JFE JOURNAL OF FLOOD ENGINEERING
1(1) January–June 2010; pp. 37–50

# CALCULATION OF BIVARIATE DOUBLE GUMBEL PROBABILITY DENSITY FUNCTION VIA A GENETIC ALGORITHM: APPLICATION TO HUITES DAM BASIN

**R. Domínguez M.[1], O. A. Fuentes M.[1], M. L. Arganis J.[2] & A. Mendoza R.[3]**

[1] *Instituto de Ingeniería y Posgrado de Ingeniería, Universidad Nacional Autónoma de México, México*

[2] *Instituto de Ingeniería y Facultad de Ingeniería, Universidad Nacional Autónoma de México, México*

[3] *Instituto de Ingeniería, Universidad Nacional Autónoma de México, México*

**ABSTRACT:** In this paper, the analytic structure of a Bivariate Double Gumbel probability density function (BDGDEF) is developed. Its application becomes important in estimating design events of hydraulic projects with a large regulating capacity when the main random variables are the maximum annual inflow rate and the maximum annual flow volume. A genetic algorithm was applied to obtain the set of parameters of the Bivariate Double Gumbel probability function, in order to maximize the likelihood function (L) for the recorded data in a hydrometric station. Results were then compared with those obtained from a deterministic algorithm.

***Keywords:*** Bivariate Double Gumbel, Likelihood Function, Huites, Genetic Algorithm.

## 1. INTRODUCTION

Dimensioning of many hydraulic works depends on a large extent on the design flood; in addition, the update of design inflow rates applicable to existing hydraulic projects is important for any country for safety reviews of the structures. At dams with a large regulating capacity, the performance of the spillway works is governed by both the maximum annual inflow rate and the maximum annual flow volume; therefore, some design methods take into account the statistical analysis of historical runoff associated to different durations, (Dominguez, 1980), others, such as those proposed by (Hiemstra, 1979), (Rivera, 1999), (Ramirez, 2000), and (Jiménez Espinoza, 2000), consider the effects of both the instantaneous peak flow rates and flood volume through the use of bivariate analysis.

On the other hand, in Mexico and other countries inflow rates data and maximum annual volume data for a given frequency are derived from two populations (González Villarreal, 1970; Rossi *et al.*, 1984), the first population is associated to conventional events and the second population could be either due to events of hurricane type or to winter phenomena (such as in the case of northwestern Mexico).

Unfortunately, Double Gumbel function has 5 parameters, and therefore, Bivariate Double Gumbel function (BDGF) has 11 parameters; such a number of parameters makes unpractical the application of least squares method or even the moments method to find them, because of the increase in the number of moments needed and in the number of non linear equations to solve. The maximization of the likelihood function (*L*) is recommended in such cases.

Because the computation of the likelihood function (*L*) requires the calculation of the probability density function and the structure for the BDGF was not found in the technical literature, this study presents the analytical calculation of the BDGDEF based on its probability distribution function.

Once such distribution is obtained, it becomes possible to obtain its parameters by maximizing the likelihood function (*L*) through applying a traditional optimization algorithm (deterministic) or the optimization technique, based on evolutionary computation, (Goldberg, 1989). A comparison of the advantages and disadvantages of these two optimization methods is discussed in this paper using as example data recorded at Huites station on the northwestern México.

## 2. METHODOLOGY

### 2.1. Bivariate Distribution Function

The general equation for bivariate extreme value distributions functions is given by Escalante Sandoval (2007).

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

Where $F(x, y)$ is the probability of having simultaneously, $X \le x$ and $Y \le y$; $F(x)$ and $F(y)$ are marginal distributions and $m$ is an association parameter that depends on the degree of interdependence between the random variables $X$ and $Y$.

The appearance of the marginal distributions can be regarded as Gumbel, as general of extreme values or as two-population Gumbel.

### 2.2. Bivariate Density Function

The probability density function $f(x, y)$ of a bivariate function is obtained from its distribution function $F(x, y)$ as:

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial y \partial x} \tag{2}$$

If a function $g(x, y)$ is defined as:

$$g(x,y) = -\left[(-\ln F(x))^m + (-\ln F(y))^m\right]^{1/m} \tag{3}$$

Taking into account the derivative of the function of a function and partially deriving, the probability density function can be expressed as (Appendix I):

$$f(x,y) = e^{g(x,y)} \left[ \frac{\partial g(x,y)}{\partial x} \frac{\partial g(x,y)}{\partial y} + \frac{\partial^2 g(x,y)}{\partial y \partial x} \right] \tag{4}$$

With the partial derivatives of *g(x, y)* given in Appendix II.

With a different approach and defining function *v(x, y)* as:

$$v(x,y) = (-\ln F(x))^m + (-\ln F(y))^m \tag{5}$$

The value of *f(x, y)* can be obtained, also analytically as:

$$f(x,y) = \frac{m}{F(x)F(y)} \frac{\partial F(x)}{\partial x} \frac{\partial F(y)}{\partial y} \left[(-\ln F(x))(-\ln F(y))\right]^{m-1} e^{-v^{m-1}} v^{\frac{1}{m}-2} \left[ \left(1 - \frac{1}{m}\right) + \frac{1}{m} v^{\frac{1}{m}} \right] \tag{6}$$

### 2.3. Bivariate Double Gumbel Function

The Bivariate Double Gumbel distribution function (BDGDIF) establishes the relationship between two random variables, each of them constituted by two populations in which one includes data of considerably higher magnitude identifying extreme conditions.

