ASSESSMENT OF AN L-KURTOSIS-BASED CRITERION FOR QUANTILE ESTIMATION

By M. D. Pandey,[1] P. H. A. J. M. van Gelder,[2] and J. K. Vrijling[3]

ABSTRACT: The estimation of extreme quantiles corresponding to small probabilities of exceedance is commonly required in the risk analysis of flood protection structures. The usefulness of L-moments has been well recognized in the statistical analysis of data, because they can be estimated with less uncertainty than that associated with traditional moment estimates. The objective of the paper is to assess the effectiveness of L-kurtosis in the method of L-moments for distribution fitting and quantile estimation from small samples. For this purpose, the performance of the proposed L-kurtosis-based criterion is compared against a set of benchmark measures of goodness of fit, namely, divergence, integrated-square error, chi square, and probability-plot correlation. The divergence is a comprehensive measure of probabilistic distance used in the modern information theory for signal analysis and pattern recognition. Simulation results indicate that the L-kurtosis criterion can provide quantile estimates that are in good agreement with benchmark estimates obtained from other robust criteria. The remarkable simplicity of the computation makes the L-kurtosis criterion an attractive tool for distribution selection.

INTRODUCTION

Background

The estimation of extreme quantiles corresponding to small probabilities of exceedance (POE) is commonly required in the risk analysis of hydraulic structures. Such extreme quantiles may represent design values of environmental loads (wind, waves, snow, and earthquakes), river discharge, and flood level specified by design codes and regulations (Technical 1990).

It is desirable that the quantile estimate be unbiased; i.e., its expected value should be equal to the true value. It is also desirable that an unbiased estimate be efficient; i.e., its variance should be as small as possible. The problem of unbiased and efficient estimation of extreme quantiles from small samples is commonly encountered in civil engineering practice. For example, annual flood discharge data may be available for the past 50–100 years, and on that basis one may have to estimate a design flood level corresponding to a 1,000–10,000 year return period (Technical 1990).

The first step in quantile estimation involves fitting an analytical probability distribution to represent adequately the sample observations. To achieve this, the distribution type should be judged from the data, and then parameters of the selected distribution should be estimated. Since the bias and efficiency of quantile estimates are sensitive to the distribution type, the development of simple and robust criteria for fitting a representative distribution to small samples of observations has been an active area of research.

Certain linear combinations of expectations of order statistics, also referred to as L-moments by Hosking (1990), have been shown to be very useful in statistical parameter estimation as compared to other standard methods, e.g., the method of moments, maximum likelihood, and least squares. The main advantage of L-moments is that, being a linear combination of data, they are less influenced by outliers, and the bias of their small sample estimates remains fairly small.

Hosking and Wallis (1997) proposed a simple but effective approach to fit three parameter distribution functions (DFs). The approach involves the computation of four L-moments from a given sample. By matching the first three L-moments, a set of three-parameter DFs (Table A1) can be fitted to the sample data. To identify the best-fit DF from this set, the use of a normalized fourth L-moment, i.e., L-kurtosis (L-K), was proposed by Hosking and Wallis (1997). The writers propose that the distribution with its L-kurtosis value closest to that of the sample value can be taken as the most acceptable DF, which should be used for quantile estimation. This criterion is referred to as the L-K criterion in the present paper. In essence, the general idea is to consider the L-kurtosis as an empirical summary measure of the distribution tail weight or shape.

In the theory of mathematical statistics, several measures of ‘‘distance’’ between the data and a DF have been developed to quantify the goodness of fit for inference purposes (Beran 1984), such as the chi-square (CS) and Kolmogorov-Smirnov (KS) tests. In modern information theory (Kullback 1959; Jumarie 1990), mathematically sound measures of probablistic distance or discrimination, namely, cross-entropy and divergence, have been widely used to quantify the degree of resemblance between two DFs, or, conversely, to select the closest possible posterior distribution given a suitable prior. Such measures can be used to develop benchmark criteria for evaluating the effectiveness of any empirical distribution selection criterion.

The objective of the paper is to evaluate the robustness of the L-K criterion for distribution selection and extreme quantile estimation from small samples. The robustness is evaluated against a set of benchmark estimates obtained from the information theoretic measure, namely, the divergence, and other distance measures used in statistics, namely, CS, probability-plot correlation (PPC), and integrated-square error (ISE). For the evaluation of robustness, a series of Monte Carlo simulation experiments were designed, in which DFs were fitted, based on L-K and benchmark criteria, to random samples generated from specified parent DFs, and the accuracies of quantile estimates were compared. The simulation study reveals that the L-K criterion is fairly effective in quantile estimation.

Organization

The next section reviews definitions of expectations of order statistics, probability-weighted moments (PWMs), and L-moments. Statistical distance measures and their properties are discussed in an upcoming section. In the information theory concepts section, general concepts of information theory are briefly discussed and measures of probabilistic distance are

[1]Asst. Prof., Univ. of Waterloo, Waterloo, ON, Canada N2L 3G1.
[2]Asst. Prof., Tech. Univ. of Delft, Delft, The Netherlands 2600 GA.
[3]Prof., Tech. Univ. of Delft, Delft, The Netherlands 2600 GA.

Note. Discussion open until January 1, 2002. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on May 11, 1999; revised September 14, 2000. This paper is part of the ***Journal of Hydrologic Engineering***, Vol. 6, No. 4, July/August, 2001. ©ASCE, ISSN 1084-0699/01/0004-0284–0292/$8.00 + $.50 per page. Paper No. 20878.

introduced. The distribution types and quantile estimation section discusses the effect of distribution shape and tail weight on quantile estimation. The assessment of robustness of L-kurtosis section evaluates the robustness of the L-K criterion via Monte Carlo simulations, and illustrates an application of the L-K criterion. The main conclusions of the study are then summarized, and Appendix I summarizes various probability distributions used in the paper.

