Nordic Hydrology, 25, 1994, 279-300

PDS-Modelling and Regional Bayesian Estimation of Extreme Rainfalls

H. Madsen and D. Rosbjerg
Inst. of Hydrodynamics and Hydraulic Engineering
Technical University of Denmark, DK-2800 Lyngby

P. Harremoës
Inst. of Environmental Science and Engineering,
Technical University of Denmark, DK-2800 Lyngby

Since 1979 a country-wide system of raingauges has been operated in Denmark in order to obtain a better basis for design and analysis of urban drainage systems. As an alternative to the traditional non-parametric approach the Partial Duration Series method is employed in the modelling of extreme rainfalls. The method is applied to two variables: the total precipitation depth and the maximum 10-minute rain intensity of individual storms. On the basis of the at-site modelling a regional analysis is carried out. It is shown that the previous assumption of spatial homogeneity of extreme rainfalls in Denmark cannot be justified. In order to obtain an estimation procedure at non-monitored sites and to improve at-site estimates a regional Bayesian approach is adopted. The empirical regional distributions of the parameters in the Partial Duration Series model are used as prior information. The application of the Bayesian approach is derived in case of both exponential and generalized Pareto distributed exceedances. Finally, the aspect of including economic perspectives in the estimation of the design events is briefly discussed.

Introduction

In the design of urban drainage systems the primary objective is to evaluate the effects of floods and pollution. It is well known that these effects occur during rainfalls of short duration whereas cumulative precipitation (*e.g.* monthly or yearly amounts) is of minor importance in this context. Thus the design has to be evaluated on the basis of extreme rainfall statistics. In particular the precipitation depth and the maximum rain intensity of specified duration during heavy storms have

specific relevance to the design of, respectively, detention basins and pipe systems. The design level is usually expressed in statistical terms as the event corresponding to a specified exceedance probability or, equivalently, in terms of the T-year event, *i.e.* the level which on the average is exceeded once in T years.

In Denmark only very few rainfall records have been available as a basis for the design of sewer systems. It has been assumed that spatial homogeneity exists for extreme rainfalls. Thus a country-wide intensity-duration-frequency curve has been authorized for the design (Danish Water Pollution Committee 1974). In order to obtain a better design basis a new system of recording raingauges was introduced in 1979. The recorded series consist of measurements of cumulative precipitation depth within each storm with a time resolution of one minute. The separation of storms is defined as periods exceeding one hour without precipitation. The raingauge system and the quality control of the data are thoroughly described by Arnbjerg-Nielsen *et al.* (1994). In the present analysis extreme value modelling is applied to the total precipitation depth and the maximum 10-minute rain intensity of individual storms, respectively. The analyzed data consist of measurements from 41 stations with recording periods ranging from 10 to 14 years.

The main objective of the extreme value modelling is the estimation of a design event. Traditionally, intensity-duration-frequency curves based on non-parametric statistics have been applied in Denmark. However, two major problems exist in this approach: i) the uncertainty of the estimate is not quantified, and ii) no direct method is applicable for extrapolation, *i.e.* estimation for return periods greater than the observation period is not possible. In order to circumvent the former problem Arnbjerg-Nielsen *et al.* (1994) introduced resampling techniques based on the Bootstrap method. In this paper a parametric approach based on the Partial Duration Series (PDS) model (also denoted the Peak Over Threshold model) is employed. A parametric model allows for quantile estimation outside the range of observations and direct assessment of the estimation errors.

The classical PDS model comprises the assumptions of Poisson-distributed annual number of threshold exceedances and independent, exponentially distributed exeedance magnitudes (Shane and Lynn 1964; Todorovic and Zelenhasic 1970). In the at-site analysis the classical model assumptions are examined in detail. In addition, the trade-off between this parameter-parsimonious model and a more accurate description of the rainfall process taking non-exponential exceedances into consideration is carried out. On basis of the at-site modelling a regional analysis is performed and the assumption of spatial homogeneity of extreme rainfalls is examined. A regional extreme value model based on Bayesian decision theory is introduced in order to improve at-site estimates and to make inferences possible at non-monitored sites. Finally, the inclusion of economic perspectives in the estimation of the design event is briefly discussed. Taking both the uncertainty of the T-year event and the economic loss into account an optimal decision rule is formulated.

The Partial Duration Series Model

By introducing a threshold level in a hydrological series and considering only peaks above this level, a PDS is obtained. In order to describe the series, modelling of exceedances in both the time and the magnitude domain is required. The basic PDS model assumes that the occurrence of peaks is Poisson-distributed with constant or one-year periodic intensity. Denoting by N the number of exceedances in t years the probability density function (PDF) of N becomes

$$P\{N=n\} = \frac{(\lambda t)^n}{n!} \exp(-\lambda t) \quad , \quad n = 0,1,2... \tag{1}$$

where λ equals the expected number of exceedances per year. The exceedance magnitudes, X_i, are assumed to be independent and identically distributed following the exponential distribution (ED). The ED has the PDF

$$f(x) = \frac{1}{\alpha} \exp\left(-\frac{x}{\alpha}\right), \; x \geq 0 \tag{2}$$

with mean and variance

$$E\{X\} = \alpha \quad , \quad \mathrm{Var}\{X\} = \alpha^2 \tag{3}$$

The T-year event, x_T, is defined as the (1-1/λT)-quantile in the distribution of the exceedances (*e.g.* Rosbjerg 1985), *i.e.*

$$x_T = F^{-1}\left(1 - \frac{1}{\lambda T}\right) = \alpha \ln(\lambda T) \tag{4}$$

Replacing α and λ in Eq. (4) with the maximum likelihood estimators

$$\hat{\alpha} = \frac{1}{N} \sum_{i=1}^{N} x_i \quad , \quad \hat{\lambda} = \frac{N}{t} \tag{5}$$

the T-year event estimator $\hat{x}_T$ is obtained. The stochastic properties of $\hat{x}_T$ were evaluated by Rosbjerg (1985) and Rasmussen and Rosbjerg (1989). In case of high quantile estimation the main contributor to the uncertainty of the T-year event is the sampling variance of $\hat{\alpha}$. For lower quantile estimation, however, the variances of both $\hat{\alpha}$ and $\hat{\lambda}$ are significant.

The basic PDS model has been widely used in the modelling of extreme hydrologic events. In some cases, however, it may be difficult to justify the underlying assumptions. For a Poisson-distributed variable the mean and the variance are identical. In a study of 26 gauging stations in Great Britain, Cunnane (1979) found a tendency for the variance of the annual number of exceedances being significantly greater than the mean. He introduced the negative binomial distribution in order to account for this variability but it did not seem to offer any satisfactory improvement. In fact, the Poisson hypothesis is rather robust (Rosbjerg 1977) and a misspecification of the distribution of the annual number of exceedances is not crucial. The most important element in the PDS approach is the modelling of the excee-

dance magnitudes. In order to justify the exponential assumption it is in some cases necessary to use a very high threshold level. In these cases an alternative exceedance model combined with a lower threshold level may be appropriate in order to obtain more efficient estimates of the T-year event. The generalized Pareto distribution (GPD), which was introduced by Pickands (1975), has recently been used in a number of hydrological studies (*e.g.* Van Montfort and Witter 1986; Fitzgerald 1989). The PDF of the GPD with the shape parameter κ and the scale parameter α^* reads

$$f(x) = \begin{cases} \frac{1}{\alpha^*} \exp\left(-\frac{x}{\alpha^*}\right) & , \kappa = 0 \\ \frac{1}{\alpha^*} \left(1 - \kappa \frac{x}{\alpha^*}\right)^{1/\kappa - 1} & , \kappa \neq 0 \end{cases} \tag{6}$$

