Regional frequency analysis of extreme rainfalls

V-T-V. Nguyen*, T-D. Nguyen* and F. Ashkar**

*Department of Civil Engineering and Applied Mechanics, McGill University, 817 Sherbrooke Street West, Montreal, Quebec, Canada H3A 2K6

**Department of Mathematics, University of Moncton, Moncton, New Brunswick, Canada E1A 3E9

Abstract This study proposes two alternative methods for estimating the distribution of extreme rainfalls for sites where rainfall data are available (gaged sites) and for locations without data (ungaged sites). The first method deals with the estimation of short-duration rainfall extremes from available rainfall data for longer durations using the "scale-invariance" concept to account for the relationship between statistical properties of extreme rainfall processes for different time scales. The second method is concerned with the estimation of extreme rainfalls for ungaged sites. This method relies on a new definition of homogeneous sites. Results of the numerical application using data from a network of 10 recording rain gauges in Quebec (Canada) indicate that the proposed methods are able to provide extreme rainfall estimates that are comparable with those based on observed at-site rainfall data.

Keywords Extreme rainfalls; frequency analysis; generalized extreme value distribution; regionalization; scale invariance; statistical modelling

Water Science and Technology **Vol 45 No 2 pp 75–81** © IWA Publishing 2002

Introduction

Rainfall frequency analyses are commonly used for the design of various hydraulic structures. More specifically, rainfall frequency analysis studies are necessary for the development of a "design storm"; that is, a rainfall temporal pattern used in the design of a hydraulic structure. The objective of rainfall frequency analyses is to estimate the amount of rainfall falling at a given point or over a given area for a specified duration and return period. For a site for which sufficient rainfall data are available (a gaged site), a frequency analysis can be performed. The precipitation data used for frequency analysis are typically available in the form of annual maximum series (AMS) (or converted to this form using continuous records of hourly or daily rainfall data). These series contain the largest rainfall in each complete year of record. An alternative data format for rainfall frequency studies is "partial duration series" (PDS) (also referred to as peaks over threshold data) which consist of all large precipitation amounts above certain thresholds selected for different durations. Arguments in favor of either of these techniques are well described in the literature (National Research Council of Canada, 1989; Stedinger *et al.*, 1993). Due to its simpler structure, the AMS-based method is more popular in practice.

Several probability models have been developed to describe the distribution of extreme rainfalls at a single site (Wilks, 1993). Unfortunately, these models are accurate only for the specific time frame associated with the data used. Hence, it was argued that the usefulness of a model should lie in its potential ability to adequately describe the rainfall process at time scales that are not included in the building of its mathematical structure. It has necessitated the need for formulating models whose mathematical structure will follow the main statistical features of the past history through a continuum of levels of aggregation. This formulation implies that the suggested model should statistically and simultaneously match various properties of the rainfall process at different levels of aggregation, whether or not these properties are included in the model. The most important practical implication of such models is that, from a higher aggregation model we could infer the statistical properties of the process at the finer resolutions that may not have been observed. Another major advantage of such a procedure involves the parsimonious parameterisation since these models would normally require a much smaller number of parameters, while traditional models need different sets of parameters for each particular time scale of the rainfall series considered (Burlando and Rosso, 1996; Kottegoda and Rosso, 1997).

Furthermore, in most practical applications, extreme rainfall records at the location of interest are often unavailable (an ungaged site). Regionalization methods (e.g. Schaefer, 1990) are frequently used to transfer rainfall information from one location to the other where the data are needed but not available, or to improve the accuracy of rainfall estimates where available records are too short. Nevertheless, traditional regionalization techniques are often criticised for their obvious subjectivity, in particular in the definition of hydrologically similar sites (or hydrologically homogeneous regions), and the lack of physical justifications.

In view of the above problems, the main objective of the present study is to propose two alternative methods for estimating the distribution of extreme rainfalls at gaged and ungaged sites, respectively. More specifically, the first method was developed for determining the distribution of short-duration (e.g. one hour) extreme rainfalls using available at-site rainfall data for longer durations. This procedure was based on the "scale-invariance" (or "scaling") concept that is currently a popular tool in the modelling and analysis of various geophysical processes (e.g. Gupta and Waymire, 1990). In this study, the scale invariance implies that statistical properties of extreme rainfall processes for different time scales are related to each other by an operator involving only the scale ratio. The second method was proposed for estimating the distribution of annual maximum daily rainfalls for sites without data. This method relied on the consideration of similarity of rainfall conditions at different sites within a homogeneous region. Extreme rainfall data from a network of 10 rain gauges in Quebec (Canada) were used to illustrate the application of the proposed methods. The Generalised Extreme-Value (GEV) distribution was used to estimate the extreme rainfall quantiles. Results of this numerical application have indicated that the proposed techniques can provide extreme rainfall estimates that are comparable with those obtained directly from historical data.

