Title
Analysis of extreme hydrologic events with Gumbel distributions: marginal and additive cases

Journal
Stochastic Environmental Research and Risk Assessment, 13(4)

ISSN
1436-3240

Authors
Loaiciga, HA
Leipnik, RB

Publication Date
1999-08-01

DOI
10.1007/s004770050042

Peer reviewed

Stochastic Environmental Research and Risk Assessment 13 (1999) 251–259 © Springer-Verlag 1999

Analysis of extreme hydrologic events with Gumbel distributions: marginal and additive cases

H. A. Loaiciga, R. B. Leipnik

Abstract The importance of the Gumbel probability distribution for the description of extreme hydrologic events is examined in this article. The key findings of this work are: (1) an iterative method of least squares was developed and found to be well-suited for the efficient fitting of the two-parameter Gumbel distribution to hydrologic extremes; (2) negative truncation is necessary to adequately describe hydrologic minima (non-negative) data, while the standard Gumbel distribution for maxima is well-suited for modeling extreme (large) hydrologic events; (3) the distribution function of the sum of two independent Gumbel variables, of importance in hydrology, has been derived and successfully applied to spring flow data. Several examples that involve the modeling of hydrologic extremes are presented and analyzed.

Introduction

The study of extreme hydrologic events has received considerable attention in the hydrology literature. A significant portion of past research on hydrologic extremes has focused on the probability distribution of precipitation and streamflow (National Research Council, 1988; Loaiciga and Marino, 1991). That research has been largely motivated by the importance of extreme (very large) hydrologic events in the planning and design of flood control structures. The opposite end of the spectrum, that is, low values, has also been the subject of long-standing

H. A. Loaiciga (✉)
Department of Geography,
University of California,
Santa Barbara, California 93106, USA
Tel.: 805-893-8053
Fax: 805-893-3146
e-mail: hugo@geog.ucsb.edu

R. B. Leipnik
Department of Mathematics,
University of California,
Santa Barbara, California 93106, USA
e-mail: leipnik@math.ucsb.edu

Partial support for this work was provided by United States Environmental Protection Agency Cooperative Agreement CR 824540-01-0 and by grant ATM-9711491 from the United States National Science Foundation. The findings and conclusions of this work represent the views of the authors and do not imply endorsement by the named agencies.

interest in hydrology, mostly in association with studies of low-flow for water supply, hydroelectric generation, navigation, and stream habitat support (Loaiciga and Marino, 1988).

The Gumbel distribution (Gumbel, 1958; Galambos, 1978; Leipnik, 1998) has been proposed, and used, to describe extreme phenomena in various disciplines, hydrology included. Unlike previous research on the distribution of extremes, however, this article deals with the sum of two Gumbel variables, either maxima or minima, rather than with the analysis of a single variable. In hydrology, the rationale for studying the sum of two Gumbel variables, which are assumed to be statistically independent herein, is justified on several grounds. For example, streamflow from two rivers, or river basins, may combine to provide water supply to a single region. Such is the case of the Colorado river (which drains through seven western states in the United States) and the Sacramento river of California, whose combined flow determines a significant fraction of the water supply for agriculture, the environment, and urban centers (Loaiciga et al., 1992a; 1993). Notice that the streamflows in the aforementioned rivers do not physically combine; yet, water supply on a regional basis is determined by their combined streamflow (California Department of Water Resources, 1998). In this particular case, the occurrence of simultaneous low flow is of utmost importance given the regional water supply reliance on combined available streamflow. Another relevant situation that involves the sum of hydrologic extremes concerns the confluence of two rivers whose combined flows creates flood hazards. The Sacramento and American rivers in northern California illustrate the latter situation. Other examples are presented below.

The Gumbel distribution has been found to provide excellent goodness-of-fit to extreme variables in many cases (Loaiciga et al., 1992b). A few examples will be given below to reaffirm this argument. Moreover, the distribution function of the sum of Gumbel variables is more tractable mathematically than those of competing models, such as the log-normal distribution (Leipnik, 1991). Among alternative extreme probability models, the Gumbel distribution has been found by the authors to be the only one that yields a distribution function for the sum of two variables which is prone to calculation without excessive computational effort, as shown below.

The remainder of this article presents evidence to support the choice of the Gumbel distribution as a suitable model of hydrologic extremes, demonstrates the mathematical procedure for deriving the distribution function of the sum of two Gumbel variables (maxima and minima), and provides an application example that concerns the sum of spring flow maxima. Parameter estimation, data fitting and truncation, and computational issues are also examined along the way.