For $\kappa=0$ the GPD yields the ED as a special case, whereas for $\kappa<0$ the Pareto distribution is obtained. The mean and the variance in the GPD are

$$E\{X\} = \frac{\alpha^*}{1+\kappa} \quad , \quad \mathrm{Var}\{X\} = \frac{\alpha^{*2}}{(1+\kappa)^2(1+2\kappa)} \tag{7}$$

While in the *ED* the scale parameter is identical to the mean value, *cf.* Eq. (3), it is seen from Eq. (7) that the mean value in the GPD differs from the scale parameter when $\kappa \neq 0$. In the present regional analysis it is, however, important to maintain the scale parameter equal to the mean value when generalizing the ED by introducing a shape parameter. This can be obtained by a reparameterization of the GPD. Consider the transformation $\alpha^*=(1+\kappa)\alpha$ in the GPD, then the PDF (now denoted the generalized exponential distribution, GED) becomes

$$f(x) = \begin{cases} \frac{1}{\alpha} \exp\left(-\frac{x}{\alpha}\right) & , \quad \kappa = 0 \\ \frac{1}{\alpha(1+\kappa)} \left(1 - \kappa \frac{x}{\alpha(1+\kappa)}\right)^{1/\kappa - 1} & , \quad \kappa \neq 0 \end{cases} \tag{8}$$

with mean and variance given by

$$E\{X\} = \alpha \quad , \quad \mathrm{Var}\{X\} = \frac{\alpha^2}{1+2\kappa} \tag{9}$$

The reparameterization of the GPD makes the Bayesian approach (see the next section) more mathematical convenient because the prior distribution of α is preserved when generalizing the ED to a GED. If the GPD formulation is used, the α-parameter has to be assigned a new prior distribution.

For $\kappa \leq 0$ the GED is defined in the interval $0 \leq x < \infty$, whereas an upper bound exists for $\kappa>0$: $0 \leq x \leq \alpha(1+\kappa)/\kappa$. When $\kappa \to -0.5$, the variance of X tends to infinity. For $\kappa=0.5$ the PDF of X is the triangular distribution. Thus, for any practical application, it seems only relevant to consider the GED in the interval $-0.5<\kappa<0.5$. In case of GED exceedances the T-year event is given by

$$x_T = \frac{\alpha(1+\kappa)}{\kappa}\left\{1 - \left(\frac{1}{\lambda T}\right)^{\kappa}\right\} \tag{10}$$

Replacing α and $\varkappa$ in Eq. (10) with the moment estimators

$$\hat{\alpha} = \hat{\mu}\ , \quad \hat{\kappa} = \frac{1}{2}\left(\frac{\hat{\mu}^2}{\hat{\sigma}^2} - 1\right) \tag{11}$$

where $\hat{\mu}$ and $\hat{\sigma}^2$ are, respectively, the estimated mean value and the estimated variance of the exceedances, and utilizing that $\hat{\lambda}=N/t$, *cf.* Eq. (5), the T-year event estimator $\hat{x}_T$ is obtained.

Bayesian Theory

A major problem in the estimation of extreme hydrological events concerns the lack of sufficient data. It is often necessary to assess the risk corresponding to return periods much longer than the length of the observed record, and this can produce unreliable at-site estimates of the T-year event. When no at-site data at all exist, the problem of estimation is even greater. In order to address these problems a regional estimation procedure is usually applied. In the present study a regional Bayesian approach is carried out to improve at-site estimates and to make inferences possible at non-monitored sites.

In the Bayesian analysis parameters are treated as stochastic variables in order to account for the imperfect knowledge of their exact values. This requires that beliefs or knowledge about the parameter $\underline{\theta}$ (which may be a vector or a scalar) are expressed by means of a prior distribution. Combining the prior knowledge and the sample information using Bayes' theorem the posterior (or updated) distribution of $\underline{\theta}$, based on t years of observation, is given by

$$f^t(\underline{\theta}) = f^t(\underline{\theta}\,|\,\underline{Y}) = \frac{f(\underline{\theta})\,l(\underline{\theta}\,|\,\underline{Y})}{\int f(\underline{\theta})\,l(\underline{\theta}\,|\,\underline{Y})\,d\underline{\theta}} \tag{12}$$

where $f(\underline{\theta})$ is the prior PDF of $\underline{\theta}$ obtained on basis of regional information, $l(\underline{\theta}\,|\,\underline{Y})$ is the sample likelihood function of $\underline{\theta}$, and $\underline{Y}$ is a set of sufficient statistics for $\underline{\theta}$.

In case of *ED exceedances* the Bayesian approach has previously been described by Rasmussen and Rosbjerg (1991). To obtain the posterior distribution of the T-year event they used a non-informative prior of the λ-parameter, whereas prior information of the α-parameter was made available using a regional regression model. The derivation given below allows the use of both informative and non-informative priors. For mathematical convenience conjugate priors are employed to describe the regional variation of the PDS parameters. For the ED the conjugate prior is an inverse gamma distribution. The prior density and the sample likelihood function of α are

$$f_\alpha(a) = \frac{1}{\theta\Gamma(\beta)}\left(\frac{\theta}{a}\right)^{\beta+1}\exp\left(-\frac{\theta}{a}\right),\quad l_\alpha(a) = \left(\frac{1}{a}\right)^N \exp\left(-\frac{S}{a}\right),\quad S = \sum_{i=1}^{N} x_i \tag{13}$$

where $\Gamma(\cdot)$ is the gamma function, and (β,θ) are the prior parameters. The posterior distribution is also an inverse gamma distribution with the updated parameters $\beta_t=\beta+N$ and $\theta_t=\theta+S$. For the Poisson distribution the conjugate prior is a gamma distribution. Hence, the prior density and the sample likelihood function of λ are

$$f_\lambda(\ell) = \frac{\tau}{\Gamma(\nu)}(\ell\tau)^{\nu-1}\exp(-\ell\tau),\quad l_\lambda(\ell) = \frac{(\ell t)^N}{N!}\exp(-\ell t) \tag{14}$$

where (ν,τ) are the prior parameters. The updated parameters in the posterior gamma distribution become: $\nu_t=\nu+N$ and $\tau_t=\tau+t$. The distribution of x_T can be deduced by change of variables (Rouselle and Hindie 1976). Assuming α and λ to be independent the prior PDF is found to be

$$f_{x_T}(x) = \int_0^\infty f_\alpha(a) f_\lambda(\ell) \left|\frac{dg}{dx}\right|\Big|_{a=g(x)} d\ell \tag{15}$$

where the transformation $a=g(x)=x/\ln(\ell T)$ is obtained from Eq. (4). The integral in Eq. (15) has to be solved numerically. The prior distribution of x_T is then used to make inferences at non-monitored sites.

If sample information is available, a posterior distribution of x_T may be calculated from Eq. (15) by substituting the prior parameters with the updated parameters ($\beta_t=\beta+N$, $\theta_t=\theta+S$, $\nu_t=\nu+N$, $\tau_t=\tau+t$) in $f_\alpha(a)$ and $f_\lambda(\ell)$. Note that the Bayesian approach also yields information about the T-year event in the case where no prior knowledge is available. This is obtained by using non-informative priors of the PDS-parameters. In this case the posterior distribution of x_T is obtained from Eq. (15) simply by setting $\beta=\theta=\nu=\tau=0$.