The marginal functions of a double Gumbel function can be expressed as:

$$F(x) = \exp\left\{ -\exp\left( -\frac{x + a_1}{c_1} \right) \right\} p_x + \exp\left\{ -\exp\left( -\frac{x + a_2}{c_2} \right) \right\} (1 - P_x) \tag{7}$$

$$F(y) = \exp\left\{ -\exp\left( -\frac{y + a_3}{c_3} \right) \right\} p_y + \exp\left\{ -\exp\left( -\frac{y + a_4}{c_4} \right) \right\} (1 - P_y) \tag{8}$$

where $a_1$, $a_2$, $a_3$ and $a_4$ are location parameters; $c_1$, $c_2$, $c_3$ and $c_4$ are scale parameters; $p_x$ and $p_y$ are segregation parameters of variables $x$ and $y$, respectively.

The derivatives of the marginal functions $F(x)$ and $F(y)$ of a Double Gumbel are shown in Appendix III.

Because of the importance of precision in the calculations of the BDGDEF, $f(x, y)$ for the evaluation of the likelihood function ($L$), in Appendix IV are shown some tests made for equations 4 and 6, first using the Bivariate Gumbel function (BGF) and then the more complex BDGF.

Results show that both eq. 4 and eq. 6 lead to results with enough accuracy, since equation 4 is easier to use, it will be employed in this study.

### 2.4. Method of Maximum Likelihood

It is established in this method that for a random variable $x$ having a probability density function $p(x;\ \alpha_1, \alpha_2, \ldots, \alpha_n)$ where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are the parameters of such function, the probability of obtaining a given value $x_i$, is proportional to $p(x_i;\ \alpha_1, \alpha_2, \ldots, \alpha_n)$ whereas the joint probability of obtaining a sample of $n$ values $x_1, x_2, \ldots x_n$ is proportional to the product:

$$L = \prod_{i=1}^{n} p(x_1, \alpha_1, \alpha_2, ..., \alpha_n) \tag{9}$$

This function is known as likelihood (Kite, 1988; Escalante Sandoval, 2007). The method of maximum likelihood implies the calculation of parameters $\alpha_1, \alpha_2, \ldots, \alpha_n$ such that $L$ becomes maximum; traditionally this can be obtained by maximizing the logarithm of the $L$ function:

$$\log L = \sum_{i=1}^{n} \log p(x_1; \alpha_1, \alpha_2, ..., \alpha_n) \tag{10}$$

Its derivatives with respect to each of its parameters are:

$$e_i = \frac{\partial \log L}{\partial \alpha_i} \tag{11}$$

To obtain the extreme values of log $L$ these derivatives are set to zero.

Sometimes, maximization of the logarithm of $L$ through the use of derivatives may lead to local minimums (Rao, 2000; Smith, 1988; Horbelt, 2002; Myiung, 2003) rather than to the maximum global value.

In this report, it is proposed to take advantage of the capability of the genetic algorithms to directly obtain the parameters to maximize the function $L$ or its logarithm with no further need to calculate the derivatives of equation 10, and set them to zero.

### 2.5. Simple Genetic Algorithm

The traditional genetic algorithm (Holland, 1975; Goldberg, 1989) generates an initial population of $n$ individuals (in this case, parameters of the models); its fitness is evaluated with an objective function. A selection is made of the best fitted individuals with the universal stochastic method or with the roulette procedure. Those individuals are subjected to the crossover and mutation operators and a new population is created, with $n$ individuals that define the next generation; the fitness evaluation process, the selection of the most suitable individuals, crossover and mutation operations and the creation of another population are iterated until a number of generations previously established is achieved.

In this paper, the simple genetic algorithm implemented in Matlab's toolbox (MathWorks, 1992) was applied, whose structure is similar to all evolution-based algorithms, (Bäck, 1996).

There were two objective functions applied in the genetic algorithm to determine the fitness of the individuals: a) the direct maximization of the likelihood function $L$ and b) the maximization of the logarithm of $L$.

## 3. APPLICATION

### 3.1. Example of Application to Huites Station

Values of $n$ = 52 maximum annual instantaneous flow rates and of volumes recorded at *Huites* station in the State of Sinaloa, shown in Table 1, were used; the station was located in the northwestern part of Mexico and it stopped operating in 1992, the year when Huites dam was built. This region is characterized by its important runoff during the winter season of several years (highlighted in Table 1), therefore, data of maximum annual flow rates and volumes are generally constituted by two different populations, for which the Double Gumbel distribution function is the most suitable for better fitting purposes.

