The joint probability of rainfall and runoff deficits in the Netherlands

Jules J. Beersma and T. Adri Buishand

Royal Netherlands Meteorological Institute (KMNI), PO Box 201, 3730 AE De Bilt, the Netherlands; email: beersma@knmi.nl

Abstract

The Netherlands are situated at the downstream end of the Rhine River. A large part of the country can be supplied with water from the river in the case of precipitation deficits. For the assessment of the economical damage due to drought it is necessary to consider the rainfall and river inflow simultaneously.

Transformed normal distributions as well as Gumbel distributions have been fitted to the observed precipitation and discharge deficits. The sensitivity of joint probabilities to the choice of the marginal distributions, the dependence structure and the 'failure region' is investigated. It is found that the bivariate normal distribution underestimates the probability that both the rainfall and runoff deficit are extreme due to its asymptotic independence.

Introduction

A large part of the Netherlands is situated in the delta of the Rhine River, the largest river in northwestern Europe (drainage area 185 000 km^2). The Rhine rises in the Swiss Alps and flows via France and Germany to the Netherlands, where it divides a number of times. As a result, large parts of the country can be supplied with water from the river in the case of precipitation deficits. The Rhine plays a major role in the overall water balance of the Netherlands; the amount of Rhine water that flows through the Netherlands is on average twice as large as the amount of water that the country receives as precipitation.

This paper considers the probability of drought in the Netherlands. The water balance indicates that it is important to consider the precipitation and the fresh water inflow from the Rhine simultaneously. Therefore particular attention is given to the joint distribution of the precipitation deficit (a measure of the local drought) and the discharge deficit (a measure of the lack of fresh water inflow).

To describe the joint distribution of precipitation and discharge deficits, bivariate probability distributions are fitted to 95 years of historical data. Univariate probability distributions are fitted first to the precipitation and discharge deficits separately. Subsequently those univariate distributions are combined into bivariate probability distributions. Particular attention is given to the choice between the dependence structure of the bivariate normal distribution and that of a (limiting) bivariate Gumbel distribution.

For a number of extreme years in the historical record, the return period of joint exceedances of the observed precipitation and discharge deficit is estimated with different bivariate distributions. These estimates are compared with the return periods obtained from a failure region based on the economical damage.

Drought characteristics

Two drought characteristics are considered; the precipitation deficit in the Netherlands and the discharge deficit of the Rhine River in the Netherlands. The precipitation deficit is defined as the cumulative difference between precipitation and grass reference evaporation, from April, 1 onward. When the precipitation deficit becomes negative it is reset to zero. The annual maximum precipitation deficit is the largest precipitation deficit that occurs between April, 1 and October, 1. Both for precipitation and evaporation daily values were available for the period 1906–2000, giving 95 independent annual maximum precipitation deficits. For practical reasons the daily data were transformed to decad data prior to the analysis. Decad data were obtained by dividing each calendar month into three decads; the first two decads of a month always consist of 10-day values and the third decad covers the remaining days.

Average precipitation for the Netherlands was obtained by averaging the precipitation sums from 13 stations spread over the country. The grass reference evaporation was derived from temperature and sunshine duration at a representative station.

The discharge deficit of the Rhine River was based on discharge measurements at the German-Netherlands border. Only decads for which the discharge was below 1800 m^3/s contribute to the discharge deficit. The discharge deficit was also calculated for the period April, 1 until October, 1 and was available for the same period (1906–2000) as the precipitation deficit.

Probability distributions for the precipitation deficit

Two distributions were fitted to the largest precipitation deficit in each year; the Gumbel distribution and the lognormal distribution. The parameters of these distributions were estimated by the maximum likelihood (ML) method.

Figure 1 presents Gumbel plots of the historical maxima, and the fitted distributions. Compared to the Gumbel distribution the lognormal distribution has a longer upper tail. The fitted distributions were subjected to the Anderson-

