ENVIRONMETRICS
Environmetrics 2002; **13**: 811–819 (DOI: 10.1002/env.483)

The bivariate lognormal distribution for describing joint statistical properties of a multivariate storm event

Sheng Yue*,†

Meteorological Service of Canada-Ontario Region, 75 Farquhar Street, Guelph, Ontario, Canada N1H 3N4

SUMMARY

The bivariate lognormal distribution is proposed as a model for the joint distribution of storm peak (maximum rainfall intensity) and storm amount. Using the marginal distributions, the joint distribution, the conditional distributions, and the associated return periods are derived. The model is found appropriate for representing multiple episodic storm events at the Motoyama meteorological observation station in Japan. Copyright © 2002 John Wiley & Sons, Ltd.

KEY WORDS: storm frequency analysis; lognormal distribution; bivariate lognormal distribution; joint probability distribution; marginal distribution; conditional distribution

1. INTRODUCTION

A storm event may appear to be a multivariate event that is characterized by its peak and total amount, which might be mutually correlated. The severity of such a storm is a function of both its peak and total amount. For example, during the monsoon season in Japan, standing rainy fronts and hurricanes or typhoons cause the annual maximum storm both in storm peak (maximum rainfall intensity) and total amount. The damage induced by this storm is determined by both its peak and total amount. However, storm frequency analysis has often concentrated on storm peak or storm amount analysis only. Single-variable storm frequency analysis provides a limited assessment of the severity of a multiple episodic storm event. Effective hydrological engineering planning, design, and management require a better understanding of such a storm event. Some efforts have been made on the analysis of joint statistical properties with regard to a multivariate storm event (see, for example, Crovelli, 1973; Hashino, 1985; Singh and Singh, 1991; Bacchi *et al*. 1994; Kelly and Krzysztofowicz, 1997; and others).

Crovelli (1973) proposed a special bivariate gamma distribution with particular gamma marginals and used it to model joint distribution of storm depths and durations. Hashino (1985) generalized the Freund bivariate exponential distribution (Freund, 1961) and employed it to represent the joint

*Correspondence to: Sheng Yue, Meteorological Services of Canada, Environment Canada CCIW, 867 Lakeshore Road, P.O. Box 5050, Burlington, Ontario, L7R 4A6, Canada.
†E-mail: sheng.yue@ec.gc.ca

Received 22 October 1999
Accepted 22 December 2000

probability distribution of rainfall intensities and the corresponding maximum storm surges at the Osaka bay, Japan. Singh and Singh (1991) derived a bivariate probability density function with exponential marginals and used it to describe the joint distribution of rainfall intensity and the corresponding depth. Bacchi *et al.* (1994) proposed another bivariate exponential model with exponential marginals and applied it to analyze the joint distribution of rainfall intensities and durations. Kelly and Krzysztofowicz (1997) developed a bivariate meta-Gaussian model for hydrological frequency analysis.

Many hydrologic events such as storm peak and storm amount are positively skewed and may be represented by a lognormal distribution. Thus, it will be useful for hydrologic engineers to implement the bivariate lognormal distribution to analyze the joint probability distribution of two correlated random variables with lognormal marginals. This article presents a procedure for using this bivariate distribution to describe joint probabilistic behavior of correlated storm peak and storm amount. The usefulness of the distribution is illustrated using storm events observed at the Motoyama meteorological station, Japan.

2. BIVARIATE LOGNORMAL DISTRIBUTION

If two continuous random variables X_1 and X_2 are lognormally distributed, then the joint distribution of these variables may be represented by a bivariate lognormal distribution. The probability density function (pdf) of the bivariate lognormal distribution is given by (Aitchison and Brown, 1957)

$$
\begin{aligned}
&f(x_1,x_2) = \frac{1}{2\pi x_1 x_2 \sigma_{Y_1}\sigma_{Y_2}\sqrt{1-\rho^2}}\exp\left(-\frac{q}{2}\right)\\
&q = \frac{1}{1-\rho^2}\left[\left(\frac{\ln x_1 - \mu_{Y_1}}{\sigma_{Y_1}}\right)^2 - 2\rho\left(\frac{\ln x_1 - \mu_{Y_1}}{\sigma_{Y_1}}\right)\left(\frac{\ln x_2 - \mu_{Y_2}}{\sigma_{Y_2}}\right) + \left(\frac{\ln x_2 - \mu_{Y_2}}{\sigma_{Y_2}}\right)^2\right]\\
&x_1 > 0,\ i = 1,\ 2,\quad -1 < \rho < +1
\end{aligned}
\tag{1}
$$