**Table 1**
**Historic Record of Maximum Flow Rates and Volumes. *Huites* Station**

| *Year* | *Q(m³/s)* | *V(10⁶ m³)* | *Year* | *Q(m³/s)* | *V(10⁶ m³)* |
|---|---|---|---|---|---|
| **1941** | **2085** | **458** | 1968 | 1534 | 1706 |
| 1942 | 2531 | 1302 | 1969 | 1508 | 837 |
| **1943** | **14376** | **1928** | 1970 | 1558 | 1001 |
| 1944 | 2580 | 871 | 1971 | 2200 | 905 |
| 1945 | 1499 | 684 | 1972 | 2225 | 442 |
| 1946 | 1165 | 720 | **1973** | **7960** | **1250** |
| 1947 | 1127 | 435 | **1974** | **3790** | **607** |
| 1948 | 3215 | 344 | 1975 | 1095 | 1768 |
| **1949** | **10000** | **2966** | 1976 | 2677 | 565 |
| **1950** | **3229** | **644** | 1977 | 1135 | 601 |
| 1951 | 677 | 111 | 1978 | 4790 | 1245 |
| 1952 | 1266 | 474 | **1979** | **6860** | **986** |
| 1953 | 1025 | 163 | 1980 | 1496 | 1076 |
| 1954 | 955 | 596 | 1981 | 4828 | 992 |
| **1955** | **4780** | **787** | 1982 | 2450 | 351 |
| 1956 | 696 | 513 | 1983 | 8275 | 1625 |
| 1957 | 593 | 69 | **1984** | **5580** | **1258** |
| 1958 | 3010 | 740 | **1985** | **3585** | **1092** |
| 1959 | 1908 | 352 | 1986 | 1349 | 1185 |
| **1960** | **15000** | **1842** | 1987 | 1429 | 766 |
| 1961 | 1396 | 689 | 1988 | 1866 | 683 |
| 1962 | 1620 | 437 | 1989 | 1868 | 428 |
| **1963** | **2702** | **885** | **1990** | **11558** | **2930** |
| 1964 | 1319 | 384 | 1991 | 2563 | 653 |
| 1965 | 1944 | 305 | 1992 | 2025 | 601 |
| 1966 | 2420 | 2716 | | | |
| 1967 | 2506 | 593 | | | |

In Ramírez and Aldama, 2000, the obtained parameters of the BDGDIF, are maximizing the likelihood function $L$ using eq.'s 10 and 11 along with the deterministic optimization method proposed by Rosenbrock, 1960. Their results and those obtained by means of a genetic algorithm are presented herein; finally it is made a comparison with respect to an empirical distribution function.

The genetic algorithm performed a random search of the BDGF parameters (eq. 4) based on two criteria: the direct maximization of the likelihood function $L$ and the maximization of the logarithm of $L$. In a general case $p_x$ and $p_y$ take values between 0.5 and 1 and the value of $m$ can be estimated around the value given by eq. 12 (Ramírez and Aldama, 2000):

$$m \approx \sqrt{\frac{1}{1-r_{QV}}} \tag{12}$$

Where $r_{QV}$ is the correlation coefficient between $Q$ and $V$. In order to give a variation interval for the rest of the parameters, an analysis of the marginal distribution function can be made using the moments method. So the genetic algorithm obtains a total of 11 parameters. Five thousand generations and populations of 200 individuals were considered; the results thus obtained and the corresponding values by Ramírez and Aldama, 2000 are presented in Table 2.

**Table 2**
**Comparison of Values of the Bivariate Double Gumbel Function Parameters**

| | | *Genetic Algorithm (GA)* | |
|---|---|---|---|
| *Parameter* | *Ramírez & Aldama* | *Max ( L)* | *Max (log L)* |
| $a_1$ | -1604.57 | -1853.79 | -1516.39 |
| $c_1$ | 740.66 | 767.93 | 680.94 |
| $a_2$ | -6669.27 | -6440.17 | -5729.79 |
| $c_2$ | 3071.53 | 3115.82 | 3140.60 |
| $a_3$ | -531.94 | -307.25 | -560.35 |
| $c_3$ | 304.02 | 259.43 | 314.77 |
| $a_4$ | -1324.47 | -1619.20 | -2000.00 |
| $c_4$ | 728.61 | 862.94 | 686.07 |
| $p_x$ | 0.7618 | 0.7694 | 0.7383 |
| $p_y$ | 0.8101 | 0.6668 | 0.9056 |
| $m$ | 1.6021 | 1.1134 | 1.6668 |

Results in Table 2, show that when maximizing $L$ there is a large variation in parameters $p_y$ and $m$ against the values obtained by Ramírez and Aldama, 2000.

When the log $L$ is maximized, some variations are presented on $p_y$ and $a_4$ (with respect to Ramírez and Aldama); such differences can be attributed to the methods of optimization applied (deterministic and random, respectively) and its implications will be shown in the next section.