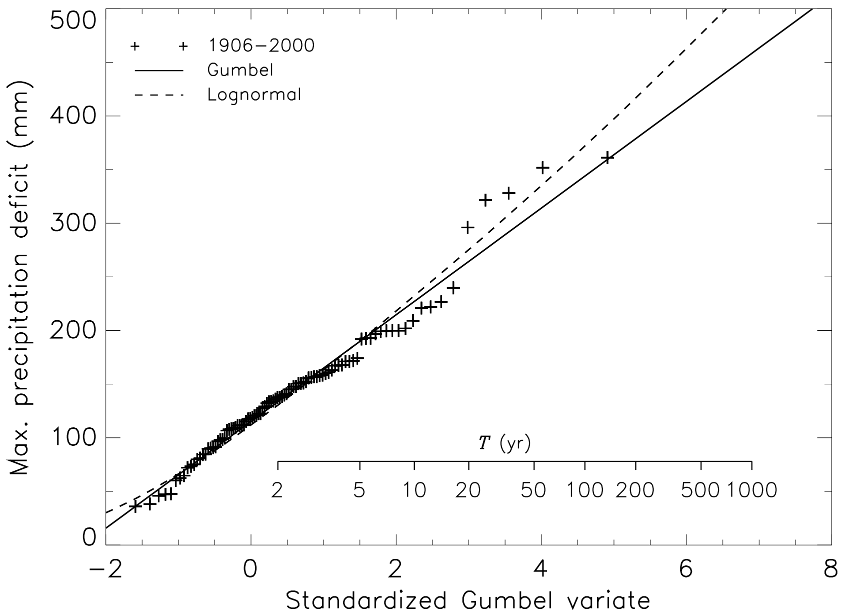

Figure 1: Ordered historical annual maximum precipitation deficits, and the fitted Gumbel and lognormal distributions

Darling (A-D) test (as in Stephens, 1986a) and the 'probability plot correlation coefficient' (ppcc) test (Vogel, 1986). These tests are sensitive to deviations in the upper tail. Both tests gave for the lognormal distribution a significant result at the 5%-level, but not at the 1%-level, while the Gumbel distribution passes both tests at the 5%-level. Thus, although the lognormal distribution seems to fit better in the tail of the distribution these tests indicate that over the whole domain the lognormal distribution does not properly fit the data while the Gumbel distribution does.

Probability distributions for the discharge deficit

The Gumbel distribution was also fitted to the annual discharge deficits. Instead of a lognormal distribution it is now assumed that the square root of the data are normally distributed.

To avoid a large influence of small values of the discharge deficit on the estimated parameters the sample was censored at a low threshold of 0.03×10^9 m^3 in the fit of the sqrt-normal distribution and at 0.6×10^9 m^3 in the case of the Gumbel distribution. For data below the threshold only the information that they are smaller than the threshold is used rather than their actual values. The parameters were estimated by the ML method, see e.g. Shumway et al. (1989) for a transformed normal distribution and Leese (1973) for the Gumbel distribution.

In Figure 2 the probability distributions of the historical data and the fitted distributions are presented. For values larger than 1.0×10^9 m^3, the fitted Gumbel

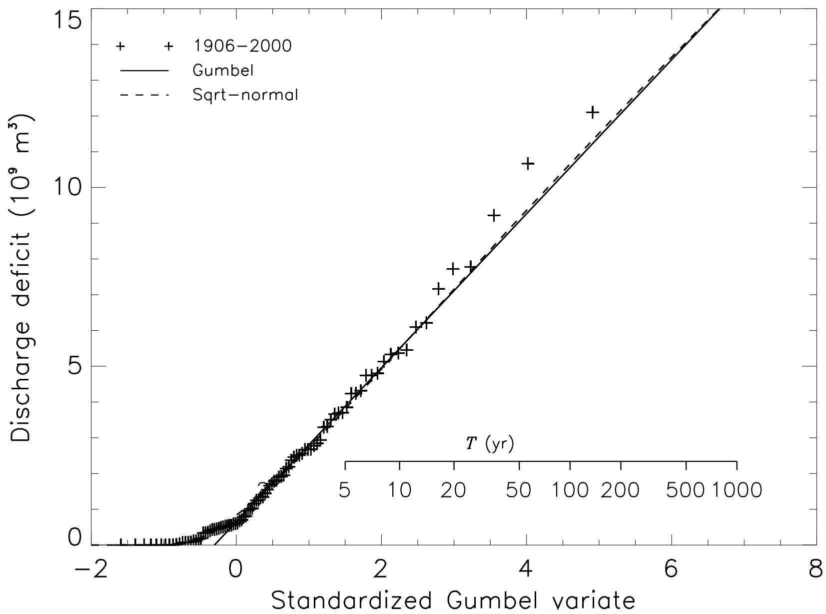

Figure 2: Ordered historical annual discharge deficits, and fitted Gumbel and sqrt-normal distributions

and sqrt-normal distributions are nearly indistinguishable. Because of the censoring the goodness-of-fit tests used in the previous section can not be applied. Both the Gumbel and the sqrt-normal distribution pass the adapted ppcc test for censored data in Stephens (1986b) at the 5%-level.