ORDER STATISTICS AND L-MOMENTS

Expectations of Order Statistics

If n observations of a random sample, $\{X_1, X_2, \ldots, X_n\}$, are rearranged in an increasing order of magnitude, $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$, then any rth member ($X_{r:n}$) of this new sequence is called the rth order statistic. Using the density function of $X_{r:n}$ presented by Kendall and Stuart (1977) along with a transformation $u = F(x)$, with $F(x)$ being the distribution function, the expectation of order statistics can be expressed in terms of the quantile function (QF), $x(u)$, as

$$E(X_{r:n}) = f\binom{n}{r}\int_0^1 x(u)u^{r-1}(1-u)^{n-r}\,du \tag{1}$$

Expectations of the maximum and minimum of a sample of size n can be easily obtained from (1) by setting $r = n$ and $r = 1$, respectively

$$E(X_{n:n}) = n\int_0^1 x(u)u^{n-1}\,du \quad \text{and} \quad E(X_{1:n}) = n\int_0^1 x(u)(1-u)^{n-1}\,du \tag{2}$$

Probability-Weighted Moment

The PWM of a random variable was formally defined by Greenwood et al. (1979) as

$$M_{i,j,k} = E[X^i u^j(1-u)^k] = \int_0^1 x(u)^i u^j(1-u)^k\,du \tag{3}$$

The following two forms of PWMs are particularly simple and useful:

$$\text{Type 1: } \alpha_k = M_{1,0,k} = \int_0^1 x(u)(1-u)^k\,du; \quad (k = 0, 1, \ldots, n) \tag{4}$$

$$\text{Type 2: } \beta_k = M_{1,k,0} = \int_0^1 x(u)u^k\,du; \quad (k = 0, 1, \ldots, n) \tag{5}$$

Comparing (2) with (4) and (5), it can be seen that α_k and β_k, respectively, are related to expectations of the minimum and maximum in a sample of size k

$$\alpha_{k-1} = \frac{1}{k}E(X_{1:k}); \quad \beta_{k-1} = \frac{1}{k}E(X_{k:k}); \quad (k \ge 1) \tag{6}$$

In essence, PWMs are the normalized expectations of the maximum/minimum of k random observations; the normalization is done by the sample size (k) itself.

L-Moments

L-moments are certain linear combinations of PWMs that are analogous to ordinary moments in the sense that they also provide measures of location, dispersion, skewness, kurtosis, and other aspects of the shape of probability distributions or data samples. An rth order L-moment is mathematically defined as

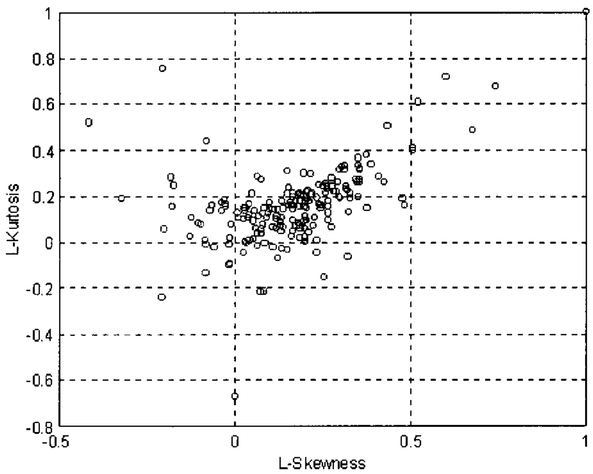

FIG. 1. Variation of L-Skewness and L-Kurtosis in a European River Discharge Database

$$\lambda_{r+1} = \sum_{k=0}^{r} p^*_{r,k}\beta_k \tag{7}$$

where $p^*_{r,k}$ represents the coefficients of shifted Legendre polynomials (Hosking 1990). Normalized forms of higher-order L-moments, $\tau_r = \lambda_r/\lambda_2$, ($r = 3, 4, \ldots$) are convenient to work with due to to their bounded variation, i.e., $|\tau_r| < 1$. Hosking and Wallis (1997) illustrated that L-moments are very efficient in estimating parameters of a wide range of distributions from small samples. The required computation is fairly limited as compared to other traditional techniques, such as the maximum likelihood and least squares.

To illustrate typical ranges of variation of L-skewness (τ_3) and L-kurtosis (τ_4), a database of river discharges was analyzed in detail by Van Gelder et al. (1999a). The database contains records of annual maximum discharges of 194 European rivers over periods ranging from 10 to 200 years. The L-moment diagram (τ_3 versus τ_4) shown in Fig. 1 suggests the typical range of L-moment as $0 \le \tau_3 \le 0.5$ and $0 \le \tau_4 \le 0.4$. In the L-moment diagram, the data set corresponding to point (1, 1) consists of 12 out of 13 identical values, which were recorded at Chelmer, Rushes Lock Station in the United Kingdom (1977–89). The data set for the point (0, -0.67) consists of only five values recorded at Alzette Esch Station in Luxembourg (1983–87).

STATISTICAL MEASURES OF DISTANCE

In mathematical statistics, the notion of minimum distance (M-D) between the data and fitted models has been commonly utilized to determine whether or not a selected probability distribution can satisfactorily represent the available information (Beran 1984). The distance can be interpreted as a general measure of the difficulty of discriminating a distribution $q(x)$ from $p(x)$; the larger the value of distance, the worse the resemblance between the two distributions under consideration.

The distance or divergence between two distributions can be measured in several ways. The geometric (Euclidean) or metric distance between two functions, $p(x)$ and $q(x)$, is the most natural measure, given as

$$d(p, q) = \sqrt{\int [p(x) - q(x)]^2\,dx} \tag{8}$$

The metric distance satisfies some fundamental properties—(1) positivity, $d(p, q) \ge 0$; (2) identity, $d(p, q) = 0$ if $p = q$; (3) symmetry, $d(p, q) = d(q, p)$; and (4) triangular inequality, $d(p, q) + d(q, r) \ge d(p, r)$.

A general class of measures having some or all of the properties of geometric distance can be defined as (Kapur and Kesavan 1992)

$$d(p, q) = \int q(x)\phi\left[\frac{p(x)}{q(x)}\right] dx \tag{9}$$

where $\phi(z)$ = twice-differentiable convex function with $\phi(1) = 0$ [note that $z = p(x)/q(x)$]. Using this expression, some specific measures can be defined, such as the variational distance when $\phi(z) = |1 - z|/2$

$$d(p, q) = \frac{1}{2}\int |q(x) - p(x)|\, dx \tag{10}$$

and the Hellinger distance when $\phi(z) = (\sqrt{z} - 1)^2/2$

$$d(p, q) = \frac{1}{2}\int [\sqrt{p(x)} - \sqrt{q(x)}]^2\, dx \tag{11}$$

Several other measures are listed by Sobczyk and Spencer (1992).

In modern information theory, a logarithmic family of measures of discrimination has been extensively utilized, especially in communication theory and pattern recognition analysis. The cross-entropy [when $\phi(z) = -\log z$] and divergence [when $\phi(z) = -(z - 1)\log z$] are two such measures, which have elegant mathematical properties and satisfy a set of consistency requirements. This topic will be discussed further in the next section.

Similar to the metric distance, (8), the ISE is occasionally used in hydrology to measure the distance between two quantile functions, $x_p(u)$ and $x_q(u)$, as

$$\mathrm{ISE}(p, q) = \int_0^1 [x_p(u) - x_q(u)]^2\, du \tag{12}$$

It should be emphasized that the knowledge of both prior [say, $p(x)$] and posterior DFs [$q(x)$] is necessary for calculating the previously discussed distance measures.

