Nordic Hydrology, 25, 1994, 267-278

Non-Parametric Statistics on Extreme Rainfall

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

H. Spliid
Inst. of Math. Modelling, Technical University of Denmark

This article focuses on rain as input data to problems related to urban storm drainage. The rain data originate from a monitoring program consisting of 56 gauges in Denmark. The gauges have observation periods ranging from 2 to 14 years. The gauges sample rain in a quantity of 0.2 mm with a resolution in time of 1 minute.

Two variables have been investigated: peak intensity and depth. Design values for return periods in the range from 0.1 to 2 years have been estimated for each gauge separately by means of the bootstrap resampling method. The estimation includes expected value, standard deviation and confidence intervals of the design value.

For large return periods the uncertainty of the estimates prevents a distinction by gauge between different statistical populations. However, for small return periods a test shows significant variation between gauges, *i.e.* the uncertainty of the estimates may not be assumed to be due to sampling variability only.

Introduction

During the last 25 years the legislation in Denmark has changed with regard to monitoring and controlling creeks, rivers, lakes and groundwater. The object is to ensure recreational amenities as well as drainage.

This has led to evaluation of the loadings of the small creeks and other sensitive recipients. The amount of water needing transportation out of an urban area increases rapidly when rainfall occurs. To avoid flooding overflows have been designed, often with outlets at small creeks sensitive to chock loadings of pollutants. Information on rainfall is crucial for the design of such outlets.

This paper focuses on rain as a random process. The data is examined statistically with focus on estimation of design values for a given return period. The return periods examined in this article focus on return periods much smaller than the observation period of the gauges, *i.e.* where estimation may take place by estimating directly from the empirical distribution function. Only two quantities of the rain events have been analyzed, the peak intensity and the total volume. The two variables represent two extreme cases of urban storm drainage. With small runoff time in the upper part of the catchment, peak intensity may cause flooding. In the lower part of the sewer system detention basins are of interest, and hence the total volume of rain contained in a rain event.

Data

The data come from a large monitoring program in Denmark. This program was started on January 1, 1979. In the present context data starting from this date till September 29, 1992 are investigated, yielding a total observation period of 13.74 years. Gauges have been situated at 70 locations. Of these, 61 gauges are active today. A total of 56 gauges have been active for more than 2 years. Only data from these gauges are considered in the present context. The position of the gauges is shown in Fig. 1.

All rain gauges are of the tipping bucket type and calibrated to sample rain in a quantity of 0.2 mm rain depth. Having accumulated this amount of precipitation, the gauge is emptied and a pulse is given. This pulse is registered with an accuracy of one minute.

Definition of a Rain Event

The pulses are converted to rain events, in accordance with the guidelines for civil engineering in Denmark (Spildevandskomitéen 1984). One rain event consists of at least two pulses. Consecutive pulses separated by more than 59 minutes are assigned to different rain events. By this definition the minimum rain depth registered will then be 0.4 mm, see also Harremoës and Henze (1981). An intensity is calculated for each minute of the rain event, *i.e.* the first minute is assigned the first pulse and all other minute intensities are calculated as the average intensity between two pulse registrations.

When considering urban storm drainage, the two main issues are flooding and pollution. Flooding mainly depends on peak intensity, whereas pollution of receiving waters mainly depends on the depth of the rain. Therefore only two variables, rain depth and maximum intensity, will be investigated. In the following we shall refer to these two quantities as the depth and the intensity.