Bivariate probability distributions

So far univariate probabilities were considered. In the introduction it was already noted that from a drought impacts point of view it is much more interesting to look at joint exceedance probabilities. Drought events that have the largest economical impact are those events that have both a large precipitation deficit and a large discharge deficit. The latter makes compensation of the local water shortage by water from elsewhere in the Rhine basin very difficult.

A logical way to proceed is to combine the univariate (marginal) probability distributions into a bivariate probability distribution. In the case that the maximum precipitation deficit is described by a lognormal distribution and the discharge deficit by a sqrt-normal distribution it would be natural to consider the bivariate normal distribution. The joint density of the standardized transformed precipitation and discharge deficits is then given by:

$$\phi_2(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right] \qquad (1)$$

where ρ is the correlation coefficient of the transformed values.

A family of bivariate extensions of the Gumbel distribution is provided by the theory of multivariate extremes (e.g. Coles, 2001). A popular model in this family is the logistic model:

$$F(x, y) = \Pr(X \leq x, Y \leq y) = \exp\left[-(e^{-x/\alpha} + e^{-y/\alpha})^{\alpha}\right] \tag{2}$$

where α characterizes the strength of the dependence between X and Y; $\alpha = 1$ corresponds with independence and $\alpha = 0$ with perfect positive dependence.

The dependence structure of the bivariate normal distribution is quite different from that of the bivariate Gumbel distribution in Eq. (2). A classical result for the bivariate normal distribution with $\rho < 1$ is that its components are asymptotically independent (Sibuya, 1960), i.e.

$$\lim_{u\to\infty} \Pr(Y > u \mid X > u) = 0. \tag{3}$$

For the bivariate Gumbel distribution, however, $\Pr(Y > u \mid X > u)$ tends to $2 - 2^{\alpha}$, and this distribution is therefore asymptotically dependent if $\alpha < 1$. Note that asymptotic dependence holds for all limiting bivariate extreme value distributions.

Dependence structure. Dependence measures for bivariate extremes have been discussed by Coles et al. (1999). To remove the influence of the marginal distributions the variables X and Y are transformed to standard uniform variables, via $U = F_X(X)$ and $V = F_Y(Y)$. For the data (x_i, y_i), $i = 1, \ldots, N$ this can be achieved in a similar way using the empirical distribution functions:

$$u_i = \hat{F}_X(x_i) = \frac{\#\, x_j\,\text{'s} \leq x_i}{N+1} \text{ and } v_i = \hat{F}_Y(y_i) = \frac{\#\, y_j\,\text{'s} \leq y_i}{N+1}\ . \tag{4}$$

One measure of dependence suggested by Coles et al. (1999) is the quantity $\chi(u)$ defined by:

$$\chi(u) = 2 - \frac{\ln \Pr(U < u,\ V < u)}{\ln \Pr(U < u)} \text{ for } 0 < u \leq 1. \tag{5}$$

Independence corresponds with $\chi(u) = 0$ and complete positive dependence with $\chi(u) = 1$. For the bivariate logistic Gumbel distribution in Eq. (2), $\chi(u) = 2 - 2^{\alpha}$. Further, for sufficiently large u:

$$\chi(u) \sim \Pr(V > u \mid U > u) \tag{6}$$

and thus $\chi(u) \to 0$ as $u \to 1$ for asymptotically independent distributions like the bivariate normal distribution. $\chi(u)$ is not influenced by a monotonic increasing transformation of the data such as the log and sqrt transformation applied to the precipitation and discharge deficits to achieve normality.