For *GED exceedances* the Bayesian procedure requires that also the additional parameter $\varkappa$ is assigned a prior distribution. Since $\varkappa$ is restricted to the interval $-0.5<\varkappa<0.5$, the beta distribution was found appropriate to express the prior knowledge of $\varkappa$. The prior PDF reads

$$f_\kappa(k) = \frac{\Gamma(\xi+\eta)}{\Gamma(\xi)\Gamma(\eta)}(k+0.5)^{\xi-1}(0.5-k)^{\eta-1},\quad -0.5 < k < 0.5 \tag{16}$$

where (ξ,η) are the prior parameters. Assuming that α and $\varkappa$ are independent the prior distribution of the T-year event becomes

$$f_{x_T}(x) = \int_0^\infty \int_{-\frac{1}{2}}^{\frac{1}{2}} f_\alpha(a) f_\kappa(k) f_\lambda(\ell) \left|\frac{dg}{dx}\right|\Big|_{a=g(x)} dk\, d\ell \tag{17}$$

where the transformation

$$a = g(x) = \frac{kx}{(1+k)\left(1-(1/\ell T)^{k}\right)} \tag{18}$$

is obtained from Eq. (10). Using the updated parameters ($\nu_t=\nu+N$, $\tau_t=\tau+t$) in $f_\lambda(l)$ and substituting $f_\alpha(a)f_\kappa(k)$ with the posterior joint density

$$f^{t}_{\alpha,\kappa}(a,k) = \frac{f_\alpha(a)\, f_\kappa(k)\, l_{\alpha,\kappa}(a,k)}{\int_0^\infty \int_{-\frac{1}{2}}^{\frac{1}{2}} f_\alpha(a)\, f_\kappa(k)\, l_{\alpha,\kappa}(a,k)\, dk\, da} \tag{19}$$

where the sample likelihood function of (α,κ) is given by

$$l_{\alpha,\kappa}(a,k) = \prod_{i=1}^{N} \frac{1}{a(1+k)} \left(1 - k\frac{x_i}{a(1+k)}\right)^{1/k-1} \tag{20}$$

the posterior PDF of x_T can be calculated from Eq. (17). Note that for $k>0$ the GED has an upper bound, and hence $l_{\alpha,\kappa}(a,k)$ is given by Eq. (20) only if $\forall x_i$: $x_i \leq a(1+k)/k$. Otherwise $l_{\alpha,\kappa}(a,k)$ is equal to zero.

Preliminary Analysis

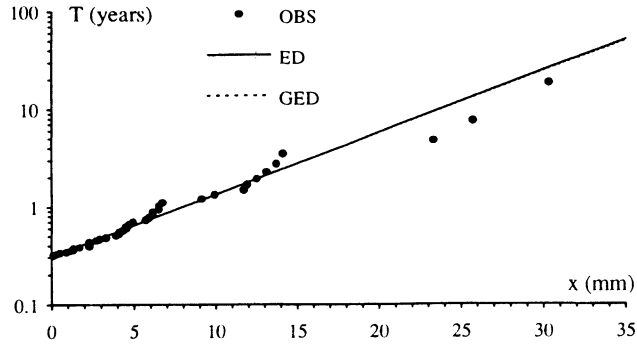
The first step in the analysis is the extraction of peaks from the historical records, *i.e.* determination of the threshold level. Different methods for the choice of threshold were discussed by Rosbjerg and Madsen (1992). They proposed an objective formulation of the threshold level, q_0, which reads

$$q_0 = E\{Q\} + cS\{Q\} \tag{21}$$

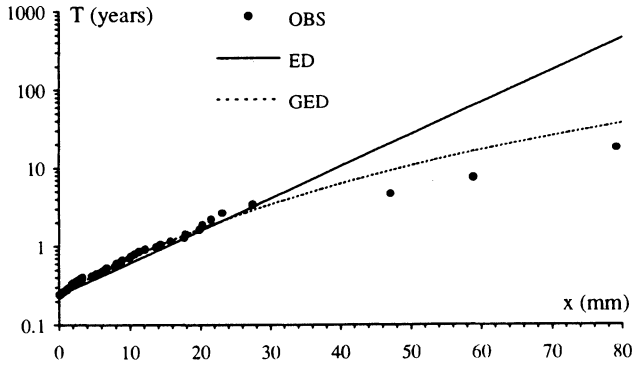
where c is a frequency factor, and $E\{Q\}$ and $S\{Q\}$ are the mean and the standard deviation, respectively, of the raw data. In a regional context, however, the choice of threshold yields a problem. To estimate the threshold at non-monitored sites the regional variation of q_0 must be taken into account. However, since q_0 and λ are strongly dependent, it was decided to use the same threshold at all sites, thus embedding the regional variation of q_0 in the variation of λ. The threshold was chosen as

$$q_{0,\mathrm{REG}} = \frac{1}{M} \sum_{i=1}^{M} q_{0,i} \tag{22}$$

where $q_{0,i}$ is calculated from Eq. (21) with c given the same value at all M stations. Rosbjerg and Madsen (1992) recommended to use a threshold corresponding to a frequency factor about 3. In the present analysis a threshold for both variables (precipitation depth and maximum 10-minute rain intensity) was chosen as the level corresponding to c=3.5. At this level the mean annual number of exceedances, λ, at the 41 stations ranges from 1.7 to 4.8 years^{-1}.



(a) Station No. 20211.



(b) Station No. 30315.

Fig. 1. The empirical distribution of exceedances (precipitation depth) compared to the estimated ED and GED, respectively.

After defining the exceedance series adequate test statistics are employed in order to test the goodness-of-fit of the basic model assumptions. A test of the Poisson assumption is based on the Fisher dispersion test statistic (Cunnane 1979). In order to verify the overall performance of the ED assumption the Kolmogorov-Smirnov test is adopted. When the parameter α is estimated from the sample, modified critical values of the test statistic must be used (Lilliefors 1969). For high quantile estimation the correct tailing-off of the right tail may be crucial. Thus, a test for detecting outliers in exponential data is conducted by considering the ED as a special case of the GPD (Van Montfort and Witter 1985).

The conclusion made from the applied tests was that the Poisson hypothesis can generally be accepted. The ED assumption was, however, rejected at a significant number of the stations due to outliers. At a 5% significance level the hypothesis of ED exceedances for the precipitation depth and the maximum 10-minute rain intensity was rejected at 12 and 7 stations, respectively. In Fig. 1 the empirical distributions of the exceedances at two typical stations in the region are shown and compared to the ED and the GED, respectively. At stations where the shape parameter $\varkappa$ is not significantly different from zero (Fig. 1a) the ED seems to give a

good fit to the observed exceedances. At stations with outliers (Fig. 1b) the ED implies a serious lack of fit for large return periods. In this case the GED implies less bias although it is not perfect. Note that for lower quantile estimation the ED and the GED are almost similar.

Regional Analysis

The previous assumption of spatial homogeneity of extreme rainfalls in Denmark was based on measurements from only 6 locations in Denmark (Danish Water Pollution Control Committee 1974). With the new system of raingauges the information of extreme rainfalls has been significantly improved and the assumption of spatial homogeneity can be examined in more detail. A measure of the spatial variability may be quantified by the regional variation (intersite variance) of the PDS parameters. The observed regional variation is, however, distorted due to sampling uncertainty. Following the approach given in Appendix A the estimate of a PDS parameter at station i, $\hat{\theta}_i$, is assumed to be governed by i) a systematic deviation δ_i from the true regional mean value θ and ii) a random sampling error ε_i, *i.e.*

$$\hat{\theta}_i = \theta + \delta_i + \varepsilon_i \quad , \quad i = 1,2,\ldots,M \tag{23}$$

At the present stage it has not been possible to describe the regional variability of δ deterministically, and it is assumed that the systematic deviations are randomly distributed within the region, *i.e.* no spatial trend or clustering in the PDS-parameters are assumed to be present. Since the variation of the PDS-parameters within subregions is of the same order of magnitude as the variation within the whole region, this assumption seems to be fulfilled. The Bayesian approach can, however, also be applied in the case where the nature and the magnitudes of δ_i can be determined from *e.g.* meteorological and topographical factors. In this case a regional regression model for the PDS-parameters could be developed, *cf.* Rasmussen and Rosbjerg (1991), in order to determine the prior information.