The intensity is defined as the maximum average intensity calculated over 10

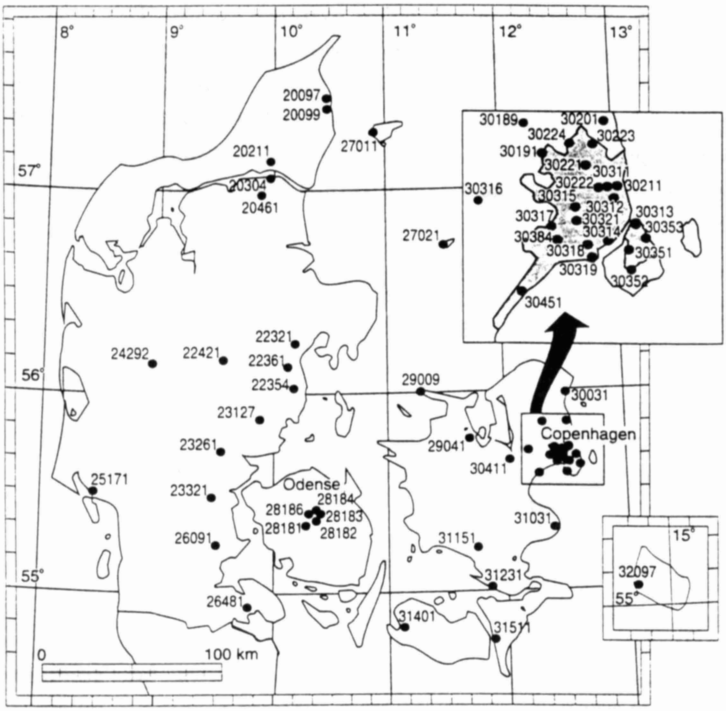

Fig. 1. Position of rain gauges.

minutes of the rain event. The depth is defined as the total volume of a rain event. Fig. 2(a) shows the empirical distribution function of depth for one of the gauges, 30031.

Preliminary Data Check

The data are modified in the following situations:

1) No data collection from gauge. The reason may be malfunction of the electrical system or Danish Institute of Meteorology (DMI) being unable to access the gauge by telephone.
2) The data collected are obviously wrong. An example is a sudden burst of pulses caused by the gauge being dirty where the rain suddenly breaks through the dirt.

The first case is accounted for by a reduction of the observation period and in the second case the rain event is excluded and the observation period is modified.

Observation Period

All the gauges have been not functioning for one or more time periods. These drop out periods may correspond to as much as 25% of the total observation period. Empirically it is known, that there is a seasonal variation over the year in the distribution of extreme events. The observation periods for each gauge have been weighted in accordance with the occurrence rates of extreme events. The weights have been calculated on a monthly basis for the two variables considered and for all return periods. The events used in the weighting are found by means of the unweighted observation period at each gauge. The unweighted observation period at gauge 30031 is 13.29 years and therefore this gauge contributes with the 133 largest events when calculating the weight for the 0.1-year return period and so forth. The weight is calculated as

$$\text{Weight}_{m,rp} = \frac{\sum^{m,g} \text{obsper}_{m,g}}{\sum^{g} \text{obsper}_{m,g}} \cdot \frac{\text{events}_{m,rp}}{\sum^{m} \text{events}_{m,rp}}$$

where $\text{Weight}_{m,rp}$ – Weight of month m for return period rp
$\text{obsper}_{m,g}$ – Unweighted observation period for month m at gauge g
$\text{events}_{m,rp}$ – Number of design events observed in month m at return period rp of variable

The weighted observation period, $W_{rp,g}$, for a return period rp at gauge g is then

$$W_{rp,g} = \sum^{m} \text{Weight}_{m,rp} \; \text{obsper}_{m,g}$$

Extreme events for both depth and intensity mainly occur in the late summer months. For large return periods, the occurrences are almost the same, whereas the occurrences for smaller return periods differ. It is observed that extreme events for depth are more uniformly distributed over the year than extreme events for intensity. Extreme intensities only occur in the months June, July, August and September with a few exceptions. The seasonal variation as such is not considered in the present study and hence is not modelled.

Quality Check

The quality check of the data was carried out at DMI. It was done by checking all rain events containing data with a one-minute intensity of 33 μm/s or more. The checking was done on the orginal database at DMI, *i.e.* before the rain events were extracted from the data.

The check was based on these three concepts: 1) weather-maps, 2) the shape of the rain events and finally 3) a comparison with other rain gauges positioned nearby.

