ABSTRACT Recently, studies have commenced on the application of the multivariate approach to extreme value distributions towards its application for the solution of hydrological problems. The theoretical foundations of such an approach were made some 20 years ago, but they have had very little impact on the field of hydrology. In this paper, a classification of such models is made and in the particular case of the differentiable models, characteristics, properties and parameter estimation procedures for the method of maximum likelihood are given for the bivariate and trivariate distributions. Finally some applications of those distributions are described, mainly in the case of bivariate distributions.

Distributions multivariates des valeurs extrêmes dans les analyses hydrologiques

RESUME Quelques études on été entreprises récemment sur l'application de l'approche multivariate à des distributions des valeurs extrêmes en vue de résoudre des problèmes hydrologiques. Les fondements théoriques d'une telle approche on été établis il y a 20 ans mais ils non pas eu d'impact dans le domaine de l'hydrologie. Dans cette communication, une classification de ces modèles est présentée et dans le cas particulier des modèles différentiables, les caractéristiques, propriétés et les méthodes pour l'estimation des paramètres par le maximum de vraisemblance sont présentées pour les distributions bivariées et trivariées. Finalement, certaines applications de ces distributions sont décrites principalement pour le cas de distributions bivariées.

NOTATION

\(a\) association parameter of the mixed model
\(F(\cdot)\) probability distribution function of \(\cdot\)
\(f(\cdot)\) probability density function of \(\cdot\)
INTRODUCTION

The use of extreme value distributions has grown steadily in the last 50 years in the field of hydrology, mainly in performing frequency analyses of annual maxima or minima. Certain univariate distributions have arisen as particular solutions to the Stability Postulate, which any extreme must satisfy. Jenkinson (1955) found the general solution to the Stability Postulate, and some authors after him have called that solution the general extreme value (GEV) distribution. This distribution is able to represent the extreme value distributions of types II and III directly, and when one takes the limit of the shape parameter to zero, the result becomes type I extreme value distribution. An excellent reference on this topic is Natural Environment Research Council (1975).

In pioneer papers Finkelstein (1953), Gumbel (1958) and Tiago de Oliveira (1958) gave the foundations for the multivariate approach to extreme value distributions. After these works, several models for bivariate extreme value distributions began to appear in the literature. Two classes of them are now known: the differentiable and the nondifferentiable models. Among the latter class are (Tiago de Oliveira, 1982): the biextremal, Gumbel and Natural models.

Since these models do not have explicit probability density functions, which makes parameter estimation difficult, and given that they have been developed for the particular case when both marginals are Gumbel distributions, they will not be described here any further. The reader is referred to Tiago de Oliveira (1982) for any additional details.

The only two known differentiable models for bivariate extreme value distributions are: the logistic and mixed models. Such names for the models have been coined (Tiago de Oliveira, 1982) from the fact that the reduced difference, when both marginals are Gumbel distributions, has the standard Logistic distribution in the first case. In the second case, the model has a dependence function coming from a mixture of such functions for the cases of dependence and independence, when the marginals are Gumbel distributions. General characteristics and multivariate extensions of these models will be given in the following sections.
DIFFERENTIABLE MODELS FOR BIVARIATE EXTREME VALUE DISTRIBUTIONS

There are only two known differentiable models: the logistic and the mixed models.

The general form of the logistic model for bivariate extreme value distributions is (Gumbel, 1960a):

$$F(x,y,m) = \exp\left\{\left[\left(-\ln F(x)\right)^m + \left(-\ln F(y)\right)^m\right]^{-\frac{1}{m}}\right\}$$  \hspace{1cm} (1)

where $m$ is the association parameter, $F(u)$ is the marginal distribution function of $u$, $m \geq 1$ and $0 \leq \rho \leq 1$, where $\rho$ is the population product-moment correlation coefficient.

