced variate Y, such as:

$$Y = (Z - \Theta_2(\sigma,\mu))/\Theta_1(\sigma,\mu),$$

and reciprocally:

$$Z = Y \cdot \Theta_1(\sigma,\mu) + \Theta_2(\sigma,\mu),$$

where μ and σ are the mean and the variance of the random variable.

The percentiles of the point rainfall and of the areal rainfall corresponding to the probability, P, of non exceedance are:

$$Z(P) = Y_P \cdot \Theta_{1Z} + \Theta_{2Z},$$

and

$$Z_A(P) = Y_P \cdot \Theta_{1A} + \Theta_{2A}.$$

The areal reduction factor is a function of both P and A:

$$K(P,A) = Z_A(P)/Z(P),$$

or

$$K(P,A) = \Theta_{1A}/\Theta_{1Z}[(Y_P + \Theta_{2A}/\Theta_{1A})/(Y_P + \Theta_{2Z}/\Theta_{1Z})],$$

which can be written as,

$$K(P,A) = [(\Theta_{1A}/\Theta_{1Z}) \cdot Y_P + \Theta_{2A}/\Theta_{2Z}]/[Y_P + \Theta_{2Z}/\Theta_{1Z}] \qquad (15)$$

4.2 Application to particular distributions

a. The Normal distribution. In this case we have:

$\Theta_1 = \sigma$ and $\Theta_2 = \mu$.

Thus:

$\Theta_{1A}/\Theta_{1Z} = \sigma_A/\sigma_Z,$

or:

$\Theta_{1A}/\Theta_{1Z} = r(A)$

with:

$$r(A) = \left\{1 - \frac{1}{A^2\sigma_Z^2} - \iint_A \gamma(u,u')\, du du'\right\}^{1/2}, \tag{16}$$

and

$\Theta_{2A}/\Theta_{1Z} = \Theta_{2Z}/\Theta_{1Z} = C_V^{-1},$

where C_V is the coefficient of variation of Z.

This leads to the following expression for the areal reduction factor:

$$K(P,A) = \frac{C_V^{-1} + r(A)(Y_P)}{C_V^{-1} + (Y_P)}. \tag{17}$$

b. The Gumbel distribution. The parameters Θ_1 and Θ_2 are such as:

$\Theta_1 = 0.78\sigma$ and $\Theta_2 = \mu - 0.577\Theta_1,$

or:

$\Theta_{2Z}/\Theta_{1Z} = 1.28C_V^{-1} - 0.577$

$\Theta_{2A}/\Theta_{1Z} = 1.28C_V^{-1} - 0.577r(A),$

and expression (15) becomes:

$$K(P,A) = \frac{C_V^{-1} + r(A)(0.78Y_P - 0.45)}{C_V^{-1} + (0.78Y_P - 0.45)} \tag{18}$$

4.3 Assessment of $r(A)$

Function $r(A)$ is a function of both the surface A and the variogram γ. The question of infering a variogram has been widely addressed in the literature. Several practitioners of geostatistics have pointed out that, in rainfall studies, a good estimation of the spatial correlation at the distances used in the interpolation process is a major factor, the functional type of the variogram being of less importance. It is noticeable that Rodriguez-Iturbe and Mejia (1974) came to the same conclusion when adjusting a spatial correlation function for the purpose of estimating long term areal mean rainfall. Among the variogram models corresponding to relations (5a) and (5b), two are very often used:

1) the spherical model:

$$\begin{aligned} &\gamma = \sigma_Z^2[3/2(h/\beta) - 1/2(h/\beta)^3] && 0 < h < \beta \\ &\gamma = \sigma_Z^2 && h > \beta \end{aligned} \tag{19}$$

where β is the range or the decorrelation distance, and

2) the exponential model:

$$\gamma = \sigma_Z^2(1 - e^{-h/\beta}). \tag{20}$$

The analytic treatment of Eq. (16) is possible for these two models, when A is a simple geometrical form (such as a rectangle, see Serra (1976)). As long as the surface area is small compared to β^2, $[1 - r(A)]$ is well approximated by a power law function of A (Fig. 5):

For spherical model:

$$1 - r(A) = 0.399(A^{1/2}/\beta) \quad (r_s = 1)$$

$$1 - r(A) = 0.434(A^{1/2}/\beta) \quad (r_s = 2)$$

$$1 - r(A) = 0.484(A^{1/2}/\beta) \quad (r_s = 3)$$

For exponential model

$$1 - r(A) = 0.245(A^{1/2}/\beta) \quad (r_s = 1)$$

$$1 - r(A) = 0.265(A^{1/2}/\beta) \quad (r_s = 2)$$

$$1 - r(A) = 0.295(A^{1/2}/\beta) \quad (r_s = 3)$$

where r_s is the side ratio of the rectangle.