On this basis, the minute intensity was either accepted or rejected. If one minute intensity is rejected, the entire rain events is rejected. Thus, correct information can be rejected. This is the case when the gauge has collected some dirt and the water suddenly breaks through the dirt. Such an event causes very high intensities for one or two minutes, but the remaining parts of the rain event are correct. The checking procedure led to an exclusion of 152 rain events out of a total of more than 100,000 events observed.

Statistical Methods

The basic idea of using statistics in urban storm drainage is to be able to assess the probability of extreme events to occur. The return periods used in design often vary form 0.1 to 2 years. This means, that the return periods of interest are much shorter than the observation periods at the gauges. For the gauges the estimates of a design value of a prescribed return period may be based on the calculation of a quantile in the observed rain events, *i.e.* no model of rainfall properties is assumed. If resampling methods are used the uncertainty of the estimates may also be assessed.

Test of Dependence and Trends in Data

The data have been tested for time-wise dependencies between observations within each gauge. The result of the tests implies, that there is no long-term trend in the data, *i.e.* the rain properties may be assumed time independent. No dependencies between consecutive rain events were found. This result is also valid for consecutive events exceeding the design value corresponding to a return period of 0.1 year. The testing is described in detail in Arnbjerg-Nielsen (1993).

Estimation of a Design Value of a Prescribed Return Period

We want to estimate quantiles in the true distribution function of the variable considered. From the true distribution it is possible to obtain the probability, p_X, of a rain event characterized by the variable to exceed some design value, x

$$p_x = 1 - P\{X \leq x\} = 1 - F(x) \tag{1}$$

where p_X – The probability of x being exceeded in one event
X – The random variable considered
x – The design value
$F(\cdot)$ – The distribution function for the random variable, X

The probability in Eq. (1) is converted to a return period, that is the mean time between exceedances of x. Following Rosbjerg (1977)

$$T = T(x) = \frac{1}{\kappa\, p_x} \tag{2}$$

where T – The return period corresponding to x
$\varkappa$ – The mean number of events per time period (year)

Throughout the paper reference will be made frequently to the design value corresponding to a return period as given in Eq. (2). Therefore the notation will be abbreviated. In the following "the design value of the variable depth corresponding to a return period of 1 year" is referred to as depth (1 year) and so forth.

Denote by $x_{(i)}$ the ith ordered observation and let $T_i = T(x_{(i)})$, where $x_{(i)}$ is the largest observation in the sample. T_i is calculated by substituting $\varkappa$ and p_x with their estimates. Various estimates for $\varkappa$ and p_x have been proposed. In this paper the estimates given by Rosbjerg (1977, 1988) are used

$$\hat{T}_i = \hat{T}(x_{(i)}) = \frac{1}{\hat{\kappa}\,\hat{p}_x}\ , \quad \hat{\kappa} = \frac{n}{W_{rp,g}}\ , \quad \hat{p}_x = \frac{i-0.3}{n+0.4} \tag{3}$$

where $\hat{\varkappa}$ – Estimate of $\varkappa$
$\hat{p}_x$ – Estimate of p_x
n – Total number of events observed at rain gauge
i – Rank of ith sorted observation

T_i is frequently calculated by the more simple formula

$$\tilde{T}_i = \tilde{T}(x_{(i)}) = \frac{1}{\tilde{p}_x}\ , \quad \tilde{p}_x = \frac{i}{\sum^m \mathrm{obsper}_{m,g}} \tag{4}$$

The numerical difference between the two statistics is negligible when the return period is less than 10% of the observation period. The main reason for choosing another estimator for $\hat{p}_x$ is that the objective of this investigation is to use a non-parametric approach and thus use distribution-free statistics. $\hat{p}_x$ is the median-based statistic both in the probability space and in the order statistics space and thus distribution free. Further discussions on this topic may be found in David (1981) and Rosbjerg (1988).