For $m = 1$, the independent case, the bivariate distribution function splits into the product of its marginals:

$$F(x,y,1) = F(x) \cdot F(y)$$  \hspace{1cm} (2)

When $m = \infty$, the bivariate distribution function is:

$$F(x,y,\infty) = \min[F(x), F(y)]$$  \hspace{1cm} (3)

as was shown by Johnson & Kotz (1970). This is the diagonal case and for the case when both marginals are Gumbel distributions, this point is the only one that does not have a planar density. In this case, it is possible to obtain an analytical expression for the population product-moment correlation coefficient (Gumbel, 1967):

$$\rho = (1 - 1/m^2)$$  \hspace{1cm} (4)

In the case of the mixed model, its general form is (Gumbel, 1960a):

$$F(x,y,a) = F(x) \cdot F(y) \exp\left\{a \left[\frac{1}{(-\ln F(x))} + \frac{1}{(-\ln F(y))}\right]^{-1}\right\}$$  \hspace{1cm} (5)

where $a$ is the association parameter, $0 \leq a \leq 1$ and $0 \leq \rho \leq 0.66$. For the case $a = 0$, the bivariate distribution splits into the product shown in equation (2), and this is the independent case. When $a = 1$, complete dependence is observed but this is not the diagonal case. This type of model has planar density in all points when its marginals are both Gumbel distributions. When the last condition is met, it is possible to obtain an analytical expression for the population product-moment correlation coefficient (Tiago de Oliveira, 1982):

$$\rho = \frac{6}{\pi^2} \arccos\left(1 - \frac{a}{2}\right)^2$$  \hspace{1cm} (6)

From the characteristics and properties of the differentiable models for bivariate extreme value distributions, Tiago de Oliveira (1982) stated: "It is, then, intuitive that the distance between independence ($m = 1$) and the assumed model for $m > 1$ is small in general, and for small samples it will probably be impossible to distinguish them", in the case of the logistic model. For the mixed
Jose A. Raynal-Villasenor & Jose D. Salas

model, he stated: "The smaller variation of the correlation coefficient and of the distance shows that the deviation from independence is smaller and most difficult to detect".

From these statements, the logistic model has been chosen, due to its greater flexibility and wider applicability, to be pursued further.

MULTIVARIATE EXTENSION OF THE LOGISTIC MODEL

For the case of the logistic model for bivariate extreme value distributions, the following multivariate extension was given by Gumbel (1960b):

\[ F(x,y,\ldots,m) = \exp\{-[(-\ln F(x))^m + (-\ln F(y))^m + \ldots]^{1/m}\} \] (7)

and Tiago de Oliveira (1975a,b) provided the following inequalities for the model in equation (7):

\[ F(x_1) F(x_2) \ldots F(x_n) \leq F(x_1,x_2,\ldots,x_n) \leq \min[F(x_1),F(x_2),\ldots,F(x_n)] \] (8)

\[ \prod_{i \neq j} F(x_i,x_j)^{2(n-1)} \leq F(x_1,x_2,\ldots,x_n) \leq \prod_{i \neq j} F(x_i,x_j)^{2(n-1)/3} \]

\[ \prod F(x_i) \] (9)

MAXIMUM LIKELIHOOD ESTIMATION OF THE PARAMETERS

For the case of bivariate distribution functions, the samples usually do not have an equal length of record, so it is necessary to have a formulation flexible enough to cover all possible combinations of arrangements of data. Such a formulation has been proposed by Raynal-Villasenor (1985), based on the generalization given by Anderson (1957):

\[ L(x,y,\theta) = \prod_{i=1}^{n_1} f(s_i,\theta_1) \prod_{i=1}^{n_2} f(x,y,\theta_2) \prod_{i=1}^{n_3} f(t_i,\theta_3) \] (10)

where \( L(.) \) is the likelihood function of \( (.) \), \( n_1, n_2, n_3 \) are the lengths of record before, during and after the common period \( n_2 \), respectively. Then \( s \) and \( t \) are the variables with records before and after the common period, respectively. \( I_1 \) are indicator numbers such that \( I_1 = 1 \) if \( n_1 > 0 \) and \( I_1 = 0 \) if \( n_1 = 0 \). \( \theta_1 \) is the vector of parameters.