where μ_{Y_i} and σ_{Y_i} are the population mean and standard deviation of $Y_i = \ln X_i,\ i = 1, 2$, and can be given by (Stedinger *et al.*, 1993):

$$
\sigma_{Y_i} = \left[\ln\left(1 + \frac{\sigma_{X_i}^2}{\mu_{X_i}^2}\right)\right]^{1/2} \tag{2a}
$$

$$
\mu_{Y_i} = \ln(\mu_{X_i}) - \frac{\sigma_{Y_i}^2}{2} \tag{2b}
$$

and where μ_{X_i} and σ_{X_i} are the population mean and standard deviation of $X_i,\ i = 1,\ 2$. ρ is the population product-moment correlation coefficient of Y_1 and Y_2. In practice, the mean (μ_{X_i}), standard deviation (σ_{X_i}), and correlation coefficient (ρ) are estimated by the method of moments using sample data.

As the cumulative distribution function (cdf) $F(x_1, x_2)$ of the bivariate lognormal distribution is not analytically attainable, it is computed by numerical integration. The conditional distributions ($F(x_1 \mid x_2)$ and $F(x_2 \mid x_1)$) of X_1 given $X_2 = x_2$ and X_2 given $X_1 = x_1$ are also lognormally distributed with different means and standard deviations.

The return period of the event $X_i > x_i,\ i = 1, 2$, is presented as follows:

$$T_{x_i} = \frac{1}{1 - F(x_i)} \ (F(x_i) = \Pr[X_i \le x_i]) \tag{3a}$$

On the basis of the same principle, the joint return period $T(x_1, x_2)$ of X_1 and X_2 associated with the event that at least one value of x_1 and x_2 is exceeded ($X_1 > x_1$ or $X_2 > x_2$, or $X_1 > x_1$ and $X_2 > x_2$) can be represented by

$$T(x_1, x_2) = \frac{1}{1 - F(x_1, x_2)} \tag{3b}$$

The conditional return period $T_{X_1|X_2}$ of X_1 given $X_2 = x_2$ is given by

$$T_{X_1\,|\,X_2} = \frac{1}{1 - F(x_1 \mid x_2)} \tag{3c}$$

The conditional return period $T_{X_2\,|\,X_1}$ of X_2 given $X_1 = x_1$ can be given by an equivalent formula.

3. APPLICATION

Observed 96-year daily rainfall data from 1896 to 1993 (except the years 1939 and 1940) at the Motoyama meteorological observation station in Japan were employed to demonstrate the usefulness of the bivariate distribution. Application results are presented in the following subsections.

3.1. *Definition of a storm sequence*

As recorded rainfall data are in the form of averages over a period of one day, one storm is defined as continuous daily rainfalls, as shown in Figure 1. Let the storm peak I (mm/day) be the maximum daily rainfall in a year and the corresponding amount A (mm) of total rainfall be given by

$$A = \sum_{j=1}^{D} u_j \tag{4}$$

where u_j is the jth daily rainfall amount (mm/day), and D is the storm duration (day).

The maximum daily rainfall (storm peak) for each year was selected first; then the corresponding storm amount was computed using Equation (4). The means and standard deviations of the storm peak and storm amount were estimated from their sample data, respectively. Parameters of the bivariate lognormal distribution were calculated using the method of moments as given by Equations. (2a) and (2b). These statistics are presented in Table 1. The product-moment correlation coefficient between the storm peak and amount was estimated using Equation (2c), and is equal to 0.757. This indicates that the storm peak and amount are strongly positively correlated.

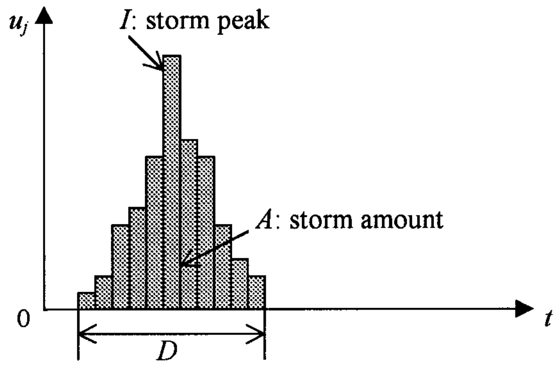

Figure 1. Characteristic values of a storm event.