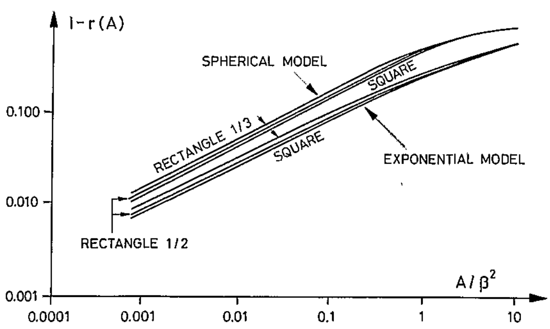
For a square surface, these expressions reduce to:

$$1 - r(A) \approx 0.4A^{1/2}\beta^{-1} \tag{21a}$$

and

$$1 - r(A) \approx 0.25A^{1/2}\beta^{-1} \tag{21b}$$

When the above range of the spherical model is equal to 1.6 *β (exponential model), the behaviour of the two models at small distances is equivalent (Fig. 6), and the above expressions are equal. The increase of $[1 - r(A)]$ remains lower than 10% for rectangles



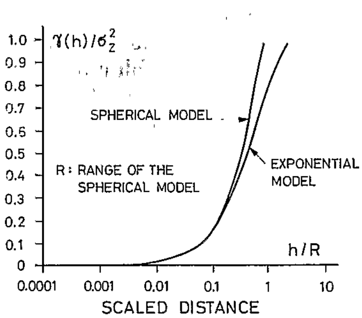


Figure 5 Figure 6

Figure 5. Evolution of the correction factor $1 - r(A)$ with the scaled area A/β^2

Figure 6. Comparison of the spherical and exponential models having same sills. The range of the spherical model is equal to 1.6 times the shape parameter β of the exponential model

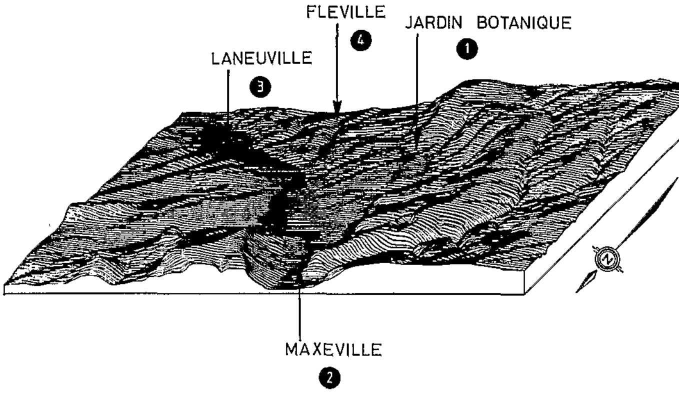


Figure 7. Data location and topography of the Nancy region

of which the side ratio is less than 2. For the Gumbel distribution, a rough assessment of $K(P,A)$ may thus be obtained using a first guess of the decorrelation distance of the process and expression (21a) which leads to:

$$\hat{K}(P,A) = \frac{C_V^{-1} + (0.4\sqrt{A/P})(0,78Y_P - 0.45)}{C_V^{-1} + (0.78Y_P - 0.45)} \qquad (22)$$

Equation (22) is a function of C_V and β only and holds for many of the shapes normally found in real life applications. The estimated value of C_V comes from the analysis of long term series, while the study of the spatial variability of W (see Eq. (6)) provides the value of β. This latter study also permits the fitting of a variogram model possibly better suited to the data than the spherical one. In such a case a better estimate of $[1 - r(A)]$ is obtained by a numerical integration of the variogram over the surface A, and used in Eqs. (17), (18) to compute $K(P,A)$ more precisely.