The Gumbel distribution in extreme hydrologic characterization

The Gumbel distribution is a two-parameter model. For maxima, the probability density function takes the following form:

$$f(x) = \frac{1}{b}\, e^{-\left(\frac{x-a}{b}\right)}\, e^{-e^{-\left(\frac{x-a}{b}\right)}} \qquad -\infty \leq x \leq \infty \tag{1}$$

in which a and b are parameters, $b > 0$. From Eq. (1) the Gumbel distribution function for maxima is:

$$P(X \leq x) \equiv F(x) = e^{-e^{-\left(\frac{x-a}{b}\right)}} \qquad -\infty \leq x \leq \infty \tag{2}$$

The Gumbel probability density function for minima is given by the following expression:

$$g(x) = \frac{1}{b}\ e^{\left(\frac{x-a}{b}\right)}\ e^{-e^{\left(\frac{x-a}{b}\right)}} \qquad -\infty \le x \le \infty \tag{3}$$

From which the Gumbel distribution function for minima is obtained:

$$P(X \le x) \equiv G(x) = 1 -\ e^{-e^{\left(\frac{x-a}{b}\right)}} \qquad -\infty \le x \le \infty \tag{4}$$

Fitting of the Gumbel distribution function for maxima, $F(x)$, to data is most conveniently done by the method of least squares. By taking natural logarithms twice in Eq. (2), assisted by proper algebraic manipulation, the following linear transformation of the Gumbel distribution arises:

$$y = -\frac{x}{b} + \frac{a}{b} = b^{*}\ x + a^{*} \qquad -\infty \le x \le \infty \tag{5}$$

in which $y = \ln\{\ln(1/F(x)\}$. For a sample of size n, the cumulative probability $F(x)$ is first estimated by the plotting positions $m/(n+1)$, $m = 1, 2, \ldots, n$ in which m is the rank of the observations in the sample after sorting them in descending order. The parameters a^{*} and b^{*} are initially estimated by least squares, from which the distribution parameters are obtained at once from $b = -1/b^{*}$ and $a = -a^{*}/b^{*}$. Once the parameters a, b are estimated, the probability $F(x)$ is recalculated analytically (by Eq. (2)) and the least-squares method is applied again to Eq. (5) to obtain updated a, b parameters.

Figure 1 shows the results of applying the described least squares fitting procedure to annual maximum 1-hr rainfall depths measured at the Chicago Airport (Dingman, 1994) from 1949 to 1972. The fitted linear equation was $y = -0.984x + 2.997$ (R^2 = 0.97), where x is in cm; therefore, a = 3.046, b = 1.0163. It is straightforward to verify that the fitted distribution of annual maximum 1-hr rainfall depth is such that $P(X > 0) \to 1$, i.e., the mass of the distribution function lies almost entirely above zero. This means that the fact that the distribution's domain includes negative real numbers, which are evidently infeasible to represent rainfall, introduces only negligible bias in the estimation of $P(X < x)$, when $x > 0$.

Figure 2 shows the least-squares fitted Gumbel distribution to annual maximum spring flow (in $m^3\ hr^{-1}$) in Comal Springs, Texas, for the years 1979 to 1989.

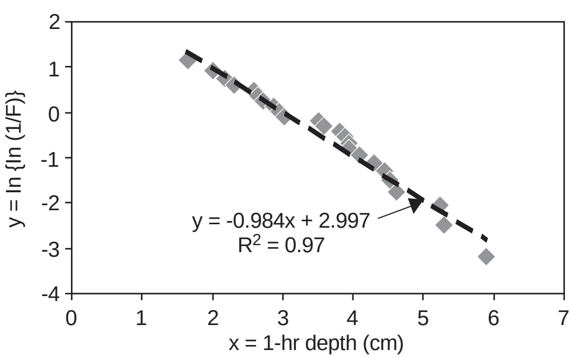

Fig. 1. Fitted distribution to 1-hr annual maximum depths at Chicago Airport (data from Dingman, 1994)

In this case, the linearly transformed Gumbel distribution was estimated to be $y = -8.34 \cdot 10^{-5}x + 2.78$ ($R^2 = 0.96$), from which the distribution parameters are $a = 33{,}333.3$, $b = 11{,}990.0$. The probability $P(X > 0) \rightarrow 1$ in this case also, which discards estimation bias that could otherwise be introduced by the distribution's negative domain.