The variance of the estimate Eq. (3) may be computed in two different ways. One method is to assume a model for the rain process which generates the observed rain events. Based on this model computation of the distribution and the variance of the statistic may be performed. An example of this type of investigation on the same data can be found in Madsen *et al.* (1994). The observations are assigned a return period according to the ranks of the observations as given in Eq. (3). The design value corresponding to a prescribed return period is then estimated by linear interpolation between the two observations with return periods immediately higher and lower than the return period of interest. For example, at gauge 30031 a total of 2910 rain events have been recorded and the weighed observation period is 13.13 years. This implies that rank 13 is assigned the return period 1.034 years and rank 14 is assigned the return period 0.959 years. The estimate of depth (1 year) is then

$$\hat{\theta} = \widehat{\mathrm{depth}}(1\ \mathrm{year}) = \mathrm{depth}_{(13)} - \frac{1.034-1}{1.034-0.959}\,(\mathrm{depth}_{(13)} - \mathrm{depth}_{(14)}) \tag{5}$$

where $\widehat{\mathrm{depth}}$ (1 year) – Design value for sample
$\mathrm{depth}_{(i)}$ – Observed value of $\mathrm{depth}_{(i)}$, i.e. the value of the ith largest depth observed

The estimates of design values for intensity are found by using the same approach.

Using theory from order statistics exact confidence intervals can be constructed based on the binomial distribution (Efron 1982; David 1981). However, by this method central confidence intervals can only be constructed for certain levels of confidence depending on the sample. Briefly, the construction of such intervals is based on the computation of the probability that any observation can represent the quantile in question, considering the ordered sample. The endpoints of the interval will be two of the observations and the confidence level will only approximately equal to the desired level, depending on the actual sample size.

Alternatively, confidence intervals can be constructed using a resampling method. The intervals obtained by means of the bootstrap resampling method has been compared to intervals obtained using the binomial distribution in Efron (1982). He concludes that for confidence levels not very close to 1 the bootstrap yields the same confidence level when using a confidence interval constructed from the binomial method. In addition to this the bootstrap method can be used to construct confidence intervals with other confidence levels than possible when using the binomial distribution.

Bootstrap Resampling Method

The bootstrap method is very simple to apply and it is based on Monte Carlo simulations. Given N realizations of a random variable the empirical distribution function is formed. From the empirical distribution function a new sample of size n is drawn. Essentially this new bootstrap sample is formed by randomly selecting n observations from the original sample with replacement. Most often $n = N$.

From this artificial sample an estimate, $\hat{\theta}$, of the parameter of interest, θ, is calculated. In our case, the parameter of interest is the design value given by Eq. (1). This procedure is carried out a large number of times, and each time θ is calculated. Finally an empirical distribution function of the estimate of the parameter is constructed. Based on this distribution function estimates of the expectation and the variance and confidence intervals or other statistics corresponding to $\hat{\theta}$ are computed.

The modus of the bootstrap is given by Efron (1982)

1) Having observed N observations of X, form the empirical distribution function, $\hat{F}$

   $\hat{F}$: mass $\frac{1}{N}$ at x_i , $i = 1,2\ldots,N$

2) Draw a bootstrap sample

   $X_1^*, X_2^*, \ldots X_n^* \in \hat{F}$

   with the realizations $x_i^* = X_i^*$ and calculate the statistic $\hat{\theta}^* = \hat{\theta}(x_1^*, x_2^*, \ldots, x_n^*)$. Thus $\{X_i^*\}$ is a random sample with n observations drawn with replacement from the observations $x_1, x_2, \ldots, x_n$. The properties of $\hat{\theta}$ from a sample of N observations are wanted, and hence $n = N$ is used.

3) Repeat step 2 B times (B being a large number). This procedure yields B independent bootstrap replications of the sample function, *i.e.*, $\hat{\theta}^{*1}$, $\hat{\theta}^{*2}$, ..., $\hat{\theta}^{*B}$. The empirical distribution of $\hat{\theta}^*$ found by this method is called the bootstrap distribution of $\hat{\theta}^*$, defined by

$$\hat{G}(s) = P_*\{\hat{\theta}^* < s\} \tag{6}$$

$P_*\{...\}$ denotes the probability computed according to the bootstrap distribution of $\hat{\theta}^*$.