Table 1. Statistics of storm peaks (I) and amounts (A)

	μ_X	σ_X	μ_Y	σ_Y
I (mm/day)	237.865	98.822	5.392	0.399
A (mm)	387.927	233.185	5.807	0.555

3.2. *Marginal distributions of storm peaks and amounts*

The non-exceedance empirical probability was estimated using the Weibull formula (Weibull, 1939; Chow, 1953),

$$P_k = \frac{k}{N+1} \tag{5}$$

where P_k is the cumulative frequency, the probability that a given value is less than the kth smallest observation in the data set of N observations.

The empirical probabilities and fitted lognormal distributions of the storm peak and storm amount are depicted on normal paper in Figures 2(a) and 2(b), respectively. There is no significant difference between the empirical and theoretical probabilities. Thus both the storm peak and amount can be represented by the lognormal distributions.

3.3. *Joint distribution of the storm peak (I) and amount (A)*

3.3.1. *Validity of the proposed model*. Empirical joint probabilities are computed based on the same principle as in the case of a single variable. A two-dimensional table is first constructed in which the variables I and A are arranged in ascending order. The element in row m and column l of the table is defined as the joint frequency function of the two random variables and is estimated by

$$f(i_m,\ a_l) = \Pr(I = i_m,\ A = a_i) = \frac{n_{ml}}{N+1} \tag{6}$$

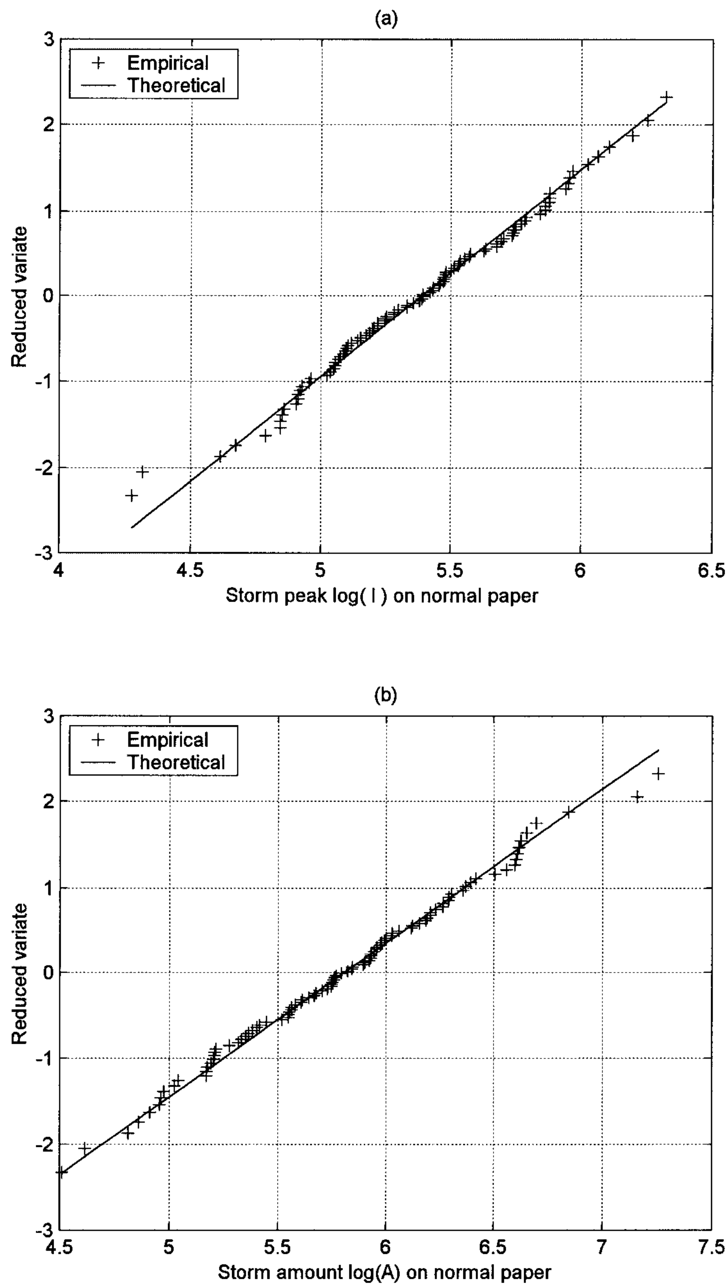

Figure 2. Distribution of (a) and storm peak (b) storm amount.