An additional example of a fitted Gumbel distribution is presented in Fig. 3, in which the fitted data correspond to annual maximum spring flow in San Marcos Springs, Texas. The estimated least-squares transformed distribution was $y = -2.76 \cdot 10^{-4}x + 3.56$ ($R^2 = 0.95$), from which the distribution parameters were obtained: $a = 12{,}898.2$, $b = 3{,}623.2$.

The fitted Gumbel distributions to annual maximum spring flows at Comal and San Marcos Springs will be used below to illustrate some of the results concerning the sum of two Gumbel variables. These two spring flows merge downstream from their source areas in the discharge zone of the Edwards Balcones Fault Zone aquifer in south central Texas (Loaiciga et al., 1998). The Edwards aquifer is a large karstified formation spanning 15,600 km^2 in surface area. The San Marcos Springs and the Comal Springs derive their spring flow from two different source areas in the aquifer. Thus, there is a degree of independence in the magnitude of these spring flows, although the theoretical assumption of statistical independence is only approximate for reasons to be elaborated upon later.

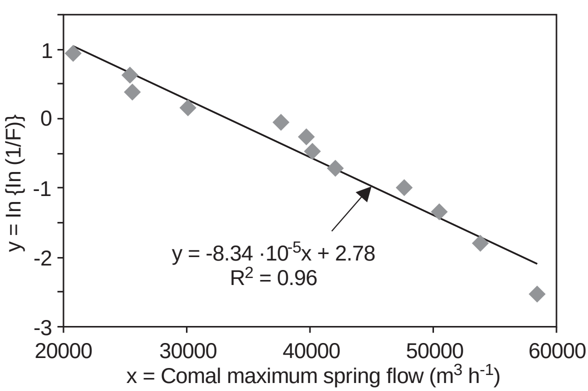

Fig. 2. Fitted distribution to annual maximum spring flow at Comal Springs, Texas

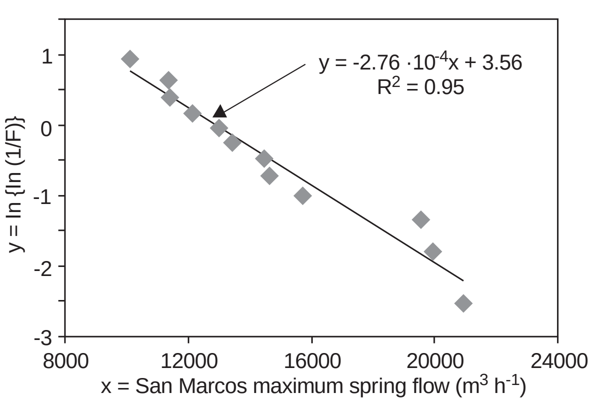

Fig. 3. Fitted distribution to annual maximum spring flow at San Marcos Springs, Texas

Truncation of the Gumbel distribution function in the case of minimum variables

Figure 4 shows the fitted Gumbel distribution (in its linearly transformed form) to the distribution of annual minimum spring flow (in $m^3\ hr^{-1}$) in Comal Springs, Texas, from 1947 through 1959. Truncation was needed in this case, as the standard distribution of Gumbel minima given in Eq. (4) did not provide an adequate fit to the data. In general, when dealing with hydrologic minima, it is necessary to impose the physical constraint that the variable of interest, be it rainfall, streamflow, or something else, must be positive. Thus, the truncation constraint $X > 0$ is imposed to yield the following truncated Gumbel distribution function for minima (Loaiciga et al., 1992):

$$P(X \leq x | X > 0) \equiv G_T(x) = \frac{e^{-e^{-\frac{a}{b}}} - e^{-e^{\left(\frac{x-a}{b}\right)}}}{e^{-e^{-\frac{a}{b}}}} \quad x > 0 \tag{6}$$