The bootstrap estimates of the expectation and the variance of θ are proposed by Efron (1982) to be the usual estimates

$$\hat{\theta}_{BOOT} = \hat{\theta}^{*(\cdot)} = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^{*b}$$
$$\hat{\sigma}^2_{BOOT} = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\theta}^{*b} - \hat{\theta}^{*(\cdot)}\right)^2 \tag{7}$$

where $\hat{\theta}_{BOOT}$ – The bootstrap estimate of θ
$\hat{\sigma}^2_{BOOT}$ – The bootstrap estimate of the variance of $\hat{\theta}$
$\hat{\theta}^{*b}$ – Estimate based on the bth bootstrap sample
$\hat{\theta}^{*(\cdot)}$ – Average of all B bootstrap estimates

Efron and Tibshirani (1986) discuss the estimation of confidence intervals for θ using the bootstrap. They demonstrate that any quantile in the distribution function of $\hat{\theta}$ can be estimated, still without assuming any model for the data. This is performed using $\hat{G}$(s) given by Eq. (6). The most simple result is to use $\hat{G}$(s) directly. However, in general the bootstrap distribution function of $\hat{\theta}^*$ may be skewed which results in some violation of the coverage probability for θ. In this situation an improvement on this coverage probability can be obtained by using the socalled bias-corrected method.

The estimate of any quantile in the distribution function of $\hat{\theta}$ is calculated following Efron and Tibshirani (1986)

$$\theta(\alpha) = \hat{G}^{-1}\left(\Phi\{2z_0 + z^{(\alpha)}\}\right) \; ; \qquad z_0 = \Phi^{-1}\{\hat{G}(\hat{\theta})\} \tag{8}$$

where $\theta[\alpha]$ – The α-quantile in the distribution function of $\hat{\theta}$
$z^{(\alpha)}$ – 100*α percentage point of the standard normal distribution
$\hat{\theta}$ – Estimate of parameter
$\Phi\{\ \}$ – The standard normal distribution function

If z_0=0 the estimate of median and average are identical and the bias-corrected method reduces to the most simple procedure, *i.e.*, to use the quantiles of $\hat{G}$(s) as estimates of the quantiles in the distribution function of $\hat{\theta}$.

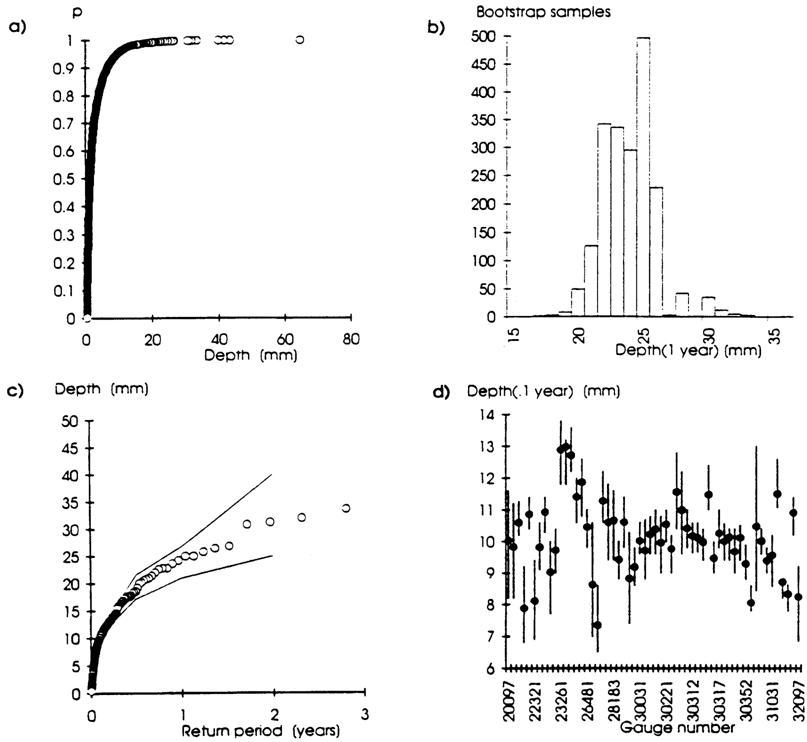

Fig. 2. The estimation procedure. (a) The emperical distribution function at gauge 30031 for depth. (b) The histogram for depth (1 year) estimated by means of the bootstrap resampling method used on the observed rain events. (c) The return periods and the estimated 90% control confidence intervals. (d) The estimated 90% central confidence intervals for all 56 gauges for depth (0.1 year).