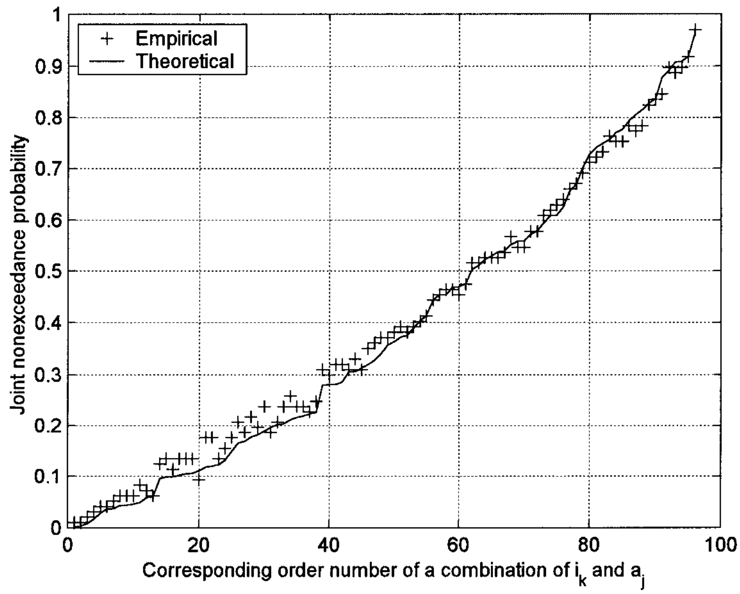

Figure 3. Comparison of empirical and theoretical probabilities of storm peaks and amounts.

where N is the total number of observations ($N = 96$), and n_{ml} is the number of occurrences of the combinations of i_m and a_l. The joint cumulative frequency (non-exceedance joint empirical probability) is then given as

$$F(i,\, a) = \Pr(I \leq i_k,\; a \leq a_j) = \sum_{m=1}^{k} \sum_{l=1}^{j} f(i_m,\, a_l) = \frac{\sum_{m=1}^{k} \sum_{l=1}^{j} n_{ml}}{N+1} \tag{7}$$

Theoretical joint probabilities of the real occurrence combinations of i_k and a_j are estimated by numerically integrating Equation (1). The empirical and theoretical joint probabilities are illustrated in Figure 3, in which the solid line represents the theoretical joint probabilities of storm peaks and volumes that are arranged in ascending order. The corresponding empirical joint probabilities are indicated by the plus sign. The x-axis is the corresponding order number of a combination of i_k and a_j. It is evident that no significant difference can be detected. It is therefore concluded that the model is suitable for representing the joint distribution of the correlated storm peak and amount.

3.3.2. *Joint cdf and joint return period of I and A.* The contours of the joint cdf and the joint return period are displayed in Figures 4(a) and 4(b), respectively. These contours indicate that, given an occurrence probability or a return period of a storm event, one can obtain various occurrence combinations of storm peaks and amounts, and vice versa. Such results cannot be obtained by marginal analysis.

3.3.3. *Conditional return periods.* The conditional return period $T_{I|A}$ of storm peak (I) given storm amount (A) and the conditional return period $T_{A|I}$ of storm amount (A) given storm peak (I) are presented in Figures 5(a) and 5(b), respectively. These allow one to obtain information concerning the occurrence return periods of storm peaks under the condition that a given storm amount occurs, and vice versa, which also cannot be provided by single-variable frequency analysis.