Given the property that the maximum of a function and its logarithm occur at the same point and due to this fact the expressions provided by the logarithm of equation (10) are much easier to handle than those produced by such an equation, the
log-likelihood function will be used instead of its natural version. So, equation (10) is transformed in, Raynal-Villasenor (1985):

$$LL(x,y,\theta) = I_1[\sum_{i=1}^{n_1} \ln f(s_i,\theta_1)] + I_2[\sum_{i=1}^{n_2} \ln f(x,y,\theta_2)] + I_3[\sum_{i=1}^{n_3} \ln f(t,\theta_3)]$$

(11)

An extension of equation (10) for the case of trivariate distribution functions is:

$$L(x,y,z,\theta) = \left[\prod_{i=1}^{n_1} f(s_i,\theta_1)\right]^{I_1} \cdot \left[\prod_{i=1}^{n_2} f(s_i,t_i,\theta_2)\right]^{I_2} \cdot \left[\prod_{i=1}^{n_3} f(x,y,z,\theta_3)\right]^{I_3} \cdot \left[\prod_{i=1}^{n_4} f(u_i,v_i,\theta_4)\right]^{I_4} \cdot \left[\prod_{i=1}^{n_5} (u_i,\theta_5)\right]^{I_5}$$

(12)

and the corresponding equation to equation (11) for this level is:

$$LL(x,y,z,\theta) = I_1[\sum_{i=1}^{n_1} \ln f(s_i,\theta_1)] + I_2[\sum_{i=1}^{n_2} \ln f(s_i,t_i,\theta_2)] + I_3[\sum_{i=1}^{n_3} \ln f(x,y,z,\theta_3)] + I_4[\sum_{i=1}^{n_4} \ln f(u_i,v_i,\theta_4)] + I_5[\sum_{i=1}^{n_5} \ln f(u_i,\theta_5)]$$

(13)

Since the parameter values which maximize equation (11) are the maximum likelihood estimators of the parameters of such distribution functions and given that equations (11) and (13) are very suitable for solution through optimization procedures, like the Rosenbrock method with multiple constrained variables, this is the approach employed to estimate the parameters of bivariate and trivariate extreme value distributions.

APPLICATION OF MULTIVARIATE EXTREME VALUE DISTRIBUTIONS TO HYDROLOGY

Four main fields of application to multivariate extreme value distributions have been detected:

(a) identification of the type of extreme value distributions,
(b) improvement in parameter estimation,
(c) transfer of extreme value information,
(d) flood frequency analysis downstream of river junctions.

They will be described as follows.

(a) Identification of the type of extreme value distributions. Using the property of the logistic model that the difference is logistic distributed when both marginals are Gumbel distributions, the following identification procedure has been devised for extreme value distributions (Raynal-Villasenor, 1985):

(i) estimate the parameters through a bivariate maximum likelihood scheme (equation (11)), assuming both marginals are General Extreme Value (GEV) distributions. This step allows the shape parameter to take any value and no specific extreme value
distribution type is assumed beforehand.

(ii) test the difference against the logistic distribution. If the result is positive, then both marginals are Gumbel distributions. If not, continue to the next step.

(iii) transform one at a time each of the variables to be Gumbel distributed, as suggested by the Natural Environment Research Council (1975):

\[ y = -\ln \left[ 1 - \left( \frac{x - \mu}{\alpha} \right)^{1/\beta} \right] \]  

where if \( x \) is GEV distributed, then \( y \) is Gumbel distributed. Then, perform the test contained in (ii). If a positive result is the outcome, then one of the marginals is GEV and the other is a Gumbel distribution. If not, continue to the next step.

(iv) transform both marginals to Gumbel distributions according to equation (14), and perform test (ii). If the result is positive, then both marginals are GEV distributions. If not, the marginals cannot be assumed to come from the family of extreme value distributions.