Results

In Fig. 2(b) $\hat{G}(s)$ at gauge 30031 for depth (1 year) is plotted. It should be noted, that the traditional non-parametrical approach uses only two observations, whereas the bootstrap resampling method uses more than 20 observations, still without assuming any model.

The bootstrap procedure has been applied to all 56 gauges considered in the present study, and for each gauge the bootstrap leads to estimation of a confidence interval. It was decided to compute 90% central confidence intervals. The intervals for the depth are visualized in Fig. 2(c). Finally 90% central confidence intervals for all 56 gauges for the depth (0.1 year) are shown in Fig. 2(d).

The bootstrap distribution function has been compared to the standard normal distribution. The distribution function of the estimates of design values resembles the normal distribution rather well for small return periods (smaller than one year). One immediate conclusion from this result is, that the z_0 given in Eq. (7) is close to zero. For larger return periods the distribution function is rather discrete as the estimates are based on smaller number of observations. Thus the bias-corrected method for large return periods (larger than one year for example) may significantly alter the confidence intervals obtained at the individual gauges and it has been used when estimating all design values.

Variability between Gauges

Based on the 90% central confidence intervals computed for each gauge separately a test of identical underlying distributions can be constructed. If the underlying design values for the gauges are assumed to be identical the variation between the corresponding estimates will be due to sampling variability only. Consequently all the confidence intervals should have a coverage probability of 90% for one common underlying design value. The true coverage probability for any possible design value may be assessed by drawing a horizontal line in Fig. 2(d) and counting the number of intervals that actually cover the line. Thus a test based on the binomial distribution can determine whether or not it is likely that the design value is the same for all gauges considered. This test assumes no spatial correlation, as positive spatial correlation will increase the probability that the hypothesis is accepted. In the present investigation the separation between the gauges in Copenhagen is small. However, the variability between the results from the Copenhagen area gauges seems to be the same as between the other gauges.

The hypothesis tested is:

H_0: At least 90% of the confidence intervals cover a common value

The result of the test is reported in Table 1. The test indicates that if an estimate for intensity (>0.5 year) or depth (>0.2 year) is wanted, the estimation may be based on the assumption that the variations of the estimates are due to sampling variability only. This implies that the design value may be assumed to be approximately the same all over Denmark. The acceptance of the hypothesis leads to one possible explanation, namely that very extreme rain events are not significantly different distributed in Denmark. Regional variation may be present, but is superseded by sampling variation when using the non-parametric approach.

For intensity (≤0.5 year) and depth (≤0.2 year) the tests indicate that the rain events from different gauges cannot be assumed to follow the same distribution. Thus, for these return periods the data indicate a regional variation. The result contradicts the assumption of Denmark being a meteorologically homogeneous area. This topic has been discussed previously by Mikkelsen and Harremoës (1993) and by Arnbjerg-Nielsen (1993).

It should be noted that the test results do not depend on the number of gauges participating in the analysis. Tests using only the 40 gauges with observation periods longer than 10 years were performed. These tests also rejected the hypothesis for small return periods.