### 3.2. Determination of the Empirical Distribution Function

To get an empirical probability $\hat{F}(x, y)$ of not exceeding neither $x$ or $y$, data were classified from higher to lower with respect to the values of flow rates, maintaining the volume of the corresponding year; subsequently, $k$ data smaller or equal than both pairs of values were counted. The value of $k$ for each pair of data was divided by the total number of data plus one ($n + 1$) so as to obtain the value of $\hat{F}(x, y)$; afterwards, the empirical probabilities were compared with the results obtained by Ramírez and Aldama, 2000, the direct maximization of $L$ and the maximization of its logarithm (in this document was applied the natural logarithm, denoted by *Ln*). Table 3 contains a sample of the results, including the mean square error, and Figures 1 to 3, show the graphical association between empirical and theoretical values for the three cases.

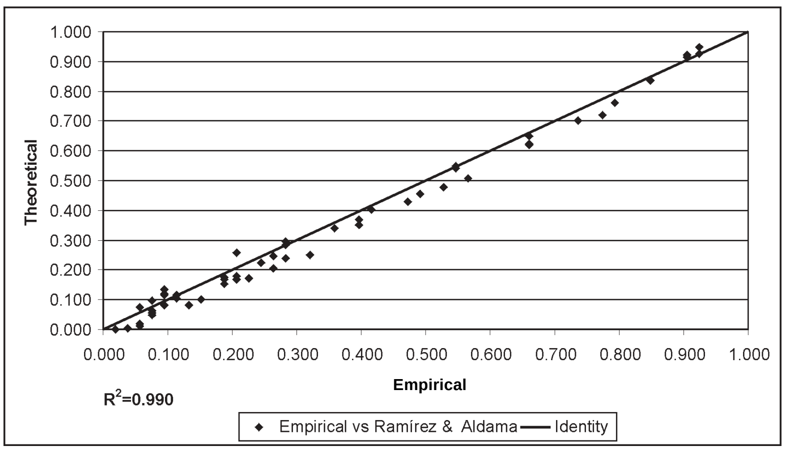


**Figure 1: Comparison between *F(x, y)* Empirical and the Function Obtained by Ramírez and Aldama (Ramirez, 2000)**

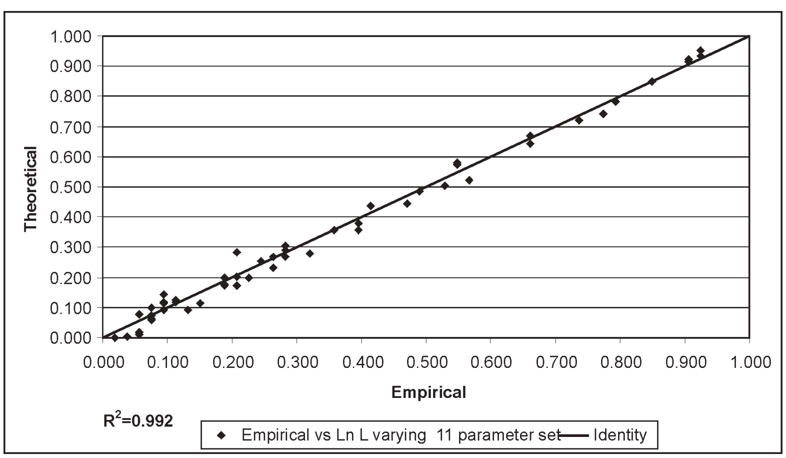


**Figure 2: Comparison between $F(x, y)$ Empirical and the Function Obtained with an GA by Maximizing log $L$**

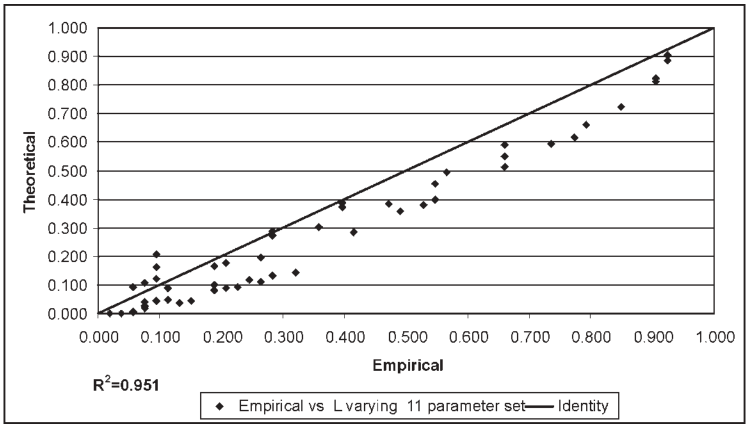


**Figure 3: Comparison between *F*(*x*, *y*) Empirical and the Function Obtained with an GA by Maximizing *L***

**Table 3**
**Estimation of the Non Exceedance Probabilities. Bivariate Double Gumbel Function. Huites Station**

| | *Ordered According to Q* | | *Values smaller or equal than Q and V* | *F Empirical* | *F Theoretical* | | |
|---|---|---|---|---|---|---|---|
| *I* | *Q(m³/s)* | *V(10⁶ m³)* | *k* | *k/(n+1)* | *Ramírez & Aldama* | *GA Max (Ln L)* | *GA Max (L)* |
| 1 | 15000 | 1842 | 48 | 0.906 | 0.912 | 0.915 | 0.811 |
| 2 | 14376 | 1928 | 48 | 0.906 | 0.920 | 0.922 | 0.821 |
| 3 | 11558 | 2930 | 49 | 0.925 | 0.949 | 0.954 | 0.903 |
| 4 | 10000 | 2966 | 49 | 0.925 | 0.927 | 0.935 | 0.886 |
| 5 | 8275 | 1625 | 45 | 0.849 | 0.835 | 0.851 | 0.723 |
| 6 | 7960 | 1250 | 42 | 0.792 | 0.761 | 0.783 | 0.661 |
| 7 | 6860 | 986 | 35 | 0.660 | 0.649 | 0.670 | 0.590 |
| 8 | 5580 | 1258 | 41 | 0.774 | 0.721 | 0.744 | 0.617 |
| 9 | 4828 | 992 | 35 | 0.660 | 0.622 | 0.643 | 0.550 |
| 10 | 4790 | 1245 | 39 | 0.736 | 0.700 | 0.722 | 0.594 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 40 | 1396 | 689 | 10 | 0.189 | 0.152 | 0.173 | 0.080 |
| 41 | 1349 | 1185 | 11 | 0.208 | 0.177 | 0.202 | 0.088 |
| 42 | 1319 | 384 | 4 | 0.075 | 0.065 | 0.072 | 0.042 |
| 43 | 1266 | 474 | 5 | 0.094 | 0.081 | 0.092 | 0.044 |
| 44 | 1165 | 720 | 8 | 0.151 | 0.100 | 0.116 | 0.044 |
| 45 | 1135 | 601 | 7 | 0.132 | 0.081 | 0.094 | 0.036 |
| 46 | 1127 | 435 | 4 | 0.075 | 0.055 | 0.063 | 0.027 |
| 47 | 1095 | 1768 | 6 | 0.113 | 0.104 | 0.118 | 0.046 |
| 48 | 1025 | 163 | 3 | 0.057 | 0.010 | 0.010 | 0.007 |
| 49 | 955 | 596 | 4 | 0.075 | 0.050 | 0.058 | 0.019 |
| 50 | 696 | 513 | 3 | 0.057 | 0.017 | 0.019 | 0.005 |
| 51 | 677 | 111 | 2 | 0.038 | 0.002 | 0.002 | 0.001 |
| 52 | 593 | 69 | 1 | 0.019 | 0.001 | 0.001 | 0.000 |
| | | | **Mean square error** | | 0.001 | 0.001 | 0.009 |