An estimate of $\chi(u)$ can be constructed by substituting empirical estimates of the probabilities in the right-hand side of Eq. (5). Figure 3 presents such an

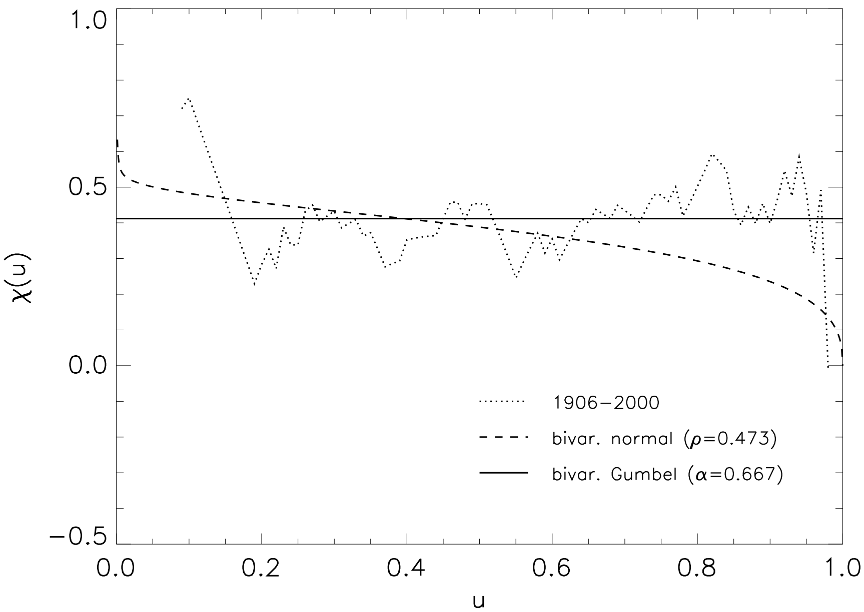

Figure 3: Dependence measure $\chi(u)$ for the historical data and the fitted bivariate distributions

estimate of $\chi(u)$ and the theoretical values for the fitted bivariate distributions. The parameters ρ and α in these distributions were estimated by the ML method, taking into account the censoring of low discharge deficits as described e.g. in Ledford and Tawn (1996). The figure shows that $\chi(u)$ is almost constant for the historical precipitation and discharge deficits. For large u, this plot of $\chi(u)$ is more in line with $\chi(u)$ for the bivariate Gumbel distribution than $\chi(u)$ for the bivariate normal distribution. For the latter $\chi(u)$ gradually decreases, but for u near 1 it abruptly drops to zero. From a physical point of view, this behavior is not very realistic since a severe drought typically extends over a large area and will thus affect the precipitation in the Netherlands as well as in the upstream Rhine catchment.

The question whether the data are asymptotically dependent or not can be investigated further by calculating for each year $T_i = \min[-1/(1-u_i), -1/(1-v_i)]$. For large z, the probability that $T_i > z$ can be approximated by the Pareto distribution (Ledford and Tawn, 1996):

$$\Pr(T_i > z) \approx cz^{-1/\eta} \tag{7}$$

where c and η are the scale and shape parameters. For the bivariate Gumbel distribution $\eta = 1$, whereas for asymptotically independent data $\eta < 1$; $\eta = (1 + \rho)/2 = 0.74$ for the bivariate normal distribution. Here η was estimated from the 70 largest values of T_i using the ML method (Hill estimator). This resulted in $\hat{\eta} = 1.12$ with a standard error of 0.13, which supports the bivariate Gumbel distribution.

Table 1: Mean return periods (yr) of joint exceedances of the observed precipitation and discharge deficits in given years for different bivariate distributions.

Year	Precipitation deficit (mm)	Discharge deficit (10^9 m^3)	Normal	Gumbel	Normal w. logistic dependence
1921	321.6	12.1	824	318	281
1976	361.1	10.7	760	296	221
1959	351.7	5.1	143	139	90
1947	296.1	7.8	142	78	65
1949	226.7	9.2	111	72	68

Return periods of joint exceedances. Both the dependence structure and the choice of the marginal distributions have an influence on the joint exceedance probabilities. Besides the bivariate normal distribution and the bivariate Gumbel distribution a third bivariate distribution is considered, namely a bivariate normal distribution with a logistic-Gumbel dependence structure. The latter is a logical combination of the other two bivariate distributions and is obtained from the bivariate Gumbel model, using the transformations:

$$\tilde{X} = \hat{H}_X^{-1}\left[\hat{G}_X(X)\right] \text{ and } \tilde{Y} = \hat{H}_Y^{-1}\left[\hat{G}_Y(Y)\right] \tag{8}$$

where $\hat{G}_X$ and $\hat{G}_Y$ are the fitted Gumbel distributions, and $\hat{H}_X$, $\hat{H}_Y$ the fitted log-normal and sqrt-normal distributions, respectively. Since these transformations are monotonic increasing, $(\tilde{X}, \tilde{Y})$ has the same logistic dependence structure as (X, Y).