Assuming that the intersite correlation is insignificant an estimator of the systematic regional variation, σ_δ^2, becomes (see Appendix A)

$$\hat{\sigma}_\delta^2 = \max\{0, s^2 - \frac{1}{M} \sum_{i=1}^{M} \hat{\sigma}_{\varepsilon_i}^2\} \tag{24}$$

where

$$s^2 = \frac{1}{M-1} \sum_{i=1}^{M} (\hat{\theta}_i - \hat{\theta})^2 \quad , \quad \hat{\theta} = \frac{1}{M} \sum_{i=1}^{M} \hat{\theta}_i \tag{25}$$

is the observed regional variance, and $\hat{\sigma}_{\varepsilon i}^2$ is the estimated sampling error variance at station i. Thus, a first indication of regional heterogeneity is a value of σ_δ^2

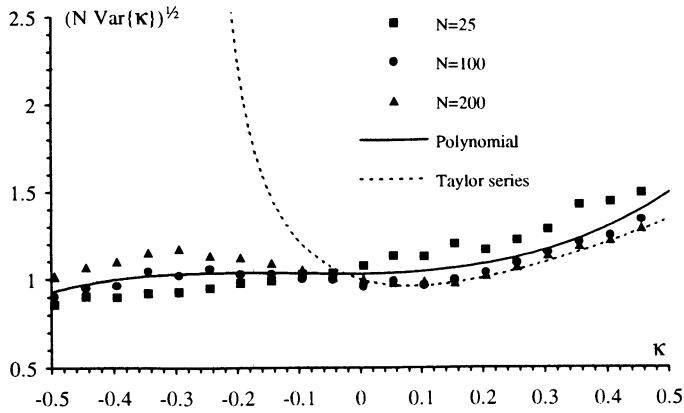


Fig. 2. Average Monte Carlo simulated values of $(N\mathrm{Var}\{\hat{\varkappa}\})^{1/2}$ compared to the assymptotic Taylor series approximation and a 3rd order polynomial approximation, respectively.

different from zero. In order to evaluate Eq. (24) the sampling error variance must be estimated. For the Poisson parameter λ the estimator of the sampling error variance reads

$$\hat{\sigma}^2_{\varepsilon_i} = \hat{\mathrm{Var}}\{\hat{\lambda}_i\} = \frac{\hat{\lambda}_i}{t_i} \tag{26}$$

The modelling of the exceedances involves estimation of the scale parameter α in case of both ED and GED exceedances. Since α is the mean of the exceedances, a robust (non-parametric) estimator of the sampling error variance is

$$\hat{\sigma}^2_{\varepsilon_i} = \hat{\mathrm{Var}}\{\hat{\alpha}_i\} = \frac{\hat{\sigma}^2_i}{N_i} \tag{27}$$

where $\hat{\sigma}_i^2$ is an unbiased estimator of the variance of the exceedances. The statistical properties of the estimator of the shape parameter $\varkappa$ has been evaluated by Rosbjerg *et al.* (1992). They deduced the asymptotic properties of $\hat{\varkappa}$ and compared with Monte Carlo simulated values. The asymptotic variance of $\hat{\varkappa}$ gives poor results for finite samples for $\varkappa<-0.1$ (see Fig. 2), and hence, it was decided to express the variance of $\hat{\varkappa}$ by fitting a polynomial to the Monte Carlo simulated values. For sample sizes in the range $25\leq N\leq 200$ the following expression was found appropriate (see Fig. 2)

$$\hat{\sigma}^2_{\varepsilon_i} = \hat{\mathrm{Var}}\{\hat{\kappa}_i\} = \frac{1}{N_i}\left(1.03 + 0.036\hat{\kappa}_i + 0.73\hat{\kappa}_i^2 + 2.12\hat{\kappa}_i^3\right)^2 \tag{28}$$

Inserting the estimated sampling variances Eqs. (26)-(28) in Eq. (24) the systematic regional variation of the PDS parameters can be calculated. The results are shown in Table 1 for precipitation depth and 10-minute rain intensity, respectively. In

Table 1 – Estimates of the systematic regional variation $\hat{\sigma}_\delta^2$, the prior mean value and the prior variance for the precipitation depth and the 10-minute rain intensity, respectively

Variable	Parameter	s^2	$\frac{1}{M}\sum_{i=1}^{M}\hat{\sigma}_{\varepsilon_i}^2$	$\hat{\sigma}_\delta^2$	Prior mean	Prior variance
Depth	λ	0.435	0.240	0.194	2.990	0.205
	α	3.474	3.336	0.137	8.586	0.222
	κ	0.0161	0.0302	0	−0.116	0.00074
Intensity	λ	0.479	0.263	0.216	3.282	0.228
	α	0.336	0.326	0.010	3.190	0.018
	κ	0.0231	0.0279	0	−0.031	0.00068

Table 2 – Test of regional homogeneity of the PDS parameters for the precipitation depth and the 10-minute rain intensity, respectively. Under the H_0-hypothesis the test statistic is distributed as $\chi^2(40)$

Variable	Test	Test statistic	Significance level
Depth	$\lambda_1=\lambda_2=\ldots=\lambda_{41}=\lambda$	68.8	0.004
	$\alpha_1=\alpha_2=\ldots=\alpha_{41}=\alpha$	81.0	<0.0005
	$\kappa_1=\kappa_2=\ldots=\kappa_{41}=\kappa$	20.0	0.996
	$\kappa_1=\kappa_2=\ldots=\kappa_{41}=0$	41.3	0.410
Intensity	$\lambda_1=\lambda_2=\ldots=\lambda_{41}=\lambda$	78.1	<0.0005
	$\alpha_1=\alpha_2=\ldots=\alpha_{41}=\alpha$	52.6	0.090
	$\kappa_1=\kappa_2=\ldots=\kappa_{41}=\kappa$	33.0	0.780
	$\kappa_1=\kappa_2=\ldots=\kappa_{41}=0$	35.7	0.660

both cases the sampling variance of the shape parameter κ more than accounts for the observed regional variability, and hence no systematic regional variation is likely to exist. For the mean of the exceedances α and the mean annual number of exceedances λ, however, the results indicate regional heterogeneity. In order to analyze this heterogeneity in more detail a statistical test was employed based on the Fisher dispersion test statistic (see Appendix B). The results of the test are shown in Table 2. For the precipitation depth the regional variation of the α and the λ parameter is significant, whereas only the variation of λ is significant for the 10-minute rain intensity. For the κ parameter the results are consistent with the indication of regional homogeneity given above. These results imply that the exceedance distribution at all sites can be described by a GED with a common shape parameter equal to the regional mean of κ. If this common shape parameter is not significantly different from zero the ED may be adequate. The results of testing $\kappa=0$ are shown in Table 2. It appears that for both precipitation depth and 10-

minute rain intensity the hypothesis of a common shape parameter equal to zero cannot be rejected. A thorough discussion of this interesting aspect is given below.

On applying the regional Bayesian approach to the PDS records prior information is expressed by means of regional distributions of the PDS parameters. Following the approach given in Appendix A the estimate of the variance in the prior distribution becomes

$$\hat{\sigma}_\theta^2 = \frac{M+1}{M}\hat{\sigma}_\delta^2 + \frac{1}{M^2}\sum_{i=1}^{M}\hat{\sigma}_{\varepsilon_i}^2 \tag{29}$$

In case of regional homogeneity ($\sigma_\delta{}^2=0$) the prior variance equals the sampling variance of the PDS parameter obtained from a regional pooling of the exceedances. Thus, if it is assumed that no regional variation is present, the Bayesian approach is consistent with a station year approach assuming a common parameter at all stations.