The scaling concept

By definition (see, e.g. Fedder, 1988), a function $f(x)$ is scaling (or scale-invariant) if $f(x)$ is proportional to the scaled function $f(\lambda x)$ for all positive values of the scale factor λ. That is, if $f(x)$ is scaling then there exists a function $C(\lambda)$ such that

$$f(x) = C(\lambda) f(\lambda x) \tag{1}$$

It can be readily shown that

$$C(\lambda) = \lambda^{-\beta} \tag{2}$$

in which β is a constant, and that

$$f(x) = x^{\beta} f(1) \tag{3}$$

Hence, the relationship between the non-central moment (NCM) of order k, μ_k, and the variable x can be written in a general form as follows:

$$\mu_k = E\left\{f^k(x)\right\} = \alpha(k) x^{\beta(k)} \tag{4}$$

in which $\alpha(k) = E\{f^k(1)\}$ and $\beta(k) = \beta k$. Notice that if the exponent $\beta(k)$ is not a linear function of k, in such cases the process is said to be "multiscaling" (Gupta and Waymire, 1990).

The scaling generalized extreme value model

Application of the GEV distribution to model the annual series of extreme rainfalls has been advocated by several researchers (see, e.g. Natural Environment Research Council, 1975; Schaefer, 1990). The cumulative distribution function, $F(x)$, for the GEV distribution is given as

$$F(x) = \exp\left[-\left(1 - \frac{\kappa(x-\xi}{\alpha}\right)^{1/\kappa}\right] \quad \kappa \neq 0 \tag{5}$$

where ξ, α and κ are respectively the location, scale and shape parameters. It can be readily shown that the k-th order NCM, μ_k, of the GEV distribution (for $k \neq 0$) can be expressed as (Pandey, 1995)

$$\mu_k = \left(\xi + \frac{\alpha}{\kappa}\right)^k + (-1)^k\left(\frac{\alpha}{\kappa}\right)^\kappa \Gamma(1+k\kappa) + k\sum_{i=1}^{k-1}(-1)^i\left(\frac{\alpha}{\kappa}\right)^i\left(\xi + \frac{\alpha}{\kappa}\right)^{k-i}\Gamma(1+i\kappa) \tag{6}$$

where $\Gamma(\cdot)$ is the gamma function. Hence, on the basis of (6), it is possible to estimate the three parameters of the GEV distribution using the first three NCMs. Consequently, the quantiles (X_T) can be computed using the following relation:

$$X_T = \xi + \frac{\alpha}{\kappa}\left\{1 - \left[-\ln(p)\right]^\kappa\right\} \tag{7}$$

in which p=1/T is the exceedance probability of interest.

Further, for a simple scaling process, it can be shown that the statistical properties of the GEV distribution for two different time scales t and λt are related as follows:

$$\kappa(\lambda t) = \kappa(t) \tag{8}$$

$$\alpha(\lambda t) = \lambda^\beta \alpha(t) \tag{9}$$

$$\xi(\lambda t) = \lambda^\beta \xi(t) \tag{10}$$

$$X_T(\lambda t) = \lambda^\beta X_T(t) \tag{11}$$

Hence, based on these relationships it is possible to derive the statistical properties of short-duration (e.g. λt = less than 1 day) extreme rainfalls using the properties of daily (t = 1 day) extreme rainfalls. The exponent β is computed based on the scaling properties of the NCMs of extreme rainfalls for various durations.