Table 3 and Figures 1 and 2, show that the results obtained by Ramírez and Aldama, 2000 and those obtained by maximizing the logarithms of function *L* by means of genetic algorithms (GA), have a very good agreement between the empirical and the calculated values, being a little bit better in the second case. On contrast, the direct maximization of *L* using the genetic algorithm produced a larger deviation between the empirical and the calculated values, as it is shown in Figure 3. This last result can be attributed to the fact that the values of the density function (eq. 4) are small by themselves; consequently, when performing the consecutive product of the 52 values to evaluate function *L*, the result tends to zero, and then, upon advancing in the number of iterations the genetic algorithm loses sensitivity in the search for the optimum set of parameters.

It is important to remember that when the first derivate of the function is set to zero, a maximum, a minimum or even an inflexion point can be obtained, so, when the first derivate is avoided to get the parameters of the function *L*, and those parameters are obtained by the direct maximization of *L*, there is not risk to fall in a minimum instead of a maximum in the function.

## 4. CONCLUSION

In this paper, an analytical determination of the density function of a Bivariate Double Gumbel distribution was made; the obtained results (eq. 4) were tested to be sure on their accuracy. The density function was used to obtain the parameters of the corresponding distribution function through the use of a genetic algorithm for the maximization of the likelihood function *L* or of the logarithm of *L*. The results thus obtained were compared with those calculated using a deterministic algorithm using empirical probabilities as parameter in order to judge its approximation. It was found that, for the case of *Huites* hydrometric station, the maximization of the logarithm of the probability function *L* using genetic algorithms provided the best correlation with respect to the empirical distribution, followed practicaly with the same accuracy by the deterministic method used by Ramírez and Aldama, 2000. The direct optimization of function *L* using also genetic algorithms did not give that good result, as it is shown in Table 1 and in Figure 3.

In this case, the use of the logarithm of *L* helped the genetic algorithm to be able to reach the set of parameters that achieved the best response to the objective function in the last generation. The shortcoming in the direct use of function *L* refers to the fact that when the density function, for each pair of values of flow rate and volume are very small, the product of such values tends to zero and the computer is no longer capable of representing it; as a consequence, the algorithm disagrees with the determination of the optimum set of parameters.

The use of genetic algorithms appears to be promising for the estimation of parameters that maximize the likelihood function *L*. However, the use of its logarithms is recommended to avoid the calculation of very small values that could produce accuracy problems. Joint application of L-moments (Eslamian and Feizi, 2007) and genetic algorithms are of interest for further investigations.

### *References*

[1] Bäck, T., *Evolutionary Algorithms in Theory and Practice,* Oxford University Press. (1996).