Another alternative is to deal directly with the available data and quantify the distance between the data and a fitted DF to determine the goodness of fit. Two of the most commonly used distance measures for this purpose are the CS deviation and the KS statistic (Benjamin and Cornell 1970). The PPC coefficient is also illustrated as a useful measure of goodness of fit (Chowdhury et al. 1991). In this test, the degree of linearity of a plot of ordered sample observations, x_i, versus corresponding quantiles, y_i, estimated from a fitted DF (F_y) is calculated as

$$\rho = \frac{1}{s_x s_y}\sum_{i=1}^{n} (x_i - m_x)(y_i - m_y); \quad (0 \le \rho \le 1) \tag{13}$$

where $y_i = F_y^{-1}[i/(n + 1)]$; and m and s denote the mean and standard deviation, respectively. Obviously, the higher the value of ρ, the better the goodness of fit. Note that several plotting position formulas are available for PPC calculation.

INFORMATION THEORY CONCEPTS

General

Information theory can be characterized as a quantitative approach to the notion of information, largely based on the probability theory and statistics. It is conceptualized that the realization of a random event (E_k) provides information about the event itself, and this information is inversely proportional to the probability of the event occurring (p_k). The self-information of the event E_k is defined as (Jones 1979)

$$S(E_k) = \log\left(\frac{1}{p_k}\right) = -\log(p_k) \tag{14}$$

The use of a logarithmic measure for information, first introduced by Hartley (1928), is intuitive in the following sense: The information provided by a deterministic event (i.e., $p_k = 1$) is zero, whereas the rarer an event ($p_k << 1$), the more information that is conveyed by its realization. It is clear from (14) that the self-information of an event increases as its uncertainty grows; i.e., the probability of occurrence reduces. In this respect, S can also be regarded as a measure of uncertainty (Jones 1979).

Entropy and Cross-Entropy

Shannon (1949) introduced the notion of entropy, similar to that used in thermodynamics and statistical mechanics, as a measure of uncertainty that prevailed before the experiment was accomplished, or as a measure of information expected from an experiment. Considering the definition of self-information from (14), Shannon's entropy can be expressed as the expectation of self-information

$$H(p) = -\sum_{k=1}^{n} p_k \ln p_k \tag{15}$$

In the context of prior and posterior probabilities, Kullback (1959) introduced the concept of cross-entropy, or directed divergence, of a posterior probability distribution $q(x)$ from a prior distribution $p(x)$

$$I(q, p) = \int_R q(x)\ln\left[\frac{q(x)}{p(x)}\right] dx \tag{16}$$

Given a prior DF, $p(x)$, and a set of constraints derived from the available data (moments or fractiles), a posterior $q(x)$ can be derived by minimizing the cross-entropy. This approach, referred to as the "minimum cross-entropy principle" (Kullback 1959), is widely applied in information theory.

The consistency is a fundamental argument in mathematical analysis; i.e., if a problem can be solved in more than one way, the results obtained by different methods must be the same. To ensure this, the inference method must satisfy some basic conditions, referred to as consistency axioms. Shore and Johnson (1980) postulated four such axioms.

1. Uniqueness: The result (i.e., posterior probability assignment) based on the available information should be unique.
2. Invariance: The result should be independent of the coordinate transformation.
3. System independence: The result should be independent of the form of information about various independent systems (i.e., consideration of marginals versus a joint density should not matter).
4. Subset independence: This is the same as axiom 3, but is applicable to independent subsets of a system. It also means that regrouping of data should not affect the inference.

The cross-entropy is a positive, additive, and convex function of probabilities that is invariant with respect to any nonlinear, nonsingular transformation. It, however, does not satisfy the symmetry and triangular-inequality properties of the metric distance. The additive property, which is necessary to satisfy the third and fourth axioms, can be explained as follows: If $q(x, y)$ and $p(x, y)$ are two bivariate distributions, with each consisting of two independent components, i.e., $p(x, y) = p_1(x)p_2(y)$ and $q(x, y) = q_1(x)\, q_2(y)$, then it can be shown that

$I(q, p) = I(q_1, p_1) + I(q_2, p_2)$. The case of distributions with correlated components is discussed elsewhere (Jumarie 1990).

Given some data-based constraints, Shore and Johnson (1980) proved that the cross-entropy minimization is the only method (''uniquely correct method'') of probabilistic inference that satisfies all of the consistency axioms. The minimization of any measure other than the cross-entropy, such as those discussed in the statistical measures of distance section, would not satisfy all of the consistency requirements. Furthermore, the cross-entropy minimization would always lead to nonnegative probabilities, whereas the same cannot be said for geometric distance type measures, e.g., (8) and (12).

Divergence

As a measure of the power of discrimination, Kullback (1959) introduced a symmetric divergence measure

$$J(q, p) = \int [q(x) - p(x)]\ln\left[\frac{q(x)}{p(x)}\right] dx = I(q, p) + I(p, q) \tag{17}$$

which shares all of the properties of cross-entropy, except that it is invariant to linear transformation only (Johnson 1979). $J(q, p)$ is also referred to as probabilistic distance, since it satisfies all of the properties of metric distance, except for the triangular inequality. The divergence is considered a comprehensive measure of probabilistic distance, as it integrates the departure of a distribution from the reference DF over the entire range of the random variable.

DISTRIBUTION TYPES AND QUANTILE ESTIMATION

General

This section is intended to evaluate the sensitivity of quantile estimates to the distribution type in a specific context of this study. Given a small random sample ($n \le 30$), a family of five distributions (DFs) can be fitted by matching the first three sample L-moments. To highlight the need for identifying the most representative DF from this set, variations in quantile bias and root-mean-square error (RMSE) with respect to these DFs are studied using Monte Carlo simulations.

The accuracy of a quantile estimate is assessed in terms of the standardized bias and RMSE, defined as

$$\text{Standardized bias} = \frac{E(\hat{z} - z)}{z} = \frac{1}{M}\sum_{i=1}^{M}\left(\frac{\hat{z}_i}{z} - 1\right) \tag{18}$$

where $\hat{z}$ denotes the sample estimate of a quantile z obtained from the parent distribution; $\hat{z}_i$ = ith sample estimate in the simulation; and M = number of simulation samples (M = 25,000). The standardized RMSE is defined as

$$\text{RMSE} = \frac{1}{z}\sqrt{E[(\hat{z} - z)^2]} \tag{19}$$

These standardized quantities are used throughout the paper.