Regional frequency analysis of extreme rainfalls

As mentioned previously, for sites where rainfall records are limited or unavailable, regional frequency analysis, which uses data from many sites, has been shown to be able to reduce the uncertainties in the estimation of extreme events. Several regional estimation methods have been proposed in the literature (Cunnane, 1988; Nguyen *et al.*, 1998), among which the index-flood procedure introduced by Dalrymple (1960) for use with AMS is the most popular and has been applied to regional estimation of extreme precipitation (Schaefer, 1990). However, one of the main difficulties in the use of this technique is related to the definition of "homogeneous" regions, which assume that data at different sites within a homogeneous group follow the same distribution except for scale. Various methods have been proposed for determining homogeneous regions, but there is no generally accepted procedure in engineering practice (Fitzgerald, 1989; Schaefer, 1990; Hosking and Wallis, 1993; Fernandez Mills, 1995).

In the present study, a homogeneous region is defined as the region in which all annual maximum rainfall series at different sites must have similar properties of rainfall occurrence. This similarity would indicate that these rainfall series are produced by the persistence in space of the same storm system. More specifically, if the occurrence of rainfalls at different rain gauges within a given concurrent period is similar, these gages are thus considered as members of a homogeneous group. For instance, if the numbers of rainy hours within a given one-day interval for two different sites are highly related, the two sites can then be considered as similar. Based on this grouping of homogeneous stations, regional relationships between the NCMs of the maximum rainfalls and the mean number of rainfall occurrences for all these stations can be developed. Finally, for a site without data, the NCMs (and the quantiles) of the extreme rainfalls can be estimated using the regional relationships of the corresponding homogeneous region and the interpolated mean number of rainfall occurrences for this ungaged location.

Numerical application

In the following, to illustrate the application of the proposed approaches, a case study is carried out using annual maximum rainfall series for durations ranging from 1 hour to 4 days from a network of 10 rain gauges in Quebec (Canada). The rainfall record lengths vary from 15 to 48 years. Firstly, the proposed GEV scaling model is used to determine the distribution of rainfall extremes for short time intervals (e.g. $\lambda t < 1$ day) from available rainfall data for longer time scales (e.g. $t \geq 1$ day). Secondly, the suggested regional procedure is adopted to determine the distribution of extreme daily rainfalls for an ungaged site.

To assess the scaling behaviour of these AMS, the log-log plots of the first three rainfall NCMs against duration are prepared for all 10 stations. For purpose of illustration, Figure 1 shows the plot for McGill station. The log-linearity exhibited in the plot indicates the power law dependency (i.e. scaling) of the rainfall statistical moments with duration (Eq. (4)). Further, the linearity of the scaling exponent $\beta(k)$ with the moment order k as shown in Figure 2 supports the assumption that the extreme rainfall series considered can be described by a simple scaling model. Hence, for a given location, it is possible to determine the NCMs and the distribution of rainfall extremes for short durations (e.g. 1 hour) using available rainfall data for longer time scales within the same scaling regime (β is known). For illustrative purposes, Figure 3 shows the comparison between the empirical (observed) and estimated GEV distributions of 1-hour rainfall extremes at McGill station. It can be seen that the estimated GEV distribution is in very good agreement with the observations.

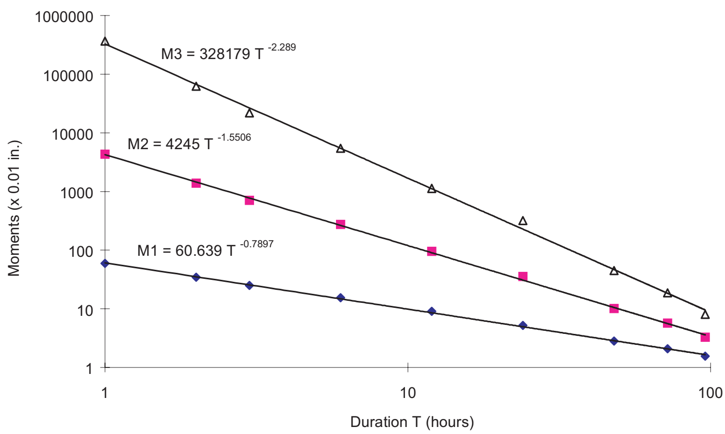

Figure 1 The log-log plot of maximum rainfall non-central moments versus rainfall duration for McGill station

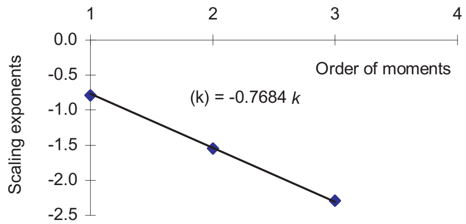

Figure 2 The plot of the scaling exponent versus the order of non-central moments for McGill station