[2] Domínguez Mora, R., Fuentes Mariles, O., and V., Franco, Avenida de diseño, Fascículo A.1.10, Manual de Diseño de Obras Civiles, Comisión Federal de Electricidad, México, 51 (1980).

[3] Escalante Sandoval, C., Application of Bivariate Extreme Value Distribution to Flood Frequency Analysis: A Case Study of Northwestern Mexico, Nat Hazards, **42** (2007) 37–46.

[4] Eslamian, S. S. and H. Feizi, Maximum Monthly Rainfall Analysis Using *L*-moments for an Arid Region in Isfahan Province, Iran, Journal of Applied Meteorology and Climatology, **46**(4) (2007) 494–503.

[5] Goldberg, D. E., *Genetic Algorithms in Search, Optimization and Machine Learning,* Addison-Wesley, USA. (1989).

[6] González Villarreal, F, Contribución al análisis de frecuencias de valores extremos de los gastos máximos en un Río, Serie azul, Instituto de Ingeniería, UNAM, México, (1970) 24.

[7] Hiemstra, L. A. V., and D. M., Francis, The Runhydrograph–Theory and Application for Flood Predictions, Water Research Commission, Pretoria. (1979).

[8] Holland, J. H., *Adaptation in Natural and Artificial Systems*, The University of Michigan Press. (1975).

[9] Horbelt, W., Timmer, J., and H. U. Voss, Parameter Estimation in Nonlinear Delayed Feedback Systems from noisy data, Physics Letters A. **299** (2002) 513–521.

[10] Jiménez Espinoza, M., Diseño Integral de vertedores, Tesis Doctoral, UNAM, México. (2000).

[11] Kite, G. W., Frequency and Risk Analyses in Hydrology, Littletown, Colorado, USA, (1988) 224.

[12] Myiung, I. J., Tutorial on Maximum Likelihood, *J. Mathematical Psych.,* **47** (2003) 90-100.

[13] Ramírez, A. I., and A., Aldama, Análisis de frecuencias conjunto para la estimación de avenidas de diseño, Avances en Hidráulica–7. IMTA, México. (2000).

[14] Rao, A. R., and K. H., Hamed, *Flood Frequency Analysis*, CRC Press, USA, (2000) *Web Site: google.books.com.*

[15] Rivera Trejo, F., and C., Escalante Sandoval, Análisis comparativo de técnicas de estimación de avenidas de diseño, Ingeniería del Agua, **6**(1) (1999) 49-54.

[16] Rosenbrock, H. H., An Automatic Method for Finding the Greatest or Least Value of a Function, *Computer Journal,* **3** (1960) 175-184.

[17] Rossi, F., Florentino, M., and P., Versace, Two-Component Extreme Value Distribution for Flood Frequency Analysis, Water Resources Research, **20**(7) (1984) 847-856.

[18] Smith, R. L., Forecasting Records by Maximum Likelihood, *J. American Statistical Association,* **83**(402) (1988) 331-338.

[19] The MathWorks, MATLAB Reference Guide, The MathWorks, Inc. (1992).

## 6. APPENDIX

### 6.1. Appendix I

The derivate of the $e^u$ function is given by:

$$\frac{d}{dx}e^u = e^u\,\frac{du}{dx} \tag{I.1}$$

Partially deriving $F(x, y)$ with respect to $x$:

$$\frac{\partial F(x,y)}{\partial x} = e^{g(x,y)} \left( \frac{\partial g(x,y)}{\partial x} \right) \tag{I.2}$$

Considering the deriving formula for a product:

$$\frac{d(uv)}{\partial x} = v\frac{du}{dx} + u\frac{dv}{dx} \tag{I.3}$$

Partially deriving eq. I.2 with respect to $y$:

$$\frac{\partial^2 F(x,y)}{\partial y \partial x} = \frac{\partial g(x,y)}{\partial x} e^{g(x,y)} \left( \frac{\partial g(x,y)}{\partial y} \right) + e^{g(x,y)} \left( \frac{\partial^2 g(x,y)}{\partial y \partial x} \right) \tag{I.4}$$

Finally:

$$\frac{\partial^2 F(x,y)}{\partial y \partial x} = e^{g(x,y)} \left[ \frac{\partial g(x,y)}{\partial x} \frac{\partial g(x,y)}{\partial y} + \frac{\partial^2 g(x,y)}{\partial y \partial x} \right] \tag{I.5}$$

### 6.2. Appendix II

Partially deriving $g(x,y)$ with respect to $x$:

$$\frac{\partial g(x,y)}{\partial x} = -\left[ (-\ln F(x))^m + \left( (-\ln F(y))^m \right) \right]^{\left(\frac{1-m}{m}\right)} \left[ \frac{(-\ln Fx)^{m-1}}{F(x)} \frac{\partial F(x)}{\partial x} \right] \tag{II.1}$$

Similarly, for the partial derivative of $g(x,y)$ with respect to $y$:

$$\frac{\partial^2 g(x,y)}{\partial y \partial x} = \left[ (-\ln F(x))^m + \left( (-\ln F(y))^m \right) \right]^{\left(\frac{1-m}{m}\right)} \left[ \frac{-\ln F(y))^{m-1}}{F(y)} \frac{\partial F(y)}{\partial y} \right] \tag{II.2}$$