4.4 Example

The data of 4 raingages in the urban area of Nancy (Eastern France) covering an approximate area of 50 km^2 (Fig. 7) were recorded for eight years. Despite the few stations available, it is possible to estimate the areal reduction factors by inferring the structural function of the rainfall process, a computation that would have been difficult to carry out by empirical methods. This inference is based on the rainfall events having yielded more than 2 mm in one hour or 5 mm in one day at one raingage at least. This selection procedure guarantees, to a certain extent, the meteorological homogeneity of the data, with no marked seasonal pattern. The resulting sample includes 500 rainfall events, with a total of 5700 hours of observations at each raingage. The rainfall depths are aggregated over 1, 2, 3, 4, 5 and 6 hour durations and the correlation coefficients are computed for the six durations and for each pair of raingages (Table 3).

Whatever the duration considered the amplitude of the mean intervals never exceeds ±8% around the mean value, and that of the variance intervals never exceeds ±11% (e.g. $z = 7.3 \pm 8\%$ and $\sigma_Z = 29.8 \pm 8\%$ in 1 hour, or $z = 12.6 \pm 6\%$ and $\sigma_Z = 30.6 \pm 11\%$ in 4 hours). It is thus assumed that the process is stationary, permitting the use of the correlogram as well as the variogram for characterizing the spatial structure of the process. Previous studies by Lebel (1984) and Laborde (1986) showed that in many regions of France, the β parameters of the models (19) and (20) vary as a power type function of the

Table 3. Correlation coefficients computed with 5700 hours of observation

Distance h	Duration (hours)						Raingages
(km)	1	2	3	4	5	6	
5.08	.57	.68	.72	.76	.79	.80	1-4
5.56	.53	.63	.67	.70	.75	.79	3-4
5.61	.50	.57	.61	.63	.66	.69	1-2
5.80	.56	.66	.71	.76	.78	.81	1-3
7.47	.50	.56	.58	.61	.65	.67	2-3
10.13	.44	.50	.51	.54	.56	.59	2-4

duration t. Selecting the exponential model and combining Eqs.s (5b) and (20), the correlogram model is thus a function of both h and t:

$$\rho(h,t) = \exp(-h/at^b)$$

The estimation of a and b is carried out by minimizing the sum of squares of the differences between the Z Fisher transform of the theoretical and experimental correlation coefficients ($Z = 1/2Ln(1-\rho)/(1+\rho)$). This yields the following model:

$$\rho(h,t) = \exp(-h/9.3t^{0.43}),$$

which is represented in Fig. 8, along with the experimental values and the 90% confidence interval based on the residual standard deviation of the correlation between Z transformed variates.

As P tends to 1 in Eqs. (17) and (18), Y_P tends to infinity and thus:

$$K(P \approx 1, A) \approx 1 - (0.25\sqrt{A})/(9.3t^{0.43})$$

or

$$K(P \approx 1, A) \approx 1 - (\sqrt{A})/(37t^{0.43})$$

It is worth noting that this expression is very similar to that resulting from an experimental study carried out in the Paris region by the French Ministry of Agriculture (1980):

$$K(P \approx 1, A) \approx 1/(1 + \sqrt{A}/(30t^{0.33}))$$

It can be seen from this example that a statistical approach permits the computation of meaningful areal reduction factors from a limited number of raingages. Present studies using a more dense network of recent telemetered stations seem to yield results in agreement with the above expressions.

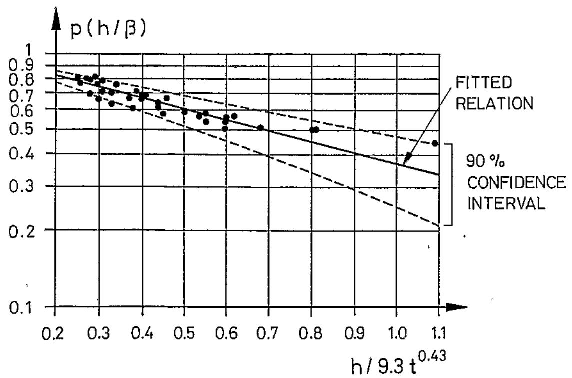


Figure 8. Correlation coefficients as a function of distance and duration

5 Conclusion

Considering the point rainfall as a spatially organized random process, the spatial correlation of which may be characterized by a structural function, it was shown that the parameters of the areal rainfall pdf can be deduced from the parameters of the point rainfall pdf's, estimated at the measurement stations. The fundamental idea is the combination of the information provided by long duration series which are scarce in space with that provided by the more dense networks which are in operation since a few recent years. In addition to classical time series analysis techniques, geostatistics provides a broad framework permitting the analysis of various types of random processes: stationary/non stationary; isotropic/anisotropic. The "climatological kriging" technique is one of the methods developed within this framework and further permits hydrologists to take advantage of the multi-realization context of most hydrometeorological studies. A preliminary step was to verify that the same model fits all the point rainfall series distributions and then to assume that this model stands for the areal value distribution as well. Given this assumption and if the process is stationary, a direct computation yields values of the unknown parameters of the areal rainfall model based on the values of the various point rainfall model parameters. On the other hand, when the process is not stationary, climatological kriging permits the derivation of a method yielding parameter estimates that were shown to be close to some reference values selected for comparison purposes. The method proved to be also suitable for the computation of areal reduction factors in a very general way.