Numerical Results

Since the generalized Pareto distribution (GPD) can simulate a wide range of tail weights in direct relation with its shape parameter (k), it was chosen as the parent in the simulation experiment. The Pareto DF is also useful as a domain of attraction of extreme values exceeding a sufficiently high threshold (Davison and Smith 1990). Note that the tail weight is equal to $-k$, such that the higher the tail weight, the longer and heavier the distribution tail (Maes 1995). The shape parameter, k, was varied from -0.4 to $+0.2$, which corresponds to the following range of L-moments: $0.2 \le \tau_2 \le 0.08$, $0.54 \le \tau_3 \le 0.25$, and $0.35 \le \tau_4 \le 0.1$. The sample size (n) is fixed at 30.

Fig. 2 shows the variation of bias with k for quantiles (POE = 10^{-3}) estimated from the five candidate DFs (Table A1). The normalized bias is minimum (<10%) for the fitted GPD, whereas it varies significantly (50% to -40%) for other DFs. It is interesting that the bias of generalized lognormal (GNO) quantile estimates is fairly close to that of the GPD for $k < 0$, whereas the similar trend is seen for the Pearson type III (PE3) estimates when $k > 0$. Fig. 3 shows that the RMSE of PE3 estimates is minimum, though the associated bias is large, as seen from the previous figure. Among the remaining four DFs, GPD and GNO estimates have a smaller RMSE.

However, in the case of a lower quantile (POE = 10^{-2}) with

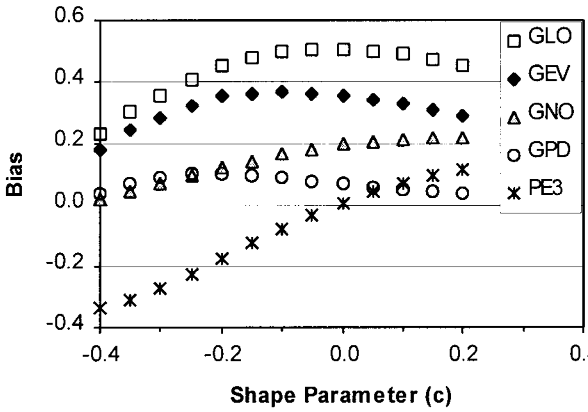

FIG. 2. Bias of Pareto Quantiles Estimated from Different DFs (POE = 10^{-3})

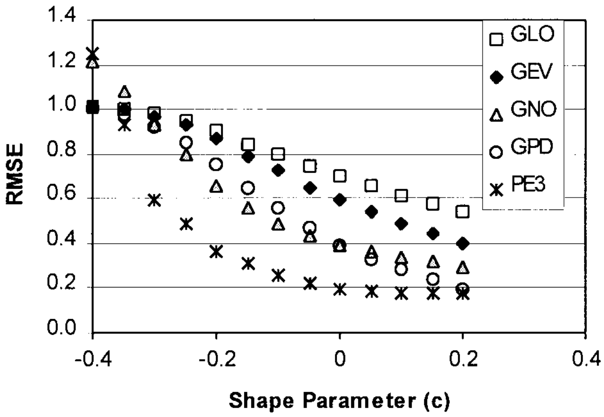

FIG. 3. RMSE of Pareto Quantiles Estimated from Different DFs (POE = 10^{-3})

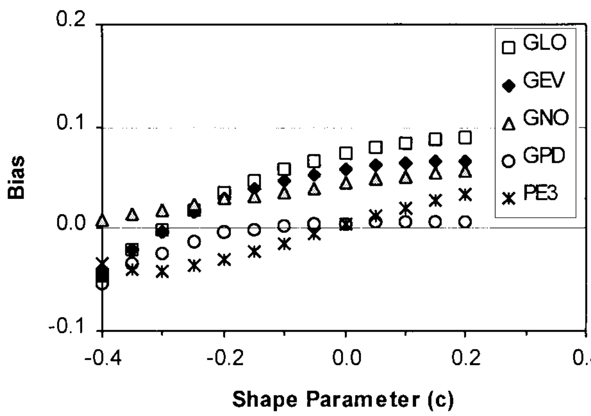

FIG. 4. Bias of Pareto Quantiles Estimated from Different DFs (POE = 10^{-2})

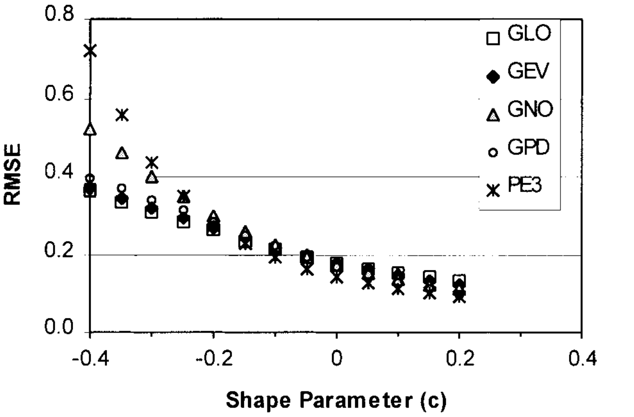

FIG. 5. RMSE of Pareto Quantiles Estimated from Different DFs (POE $= 10^{-2}$)

the same amount of data (n = 30), the magnitude of the bias and its sensitivity to the distribution shape decrease significantly, as shown in Fig. 4. In fact, all of the DFs result in a fairly small bias (<5%) for $k < 0$. The RMSE, as seen in Fig. 5, is fairly insensitive to the distribution type, and it lies in a narrow band with few exceptions.

Remarks

In the method of L-moments, the bias of an extreme quantile (POE $\approx 10^{-3}$) estimate is more sensitive than the RMSE to the distribution type fitted to the sample data. Therefore, knowledge of the correct distribution type has relevance to the minimization of bias. Despite the influence of the distribution type on quantile estimates, the dependence is not unique; i.e., more than one distribution can provide reliable estimates with a low bias and high efficiency, depending on the degree of tail equivalence between the population and the fitted DFs. In this respect, the importance of identifying the exact parent distribution from a set of DFs with identical first three L-moments is limited; any distribution reasonably close to the parent can serve the purpose. It is proposed that L-kurtosis is such a measure that can quantify approximately the degree of closeness of the sample data to a candidate distribution (Van Gelder et al. 1999b).