The mixed partial of $g(x, y)$ is obtained by partially deriving equation II.1 with respect to $y$:

$$\begin{aligned} \frac{\partial^2 g(x,y)}{\partial y \partial x} = -\Bigg\{ & \left[ (-\ln F(x))^m + \left( (-\ln F(y))^m \right) \right]^{\left(\frac{1-m}{m}\right)} \frac{\partial}{\partial y} \left[ \frac{(-\ln F(x))^{m-1}}{Fx} \frac{\partial F(x)}{\partial x} \right] \\ & + \left[ \frac{(-\ln F(x))^{m-1}}{F(x)} \frac{\partial F(x)}{\partial x} \right] \frac{\partial}{\partial y} \left[ (-\ln(F(x))^m + (-\ln F(y))^m \right]^{\frac{1-m}{m}} \Bigg\} \end{aligned} \tag{II.3}$$

Simplifying:

$$\frac{\partial^2 g(x,y)}{\partial y \partial x} = (m-1)\left[ (-\ln F(x))^m + (-\ln F(y))^m \right]^{\frac{1-2m}{m}} \left[ \frac{(-\ln F(y))^{m-1}}{F(y)} \frac{\partial F(y)}{\partial y} \right] \left[ \frac{(-\ln F(x))^{m-1}}{F(x)} \frac{\partial F(x)}{\partial x} \right] \tag{II.4}$$

### 6.3. Appendix III

The derivates of the marginal functions of a Double Gumbel function are given by:

$$\frac{\partial F(x)}{\partial x} = \frac{p_x}{c_1} e^{-e^{-\left(\frac{x+a_1}{c_1}\right)}} e^{-\left(\frac{x+a_1}{c_1}\right)} + (1-p_x)\frac{1}{c_2} e^{-e^{-\left(\frac{x+a_2}{c_2}\right)}} e^{-\left(\frac{x+a_2}{c_2}\right)} \tag{III.1}$$

$$\frac{\partial F(y)}{\partial y} = \frac{p_y}{c_3} e^{-e^{-\left(\frac{y+a_3}{c_3}\right)}} e^{-\left(\frac{y+a_3}{c_3}\right)} + (1-p_y)\frac{1}{c_4} e^{-e^{-\left(\frac{y+a_4}{c_4}\right)}} e^{-\left(\frac{y+a_4}{c_4}\right)} \tag{III.2}$$

### 6.4. Appendix IV

The accuracy in calculating the density function is very important in the application of the maximum likelihood method; because of this, several tests described in this appendix were made. First, the equivalence between eq. 4 and eq. 6, then their behavior in the particular case of the BGF and finally, the estimation of the volume under the surface in a given region are presented.

To check the equivalence of eq. 4 and eq. 6 in a BDGF, values of its 11 parameters were proposed (Table IV.1); the results obtained for some values of pairs ($x$, $y$) are presented in Table IV.2.

**Table IV.1**
**Proposal of Values for a Bivariate Double Gumbel Function**

| *Parameters* | *Value* | *Parameters* | *Value* |
|---|---|---|---|
| $a_1$ | -1240.8351 | $c_1$ | 281.07395 |
| $a_2$ | 1000 | $c_2$ | 100 |
| $a_3$ | -439832.88 | $c_3$ | 250433.49 |
| $a_4$ | 1000 | $c_4$ | 100 |
| $p_x$ | 0.8 | $p_y$ | 0.7 |
| $m$ | 1.5 | | |

**Table IV.2**
**Results Obtained for Bivariate Double Gumbel Function $f(x, y)$**

| | | $f(x, y)$ | |
|---|---|---|---|
| $x$ | $y$ | *Eq. 4* | *Eq. 6* |
| 1100 | 300000 | 9.429E-10 | 9.429E-10 |
| 1400 | 300000 | 8.385E-10 | 8.385E-10 |
| 1100 | 600000 | 7.068E-10 | 7.068E-10 |
| 1400 | 600000 | 1.058E-09 | 1.058E-09 |

Results given in Table IV.2 verify that equations 4 and 6 are equivalent to estimate the density function of a BDGF.

In order to examine equations 4 and 6 in a particular case, let us consider the BGF whose probability distribution function is given by:

$$F(x,y)=e^{-\left(e^{-m\left(\frac{x-a_1}{c_1}\right)}+e^{-m\left(\frac{y-a_3}{c_3}\right)}\right)^{\frac{1}{m}}} \tag{IV.1}$$

Partially deriving with respect to *x* and then with respect to *y*, the following simplified density function is obtained:

$$F(x,y)=\frac{e^{-m\left(\frac{x-a_1}{c_1}\right)}e^{-m\left(\frac{y-a_3}{c_3}\right)}}{c_1c_3}e^{-\left(e^{-m\left(\frac{x-a_1}{c_1}\right)}+e^{-m\left(\frac{y-a_3}{c_3}\right)}\right)^{\frac{1}{m}}} \tag{IV.2}$$

The form (IV.2) of the BGDEF is relatively easier to obtain from literature (Escalante Sandoval 2007), but it is not easy to find the form of a BDGDEF. In order to verify the equivalence of equations 4, 6 with IV.2 in some points (*x*, *y*), the parameters $a_1$, $c_1$, $a_3$, $c_3$ *and m* of Table IV.3 were suggested. The results are shown in Table IV.4.