References

Bastin, G.B.; Lorent, C.; Duque, M. 1984: Optimal estimation of the average rainfall and optimal selection of raingage locations. Water Resour. Res. 20, 463-470

Brunet-Moret, Y.; Roche, M. 1966: Etude théorique et m 'e´'thodologique del'abattement des pluies. Cah. ORSTOM, sér. hydrol. III

Chua, S.H.; Bras, R.L. 1982: Optimal estimators of mean areal precipitation in regions of Orographic influence. J. of Hydrology. 57, 23-48

Creutin, J.D.; Obled, C. 1982: Objective analysis and mapping techniques for rainfall fields: an objective comparison. Water Reour. Res. 18, 413-431

Delhomme, J.P.; Delfiner, P. 1973: Application du krigeage à l'optimisation d'une campagne pluviométrique en zone aride. Proc. Symp. on Design of water resources Projects with Inadequate Data, 2, 191-210, UNESCO, Madrid

Desbordes, M. 1987: Contribution à l'analyse et à la modélisation des mécanismes hydrologiques en milieu urbain. Thesis. Universite des Sciences et Techniques du Languedoc. Montpellier, France

Gandin, L.S. 1965: Objective analysis of meteorological fields. Leningrad, 1963. Translated from Russian by Israel Program for Scientific Translation. Jerusalem

Guillot, P.; Duband, D. 1967: The gradex to compute extreme flood probability from the observation of rainfall. In: International Hydrology Symposium. 1, 506-515, Colorado State University, Fort Collins, Colorado

Journel, A.; Huijbregts, Ch. 1978: Mining geostatistics. New York: Academic Press

Journel, A. 1986: Geostatistics: Models and tools for the Earth Sciences. Math. Geol. 18, 93-117

Laborde, J.P. 1984: Analyse des données et cartographie automatique en hydrologie. Eléments d'hydrologie lorraine. Thèse d'Etat. INPL. Nancy

Laborde, J.P. 1986: Les pluies sur les petits bassins versants: une fonction aléatoire dont on peut estimer le variogramme. Hydrologie Constinentale. 1, 3-13

Lebel, T. 1984: Moyenne spatiale de la pluie sur un bassin versant: Estimation optimale, generation stochastique et gradex des valeurs extremes. Thesis, Institut National Polytechnique de Grenoble, France

Lebel, T.; Bastin, G. 1985: Variogram identification by the mean-square interpolation error method with application to hydrology fields. J. of Hydrology. 77, 31-56

Matheron, G. 1972: Theorie des variables regionalisees. In: Traite d'informatique geologique. 306-378, Masson, Paris

Ministere de l'Agriculture. 1980: Synthese nationale sur les crues des petite bassins versants.

Obled, Ch.; Lebel, T.; Slimani, M. 1986: Estimation, regionalization and spatial averaging of extreme rainfall at small time steps. 3rd International Conference On Statistical Climatology. Vienna. Austria

Rodriguez-Iturbe, I.; Mejia, J. 1974: The design of rainfall networks in time and space. Water Resour. Res. 10, 713-727

Rodriguez-Iturbe, I.; Mejia, J. 1974: On the transformation of point rainfall to areal rainfall. Water Resour. Res. 10, 729-735

Serra, J. 1967: Echantillonnage et estimation locale des phénomènes de transition mincés. Thesis, Institut de Recherche de la Sidérurgie, Nancy

Slimani, M.; Lebel, T. 1986: Comparison of three methods of estimating rainfall frequency parameters according to the duration of accumulation. In: Singh, J.V.; Reidel, D. (eds.) Hydrologic frequency modeling; International symposium on flood frequency and risk analyses. D. Reidel Publishing Company, 277-291

U.S. Weather Bureau. 1958: Rainfall intensity-frequency regime, 2, Southeastern United States, Tech. Pap. 29, U.S. Department of Commerce, Washington D.C.

Accepted September 1, 1988