Nordic Hydrology, 20, 1989, 201-214

Regional Exceedance Probabilities

Paper presented at ICWRS-Workshop on Risk
and Uncertainty in Hydrologic Design
Oslo, Norway, February – 1989

Lars Gottschalk
Dept. of Geophysics, University of Oslo, Norway

Construction of a regional flood frequency curve is based, as a rule, on fitting this curve to representative quantiles. In a regional sample of floods the probability of extreme values corresponding to return periods, that exceed the record lengths, is much larger than that of individual series, used to determine the representative quantiles. The probabilities of exceedance of regional extremes can be calculated straightforward in case of independent data, applying the theory of order statistics. For regionally dependent data one can define an equivalent number of independent regional series and then utilize the theory for independent data. This approach is exemplified with flood data from Norway.

Introduction

A common approach to the construction of a regional frequency curve of floods is the one outlined by Dalrymple (1960), the so-called flood index method. It contains the following steps:

1) Normalization of each station flood by dividing each flood by some location estimate for that record (mean, median).
2) Rank ordering of each normalized record.
3) Estimation of a representative quantile (say, by taking a mean or median of the normalized floods by the same rank) for each rank.
4) Fit of a distribution function to the representative quantiles.

The approach assumes homogeneity with respect to flood frequency characteristics within a region. It can be expected that extreme values with high return periods are not local phenomena but have a regional scale, *i.e.* the local features of a certain basin are not too important and the regional pattern of antecedent conditions is of a relatively minor influence.

Dalrymple states that combining records regionally by the station year method is not applicable in case of flood data. It would demand that the regional floods besides being taken from a regional homogeneous population are also spatially independent. The approach has therefore been to average all regional ranked records, rather than to add short records to produce a long-term record.

In a regional sample of floods the probability of the occurrence of extreme values corresponding to return periods that widely exceed the record lengths, is much larger than that of individual series. It is therefore of interest to calculate the probabilities of exceedance of such regional extremes, and thereby utilize the regional information on extreme floods. In case of independent data the problem has a straightforward solution resulting from the theory of order statistics. For regionally dependent data the problem has been reduced in this paper to the one of finding an equivalent number of independent regional series and then applying the theory of independent data. The approach is exemplified with flood data from Norway.

Regional Probability of Exceedance in Case of Independence

In most cases the forecast of extreme values will be based on the knowledge of the parent distribution. If we deal only with their frequencies and not their sizes, the procedures are distribution free, *i.e.* they are valid for all continuous distributions (Gumbel 1958). A statistical procedure which is insensitive to departures from the underlying assumptions is called robust, a term introduced by Box (1953). In case of distribution free methods the robustness is precise and assured.

The distribution of exceedances follows the Polya distribution function (Moran 1968). Suppose that $x_1,\ldots,x_n,y_1,\ldots,y_N$ are independent random variables all with the same parent distribution function, $F(x)$, which is a continuous function. Let x_i be the m-th largest of the set $x_1,\ldots,x_n$ and p_j; $j=0,1,\ldots,N$ be the probability that j of the $y_1,\ldots,y_N$ are larger than x_i. It can then be shown that

$$p_j = \frac{m \begin{bmatrix} n \\ m \end{bmatrix} \begin{bmatrix} N \\ j \end{bmatrix}}{(n+N) \begin{bmatrix} N+n-1 \\ m+j-1 \end{bmatrix}} \; ; \; j = 0, 1, \ldots, N \qquad (1)$$

This probability is independent of the distribution function $F(x)$ as long as the latter is continuous. Clearly what matters are not the actual values of the x_i and the y_j, but their relative order. Let us consider the particular case where $N=1$, $m=1$. The problem is then equivalent to finding the probability that x_{n+1} is the largest in a sample $x_1,\ldots, x_{n+1}$ of independent random variables from some continuous distribution, or in another formulation estimating the probability that any future flow, say x_{n+1}, exceeds max $(x_1,\ldots,x_n)$. Inserting the specific values for N and m we find

$$p\{x_{n+1}>\max(x_1,\ldots,x_n)\} = \frac{1}{n+1} \tag{2}$$

which is a well known expression in hydrology.

Let us now generalize to the situation of having n sets of data each containing p elements $\{x_{ij};\ i=1,\ldots,n;\ j=1,\ldots,p\}$ (n years of data at p sites in a region). Independence is assumed. We then take out the largest value among the p available elements for each i, $x_{i(1)}$; $i=1,\ldots,n$. The $x_{i(1)}$ are ordered from the largest to the smallest $x_{(i)(1)}$; $i=1,\ldots,n$.