For 5 extreme years in the historical record the return periods of joint exceedances of the observed precipitation and discharge deficit, i.e. $T = 1/\Pr(X > x_i, Y > y_i)$ were determined. Table 1 compares the estimates of T from the different bivariate models. The return period for the most extreme years 1921 and 1976 is about 800 years for the bivariate normal distribution. These long return periods are mainly due to the asymptotic independence of this distribution. The return periods for 1921 and 1976 reduce to about 250 years if a bivariate normal distribution with logistic-Gumbel dependence structure is assumed.

Failure regions

In the previous section return periods were found for the most extreme years that are considerably longer than the length of the historical records from which they were derived. Besides lack-of-fit and sampling variability this is due to the fact that the probability that two different variables exceed some high level simultaneously is smaller than the marginal exceedance probabilities for each of the two variables. For the fitted normal distribution with logistic-Gumbel dependence the magnitude of this effect was estimated with a Monte-Carlo experiment in which 10 000 samples of 95 years from that distribution were generated. For each 95-year sample the return periods of the joint exceedances of the simulated precipitation and discharge deficits were determined. The median of the longest return period in the 95-year samples is 320 years which is very close to the 280 years for 1921 in Table 1.

In practical applications, the joint probability that X and Y lie in a 'failure region' different from the rectangle defined by $(X > x,\ Y > y)$ might be of interest. Structures e.g. often fail if a combination of the constituent variables becomes extreme. This combination then marks the boundary of the failure region. For the assessment of droughts in the Netherlands it is useful to base the failure region on the economical damage D_{E}.

The economical damage from 7 historical years (1949, 1959, 1967, 1976, 1985, 1995 and 1996) reveals that D_{E} can be approximated as:

$$D_{\mathrm{E}} = ax + by + c \tag{9}$$

with x the precipitation deficit and y the discharge deficit. The regression coefficients a, b and c were estimated by a least-squares fit. Let x_i and y_i be the observed precipitation and discharge deficits for the year of interest. Events with a precipitation and discharge deficit in the region above the line through (x_i, y_i) and with slope $-a/b$ should then be considered as more extreme in terms of economical damage. For the years 1976, 1959 and 1949, Fig. 4 compares the boundary of this failure region with the rectangle $(X > x_i,\ Y > y_i)$. The slope of the bounding line indicates that the economical damage is relatively more sensitive to the precipitation deficit, which was not unexpected. Table 2 presents, for each of the historical years in Table 1, the return periods of joint events in the failure region based on Eq. (9). The estimated return periods in Table 2 are much shorter than those in Table 1, in particular for 1921 and 1976. Using a failure region related to the economical damage gives on average the longest return period for 1976 while in Table 1 the longest return period is found for 1921. This is a result of the relatively smaller contribution of the discharge deficit to the economical damage (see Fig. 4). In Table 1 the return periods are longest for the bivariate normal distribution while in Table 2 the longest return periods are found for the bivariate Gumbel distribution. The shortest return periods are found in both tables for the bivariate normal distribution with logistic-Gumbel dependence, but the difference with the standard bivariate normal distribution is much smaller in Table 2. This is in line with results of Tawn (1988) and Coles and Tawn (1994) that the sensitivity of joint probabilities to assumptions about the dependence structure varies considerably with the type of failure region.