ASSESSMENT OF ROBUSTNESS OF L-KURTOSIS

Distribution Selection Criteria

Minimum L-Kurtosis Difference Criterion

From a set of candidate distributions having identical first three L-moments, choose that distribution whose L-kurtosis is the closest to that of the sample data; i.e., the L-kurtosis difference is minimum. It should be clarified that in the proposed L-K criterion, the sample estimate of τ_4 is used without any correction for the associated bias, which is in contrast with the proposal of Hosking and Wallis (1997). To support this, the bias and RMSE of τ_4 estimates (sample size = 30) are plotted in Fig. 6 against the L-kurtosis of the kappa parent DF used in the simulations. For the practical range of τ_4, the bias is fairly small; its maximum value is about 15%. The RMSE is almost constant with respect to τ_4. It is therefore argued that a small bias and limited sensitivity to the distribution shape, as discussed in the foregoing remarks section, provide reasonable grounds to ignore the bias correction. To evaluate the effectiveness of the L-K criterion, a set of benchmark distance measures are used in conjunction with the following principle.

Minimum Distance Criterion

From a set of candidate DFs, choose that distribution whose distance from the parent is minimum. Four measures of distance are used for comparison—(1) divergence (M-D); (2) ISE; (3) CS deviation; and (4) PPC. The KS test is not applicable here, since parameters of a fitted DF are estimated directly from the sample (Benjamin and Cornell 1970; Chowdhury et al. 1991). Although a sample can be split to apply the KS test in such cases, this approach did seem impractical in the present context due to the small sample size.

Simulation Experiment

The steps involved in the simulation experiment are briefly described as follows. A random sample was drawn from a

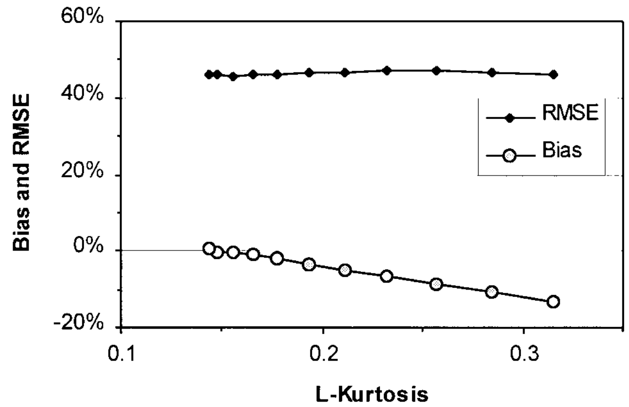

FIG. 6. Bias and RMSE of Sample Estimates of L-Kurtosis (Parent: Kappa, n = 30)

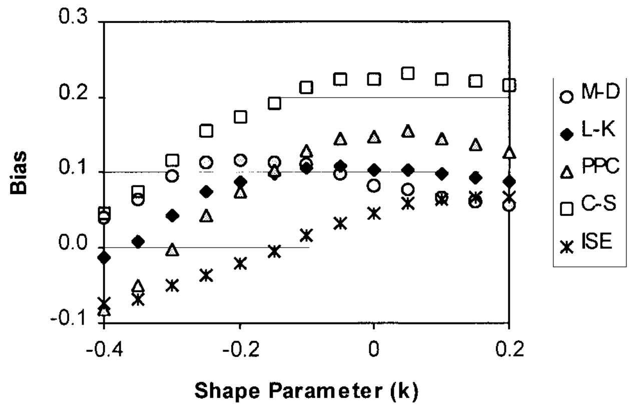

FIG. 7. Bias of Pareto Quantiles Estimated from Various DF-Selection Criteria (POE $= 10^{-3}$)

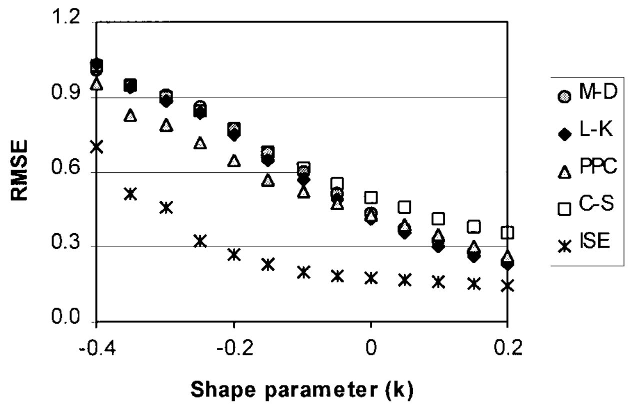

FIG. 8. RMSE of Pareto Quantiles Estimated from Various DF-Selection Criteria (POE $= 10^{-3}$)

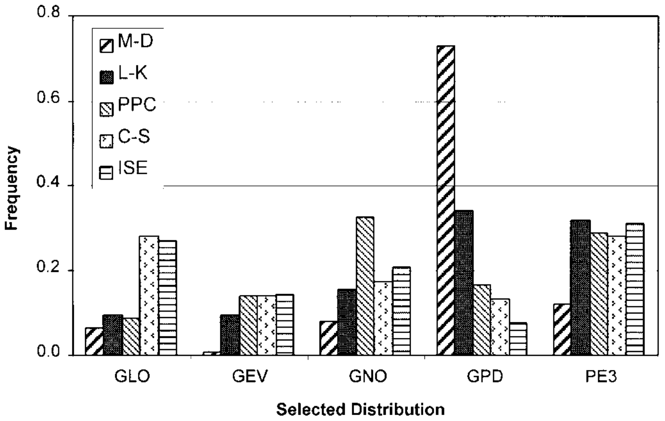

FIG. 9. Comparison of Frequencies of DF Selection from Various Criteria (Parent: Pareto)

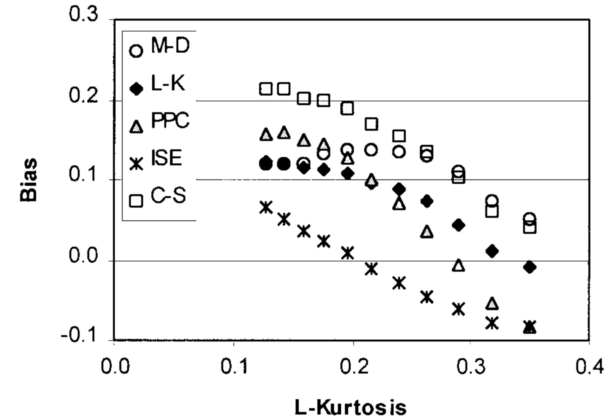

FIG. 10. Bias of Kappa Quantiles Estimated from Various DF-Selection Criteria (POE = 10^{-3})