As independence is assumed we find immediately that the probability of exceedance of the largest value $x_{(1)(1)}$ among totally np elements is

$$p\{x>\max(x_{11},\ldots,x_{1n},x_{21},\ldots,x_{2n},x_{31},\ldots,x_{pn})\} \equiv p_1 = \frac{1}{np+1} \tag{3}$$

The probability of exceedance of the second largest value is the sum of the probability for two events, namely that x is larger than $x_{(1)(1)}$ and the complementary event that x is less than $x_{(1)(1)}$ but at the same time larger than the $n(p\text{-}1)$ elements of data when the p elements belonging to the set $x_{(1)(j)}$; $j=1,\ldots, p$ are excluded. We thus find

$$p\{x>\max(x_{21},\ldots,x_{2n},x_{31},\ldots,x_{pn})\} = p_2 = \frac{1}{np+1} + \frac{np}{np+1}\,\frac{1}{n(p-1)+1} \tag{4}$$

The same principle can now be repeated to find successive probabilities of exceedance of $x_{(i)(1)}$. For the i-th value p_i the following expression is valid

$$p_i \equiv \frac{1}{np+1} + \frac{np}{np+1}\,\frac{1}{n(p-1)+1} + \ldots + \frac{np}{np+1}\,\frac{n(p-1)}{n(p-1)+1} \cdots$$
$$\cdots \frac{n(p-i+1)}{n(p-i+1)+1}\,\frac{n(p-i)}{n(p-i)+1}\;; \qquad i = 1,\ldots,n \tag{5}$$

In the following we will call the probabilities p_i for regional exceedance probabilities related to the ordered series of regional maximum values $x_{(i)(1)}$; $i=1,\ldots,n$. An expression similar to Eq. (5) is quoted by Conover and Benson (1962) without giving a reference.

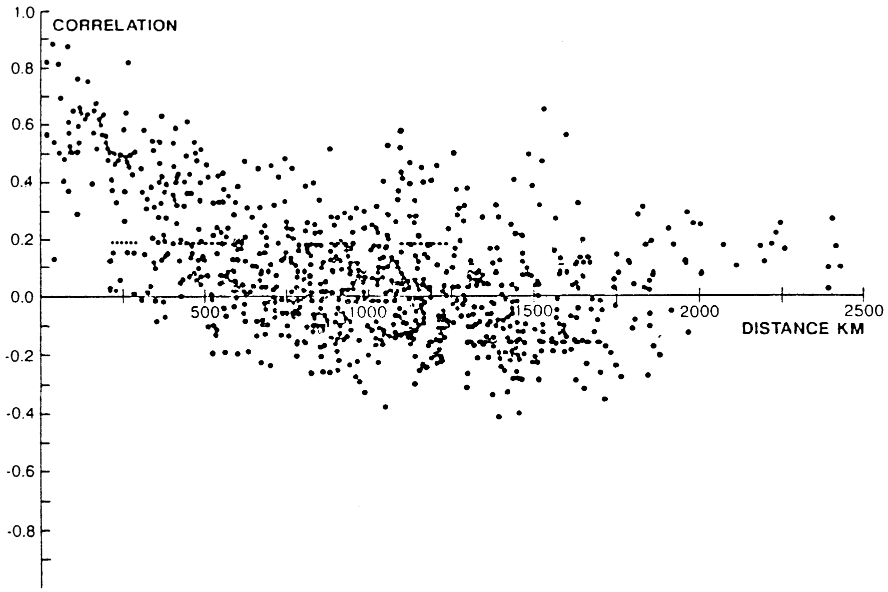

Fig. 1. Spatial correlation coefficients between pairwise observation stations as a function of the distance between them for the whole of Norway.

Table 1 – Estimated mean regional correlation coefficients.

	$\bar{\varrho}$	$\bar{\varrho}x_{10}y_{10}$	$\bar{\varrho}x_{20}y_{20}$	$\bar{\varrho}x_{100}y_{100}$	$\bar{\varrho}x_{1000}y_{1000}$
All Norway					
Normal	0.094	0.079	0.075	0.070	0.067
Gamma	0.094	0.079	0.073	0.072	0.069
Lognormal	0.094	0.074	0.070	0.066	0.063
Region 1					
Normal	0.325	0.246	0.224	0.197	0.180
Gamma	0.325	0.246	0.212	0.206	0.190
Lognormal	0.325	0.228	0.210	0.191	0.179
Region 2					
Normal	0.274	0.206	0.188	0.164	0.150
Gamma	0.274	0.206	0.177	0.171	0.158
Lognormal	0.274	0.191	0.176	0.159	0.148
Region 3					
Normal	0.436	0.338	0.311	0.278	0.257
Gamma	0.436	0.338	0.296	0.288	0.269
Lognormal	0.436	0.315	0.293	0.268	0.254
Region 4					
Normal	0.393	0.299	0.273	0.240	0.220
Gamma	0.393	0.299	0.258	0.251	0.232
Lognormal	0.393	0.277	0.256	0.233	0.219