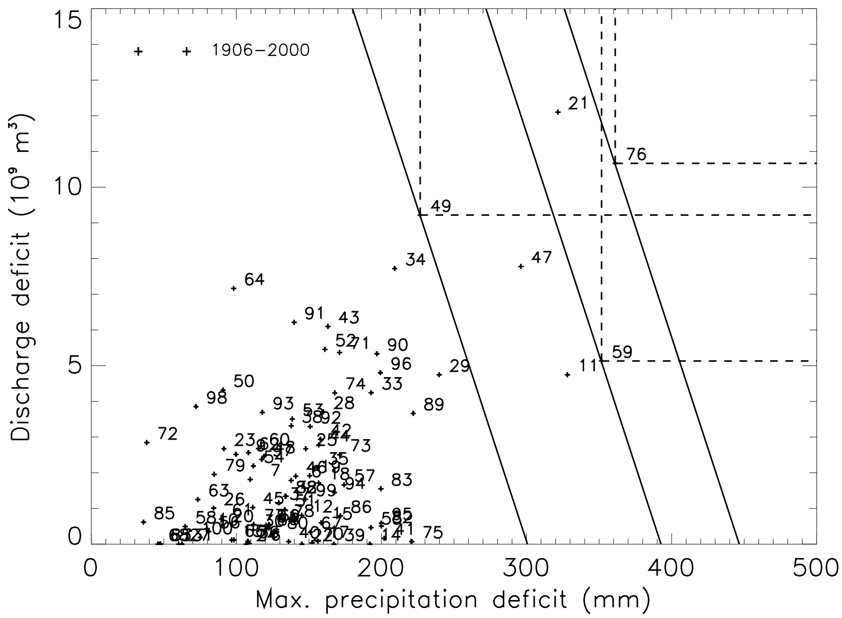

Figure 4: Failure region related to the economical damage (Eq. 9) and the rectangle ($X > x_i$, $Y > y_i$) for the historical years 1976, 1959 and 1949 (indicated as 76, 59 and 49).

Table 2: Mean return periods (yr) of situations that the precipitation and discharge deficits are more extreme than the observed deficits in the given years in terms of economical damage (Eq. 9) for the different bivariate distributions.

Year	Precipitation deficit (mm)	Discharge deficit (10^9 m^3)	Normal	Gumbel	Normal w. logistic dependence
1921	321.6	12.1	99	113	79
1976	361.1	10.7	147	172	110
1959	351.7	5.1	66	75	55
1947	296.1	7.8	41	46	36
1949	226.7	9.2	17	19	17

Discussion and conclusion

Different probability distributions were fitted to the annual maximum precipitation deficit in the Netherlands and the annual discharge deficit of the Rhine River. It was found that the degree of association between large values is too weak if the dependence structure of a bivariate normal distribution is assumed. This results in a strong underestimation of the probabilities of joint exceedances of extreme values. The joint occurrence of large values is better described by the dependence structure of a limiting Gumbel distribution. The use of this dependence function was studied with Gumbel and transformed normal marginals. The latter describes the upper tail of the precipitation deficit distribution better, leading to shorter return periods between extreme bivariate events than the Gumbel distribution. The assumption of Gumbel marginals was, however, not rejected by the Anderson-Darling and the ppcc tests. For the most extreme year in terms of economical damage, 1976, the estimated exceedance probability was once in 172 years for the bivariate Gumbel distribution and once in 110 years for the transformed normal distribution with logistic-Gumbel dependence. A study of the uncertainty of these estimates was beyond the scope of this paper.

Acknowledgement. The work was performed for the 'Drought studies of the Netherlands' in cooperation with the Institute for Inland Water Management and Waste Water Treatment (RIZA), Lelystad.

References

Coles, S.G., 2001: *An introduction to statistical modeling of extreme values.* Springer, London.

Coles, S.G., J. Heffernan, and J.A. Tawn, 1999: Dependence measures for multivariate extremes. *Extremes*, **2,** 339–365.

Coles, S.G. and J.A. Tawn, 1994: Statistical methods for multivariate extremes: an application to structural design (with discussion). *Appl. Statist.*, **43,** 1–48.

Ledford, A. and J.A. Tawn, 1996: Statistics for near independence in multivariate extreme values. *Biometrika*, **83,** 169–187.

Leese, M.N., 1973: Use of censored data in the estimation of Gumbel distribution parameters for annual maximum flood series. *Water Resour. Res.*, **9,** 1534–1542.

Shumway, R.H., A.S. Azari, and P. Johnson, 1989: Estimating mean concentrations under transformation for environmental data with detection limits. *Technometrics*, **31,** 347–356.

Sibuya, M., 1960: Bivariate extreme statistics. *Ann. Inst. Statist. Math.*, **11,** 195–210.

Stephens, M.A., 1986a: *Tests based on EDF statistics*, pages 97–193. In: Goodness-of-fit Techniques (ed. R.B. D'Agostino and M.A. Stephens). Marcel Dekker, Inc., New York and Basel.

Stephens, M.A., 1986b: *Tests based on regression and correlation*, pages 195–233. In: Goodness-of-fit Techniques (ed. R.B. D'Agostino and M.A. Stephens). Marcel Dekker, Inc., New York and Basel.

Tawn, J.A., 1988: Bivariate extreme value theory: models and estimation. *Biometrika*, **75,** 397–415.

Vogel, R.M., 1986: The probability plot correlation test for the normal, lognormal, and Gumbel distributional hypotheses. *Water Resour. Res.*, **22,** 587–590.