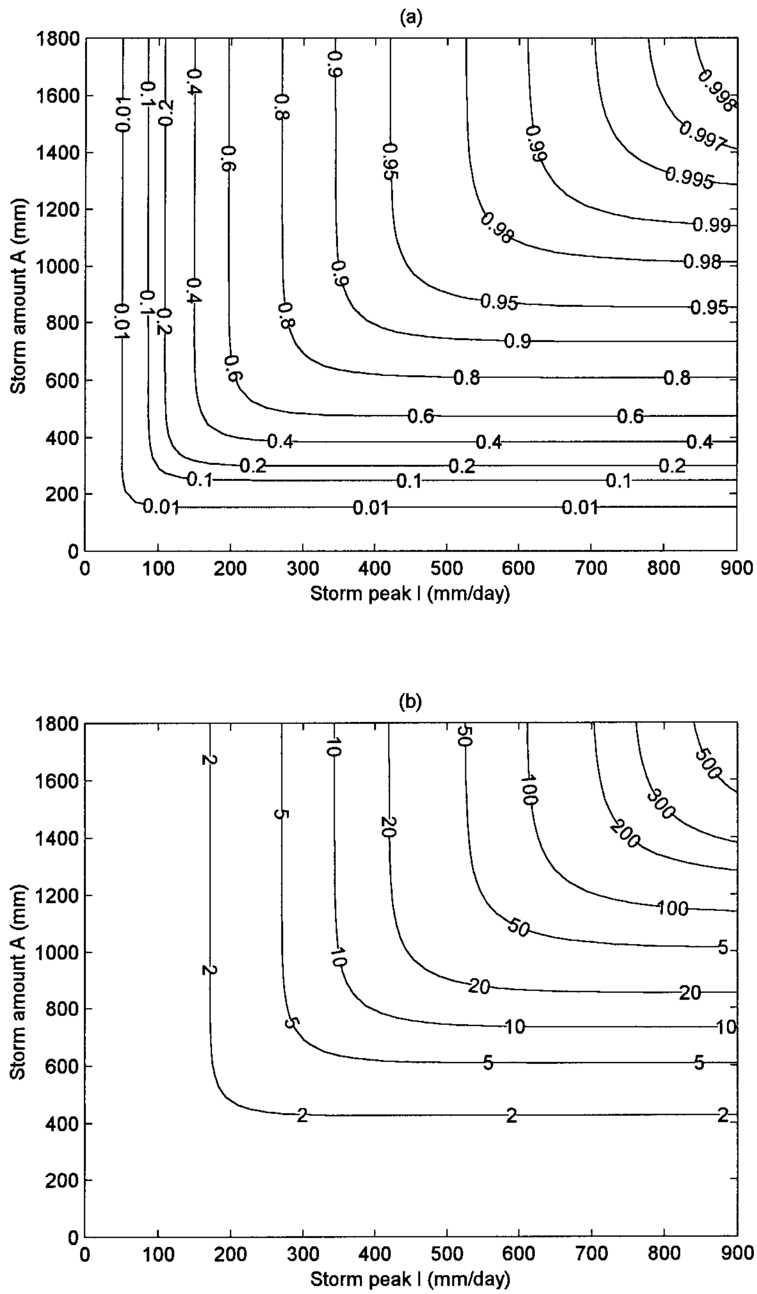

Figure 4. Contours of (a) joint cdf of storm peaks and amounts and (b) joint return period of storm peaks and amounts.