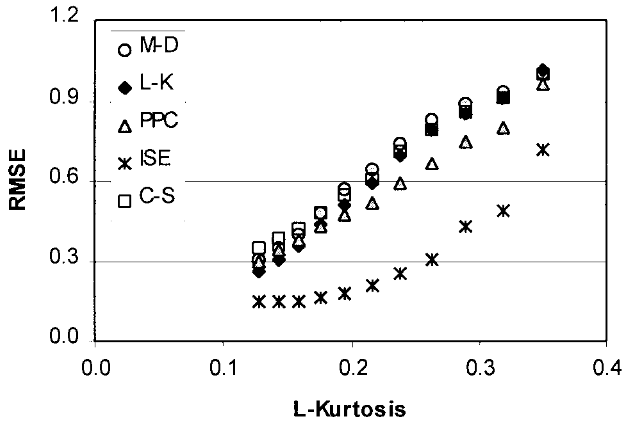

FIG. 11. RMSE of Kappa Quantiles Estimated from Various DF-Selection Criteria (POE = 10^{-3})

parent DF (Pareto and kappa) with preselected parameters. The candidate DFs (Table A1) were fitted to sample data by matching the first three sample L-moments. The divergence and ISE between the parent and the fitted DFs were computed, whereas the CS and PPC were evaluated using the simulated sample and the fitted DFs. The differences of the L-kurtosis of the fitted DFs from that of the sample were calculated to apply the L-K criterion. Using the L-K and minimum distance criteria, a representative DF was chosen to compute the required quantile. The process was repeated several times to estimate the bias and RMSE of the quantile values. Numerical computation utilized several FORTRAN routines developed by Hosking (1990).

Simulation Results: Pareto Parent

The bias and RMSE of the Pareto quantiles (POE = 10^{-3}, n = 30) were evaluated for several values of the shape parameter using 25,000 simulations in each case. As seen from Fig. 7, the quantile biases for the M-D and L-K are in fairly close agreement with the maximum of 10%. The bias for PPC increases beyond 10% when $k > 0$, and it is the highest for the CS criterion. For $k < 0$, the ISE criterion leads to the least bias ($\pm$5%), and it converges to M-D and L-K criteria for $k > 0$. Fig. 8 shows that the RMSE is somewhat insensitive to the selection criteria, with the exception of ISE, which leads to a minimum RMSE. The M-D, L-K, and PPC have an almost identical RMSE, and CS generally results in the highest value. The relatively poor performance of CS is attributed to the fact that it is more sensitive to the central shape of the distribution. To explain further, frequencies of DF selection using various criteria are plotted in Fig. 9 for $k = -0.2$; i.e., $\tau_3 = 0.43$ and $\tau_4 = 0.25$. The probability of selecting the correct DF, (i.e., Pareto), is the highest for the M-D criterion, 75%, followed by 35% for the L-K criterion. It is interesting that in the case of ISE, selection probability is the lowest for the Pareto (7%) and the highest for PE3 (30%). The remaining parameter DFs are selected in a somewhat uniform manner with 20% probability. The PPC is more inclined to select GNO (32%) and PE3 (29%), whereas CS prefers the generalized logistic (GLO) (29%) and PE3 (28%) in the DF selection.

Numerical results presented in Figs. 7 and 9 again confirm our earlier remarks about the nonuniqueness of the best-fit DF from the viewpoint of quantile estimation. This can also be interpreted as the fact that the first three sample L-moments are so effective in capturing distribution properties that minor ''fine-tuning'' is required in selecting the representative DF. This fine-tuning can be effectively accomplished by the L-K criterion.

In the comparison of the ISE and M-D criteria, it should be emphasized that they measure distance in different space. The ISE is the distance between QFs in the probability space ($0 \leq u \leq 1$), whereas M-D is the distance between probability density functions in random variable space ($0 \leq x \leq \infty$). The ISE appears to be quite effective in selecting the ''closest'' QF for minimizing the bias and RMSE, which could, however, be different than the parent density. Thus, if the objective is to identify the parent density, the M-D criterion is more effective, as expected (Fig. 9). The most notable point is that the L-K criterion seems to combine the best features of the two benchmark measures.

Simulation Results: Kappa Parent

To evaluate the accuracy of L-K criterion in a more general setting, the kappa DF was used as the parent in simulations. Shape parameters of the kappa DF were varied to cover a practical range of L-kurtosis and skewness values—$0.1 < \tau_4 \leq 0.35$, $0.2 < \tau_3 \leq 0.5$.

In Fig. 10, the variation of the quantile bias is plotted against the L-kurtosis of the kappa DF (POE = 10^{-3}, n = 30). The L-K and M-D criteria are in close agreement. The PPC and ISE tend to underestimate the quantile value when $\tau_4 > 0.2$. Similar to the Pareto case, RMSE increases with τ_4 without much sensitivity to the selection criterion (Fig. 11). The ISE criterion again results in the least RMSE.

Frequencies of DFs selected from various criteria are displayed in Fig. 12 for the case of the kappa DF with $\tau_3 = 0.35$ and $\tau_4 = 0.2$. It is interesting that both bias and RMSE resulting from the L-K and M-D criteria are fairly close, though the best-fit DFs selected by them are relatively different. The PPC and L-K criteria have a high tendency to select PE3. M-D prefers GNO, whereas ISE is more inclined to select GLO and

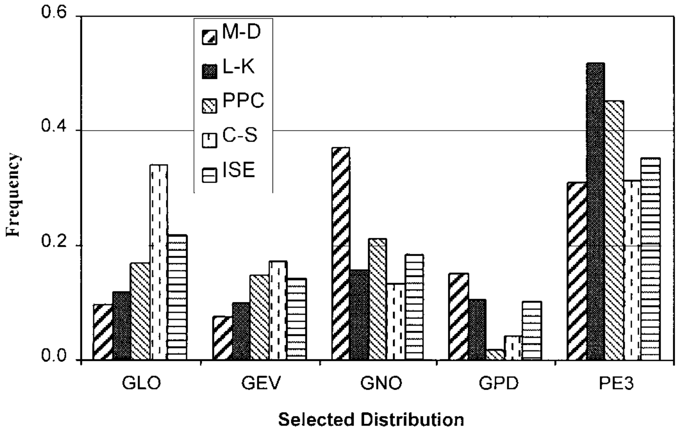

FIG. 12. Comparison of Frequencies of DF Selection from Various Criteria (Parent: Kappa)