Using the method of moments estimators of the prior parameters become

$$\hat{\nu} = \frac{\hat{\mu}_\lambda^2}{\hat{\sigma}_\lambda^2} \quad , \quad \hat{\tau} = \frac{\hat{\mu}_\lambda}{\hat{\sigma}_\lambda^2} \tag{30}$$

$$\hat{\beta} = 2 + \frac{\hat{\mu}_\alpha^2}{\hat{\sigma}_\alpha^2} \quad , \quad \hat{\theta} = \hat{\mu}_\alpha\left(1 + \frac{\hat{\mu}_\alpha^2}{\hat{\sigma}_\alpha^2}\right) \tag{31}$$

$$\hat{\xi} = \frac{(\hat{\mu}_\kappa + \frac{1}{2})^2(\frac{1}{2} - \hat{\mu}_\kappa)}{\hat{\sigma}_\kappa^2} - \hat{\mu}_\kappa - \frac{1}{2} \quad , \quad \hat{\eta} = \frac{\hat{\xi}}{\hat{\mu}_\kappa + \frac{1}{2}} - \hat{\xi} \tag{32}$$

where $\hat{\mu}$ and $\hat{\sigma}^2$ are the estimated mean values and variances in the regional samples of λ, α and κ values. The estimated prior mean values and variances are shown in Tabel 1. Based on the estimated prior parameters the prior distribution of the T-year event can be calculated from Eqs. (15) and (17) in case of ED and GED exceedances, respectively. The prior distributions of the precipitation depth for T=2,10,50 years are shown in Fig. 3. In addition the mean values and the variances of the prior distributions are shown in Table 3. For lower quantile estimation the ED and the GED approach yields almost similar prior densities. For higher quantile estimation the inclusion of the additional parameter κ implies a prior distribution with a larger mean value and variance. Although the hypothesis of a common shape parameter equal to zero could not be rejected, the difference between the ED and the GED approach is seen to be increasing for large return periods. However, for the maximum 10-minute rain intensity the difference between the ED and the GED approach is less pronounced. This indicates (which is also consistent with the results given in Tables 1 and 2) that the distribution of the 10-minute rain intensity is closer to the ED than the distribution of the precipitation depth.

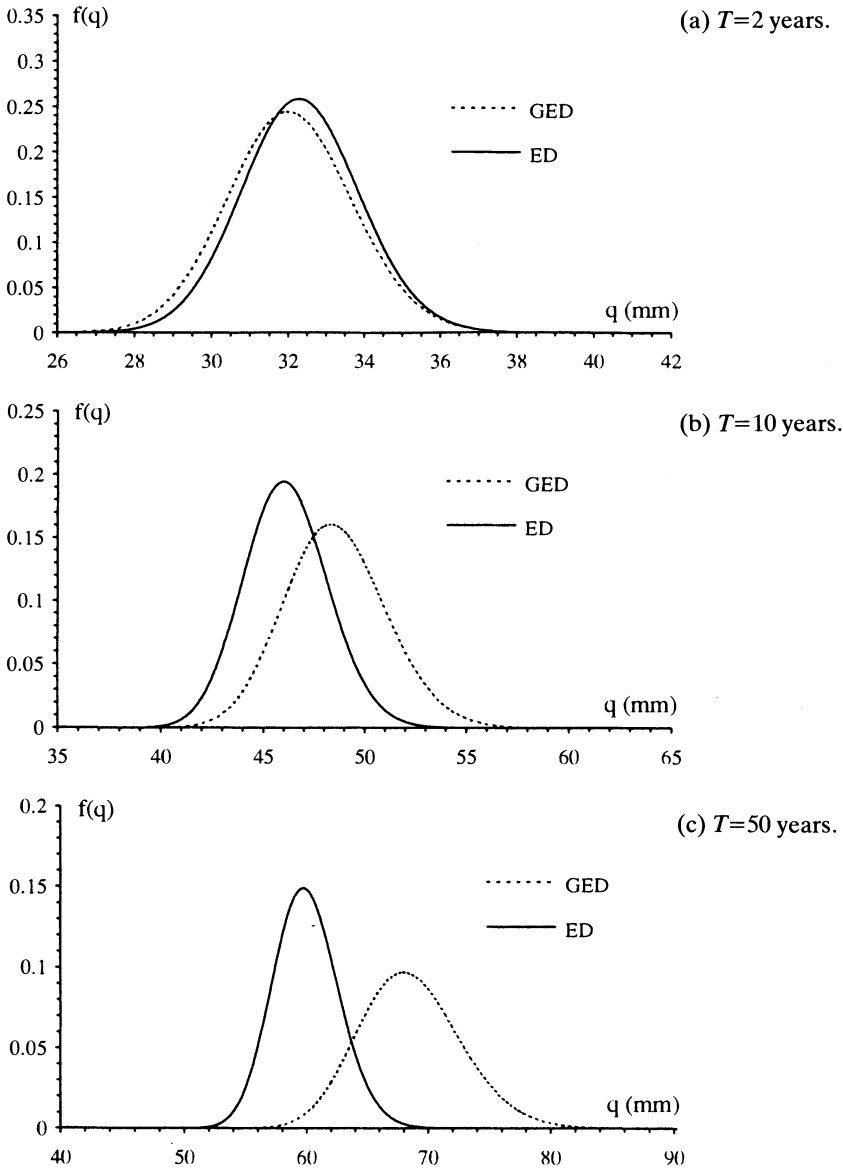


Fig. 3. Prior distribution of the T-year event for the precipitation depth ($q=q_0+x$) for ED and GED exceedances, respectively.

If measurements are available at the site in consideration a posterior distribution of the T-year event can be calculated. It should be noted that the application of Bayes' theorem demands prior knowledge and sample information to be independent. Hence, the prior information at the site being considered is obtained from the

Table 3 – Mean value $E\{x_T\}$ and variance $\mathrm{Var}\{x_T\}$ of the prior distributions and the posterior distributions at stations Nos. 20211 and 30315, respectively, for return periods T=2,10,50 years

T (years)		ED		GED	
		$E\{x_T\}$ (mm)	$\mathrm{Var}\{x_T\}$ (mm^2)	$E\{x_T\}$ (mm)	$\mathrm{Var}\{x_T\}$ (mm^2)
2	Prior	32.3	2.36	32.1	2.71
	st. 20211	32.5	1.35	32.3	1.59
	st. 30315	34.2	1.51	33.8	1.75
10	Prior	46.2	4.05	48.6	6.29
	st. 20211	46.1	2.56	48.8	4.19
	st. 30315	48.4	3.23	51.0	5.08
50	Prior	60.0	6.80	68.6	17.40
	st. 20211	59.8	4.50	68.8	13.86
	st. 30315	62.6	5.96	72.0	16.40

40 other stations in the region. The effect of combining site-specific and regional information using the ED and the GED approach, respectively, is shown in Figs. 4 and 5 for two different stations in the region. In addition the mean values and the variances of the posterior distributions are shown in Table 3. The sample information from the two stations are quite different. At station No. 20211 the ED gives a good fit, *cf.* Fig. 1a, whereas the exceedances at station No. 30315 have a significantly heavier tail than the ED, *cf.* Fig. 1b. The different nature of the exceedances has, however, no significant effect on the posterior distributions when comparing the ED and the GED approaches. The difference between the ED and the GED posterior distributions is only pronounced for higher quantile estimation, and this difference does not seem to depend on the at-site shape parameter κ. The fact that the prior information presumes regional homogeneity with respect to κ makes the updated distribution rather insensitive to the at-site shape parameter.

Generally, the variance in the updated distribution is reduced compared to both the site-specific distribution (using at-site data only) and the prior distribution. It is obvious that the improvement in precision of the updated T-year event depends on the amount of at-site data. This improvement by using the regional approach is especially pronounced if only few years of at-site data are available, but even when the whole at-site data series is used the inclusion of regional (prior) information significantly reduces the uncertainty of the T-year event. Note that the Bayesian analysis does not state a point estimator of x_T explicitly. A point estimator can, however, easily be obtained as the expected value or some appropriate quantile in the x_T-distribution. A brief discussion of optimal estimation is given in the last section of the paper.