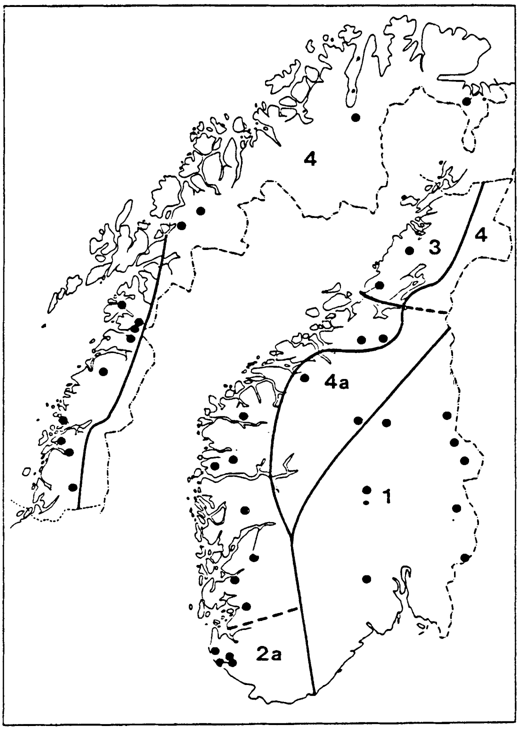

Fig. 2. Hydrological regions of Norway for annual rain floods.

Spatial Correlation of Extreme Values

Flood series are mutually correlated in space. Spatial correlation coefficients between pairwise observation stations as a function of the distance between them for rain floods are shown in Fig. 1. The correlation coefficients have been calculated for the whole of Norway as well as for four hydrological subregions (Fig. 2). The average correlation coefficients for the different data sets are given in Table 1.

The fact that two series, say x and y, observed at two different sites some distance apart are correlated also makes the statistical moments and quantiles to be correlated. It can easily be shown that the correlation between mean values $\rho_{m_x m_y}$ is equal to the correlation between the series themselves ρ

$$\rho_{m_x m_y} = \rho \quad (6)$$

In the same manner the correlation coefficients between other moments like variance s^2, standard deviation s, coefficient of variation C_v and coefficient of skewness C_s are found to be (Blochinov 1970)

$$\rho_{s_x^2, s_y^2} = \rho^2 \quad (7)$$

$$\rho_{s_x, s_y} = \rho^2 \tag{8}$$

$$\rho_{cv_x, cv_y} = \frac{\rho(\rho + 2C_{vx}C_{vy})}{[(1+2C_{vk}^2)(1+2C_{vy}^2)]^{\frac{1}{2}}} \tag{9}$$

$$\rho_{cs_x, cs_y} \equiv \rho^3 \tag{10}$$

Normality is assumed in all cases. For other distributions more complex expressions are derived.

Quantiles x_p and y_p are generally expressed in terms of frequency factors on the form

$$x_p = m_x + k_{px} s_x \quad \text{and} \quad y_p = m_Y + k_{py} s_y$$

where m_x and m_y are the estimated means and s_x and s_y – the estimated standard deviations of the series x and y, respectively. k_{px} and k_{py} are frequency factors related to the probabilities p_x and p_y. For any distribution of x and y (it is assumed that they belong to the same theoretical distribution) a relationship can be derived between the probabilities p_x and p_y and the frequency factors.

Applying a first order approximation to a two parameter distribution, the following expression is derived for the correlation $\varrho_{x_p y_p}$ between quantiles x_p and y_p

$$\rho_{x_p y_p} \equiv \frac{C(m_x, m_y) + k_{px}k_{py}C(s_x, s_y) + k_{px}C(s_x, m_y) + k_{py}C(m_x, s_y)}{\{[V(m_x) + k_{px}^2 V(s_x) + 2k_{px}C(m_x, s_x)][V(m_y) + k_{py}^2 V(s_y) + 2k_{py}C(m_y, s_y)]\}^{\frac{1}{2}}} \tag{11}$$

where $C()$ and $V()$ denote the covariance and variance, respectively. For the normal distribution we have (Kendall and Stuart 1977)

$$V(m) = \sigma^2/n \qquad V(s) \equiv \sigma^2/2n \qquad C(m_x, m_y) \equiv \rho\sigma_x\sigma_y/n$$

$$C(s_x, s_y) = \rho^2\sigma_x\sigma_y/2n$$

$$C(m_x, s_x) = C(m_y, s_y) = C(m_x, s_y) = C(m_y, s_x) = 0 \tag{12}$$

where σ^2_x and σ^2_y are the theoretical variances of x and y, respectively. Insertion into Eq. (11) yields

$$\rho_{x_p y_p} \equiv \rho\{1 + (\rho\lambda_{p_x}\lambda_{p_y})/2\} / \{[1+\lambda_{p_x}^2/2][1+\lambda_{p_y}^2/2]\}^{\frac{1}{2}} \tag{13}$$

where λ_{p_x} and λ_{p_y} are ordinates of the standard normal distribution related to the probabilities p_x and p_y respectively.