(b) It has been shown (Clarke, 1980; Rueda, 1981 and Raynal-Villasenor, 1985) that there exists an improvement in the parameter estimation phase, when the bivariate approach to distribution functions is used. Particularly, in the last two cases, related with flood information data, the improvement has been observed even in the case when both samples have the same length of record. As an example, Table 1 (Raynal-Villasenor, 1985) shows this condition. The measure of improvement is the relative information ratio, defined as (Raynal-Villasenor, 1985):

\[ I_a(\theta_i) = \frac{\text{var}_a(\theta_i)}{\text{var}_a(\theta_i,B)} \]  

where \( I_a(\theta_i) \) is the asymptotic relative information ratio, \( \text{var}_a(\theta_i) \) and \( \text{var}_a(\theta_i,B) \) are the asymptotic variances of parameter \( \theta \) obtained through univariate and bivariate maximum likelihood procedures, respectively.

The trivariate extension for flood data and the whole procedure for drought data is under study and it will be reported elsewhere in

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Relative asymptotic information ratios for the parameters of the bivariate GEV distribution (( \beta_1 = -0.10, \beta_2 = -0.15 ) and ( m = 2.0 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>( n_2 + n_3: )</td>
</tr>
<tr>
<td></td>
<td>25</td>
</tr>
<tr>
<td>( u_1 )</td>
<td>1.03</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>0.95</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>1.01</td>
</tr>
<tr>
<td>( u_2 )</td>
<td>1.03</td>
</tr>
<tr>
<td>( \alpha_2 )</td>
<td>0.95</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1.01</td>
</tr>
</tbody>
</table>
the near future.

(c) Derived directly from what has been explained in the preceding point, it is possible, using multivariate extreme value distributions, to establish a transfer of extreme value data from gauging stations with a longer record to stations with shorter records (Raynal-Villasenor, 1985). Table 2 contains a sample of several stations that have been analysed, and it shows the number of years gained with this process of transfer information.

The trivariate extension for flood data and the process of drought data are under study.

(d) Flood frequency downstream of river junctions has been attained only for the cases of complete dependence or complete independence, using non-parametric approaches (Salas, 1980; Linsley & Franzini, 1979). Using the multivariate approach to extreme value distributions, it is possible to provide an alternative method of modelling the physical process for the case of partial dependence. Such an approach is made assuming that the flood data gauged at stations upstream of the confluence may be used to represent what happens downstream of it. The form of the model is the well-known convolution equation for random variables (Woodroofe, 1975):

\[
F(t) = \int_{-\infty}^{t} \int_{-\infty}^{\infty} f(t - s, s) \, ds \, dt
\]

where \( F(t) \) is the probability distribution function of the flood data downstream of the confluence, and \( t \) represents the sum of two random variables, the samples of which have been gauged at stations upstream of the confluence. Raynal-Villasenor & Salas (1986) have analysed a case study of this kind and the results obtained proved to be satisfactory.

**CONCLUSIONS**

The multivariate approach to extreme value distributions has been presented as regards its application to the solution of hydrological problems.

It was shown that the use of multivariate extreme value distributions may improve the estimation of parameters of extreme value distributions. In addition to this feature, multivariate extreme value distributions provide a means for the identification of such distributions in the univariate form. A third application

<table>
<thead>
<tr>
<th>Station</th>
<th>Location</th>
<th>Scale</th>
<th>Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cottonwood Creek, Colorado, USA</td>
<td>3</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>Rio Fernando, New Mexico, USA</td>
<td>8</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Palo Dulce, Sin, Mexico</td>
<td>9</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Toahayana, Sin, Mexico</td>
<td>11</td>
<td>8</td>
<td>-</td>
</tr>
</tbody>
</table>
was the ability of such distributions to transfer information related with extreme value data. Finally, this kind of distribution could be applied as an alternative to the solution of specific hydrological problems, e.g. a flood frequency analysis downstream of confluences. The authors consider this a very promising field of research and they would like to quote Tiago de Oliveira (1975b) "if the bivariate extremes area is largely open we can say that the field of multivariate extremes is almost completely open".

ACKNOWLEDGEMENTS The authors wish to express their gratitude to the Engineering Graduate Studies Division, Universidad Nacional Autonoma de Mexico and to the Civil Engineering Department, Colorado State University, for the support provided to the realization of this paper.

REFERENCES

Rueda, E. (1981) Transfer of information for flood related...
variables. MSc Thesis, Colorado State University, Fort Collins, Colorado, USA.