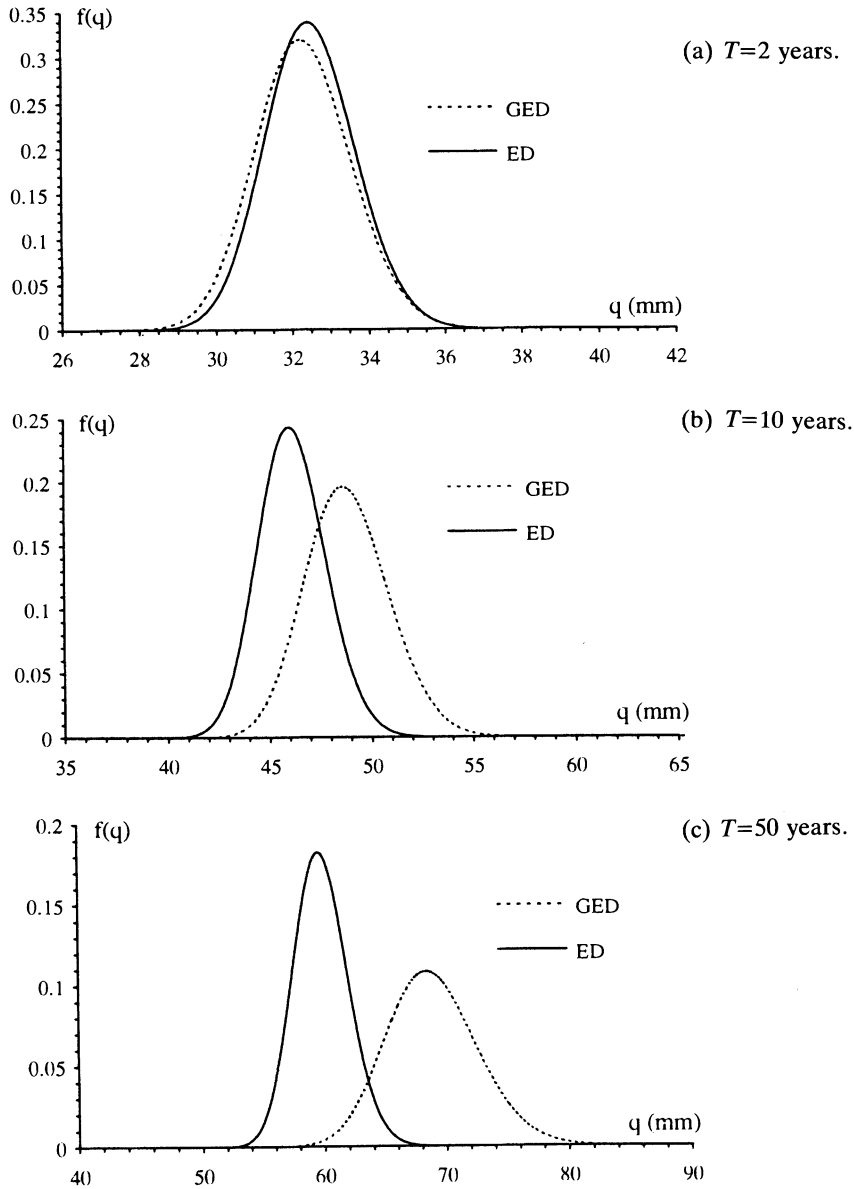


Fig. 4. Posterior distribution on the T-year event for the precipitation depth ($q=q_0+x$) at station No. 20211 for ED and GED exceedances, respectively.

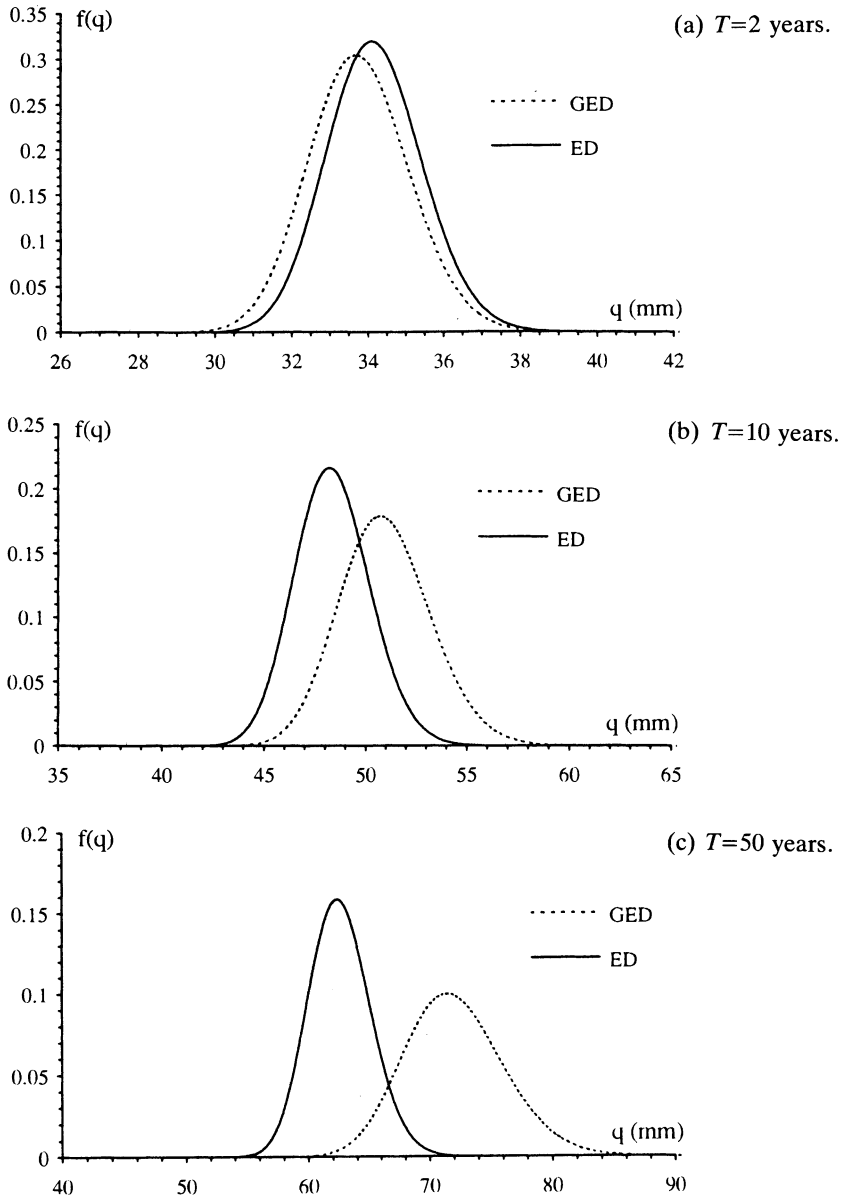


Fig. 5. Posterior distribution of the T-year event for the precipitation depth ($q=q_0+x$) at station No. 30315 for ED and GED exceedances, respectively.