In case of lognormally distributed data the quantiles and frequency factors as well as variances and covariances for the logarithms of the variables are the same as in case of the normal distribution. To avoid computation of the moments of logarithms we can base the calculations on the moments of the original series. Applying the general equation for moments about the origin to the bivariate log-normal frequency distribution, we find for the mixed moment α_{rs}

$$\alpha_{rs} = \int_0^\infty\int_0^\infty \frac{x^r y^s}{2\pi xy\,\sigma_x\sigma_y\sqrt{1-\rho_n^2}} \exp\left\{-\frac{1}{2(1-\rho_n^2)}\left[\left(\frac{lnx-\mu_{nx}}{\sigma_{nx}}\right)^2 - \right.\right.$$

$$\left.\left. - 2\rho_n \frac{(lnx-\mu_{nx})(lnx-\mu_{ny})}{\sigma_{nx}\sigma_{ny}} + \left(\frac{lny-\mu_{ny}}{\sigma_{ny}}\right)^2\right]\right\}dx\,dy \qquad (14)$$

where $\mu_{nx},\mu_{ny},\sigma_{nx},\sigma_{ny},\varrho_n$ are parameters of the distribution. Using the substitutions

$$\gamma \equiv (lnx - r\sigma_{nx}^2 - \rho_n s\sigma_{nx}\sigma_{ny} - \mu_{nx})/\sigma_{nx}$$

$$\delta = (\ ny - s\sigma_{ny}^2 - \rho_n r\sigma_{nx}\sigma_{ny} - \mu_{ny})/\sigma_{ny}$$

Eq. (10) is rewritten as

$$\alpha_{rs} \equiv \exp[r\mu_{nx}+s\mu_{ny}+r^2\sigma_{nx}^2/2+s^2\sigma_{ny}^2/2+\rho_n rs\sigma_{nx}\sigma_{ny}]*$$

$$\int_\infty^\infty\int_\infty^\infty \frac{1}{2\pi\sqrt{1-\rho_n^2}} \left\{\frac{1}{2(1-\rho_n^2)}\left[\gamma^2-2\rho_n\gamma\,\delta+\delta^2\right]\right\} d\gamma\,d\delta$$

and since the latter double integral equals unity the equation for the mixed moment is

$$\alpha_{rs} = \exp[r\mu_{nx}+s\mu_{ny}+r^2\sigma_{nx}^2/2+s^2\sigma_{ny}^2/2+\rho_n rs\sigma_{nx}\sigma_{ny}] \qquad (15)$$

The central moments are derived by applying well known relations from probability theory and the variances and covariances are found accordingly. After some tedious calculations we find the following first order approximate expression for the correlation between lognormally distributed quantiles

$$\rho_{x_p y_p} \equiv \{\rho + \frac{1}{4}k_{px}k_{py}[\rho^4C_{vx}^4C_{vy}^4+4\rho^3C_{vx}^3C_{vy}^3+\rho^4C_{vx}^4C_{vy}^2+\rho^4C_{vx}^2C_{vy}^4+$$

$$+\ \rho^2(6+\rho^2)C_{vx}^2C_{vy}^2+4\rho^3C_{vx}^3C_{vy}+4\rho^3C_{vx}C_{vy}^3+4\rho(1+\rho^2)C_{vx}C_{vy}+$$

$$+\ 4\rho^2C_{vx}^2+4\rho^2C_{vy}^2+2\rho^2]+\frac{1}{2}k_{px}[\rho^2C_{vx}^2C_{vy}+2\rho C_{vx}+\rho^2C_{vy}]+$$

$$+\frac{1}{2}k_{py}[\rho^2C_{vx}C_{vy}^2+2\rho C_{vy}+\rho^2C_{vx}]\}/(\{1+\frac{1}{4}k_{px}^2[C_{vx}^8+6C_{vx}^6+$$

$$+ 15C_{vx}^4+16C_{vx}^2+2] + k_{px}[C_{vx}^3+3C_{vx}]\}\{1+\frac{1}{4}k_{py}^2[C_{vy}^8+6C_{vy}^6+$$

$$+ 15C_{vy}^4+16C_{vy}^2+2] + k_{py}[C_{vy}^3+3C_{vy}]\})^{\frac{1}{2}} \quad (16)$$

where $k_{p\cdot}$ is the frequency factor for the lognormal distribution (Kite 1977)

$$k_{p\cdot} \equiv \{\exp[(\ln(1+C_{v\cdot}^2))^{\frac{1}{2}}\lambda_{p\cdot}-\ln(1+C_{v\cdot}^2)/2]-1\}/C_{v\cdot} \quad (17)$$

Blochinov (1970) suggested the following formula for calculation of the correlation coefficient between gamma distributed quantiles $x_p=m_xk_{px}(C_{vx})$ and $y_p=m_yk_{py}(C_{vy})$

$$\rho_{x_p y_p} = \rho\left\{k_{px}k_{py}+\frac{1}{2}\rho\left(\frac{\partial k_{px}}{\partial C_{vx}}\right)\left(\frac{\partial k_{py}}{\partial C_{vy}}\right)\right\}/$$

$$\left\{\left[k_{px}^2+\frac{1}{2}\left(\frac{\partial k_{px}}{\partial C_{vx}}\right)^2\right]\left[k_{py}^2+\frac{1}{2}\left(\frac{\partial k_{py}}{\partial C_{vy}}\right)\right]\right\}^{\frac{1}{2}} \quad (18)$$