The truncated distribution in Eq. (6) can be transformed into an equivalent linear form suitable for least squares estimation. Taking the natural logarithm twice in Eq. (6), coupled with suitable algebraic manipulation, the linear transformation of the truncated Gumbel distribution for minima becomes:

$$y = \frac{x}{b} - \frac{a}{b} = b^* x + a^* \quad x > 0 \tag{7}$$

in which

$$y = \ln\left\{\ln\left[\frac{1}{P(X > 0)(1 - G_T)}\right]\right\} \tag{8}$$

The procedure applied to fit the annual minimum spring flow data in Fig. 4 requires initial estimates of the distribution parameter a, b to approximate $P(X > 0) = \exp[-\exp(-a/b)]$. This is done by the method of moments, relying on the fact that the mean and variance of the distribution function of Gumbel minima (see Eq. (4)) are $a - bC$ and $b^2(\pi^2/6)$, respectively, in which C = 0.577215... is Euler's constant. From the data, the sample mean and variance are calculated, and from these, the parameters a and b are solved for. The truncated cumulative distribution function, G_T, which is given in Eq. (6), is

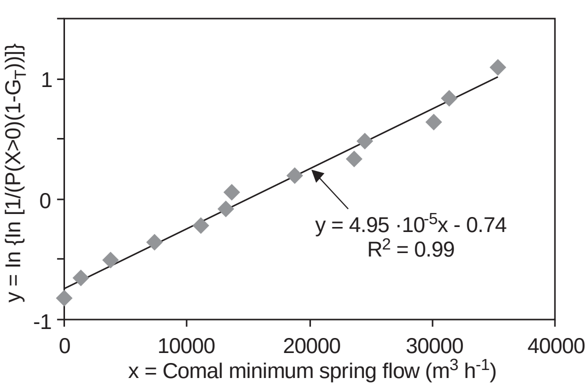

Fig. 4. Fitted distribution to truncated annual minimum spring flow at Comal Springs, Texas

initially estimated by the plotting positions $m/(n + 1)$, in a manner identical to that used in the procedure applied to fitting Gumbel maxima. Next, Eq. (7) is fitted by least squares, from which the parameters a, b are updated. $P(X > 0)$ and G_T are also updated (the update of G_T is done by means of Eq. (6)) and a new least-squares fit of Eq. (7) is carried out. Parameter updates and least-squares fitting are iterated until parameter convergence is achieved. Our experience with the data sets examined (annual minimum spring flows at Comal Springs and San Marcos Springs, Texas) indicates that convergence is achieved in three to four iterations.

Figure 4 shows that the least squares truncated Gumbel distribution fitted to the Comal Springs annual minimum spring flow was $y = 4.95 \cdot 10^{-5}x - 0.74$ ($R^2 = 0.99$), from which the distribution parameters were estimated as $a = 14{,}949.5$, and $b = 20{,}202$. In this instance, the probability $P(X > 0) = \exp[-\exp(-a/b)] = 0.62$, which implies that the mass of the non-truncated Gumbel distribution which falls below zero (i.e., $P(X < 0)$) is approximately 0.38, a non-negligible value. Since, in actuality, spring flow is a positive variable, the fitting of a non-truncated Gumbel distribution to spring flow minima would result in large errors. The same conclusion applies to other positive-valued hydrologic minima in which $P(X < 0)$ does not approach zero.

The distribution of the sum of two independent Gumbel variables

There are several formal methods that in theory could be utilized to derive the distribution function of the sum ($Z = X + Y$) of two independent Gumbel variables X and Y with probability density functions $f_X(x)$ and $f_Y(y)$, respectively. One of them involves an inverse transformation on a vector-valued function of (X,Y), which involves X + Y (see, e.g., Bickel and Doksum, 1977, p. 10 and 11). Another classical approach is by convolution of the distributions of X and Y (see, e.g., Ross, 1993, p. 53). When applied to two-parameter Gumbel variables, either minima or maxima, these two methods yield complex results that cannot be resolved in terms of a well-defined probability density function for the sum of two variables. A third approach to obtain the distribution density of two independent Gumbel variables (and exploited in Leipnik, 1998) is by means of the Fourier transform pair defined by the probability density function, f_Z, and the characteristic function, ϕ_Z, of the sum of Gumbel variables $Z = X + Y$ (Lukacs, 1960). Let t be a real variable, i be the complex number $i^2 = -1$, $E(\bullet)$ denote the statistical expectation operator. Then:

$$\phi_Z(t) \equiv E\left(e^{it(X+Y)}\right) = \int_{-\infty}^{\infty} e^{itz} f_Z(z)dz = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{it(x+y)} f_X(x) f_Y(y) dx\, dy \quad (9)$$

and

$$f_Z(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itz}\,\phi_Z(t)dt \quad (10)$$