Estimation of the Design Event

The outlined Bayesian principles quantify the uncertainty of the T-year event when all (known) relevant information is considered. When estimating the design event these uncertainties must be taken into account. In addition the design engineer should include economic perspectives in order to obtain optimal decisions. The Bayesian theory yields an optimal decision rule in which the uncertainty of the T-year event and the economic loss are considered jointly. Denoting by x^* the decision rule and by x the true T-year event, the total economic loss can be expressed as

$$L(x^*,x) = C(x^*) + D(x^*,x) \tag{33}$$

where $C(\cdot)$ denotes the construction costs, and $D(\cdot,\cdot)$ denotes loss due to damage from floods and pollution. The optimal decision rule minimizes the total expected loss, *i.e.*

$$x_T^{OPT} = \min_{x^*}\left(E\{L(x^*,x)\,|x^*\}\right) = \min_{x^*}\left(\int L(x^*,x)\, f_{x_t}(x)\, dx\right) \tag{34}$$

Although Eq. (34) seems rather simple, the decision rule may be difficult to apply in practice especially due to problems of quantifying the damages. Instead a qualitative method for evaluation of Eq. (34) may be appropriate. Assuming the loss function to be linear, *i.e.*

$$L(x^*,x) = \begin{cases} (x^*-x) & , \quad x \leq x^* \\ b(x-x^*) & , \quad x > x^* \end{cases} \tag{35}$$

where $b>0$, the optimal decision rule is the $b/(b+1)$-quantile in the distribution of the T-year event. The loss function simply states that it is b times more expensive to make an underdesign than an overdesign of the same magnitude. If b can be determined by engineering judgement, the design engineer has a practicable tool to make optimal designs of sewer systems.

Conclusions

The PDS model is employed in the analysis of extreme rainfalls from a country-wide system of raingauges in Denmark. The method is applied to two variables: the total precipitation depth and the maximum 10-minute rain intensity of individual storms.

In the at-site analysis the basic model assumptions are examined, and it is concluded that the ED assumption for the exceedances implies a serious lack of fit for large return periods due to outliers. At these stations the GED is appropriate for modelling the exceedances although it is not perfect.

Based on the estimated at-site parameters from the 41 stations a regional analysis is carried out. The previous assumption of spatial homogeneity of extreme rainfalls in Denmark is examined using the Fisher dispersion test statistics. For the precipitation depth the regional variation of the scale parameter α and the intensity parameter λ is significant, whereas only the variation of λ is significant for the 10-minute rain intensity. For both variables the test indicates regional homogeneity with respect to the shape parameter $\varkappa$ implying the exceedance distribution at all stations to be described by a GED with a common shape parameter equal to the regional mean value of $\varkappa$.

In order to make inferences possible at non-monitored sites and to improve at-site estimates a regional model based on Bayesian decision theory is introduced. The prior information is expressed by means of regional empirical distributions of the PDS-parameters. The Bayesian approach is applied for both ED and GED exceedances. For lower quantile estimation the two approaches are almost similar, whereas the GED T-year event distribution has a larger mean value and variance for higher quantile estimation. Although the hypothesis of a common shape parameter equal to zero (implying ED exceedances at all stations) could not be rejected, the difference between the two approaches is pronounced for large return periods. Thus, in order to take into account the prior information from outliers in extreme rainfalls the GED approach must be applied. The ED approach is, however, appealing because of the rather simple updating procedure, and hence it may be preferable for lower quantile estimation, say for $T<10$ years.

Finally a method is illustrated for estimating the design level where both the uncertainty of the T-year event and the economic loss are simultaneously taken into account. If the loss function is difficult to quantify, a qualitative procedure based on engineering judgement can be used.

References

Arnbjerg-Nielsen, K., Spliid, H., and Harremoës, P. (1994) Non-parametric statistics on extreme rainfalls, *Nordic Hydrology, Vol. 25(4),* pp. 267-278.

Cunnane, C. (1979) A note on the Poisson assumption in partial duration series models, *Water Resour. Res., Vol. 15(2),* pp. 489-494.

Danish Water Pollution Control Committee (1974) Determination of intensity-duration-frequency relationships (in Danish), Paper No. 16, Teknisk Forlag, Copenhagen, Denmark.

Fitzgerald, D. L. (1989) Single station and regional analysis of daily rainfall extremes, *Stochastic Hydrol. Hydraul., Vol. 3,* pp. 281-292.

Lilliefors, H. W. (1969) On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown, *J. Amer. Stat. Assoc., Vol. 64,* pp. 387-389.

Pickands, J. (1975) Statistical inference using extreme order statistics, *Ann. Stat., Vol. 3,* pp. 119-131.

Rasmussen, P. F., and Rosbjerg, D. (1989) Risk estimation in partial duration series, *Water Resour. Res., Vol. 25(11),* pp. 2319-2330.

Rasmussen, P. F., and Rosbjerg, D. (1991) Application of Bayesian principles in regional flood frequency estimation, in G. Tsakiris (ed.), *Advances in Water Resources Technology,* Balkema, 65-75.

Rosbjerg, D. (1977) Return periods of hydrological events, *Nordic Hydrology, Vol. 8,* pp. 57-61.

Rosbjerg, D. (1985) Estimation in partial duration series with independent and dependent peak values, *J. Hydrol., Vol. 76,* pp. 183-195.

Rosbjerg, D., and Madsen, H. (1992) On the choice of threshold level in partial duration series, Nordic Hydrological Conference, Alta, NHP-report No. 30, 604-615.

Rosbjerg, D., Madsen, H., and Rasmussen, P. F. (1992) Prediction in partial duration series with generalized Pareto-distributed exceedances, *Water Resour. Res., Vol. 28(11),* pp. 3001-3010.

Rouselle, J., and Hindie, F. (1976) Incertitude dans les debits de crues: Approche Bayesienne, *J. Hydrol., Vol. 30,* pp. 341-349.

Shane, R., and Lynn, W. R. (1964) Mathematical model for flood risk analysis, *J. Hydraul. Div. ASCE, Vol. 90,* pp. 1-20.

Tasker, G. D., and Stedinger, J. R. (1986) Regional skew with weighted least squares regression, *J. Water Resour. Plan. Manag., ASCE, Vol. 112(2),* pp. 225-237.

Todorovic, P., and Zelenhasic, E. (1970) A stochastic model for flood analysis, *Water Resour. Res., Vol. 6(6),* pp. 1641-1648.

Van Montfort, M. A. J., and Witter, J. V. (1985) Testing exponentiality against generalised Pareto distribution, *J. Hydrol., Vol. 78,* pp. 305-315.

Van Montfort, M. A. J., and Witter, J. V. (1986) The generalized Pareto distribution applied to rainfall depths, *Hydrol. Sci. J., Vol. 31,* pp. 151-162.

First received: 9 August, 1993
Revised version received: 11 January, 1994
Accepted: 22 March, 1994

Appendix A: Determination of the Prior Variation

Let $\hat{\theta}_i$ be an unbiased estimator of θ_i in a region of M sites. The at-site estimators are subjected to random sampling errors, *i.e.*

$$\hat{\theta}_i = \theta_i + \varepsilon_i \quad , \quad i = 1,2,\ldots,M \tag{A1}$$

where

$$E\{\varepsilon_i\} = 0 \quad , \quad \mathrm{Cov}\{\varepsilon_i,\varepsilon_j\} = \begin{cases} \sigma^2_{\varepsilon_i} & , \quad i = j \\ \sigma_{\varepsilon_i}\sigma_{\varepsilon_j}\rho_{ij} & , \quad i \neq j \end{cases} \tag{A2}$$

In Eq. (A2) ϱ_{ij} is the correlation coefficient between $\hat{\theta}_i$ and $\hat{\theta}_j$. In addition, the true values θ_i may be governed by a systematic deviation δ_i from the true regional mean value θ, *i.e.*

$$\theta_i = \theta + \delta_i \quad , \quad i = 1,2\ldots,M \tag{A3}$$

where

$$E\{\delta_i\} = 0 \quad , \quad \mathrm{Cov}\{\delta_i,\delta_j\} = \begin{cases} \sigma^2_\delta , & i = j \\ 0 , & i \neq j \end{cases} \tag{A4}$$