The formulas (13), (16) and (18) have been applied to calculate mean regional correlation between 1/10,1/20,1/100 and 1/1000 quantiles based on the correlation coefficients in Fig. 1. The results are shown in Table 1. It can be seen from the table that the coefficients are very much the same for the three distributions. The correlation decreases towards zero for more rare extremes, but the convergence is rather slow. To analyse the effect of the coefficient of variation, the calculations were repeated assuming constant C_v equal to 0.2, 0.5 and 1.0, respectively (Table 2). The values are almost equal for low and moderate C_v', while for high values of C_v the correlation coefficients are approaching zero quicker, especially for the lognormal distribution.

Table 2 – Estimated mean regional correlation coefficients in case of hypothetical constant coefficient of variation C_v

	$\bar{\varrho}$	$\bar{\varrho}x_{10}y_{10}$	$\bar{\varrho}x_{20}y_{20}$	$\bar{\varrho}x_{100}y_{100}$	$\bar{\varrho}x_{1000}y_{1000}$
All Norway					
Normal	0.094	0.079	0.075	0.070	0.067
Gamma					
C_v=0.2	0.094	0.080	0.074	0.073	0.070
C_v=0.5	0.094	0.078	0.072	0.071	0.068
C_v=1.0	0.094	0.077	0.071	0.070	0.067
Lognormal					
C_v=0.2	0.094	0.078	0.074	0.070	0.067
C_v=0.5	0.094	0.078	0.068	0.064	0.061
C_v=1.0	0.094	0.057	0.052	0.049	0.047

Equivalent Number of Independent Series

Let us consider a statistical parameter A_i (moment or quantile value) estimated from a set of observations x_{ij}; $j=1,\ldots,n$ at the observation site i. The theoretical variance of A_i is σ_{Ai}. Within a region there are totally p observation sites. We assume homogeneity, *i.e.* $E\{A_i\}=A$ and $E\{\sigma_{Ai}\}=\sigma_A$. The regional variance of A_i is then (Kritskij and Menkel 1970)

$$\mathrm{var}\{A_i\} \equiv \sigma_A^2 \frac{p-1}{p}(1-\bar{\rho}_A) \qquad (19)$$

where

$$\bar{\rho}_A \equiv \frac{2}{p(p-1)} \sum_{i<j} \rho_{A_{ij}}$$

is the mean regional correlation coefficient for the parameter A. Such mean regional correlation coefficients for Norwegian rain floods have been shown in Table 1. If the assumption of homogeneity is fulfilled then the p point records of sample size n can be pulled together to form a single record of pn observations. This method of analysis is called "the station year method", and its objective is to extract the most reliable information possible about statistical parameters of interest. In case of spatially independent data the variance of a regional mean

$$\bar{A} = \frac{1}{p}\sum_{i=1}^{p} A_i \qquad (20)$$

is equal to

$$V\{\bar{A}\} \equiv \sigma_A^2 = \frac{\sigma_A^2}{p} \qquad (21)$$

When the variables at the p observation sites are mutally correlated, this variance is

$$V\{\bar{A}\} = \frac{\sigma_A^2}{p}[1+(p-1)\bar{\rho}_A] \qquad (22)$$

A consequence of the correlation is a loss of information in the regional sample in comparison to the case of independent data. The relative information content I can be defined as the ratio of the theoretical variance $\sigma^2{}_o$ for a phenomenon and the variance σ^2 estimated from collected data. In our application the relative loss of information due to correlation in data samples in the estimation of the regional parameter A in terms of relative information content is $(1+(p\text{-}1)\bar{\varrho}_A)$, which follows when comparing Eqs. (21) and (22). Another way of expressing this is to define an effective number of independent observations p_e which has a relative information content equal to the one of the correlated sample. p_e is calculated as

$$p_e \doteq \frac{p}{1+(p-1)\bar{\rho}_A} \qquad (23)$$

We can apply this equation to the existing data set to calculate the effective numbers of observation series when A is set equal to a certain quantile in accordance with the values in Table 1. The result is shown in Table 3 where also the number of independent station years $p_e n$ for a region is given. The reduction of the number of series is rather drastic and is more sensitive to changes in $\bar{\varrho}_A$ than in p. The calculated values are rather stable for different probabilities p and also little sensitive to the underlying parent distribution for low and moderate values of C_v.

Regional Exceedance Probabilities in Case of Dependent Series

To calculate regional exceedance probabilities in case of dependent series it is suggested to use Eq. (5), when the actual number of series p is replaced by the effective number of independent series p_e. This number depends on the underlying parent distribution and the probability of exceedance (compare Table 3). The differences in p_e are, however, rather small for the three theoretical distributions investigated and these differences do not have any significant effect on the calculated exceedance probabilities.