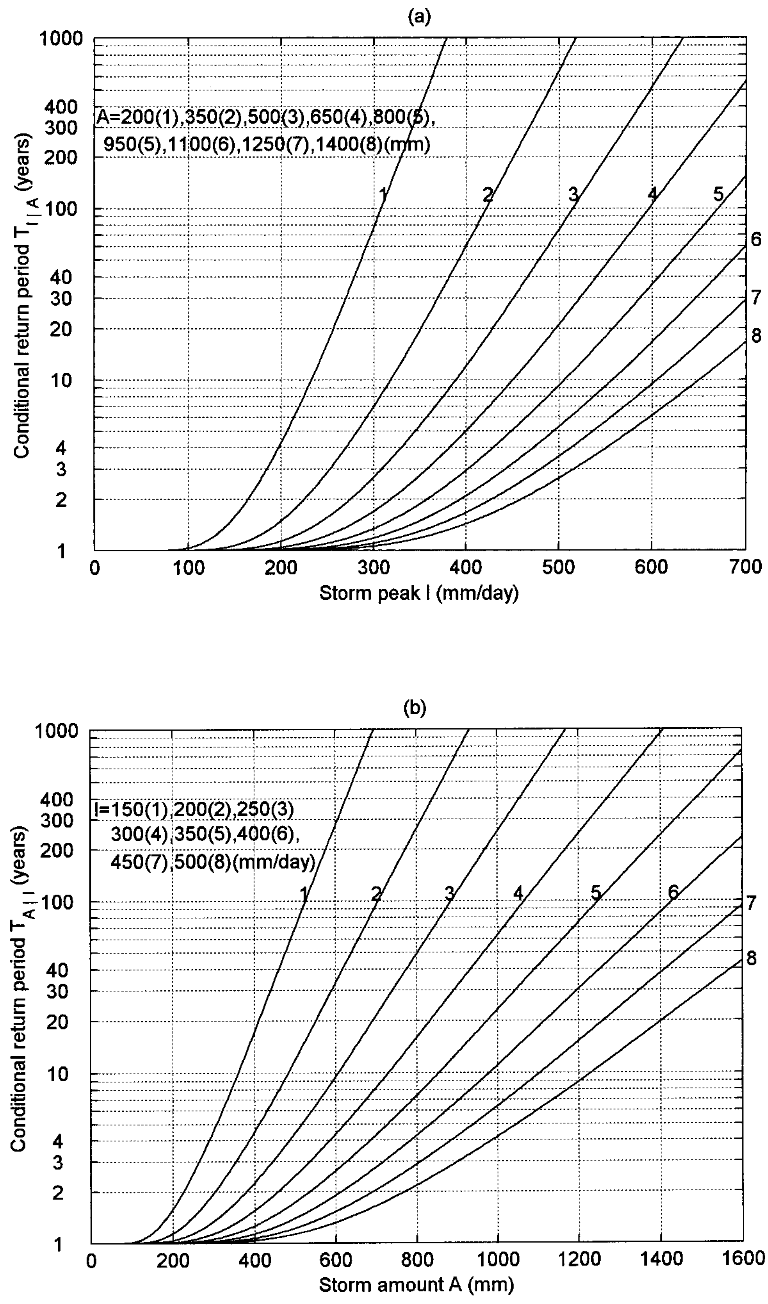

Figure 5. Conditional return period of (a) storm peaks given storm amount and (b) storm amounts given storm peak.

4. CONCLUSIONS

This study provides a procedure for using the bivariate lognormal distribution model to represent joint statistical properties of a multivariate storm event that is characterized by correlated storm peak and amount. On the basis of this model, one can obtain the joint probability distribution, the conditional distributions, and the associated return periods of two correlated random variables if their marginals are lognormally distributed.

The usefulness of the model is demonstrated by analyzing the joint probabilistic behavior of storm events observed at the Motoyama meteorological observation station in Japan. The computing results indicate that the model provides additional information which cannot be obtained by single-variable storm frequency analysis, such as the joint return periods of the combinations of storm peaks and amounts, and the conditional return periods of one variable given the other. These results should be useful for analysis and assessment of the risk associated with multivariate hydrological events.

ACKNOWLEDGEMENTS

The main work of this article was completed when the author was in INRS-EAU, University of Quebec. The partial financial support of the Natural Sciences and Engineering Research Council (NSERC) of Canada through a postdoctoral fellowship of Chair in Statistical Hydrology of INRS-EAU is gratefully acknowledged. The author wishes to thank Prof. M. Hashino of the University of Tokushima, Japan, for providing the daily precipitation data. The author also thanks Prof. B. Bobée of INRS-EAU for his kind assistance. The constructive comments and corrections made by Dr. A.H.El-Shaarawi significantly improved the quality of the paper.

REFERENCES

Aitchison J, Brown JAC. 1957. *The Lognormal Distribution*. Cambridge University Press: Cambridge.

Bacchi B, Becciu G, Kottegoda NT. (1994). Bivariate exponential model applied to intensities and durations of extreme rainfall. *Journal of Hydrology* **155**: 225–236.

Chow VT. 1953. Frequency analysis of hydrological data with special application to rainfall intensities. *University of Illinois, Eng. Exper. Stat. Bull.*: 414.

Crovelli RA. 1973. A bivariate precipitation model. Paper presented at 3rd Conference On Probability and Statistics in Atmospheric Science, American Meteorological Society, Boulder, CO., USA; 130–134.

Freund JE. 1961. A bivariate extension of the exponential distribution. *Journal of the American Statistical Association* **56**: 971–977.

Hashino M. 1985. Formulation of the joint return period of two hydrologic variates associated with a Poisson process. *Journal of Hydroscience and Hydraulic Engineering* **3**(2): 73–84.

Kelly KS, Krzysztofowicz R. 1997. A bivariate meta-Gaussian density for use in hydrology. *Stochastic Hydrology and Hydraulics* **11**: 17–31.

Singh K, Singh VP. 1991. Derivation of bivariate probability density functions with exponential marginals. *Stochastic Hydrology and Hydraulics* **5**: 55–68.

Stedinger JR, Vogel RM, Foufoula-Georgiou E. 1993. Frequency analysis of extreme events. *Handbook of Hydrology*, Maidment DR (ed.), McGraw-Hill, Inc.

Weibull W. 1939. A statistical theory of the strength of materials. Ingeniors Vetenskaps Akademiens-Handlingar (The Royal Swedish Institute for Engineering Research Proceedings), Stockholm; 151.