Combining Eqs. (A1) and (A3) gives

$$\hat{\theta}_i = \theta + \delta_i + \varepsilon_i \tag{A5}$$

Assuming ε_i and δ_i to be independent stochastic variables the covariance between $\hat{\theta}_i$ and $\hat{\theta}_j$ becomes

$$\mathrm{Cov}\{\hat{\theta}_i,\hat{\theta}_j\} = \begin{cases} \sigma^2_{\varepsilon_i} + \sigma^2_\delta & , \quad i = j \\ \sigma_{\varepsilon_i}\sigma_{\varepsilon_j}\rho_{ij} , & i \neq j \end{cases} \tag{A6}$$

The sampling variance of the at-site estimator $\hat{\theta}_i$ and the variance of the systematic regional variation are unknown and must be estimated from the observed data. The former is usually obtained using asymptotic theory or Monte Carlo simulations. The latter is estimated from the observed regional variance s^2 given by

$$s^2 = \frac{1}{M-1}\sum_{i=1}^{M}(\hat{\theta}_i - \hat{\theta})^2 \tag{A7}$$

where

$$\hat{\theta} = \frac{1}{M}\sum_{i=1}^{M}\hat{\theta}_i \tag{A8}$$

is the estimated regional mean value. Eq. (A7) may be written

$$(M-1)s^2 = \sum_{i=1}^{M}(\hat{\theta}_i - \hat{\theta})^2 = \sum_{i=1}^{M}\hat{\theta}_i^2 - M\hat{\theta}^2 \tag{A9}$$

Taking expected values of the two terms on the right side of Eq. (A9) leads to

$$E\left\{\sum_{i=1}^{M}\hat{\theta}_i^2\right\} = M\theta^2 + \sum_{i=1}^{M}\sigma^2_{\varepsilon_i} + M\sigma^2_\delta \tag{A10}$$

and

$$E\{M\hat{\theta}\} = M(\mathrm{Var}\{\hat{\theta}\} + (E\{\hat{\theta}\})^2) = M(\mathrm{Var}\{\hat{\theta}\} + \theta^2) \tag{A11}$$

where

$$\mathrm{Var}\{\hat{\theta}\} = \mathrm{Var}\left\{\frac{1}{M}\sum_{i=1}^{M}\hat{\theta}_i\right\} = \frac{1}{M^2}\left\{M\sigma_\delta^2 + \sum_{i=1}^{M}\sigma_{\varepsilon_i}^2 + 2\sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\sigma_{\varepsilon_i}\sigma_{\varepsilon_j}\rho_{ij}\right. \tag{A12}$$

Inserting Eqs. (A10)-(A12) in Eq. (A9) leads to the following expression

$$E\{(M-1)s^2\} = (M-1)\sigma_\delta^2 + \frac{M-1}{M}\sum_{i=1}^{M}\sigma_{\varepsilon_i}^2 - \frac{2}{M}\sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\sigma_{\varepsilon_i}\sigma_{\varepsilon_j}\rho_{ij} \tag{A13}$$

Thus an estimator of the variance of the systematic regional variation becomes

$$\hat{\sigma}_\delta^2 = s^2 - \frac{1}{M}\sum_{i=1}^{M}\hat{\sigma}_{\varepsilon_i}^2 + \frac{2}{M(M-1)}\sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\hat{\sigma}_{\varepsilon_i}\hat{\sigma}_{\varepsilon_j}\hat{\rho}_{ij} \tag{A14}$$

If no intersite correlation is present, *i.e.* $\varrho_{ij}=0$, Eq. (A14) is equivalent to the estimator of the model error variance of the skewness given by Tasker and Stedinger (1986). If all stations have almost equal sampling variances, Eq. (A14) may be written

$$\hat{\sigma}_\delta^2 \approx s^2 - \frac{1}{M}(1-\bar{\rho})\sum_{i=1}^{M}\hat{\sigma}_{\varepsilon_i}^2 \tag{A15}$$

where

$$\bar{\rho} = \frac{2}{M(M-1)}\sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\hat{\rho}_{ij} \tag{A16}$$

is the average interstation correlation coefficient. When applying Eqs. (A14) or (A15) one may find a negative value of σ_δ^2. In these instances the sampling variances more than account for the observed regional variability and no systematic deviations from the regional mean value is likely to be present, *i.e.* $\sigma_\delta^2=0$.

The mean and the variance of the prior distribution of θ are evaluated by considering the statistical properties of $\hat{\theta}_i$ at an arbitrary location in the region where no measurements are available. The estimator of θ_i reads

$$\hat{\theta}_i = \hat{\theta} + \delta_i \tag{A17}$$

The estimated prior mean and variance become

$$\hat{\mu}_\theta = \hat{\theta}\ , \quad \hat{\sigma}_\theta^2 = \hat{\mathrm{var}}\{\hat{\theta}\} + \hat{\sigma}_\delta^2 \tag{A18}$$

Inserting Eqs. (A12) and (A15) in Eq. (A18) leads to the following expression of the estimated prior variance

$$\hat{\sigma}_\theta^2 \approx \frac{M+1}{M}\hat{\sigma}_\delta^2 + \frac{1}{M^2}\left(1+\bar{\rho}(M-1)\right)\sum_{i=1}^{M}\hat{\sigma}_{\varepsilon_i}^2$$

$$= \begin{cases} \frac{M+1}{M}s^2 + \frac{1}{M}(2\bar{\rho}-1)\sum_{i=1}^{M}\hat{\sigma}_{\varepsilon_i}^2\ , & \text{for } \sigma_\delta^2 > 0 \\ \frac{1}{M^2}\left(1+\bar{\rho}(M-1)\right)\sum_{i=1}^{M}\hat{\sigma}_{\varepsilon_i}^2\ , & \text{for } \sigma_\delta^2 = 0 \end{cases} \tag{A19}$$

For $\bar{\varrho}=0$ (A19) reads

$$\hat{\sigma}_\theta^2 = \begin{cases} \frac{M+1}{M} s^2 - \frac{1}{M} \sum_{i=1}^{M} \hat{\sigma}_{\varepsilon_i}^2 & , \quad \text{for} \quad \sigma_\delta^2 > 0 \\ \frac{1}{M^2} \sum_{i=1}^{M} \hat{\sigma}_{\varepsilon_i}^2 & , \quad \text{for} \quad \hat{\sigma}_\delta^2 = 0 \end{cases} \tag{A20}$$

Appendix B: Test of Regional Variation

Consider the null hypothesis H_0 of no regional variation between θ_i-values, *i.e.*

$$H_0: \quad \theta_1 = \theta_2 = \ldots = \theta_M = \theta \tag{B1}$$

against the alternative

$$H_1: \quad \exists \quad (i,j): \quad \theta_i \neq \theta_j \tag{B2}$$

Under the null hypothesis the variance of $\hat{\theta}_i$ is σ_ε^2. Assuming that $\hat{\theta}_i$ is approximately normally distributed and that no intersite correlation is present then the Fisher dispersion test statistics

$$K = \sum_{i=1}^{M} \left(\frac{\hat{\theta}_i - \hat{\theta}}{\sigma_{\varepsilon_i}} \right)^2 , \qquad \hat{\theta} = \frac{1}{M} \sum_{i=1}^{M} \hat{\theta}_i \tag{B3}$$

is approximately distributed as $\chi^2(M\text{-}1)$. The hypothesis of no regional variation is rejected for large values of K implying that the regional variance is significantly greater than the sampling variance which would be expected in case of regional homogeneity.

Address:
D. Rosbjerg and H. Madsen,
Institute of Hydrodynamics and
Hydraulic Engineering, ISVA,
Technical University of Denmark,
Building 115,
DK-2800 Lyngby, Denmark.

P. Harremoës,
Inst. of Environmental Science and Engineering,
Technical University of Denmark,
Building 115,
DK-2800 Lyngby, Denmark.