Regional exceedance probabilities have been calculated for the Norwegian data on rain floods. The calculations have been performed on 6 sets of non-overlapping 10-year series, 3 sets of 20-year series, 2 sets of 30-year series and one set of 50-year series. The number of series varies from 157 for the 10-year series to 42 for the 60-year series. All series have been normalized by division of each series by its mean value. The results of the calculations are shown in terms of plots of the ordered series $x_{(i)(1)}$; $i=1,\ldots,n$ against the regional exceedance probabilities p_i; $i=1,\ldots,n$ (Fig. 3). For the shorter record lengths, 6,3 and 2 plots are shown in each diagram. The spread between the plots illustrates the uncertainty in the relations. Although data represent the whole of Norway the plotted curves are relatively stable. The largest value of the set of 60-year series, which is 6 times the mean value, has an exceedance probability of 1/704. The other data sets show lower exceedance probabilities as the number of independent station years is less.

The same type of analysis has also been performed on the data for the four subregions. Plots for the set of 60-year series are shown in Fig. 4. The discrepancies between the four plots are rather small also in this case, which should support the assumption of homogeneous data sets. The exceedance probabilities for the largest values of the regional extremes are, of course, lower in this case compared to when the whole of Norway is treated as one region. One of the regions contains a very large extreme value. One can suspect that such a value has an exceedance probability that is still smaller than the one estimated from the regional data. A possible test is to unite data into a larger regional sample, say for the whole of Norway. We see from the plot in Fig. 3 that it is the same value that is the largest for the whole of Norway which is now given a smaller probability of exceedance.

Table 3 – Estimated effective number of independent observation series p_e and station years $p_e n$ (in brackets) for 1000-year flood.

	Length of record (n) years			
All Norway	10	20	30	60
Number of series p	157	112	85	42
Normal	13.7(137)	13.3(265)	12.8(385)	11.2(673)
Gamma	13.3(133)	12.9(259)	12.5(375)	11.0(658)
Lognormal	14.5(145)	14.0(280)	13.5(405)	11.7(703)
Region 1				
p	37	24	17	9
Normal	4.9(49)	4.7(93)	4.4(131)	3.7(221)
Gamma	4.7(47)	4.5(89)	4.2(126)	3.6(214)
Lognormal	5.0(50)	4.7(94)	4.4(132)	3.7(222)
Region 2				
p	49	35	28	13
Normal	6.0(60)	5.7(115)	5.5(166)	4.6(279)
Gamma	5.7(57)	5.5(110)	5.3(159)	4.5(269)
Lognormal	6.0(60)	5.8(116)	5.6(168)	4.7(281)
Region 3				
p	52	37	29	14
Normal	3.7(37)	3.6(72)	3.5(106)	3.2(194)
Gamma	3.5(35)	3.5(69)	3.4(102)	3.1(187)
Lognormal	3.7(38)	3.6(73)	3.6(107)	3.3(195)
Region 4				
p	22	17	11	6
Normal	3.9(39)	3.8(75)	3.4(103)	2.9(171)
Gamma	3.7(37)	3.6(72)	3.3(99)	2.8(167)
Lognormal	3.9(39)	3.8(75)	3.4(103)	2.9(172)

Conclusions

Construction of a regional flood frequency curve is based, as a rule, on fitting this curve to representative quantiles. In a regional sample of floods the probability of extreme values corresponding to return periods, that exceed the record lengths, is much larger than that of individual series, used to determine the representative quantiles. In this paper it has been shown how this can be accounted for by calculating regional exceedance probabilities. These probabilities of exceedance of regional extremes can be calculated straightforward in case of independent data, applying the theory of order statistics. The statistical procedure utilized is insensitive to assumptions about parent distribution, *i.e.* robustness is precise and assured.