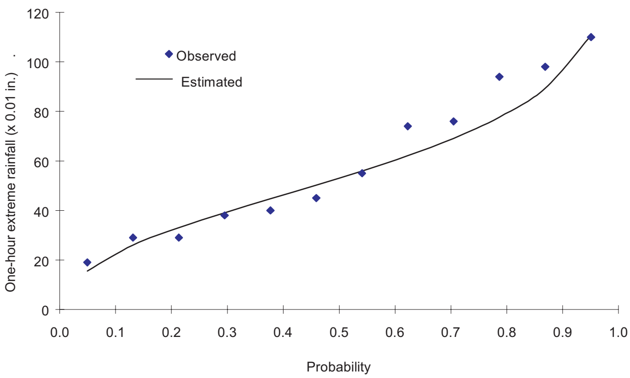

Figure 3 Empirical (observed) and estimated distributions of maximum hourly rainfalls for McGill station

As mentioned above, the present study is concerned with the estimation of the distribution of 1-day maximum rainfalls for sites where rainfall data are not available. The proposed regional estimation procedure requires the grouping of stations into homogeneous groups. The high correlation of the number of rainy hours between different sites within a concurrent period can be used as a criterion for judging the similarity of rainfall conditions at these sites. More objectively, principal component analysis is performed using the series of number of rainy hours observed at each rain gauge in order to assess the similarity of rainfall occurrences between these gages. For instance, on the basis of the proposed homogeneity criterion, four stations Brebeuf, Dorval, McGill, and St-Hubert among the 10 gages considered can be grouped into one homogeneous group. Hence, regional relationships between the first three NCMs (M_1, M_2 and M_3) and the number of rainy hours for each homogeneous group can be developed.

Further, to simulate the ungaged condition the Jacknife procedure is used in the present study. That is, one gage was removed from a homogeneous group, and then regional relationships between the first three NCMs and the mean number of rainy hours in a one-day period were developed using the data from the remaining stations in the group. For example, if the site at McGill station was assumed to be ungaged, the following regional relationships were obtained based on the available rainfall data at Brebeuf, Dorval, and St-Hubert:

$$M_1 = 0.619 \cdot N_R \tag{12}$$

$$M_1 = 02.087 \cdot e^{0.025 M_2} \tag{13}$$

$$M_3 = 200.38 \cdot e^{0.013 M_2} \tag{14}$$

in which N_R is the mean number of rainy hours during a one-day period with rain. Hence, on the basis of (12), (13), and (14) the first three NCMs and the distribution of maximum daily rainfalls for McGill can be computed if the mean number of rainy hours N_R for this ungaged

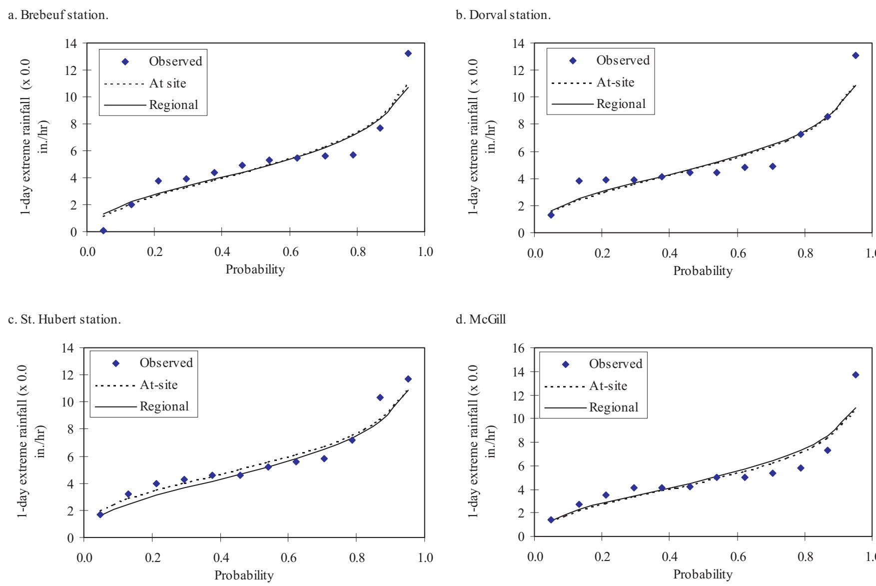

Figure 4 Empirical (observed) and estimated (at-site and regional) distributions of annual maximum daily rainfalls for Brebeuf, Dorval, St-Hubert, and McGill stations