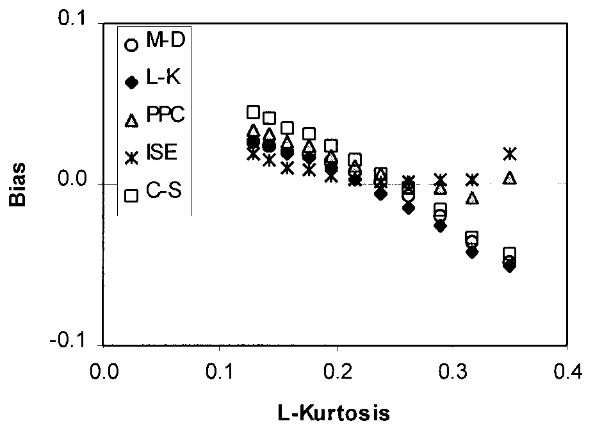

FIG. 13. Bias of Kappa Quantiles Estimated from Various DF-Selection Criteria (POE = 10^{-2})

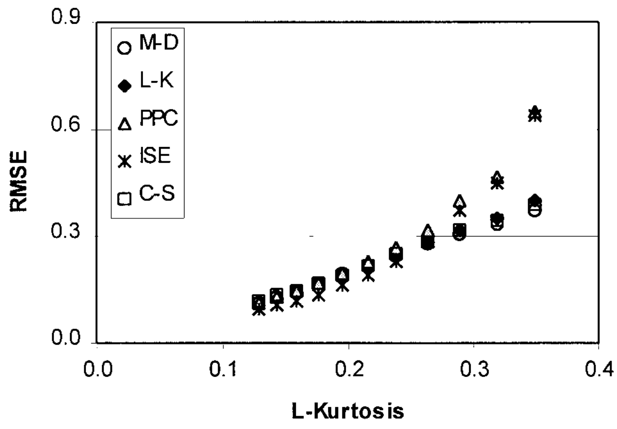

FIG. 14. RMSE of Kappa Quantiles Estimated from Various DF-Selection Criteria (POE = 10^{-2})

PE3. These observations again confirm the general conclusions of the remarks section and the simulation results: Pareto parent section.

The bias and RMSE for a lower POE of 10^{-2} are shown in Figs. 13 and 14, respectively. As expected from the results of Figs. 4 and 5, there is fairly limited sensitivity to the selection criterion. In essence, any DF selected from the candidate set of five DFs would provide adequate quantile estimates.

Application

To illustrate the application of L-K criterion, annual maximum discharge data of the Spey River collected at Boat-o-Brig Station (5751N-315W), Scotland) for the past 44 years (1952–95) were analyzed. The data were provided by the Global Runoff Data Centre, Koblanz, Germany. The sample average and coefficient of variation of the annual maximum discharge are 377 m^3/s and 43%, respectively. The sample values were normalized by the average, and five DFs were fitted by the method of L-moments. Table 1 presents the sample L-moments along with those of the fitted DFs. The last row of Table 1 shows the L-kurtosis difference, which is minimum for the GLO DF, with the following parameters: location, 0.8859; scale, 0.1708; and shape, −0.3497. Therefore, GLO is selected as the most representative DF, and is plotted along with sample data in Fig. 15. Visually, the GLO appears to fit the data quite well. Quantile values for return periods of 100, 500, and 1,000 years are obtained as 1,068, 1,767, and 2,211 m^3/s, respectively. Note that the PPC and CS criteria selected GLO and GNO, respectively.

With reference to applications of the M-D criterion, as stated earlier (statistical measures of distance section), it requires the knowledge of a prior distribution. For this reason, it has direct application in providing benchmark estimates for error estimation, as illustrated in the paper. Furthermore, in the regional frequency analysis (Hosking and Wallis 1997), the divergence between the regional parent and site-specific distributions can be calculated, and the application of the M-D criterion is anticipated to assess the goodness of fit and the degree of heterogeneity in a group of sites. This topic is an area of future investigation.

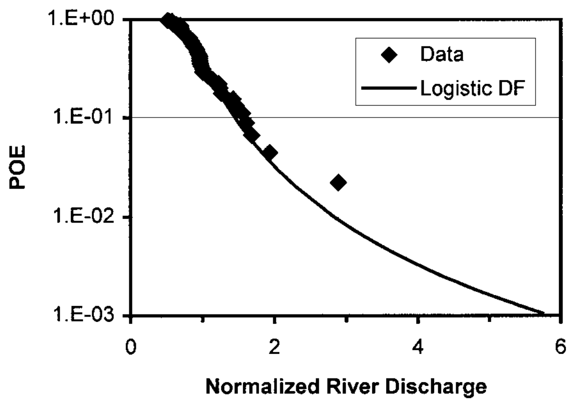

FIG. 15. Application of L-K Criterion to the Spey River Discharge Data

TABLE 1. Application of L-K Criterion to Spey River Discharges (Normalized Data)

L-moments	Sample values	L-Moments of Fitted DFs: GLO	GEV	GNO	GPD	PE3
λ_1	1.0	1.0	1.0	1.0	1.0	1.0
λ_2	0.2107	0.2107	0.2107	0.2107	0.2107	0.2107
τ_3	0.3497	0.3497	0.3497	0.3497	0.3497	0.3497
τ_4	0.2628	0.2686	0.2467	0.2195	0.1797	0.1723
[L-kurtosis difference]	$\lvert\tau_4(\text{sample} - \tau_4(\text{DF})\rvert$	0.0058	0.0161	0.0433	0.0831	0.0905

TABLE A1. Cumulative Distribution Functions of Various Distributions

Symbol	Probability distribution	$F(x)$	Range for $k < 0$
GLO	Generalized logistic	$1/(1 + e^{-y})$	$\xi + \alpha/k \leq x < \infty$
GEV	Generalized extreme value	$\exp(-e^{-y})$	$\xi + \alpha/k \leq x < \infty$
GNO	Generalized lognormal	$\Phi(y)$	$\xi + \alpha/k \leq x < \infty$
GPD	Generalized Pareto	$1 - e^{-y}$	$\xi \leq x < \infty$
PE3	Pearson type III	$G(\beta, z)/\Gamma(\beta)$	$\xi \leq x < \infty$
KAP	Kappa	$[1 - h(1 - kz)^{1/k}]^{1/h}$	$\xi + \alpha/k \leq x < \infty$ $(h < 0)$

Note: $\Phi(y)$ is the standard normal distribution. In defining the Pearson type III distributions, other notations used are as follows: $\Gamma(\beta)$ denotes the gamma function; $G(\beta, z)$ defines the incomplete gamma function, $G(\beta, z) = \int_0^z t^{\beta-1}e^{-1}\,dt$; and the mean ($\mu$), standard deviation ($\sigma$), and skewness ($\gamma$) define the remaining parameters as $\beta = 4/\gamma^2$, $\alpha = \sigma|\gamma|/2$, and $\xi = \mu - (2\sigma/\gamma)$. The bounds of the above distributions are reversed when $k > 0$. The lower bound becomes the upper bound, and the lower bound tends to be $-\infty$. The range of GPD is an exception, which varies as $\xi \leq x \leq \xi + \alpha/k$. In the case of the kappa distribution, h is the second shape parameter.