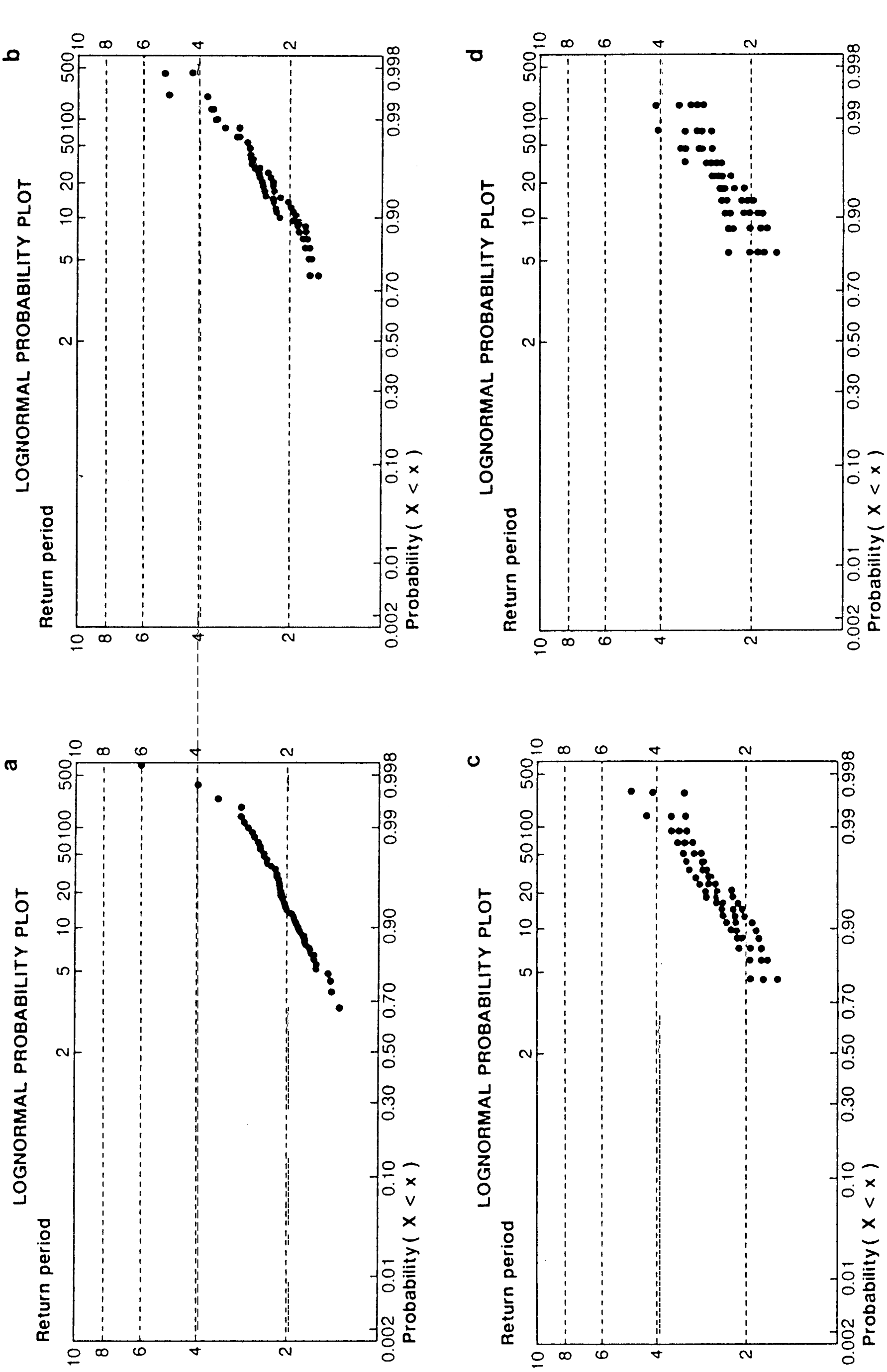

Fig. 3. Plots of estimated exceedance probabilities p_i against ordered regional series $x_{(i)}(1)$; $i=1,\ldots,n$ for all of Norway: a) 60-year series, $p=42$ b) 30-year series, $p=85$ c) 20-year series, $p=112$ d) 10-year series, $p=157$.