site is known. In this study, the value of N_R for an ungaged location is estimated by interpolation from the mean numbers of rainy hours at stations in the same homogeneous group using the inverse square distance method. Figure 4 shows the comparison between empirical (observed) and estimated distributions of 1-day rainfall extremes at Brebeuf, Dorval, St-Hubert, and McGill stations for the case where rainfall data are missing (regional curve) as well as for the case where rainfall data are available (at-site curve). The good agreement between the estimated regional and empirical distributions for 1-day rainfall extremes as shown in this case study has indicated the feasibility of the proposed regional estimation method. In addition, it can be seen that the regional estimate of daily rainfall extreme distribution is as accurate as the at-site estimate in this case.

Conclusions

The major findings of the present study can be summarised as follow.

(a) By considering the scaling of statistical properties of extreme rainfall processes, a new extrapolation method has been proposed for the estimation of short-duration extreme rainfall distribution using available rainfall data for longer time scales (partially-gaged sites). Results of an illustrative application have indicated that the proposed scaling approach could provide extreme rainfall estimates that are comparable with the empirical ones.

(b) An important step in regional estimation of extreme rainfalls for ungaged sites is the definition of hydrologically homogeneous regions. To this end, it has been shown that homogeneity of rainfall conditions at different sites can be defined based on the similarity of rainfall occurrences within a concurrent time period. Further, it has been demonstrated that the proposed regional estimation method can provide maximum daily rainfall estimates that are as accurate as the fitted at-site values.

References

Burlando, P. and Rosso, R. (1996). Scaling and multiscaling models of depth-duration-frequency curves for storm precipitation. *J. Hydrology*, **187**, 45–64.

Cunnane, C. (1988). Methods and merits of regional flood frequency analysis. *J. Hydrology*, **100**, 269–290.
Dalrymple, T. (1960). Flood frequency analysis, *Manual of Hydrology: Part 3. Flood flow techniques,* US Geological Survey Water-Supply Paper 1543-A.
Fedder, J. (1988). *Fractals*. Plenum Press, New York, 283 pages.
Fernandez Mills, G. (1995). Principal component analysis of precipitation and rainfall regionalization in Spain. *Theoretical and Applied Climatology*, **50**, 169–183.
Fitzgerald, D.L. (1989). Single station and regional analysis of daily rainfall extremes. *Stochastic Hydrology and Hydraulics*, **3**, 281–292.
Gupta, V.K. and Waymire, E. (1990). Multiscaling properties of spatial rainfall and river flow distributions. *J. Geophysical Research*, **95**(D3), 1999–2009.
Hosking, J.M.R. and Wallis, J.R. (1993). Some statistical useful in regional frequency analysis. *Wat. Resources Res*., **29**(2), 271–281.
Kottegoda, N.T. and Rosso, R. (1997). *Statistics, Probability and Reliability for Civil and Environmental Engineers*, McGraw-Hill, 735 pages.
National Research Council of Canada (1989). *Hydrology of Floods in Canada: a Guide to Planning and Design*, Ottawa, Canada, 245 pages.
Natural Environment Research Council (1975). *Flood Studies Report, Volume II: Meteorological Studies*, London, UK, 81 pages.
Nguyen, V-T-V., Nguyen, T-D. and Wang, H. (1998). Regional estimation of short duration rainfall extremes. *Wat. Sci. Tech*., **37**(11), 15–19.
Pandey, G.R. (1995). Regional estimation of floods and rainfalls for ungaged sites, Doctoral Thesis, Department of Civil Engineering and Applied Mechanics, McGill University, Montreal, 197 pages.
Schaefer, M.G. (1990). Regional analyses of precipitation annual maxima in Washington State. *Wat. Resources Res*., **26**(1), 119–131.
Stedinger, J.R., Vogel, R.M. and Foufoula-Georgiou, E. (1993). Frequency analysis of extreme events. *Handbook of Hydrology*. D.R. Maidment (ed.). McGraw-Hill, New York, 18.1–18.66.
Wilks, D.S. (1993). Comparison of the three-parameter probability distributions for representing annual extreme and partial duration precipitation series. *Wat. Resources Res.,* **29**(10), 3543–3549.