CONCLUDING REMARKS

Since L-moments of a higher order (>2) can be reliably (i.e., less bias and RMSE) estimated from small samples, their use in distribution fitting and quantile estimation is expected to improve the accuracy of quantile estimates. L-kurtosis, the normalized fourth-order L-moment, has been suggested in the literature as an effective measure of the distribution tail weight or shape.

The central objective of the paper is to assess the effectiveness of L-kurtosis in the method of L-moments for distribution fitting and extreme quantile estimation from small samples. For this purpose, the performance of the proposed L-K criterion is compared against a set of benchmark criteria developed based on the concept of minimum distance as a measure of goodness of fit. The benchmark measures used in the paper are the divergence (M-D), ISE, the chi-square test, and the probability-plot correlation. The divergence has been developed and extensively utilized in modern information theory as a mathematically comprehensive measure of probabilistic distance.

In the L-moment method, a family of five trial DFs, with three parameters each, can be fitted to a given random sample by matching the first three sample L-moments. In this context, the distribution shape has more influence on bias than does the RMSE of the quantile estimate. However, there can be more than one distribution with similar tails that can provide accurate quantile estimates, such that the importance of identifying the parent (population) distribution is somewhat diminished. Conversely, the first three sample L-moments are so effective in capturing distribution properties that they can provide multiple solutions to the representative DF for quantile estimation. Therefore, selection of the most representative DF becomes a matter of fine-tuning, which can be effectively accomplished by the L-K criterion, as illustrated in the paper.

Simulation results presented in the assessment of robustness of L-kurtosis section indicate that quantile estimates obtained from the L-K criterion are in fairly close agreement with all of the benchmark criteria for a wide range of distribution parameters. The most notable point is that the quantile bias and RMSE obtained from the L-K criterion are in agreement with those obtained from the robust criteria of the minimum divergence and ISE. It can therefore be concluded that L-kurtosis is a good indicator of distribution shape, and its use in quantile estimation is effective. The remarkable simplicity of the computation makes the L-K criterion an attractive tool for distribution fitting. The subject matter and conclusions of the present investigation would benefit from further validation.

APPENDIX I: DESCRIPTION OF PROBABILITY DISTRIBUTIONS

Three parameters commonly used to define a probability distribution are the location (ξ), scale (α), and shape (k) parameters. Various cumulative distribution functions, $F(x)$, are defined in Table A1 using the following notation: $y = -\log(1 - z)/k$, for $k \neq 0$ and $y = z$ in the case of $k = 0$, where $z = (x - \xi)/\alpha$. Further details of these distributions can be found in Hosking and Wallis (1997).

ACKNOWLEDGMENTS

We gratefully acknowledge the comments of the anonymous reviewers and appreciate a specific suggestion to include more benchmark criteria, namely, ISE and CS, which led to considerable improvements in the paper. The first writer gratefully acknowledges the financial support for this study provided by the Faculty of Civil Engineering, Technical University of Delft, The Netherlands, in the form of a visiting fellowship, and partial support from the Natural Science and Engineering Council of Canada.

REFERENCES

Benjamin, J. R., and Cornell, C. A. (1970). *Probability, statistics and decision for civil engineers*, McGraw-Hill, New York.

Beran, R. (1984). ''Chapter 30: Minimum distance procedure.'' *Non-parametric methods*, P. R. Krishnaih and P. K. Sen, eds., North-Holland, Amsterdam, 741–754.

Chowdhury, J. U., Stedinger, J. R., and Lu, L. (1991). ''Goodness-of-fit tests for regional generalized extreme value flood distributions.'' *Water Resour. Res.*, 27(7), 1765–1776.

Davison, A. C., and Smith, R. L. (1990). ''Models of exceedances over high thresholds.'' *J. Royal Statistical Soc.*, Ser. B(52), U.K., 393–442.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R. (1979). ''Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form.'' *Water Resour. Res.*, 15(5), 1049–1054.

Hartley, R. V. (1928). ''Transmission of information.'' *Bell Sys. Technol.*, 7, 535–563.

Hosking, J. R. M. (1990). ''L-moments analysis and estimation of distributions using linear combination of order statistics.'' *J. Royal Statistical Soc.*, Ser. B(52), 105–124.

Hosking, J. R. M., and Wallis, J. R. (1997). *Regional frequency analysis: An approach based on L-moments*, Cambridge University Press, London.

Johnson, R. W. (1979). ''Axiomatic characterization of the directed divergences and their linear combinations.'' *IEEE Trans. on Information Theory*, New York, 25(6), 709–716.

Jones, D. S. (1979). *Elementary information theory*, Clarendon, Oxford, England.

Jumarie, G. (1990). *Relative information: Theories and applications*, Springer, Berlin.

Kapur, J. N., and Kesavan, H. K. (1992). *Entropy optimization principles with applications*, Academic, San Diego.

Kendall, M. G., and Stuart, A. (1977). *Advanced theory of statistics*, Griffin, London.

Kullback, S. (1959). *Information theory and statistics*, Wiley, New York.

Maes, M. A. (1995). ''Tail heaviness in structural reliability.'' *Applications of statistics and probability, Proc., ICASP-7*, M. Lemaire et al., eds., Balkema, Rotterdam, The Netherlands, 2, 997–1002.

Shannon, C. E. (1949). *The mathematical theory of communication*, The University of Illinois Press, Urbana, Ill.

Shore, J. E., and Johnson, R. W. (1980). ''Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy.'' *IEEE Trans. on Information Theory*, New York, 26(1), 26–37.

Sobczyk, K., and Spencer, B. F. (1992). *Random fatigue: From data to theory*, Academic, San Diego.

Technical Advisory Committee on Water Defences (TAW). (1990). ''Probabilistic design of flood defences.'' *Rep. 141*, Centre for Civil Engineering Research and Codes, Gouda, The Netherlands.

Van Gelder, P. H., Neykov, N. M., and Neytchev, P. (1999a). ''Probability of annual maximum river discharges in north-western and central Europe.'' *Rep. Prepared for Netherlands Ministry for Transport, Public Works and Water Mgmt.*, TU Delft, Delft, The Netherlands.

Van Gelder, P. H., Pandey, M., and Vrijling, J. K. (1999b). ''The use of L-kurtosis in the estimation of extreme floods.'' *Facing the new millenium, Proc., 9th Annu. Conf. on Risk Anal.*, L. H. J. Goossens, ed., TU Delft Press, Delft, The Netherlands,182–185.