Table 1 – Result of the testing of the hypothesis of Eq. (9). The testing is performed by counting the number of gauges where the confidence intervals do not cover a common estimate of the design value. Two tests have been used: one uses the average of the estimates from the gauges (Regional mean), while the other uses the particular design value, where most of the confidence intervals cover. If the hypothesis is accepted (+) it can be assumed that the estimates describe the same underlying population. Only gauges with observation period 10 times longer than the return period have been taken into account

Variable	Return period (years)	Number of rain gauges whose confidence intervals do not cover estimate of true value of design value of return period			
		Number of gauges	Regional mean	Design value with most confidence intervals covering	Result of testing
Depth	0.1	56	26	21	–
	0.2	56	19	17	–
	0.5	43	6	5	+
	1	40	3	2	+
Intensity	0.1	56	18	18	–
	0.2	56	16	12	–
	0.5	43	8	8	–
	1	40	7	5	+

Conclusion

The non-parametrical bootstrap resampling method has been evaluated and found suitable for estimation of the variability of estimates of design values for peak intensity and for depth of a rain event for return periods up to approximately 1/10 of the total observation period at a rain gauge.

On the basis of the bootstrap estimates it was tested if the regional variation of the estimates could be assumed to be the result of sampling variations only. For return periods shorter than or equal to 0.2 years the tests rejected the hypothesis that the variation could be assumed to be due to random sampling variation for both variables. Thus a significant variation of the estimates depending on the geographical location of the gauges was detected. For the 0.5-year return period the hypothesis was accepted for the depth of a rain event, but not for the peak intensity. For higher return periods the hypothesis was accepted for both these variables.

The analysis suggests that for return periods longer than 0.5 years the same design criteria for sewer systems may be used in all regions. For return periods shorter than 0.5 years the estimates obtained at the local gauges seem to yield the best information on the design values for sewer systems in the corresponding regions.

Acknowledgements

We are indebted to climatologist, M. Sc. Stig Rosenørn at the Danish Institute of Meteorology (DMI) for his advice, help and work on assuring the quality of the data.

Data used in this investigation have kindly been made available by Spildevandskomitéen and the DMI.

References

Arnbjerg-Nielsen, K. (1993) Non-Parametric Statistics on Extreme Rainfall, M.Sc.-thesis *No. 5/93.* IMSOR & LTH, Technical University of Denmark, DK-2800 Lyngby, Denmark.

David, H. A. (1981) *Order Statistics*, 2. edition, John Wiley & Sons.

Efron, B. (1982) The Jackknife, the Bootstrap and Other Resampling Plans. Society for Industrial and Applied Mathematics, Philadelphia.

Efron, N., and Tibshirani, R. (1986) Bootstrap Methods for Standard Errors, Confidence Intervals and Other Measures of Statistical Accuracy, *Statistical Science, Vol. 1, (1),* pp. 54-77.

Harremoës, P., and Henze, A. (1981) This Century's Shower in Denmark, Registered in Denmark by new Gauge System, *Vannet i Norden, Vol. 14, (2),* pp. 43-56.

Madsen, H., Rosbjerg, D., and Harremoës, P. (1994) PDS-Modelling and Regional Baysian Estimation of Extreme Rainfalls, *Nordic Hydrology, Vol. 25 (3),* pp. 279-300.

Mikkelsen, P. S., and Harremoës, P. (1993) Uncertainties in Urban Runoff Extreme Value Calculations caused by Statistical/Geographical Variation in Rainfall Data. Accepted for the 6.th Int. Conf. on Urban Storm Drainage, Niagara Falls, 1993.

Rosbjerg, D. (1977) Return Periods of Hydrological Events, *Nordic Hydrology, Vol. 8 (5),* pp. 57-61.

Rosbjerg, D. (1988) A Defence of the Median Plotting Postition. Progress Report no. 66, pp. 17-28. ISVA, Technical University of Denmark, DK-2800 Lyngby.

Spildevandskomiteen (1984) Maksimum Discharge and Volumes of Basins from historical Rain Series, Paper No. 18. ISBN 87-88671-00-3, Danish Society of Engineering.

First received: 18 June, 1993
Revised version received: 11 January, 1994
Accepted: 14 March, 1994

Address:

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

H. Spliid,
Inst. of Mathematical Modelling,
Technical University of Denmark,
Building 321,
DK-2800 Lyngby, Denmark.