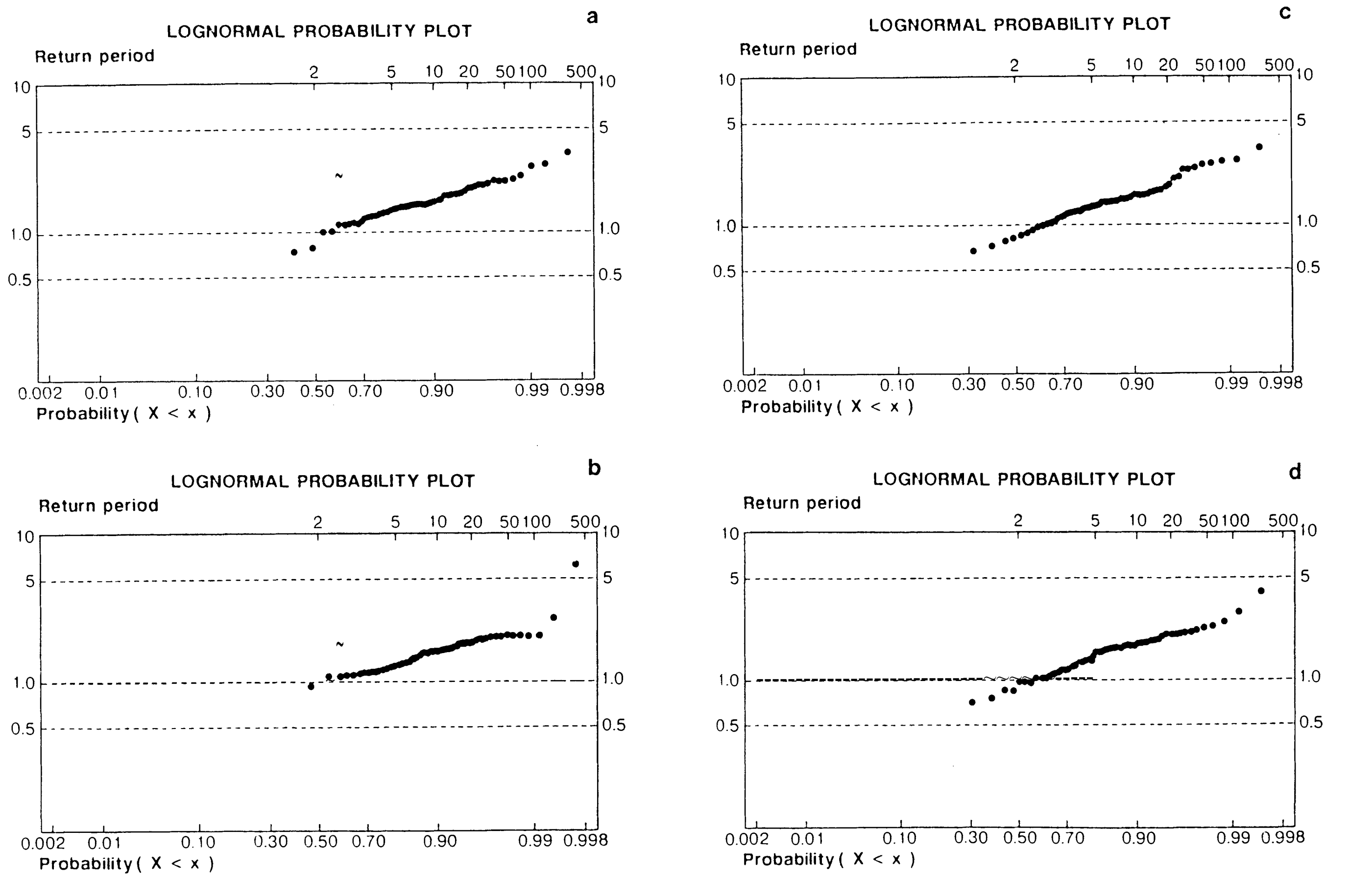

Fig. 4. Plots of estimated exceedance probabilites p_i against ordered regional series $x_{(i)(1)}$; $i=1,\ldots,n$ for subregions and 60-year series: a) region 1 b) region 2 c) region 3 d) region 4.

In case of regionally dependent data, which is the usual situation for floods, it is suggested to define an equivalent number of independent regional series and then utilize the theory for independent data. The estimation of the equivalent number of independent regional series involves regional correlation coefficients between quantiles. An assumption about parent distribution is necessary. It was shown, however, that in case of the normal, lognormal and gamma distributions the estimates are rather robust at least for small and moderate values of the coefficient of variation.

The presented approach assumes homogeneous regional samples. It can be expected that extreme values with high return periods are not local phenomena but have a regional scale, *i.e.* the local features of a certain basin are not too important and the regional pattern of antecedent conditions is of a relatively minor influence. This remains, however, to be further studied and verified. The methodology is also sensitive to outliers in the regional sample. A natural step must be to unite data into larger regional classes of data, if a regional sample contains a very large flood, like in case of the regional data from Norway exemplified in the text. This, on the other hand, can violate the homogeneity assumption. There is a need to define an »optimal« size of a regional sample in relation to the regional exceedance probability.

References

Blochinov, E.G. (1970) O svyazyakh mezhdu otsenkami parametrov raspredeleniya veroyatnosteij korrelativno zavisimykh velichin, *Trudy GGI, vyp. 180.*

Box, G.E.P. (1953) Non-normality and test on variances, *Biometrika, Vol. 40,* pp. 318-335.

Conover, W.J., and Benson, M.A. (1962) Long-term flood frequencies based on extremes of short term records, Geological Survey Research, Article 228 E159-160.

Dalrymple, T. (1960) Flood frequency analyses, USGS Water Supply Paper No 1543A pp. 1-47.

Gumbel, E.J. (1953) *Statistics of Extremes,* Columbia University Press.

Kendall, M., and Stuart, A. (1977) *The advanced theory of statistics,* Vol 1, fourth edition, Charles Griffin, London.

Kite, G.W. (1977) Frequency and risk analyses in hydrology, Water Resources Publications, Fort Collins, Colorado.

Kritskij, S.N., and Menkel, M.F. (1970) Ob otsenke parametrov raspredeleniya veroyatnostej po gruppam korrelativno svyazannykh gidrologicheskikh ryadov, *Trudy GGI, vyp, 180,* 30-49.

Moran, P.A.P. (1969) *An introduction to probability theory,* Clarendon Press. Oxford.

Received: 29 June, 1989

Address: Department of Geophysics,
University of Oslo, Box 1022 Blindern,
N-0315 Oslo 3, Norway.