Equation (9) implies that, due to the independence of X and Y, the characteristic function of the their sum Z is simply the product of the individual characteristic functions, which are known. In the case of the sum of Gumbel maxima, the characteristic function of Z is then:

$$\phi_Z(t) = \phi_X(t)\ \phi_Y(t) = e^{it(a_1+a_2)}\Gamma(1 - itb_1)\Gamma(1 - itb_2) \quad (11)$$

in which (a_1, b_1) and (a_2, b_2) are the parameters of the distributions of X and Y, respectively, and $\Gamma(\bullet)$ is the gamma function (see e.g., Gradshteyn and Ryzhik, 1994, for a treatment of the gamma function). If X and Y are truncated minima, the characteristic function of their sum is:

$$\phi_Z(t) = \frac{e^{it(a_1+a_2)}}{e^{-\left(e^{-\frac{a_1}{b_1}}+e^{-\frac{a_2}{b_2}}\right)}} \, \Gamma(e^{-\frac{a_1}{b_1}},\ 1+itb_1)\,\Gamma(e^{-\frac{a_2}{b_2}},\ 1+itb_2) \tag{12}$$

in which $\Gamma(\bullet, \bullet)$ is the incomplete gamma function.

The probability density function, $f_Z(z)$, of the sum of two independent Gumbel maxima is derived by applying the Fourier transform (10) to the characteristic function given in Eq. (11). The distribution function $F_Z(z)$ is then obtained by integration of $f_Z(z)$ by the method of residues. The result is (where $z' = z - a_1 - a_2$):

$$F_Z(z) = 1 - \pi \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(n-1)!}\left[\frac{b_2}{b_1} e^{-\frac{n}{b_1}z'} \cdot \frac{\csc(n\pi\frac{b_2}{b_1})}{\Gamma(1+n\frac{b_2}{b_1})} + \frac{b_1}{b_2} e^{-\frac{n}{b_2}z'} \cdot \frac{\csc(n\pi\frac{b_1}{b_2})}{\Gamma(1+n\frac{b_1}{b_2})}\right] \tag{13}$$

in which csc = 1/sin and the argument of the function is in radians. The distribution function of the sum of two truncated Gumbel minima can be obtained in a similar fashion as done with the sum of two Gumbel maxima, although the results involve a higher degree of complexity than that involved in deriving Eq. (13). The authors are at present developing the distribution function of the sum of two truncated Gumbel minima.

The calculation of probabilities via Eq. (13) presents difficulties when $z \rightarrow 0$ because the summation must be carried out for a large number of terms, say, $n \geq 100$ to preserve accuracy in the calculations. High-precision arithmetic is thus required to prevent overflow in the computations. To illustrate the application of Eq. (13), we have used the distribution functions fitted to the Comal Springs annual maximum spring flow (see Fig. 2) and to the San Marcos Springs annual maximum spring flow (shown in Fig. 3). Annual maximum spring flows at Comal Springs and at San Marcos Springs are not strictly independent, as theory assumes in the previous developments, in spite of the fact that their drainage basins are different. This is due to the fact that many storms that recharge their drainage basins are of a mesoscale nature, which introduces a pattern of climatic similarity in the recharge feeding the two springs. Nevertheless, annual maximum spring flows at Comal Springs and San Marcos Springs do not necessarily occur simultaneously due to different lag times between the time of recharge and the time of discharge at each spring. [Loaiciga et al. (1992) have examined various formal statistical methods to test for statistical independence of hydrologic variables.] Thus, the analysis of the sum of Comal and San Marcos annual maximum spring flows is meaningful only if it is interpreted in a broader context. Specifically, the joint occurrence of annual spring flow maxima is a temporal stochastic process, which happens in a given time period with a probability p, which can be estimated from historical records. Thus, the probability of attaining a sum of maxima less than z is given by $p \cdot F_Z(z)$.