**Table IV.3**
**Proposal of Values of a Bivariate Gumbel Function**

| *Parameters* | *Value* |
|---|---|
| $a_1$ | -1240.8351 |
| $c_1$ | 281.07395 |
| $a_3$ | -439832.88 |
| $c_3$ | 250433.49 |
| $m$ | 1.5 |

**Table IV.4**
**Comparison of the Bivariate Gumbel Values of *f*(*x*, *y*)**

| | | *F(x, y)* | | |
|---|---|---|---|---|
| *x* | *Y* | *Eq. 4* | *Eq. 6* | *Eq. IV.2* |
| 1100 | 300 000 | 2.060E-09 | 2.060E-09 | 2.060E-09 |
| 1400 | 300 000 | 1.272E-09 | 1.272E-09 | 1.272E-09 |
| 1100 | 600 000 | 1.259E-09 | 1.259E-09 | 1.259E-09 |
| 1400 | 600 000 | 1.767E-09 | 1.767E-09 | 1.767E-09 |

From Table IV.4, it can be verified that equations 4, 6 and IV.2 are equivalent for purposes of estimating the density function of a BGF.

### *6.4.1. Verification in Estimating the Bivariate Double Gumbel Density Function by the Calculation of the Volume under the Surface f(x, y)*

In order to check the analytical solution obtained for the BDGDEF, an approach to the volume under the surface *f(x, y)*, in a region such as those presented in Figure IV.1 is compared with the probability of being in the region as obtained from the distribution function (eq. 1) considering extreme *x, y* values and assuming values for the 11 parameters of the function.

The volume under the density function in a region such as that presented in Figure IV.1 can be estimated as:

$$Volume \approx (x\ \text{max} - x\ \text{min})(y\ \text{max} - y\ \text{min}) \sum_i \frac{f(xi, yi)}{n}; i = 1, 2, 3 \ldots n \qquad \text{(IV.3)}$$

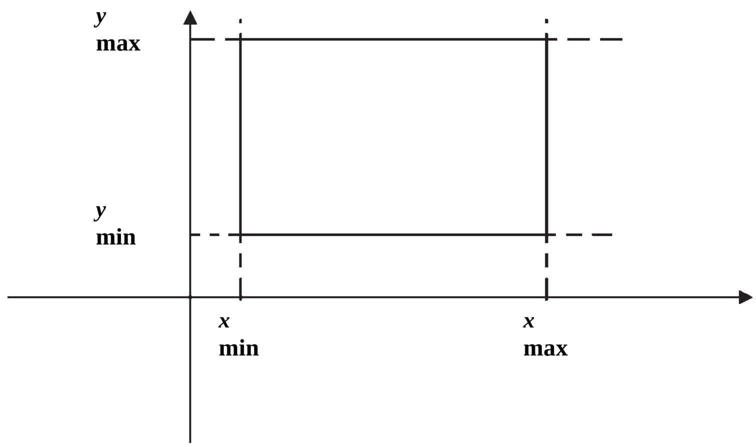


**Figure IV.1: Region under the Curve *f*(*x*, *y*)**

In another way, the probability of being in the region can be calculated from the distribution function given by eq. 1 as:

$$Volume = F(x_{max}, y_{max}) - F(x_{min}, y_{max}) - F(x_{max}, y_{min}) + F(x_{min}, y_{min}) \qquad \text{(IV.4)}$$

Taking into account Figure IV.1 and assuming values of $x_{min} = 1100$ and a $x_{max} = 1400$, with a $\Delta x = 25$, $y_{min} = 300000$ and a $y_{max} = 600000$,with a $\Delta y = 25000$ and the parameters given in Table IV.5, equations IV.3 and IV.4 are applied with the results presented on Table IV.6 :

**Table IV.5**
**Parameters Considered in Estimating the Volume under the Surface *f(x, y)***

| *Parameters* | *Value* | *Parameters* | *Value* |
|---|---|---|---|
| $a_1$ | -1240.8351 | $c_1$ | 281.07395 |
| $a_2$ | 1000 | $c_2$ | 100 |
| $a_3$ | -439832.88 | $c_3$ | 250433.49 |
| $a_4$ | 1000 | $c_4$ | 100 |
| *px* | 0.8 | *py* | 0.7 |
| *m* | 1.5 | | |

**Table IV.6**
**Results in Estimating the Volume under the Surface with eq. IV.3 and IV.4**

| *Eq. IV.3* | *Eq. IV.4* |
|---|---|
| 0.1096 | 0.1033 |

The difference between the volumes reported in Table IV.6 is about 6% and can be attributed to the approximation given by eq. IV.3.