Figure 5 shows the calculated distribution function $F_Z(z)$ of the sum of annual maximum spring flow at Comal Springs and San Marcos Springs. Equation (13) was programmed in Fortran 90 and special functions (cosecant, gamma, factorial,

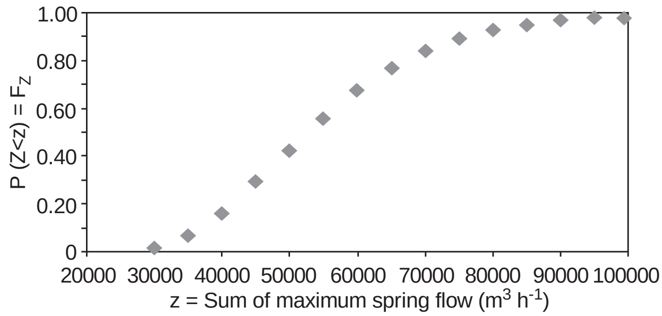

Fig. 5. The distribution function of the sum of Gumbel-distributed annual maximum spring flows at San Marcos Springs and Comal Springs, Texas

exponential) were evaluated in double precision. The International Mathematical Subroutine Library (IMSL) was used to evaluate the gamma function. It can be seen in Fig. 5 how the distribution becomes negligible below 30,000 $m^3\ hr^{-1}$. Equation (13) allows the calculation of quantiles, such as, for example, Q_{10} and Q_{100}, the 10- and 100-yr combined spring flows, which are equal to 76,000 and 105,000 $m^3\ hr^{-1}$, respectively.

Conclusions

The importance of the Gumbel probability distribution for the description of extreme hydrologic events has been examined in this paper. The key findings of this article are: (1) an iterative method of least squares was developed and found to be well-suited for the efficient fitting of the two-parameter Gumbel distribution to hydrologic extremes; (2) truncation of negative values is necessary to adequately describe hydrologic minima (non-negative) data, while the standard Gumbel distribution for maxima is well-suited for modeling extreme (large) hydrologic events; (3) the distribution function of the sum of two independent Gumbel variables, of importance in hydrology, has been derived and successfully applied to spring flow data. Future work involves the development of the distribution function of the sum of two Gumbel (truncated) minima as well as the analysis of the temporal joint occurrence of hydrologic maxima and minima.

References

Bickel, P.J. and K.A. Doksum. 1977. Mathematical Statistics. Holden-Day, Inc., San Francisco, California.

California Department of Water Resources. 1998. The California Water Plan Update, Vols. 1 and 2. Bulletin 160–98, Sacramento, California.

Dingman, S.L. 1994. Physical Hydrology. McMillan Publishing Co., New York.

Galambos, J. 1978. The Asymptotic Theory of Extreme Order Statistics. John Wiley and Sons Publishing Co., New York.

Gradshteyn, I.S. and I.M. Ryzhik. 1994. Table of Integrals, Series, and Products. Academic Press, San Diego, California.

Gumbel, E.J. 1958. Statistics of Extremes. Columbia University Press, New York.

Leipnik, R.B. 1991. On lognormal random variables. Journal of the Australian Mathematical Society, Series A, 32, 327–347.

Leipnik, R.B. 1998. Concurrent catastrophes: the sum of two independent Gumbels. Comptes Rendus Mathematical Reports of the Academy of Sciences of Canada, 20(3), 91–96.

Loaiciga, H.A. and M.A. Marino. 1988. Fitting minima of flows via maximum likelihood. Journal of Water Resources Planning and Management, 114(1), 78–90.

Loaiciga, H.A. and M.A. Marino. 1991. On the recurrence interval of geophysical events. Journal of Water Resources Planning and Management, 117(2), 260–272.

Loaiciga, H.A., Michaelsen, J., Garver, S., Haston, L., and R.B. Leipnik. 1992a. Droughts in river basins of the United States, Geophysical Research Letters, 19(20), 2051–2054.
Loaiciga, H.A., Michaelsen, J., and P.F. Hudak. 1992b. Truncated hydrologic distributions in hydrologic analysis. Water Resources Bulletin, 28(5), 853–863.
Loaiciga, H.A., Haston, L. and J. Michaelsen. 1993. Dendrohydrology and long-term hydrologic phenomena. Reviews of Geophysics, 31(2), 151–171.
Loaiciga, H.A., Maidment, D., and J.B. Valdes. 1998. Climate change impacts on the water resources of the Edwards Balcones Fault Zone Aquifer, Texas. Report ASCE/USEPA Co-operative Agreement CR824540-01-0, Department of Geography, University of California, Santa Barbara, California.
Lukacs, E. 1960. Characteristic Functions. Hafner Publishing Co., New York.
National Research Council. 1988. Estimating Probabilities of Extreme Floods: Methods and Recommended Research. National Academy Press, Washington, D.C.
Ross, S.M. 1993. Introduction to Probability Models. Academic Press, San Diego, California.