

# PROBABILITY WEIGHTED MOMENTS ESTIMATORS FOR THE GEV DISTRIBUTION FOR THE MINIMA

**Jose A. Raynal-Villasenor**

Department of Civil and Environmental Engineering, Universidad de las Americas, Puebla  
72810 Cholula, Puebla, Mexico

## ABSTRACT

The probability weighted moment (PWM) estimators for the parameters and quantiles, using the general extreme value distribution for the minima (GEVM), is presented towards its application in low flow frequency analysis. The procedures to compute the parameters and design events (quantiles) for several return periods are shown in the paper. Two measures of goodness of fit tests are contained in the paper to compare the proposed methodology with other models in competition. A full example of application is presented in the paper to show how easy is to apply the proposed methodology.

**Keywords:** *low flow, frequency analysis, parameter estimation, confidence limits, method of probability weighted moments.*

## 1. INTRODUCTION

A subject of paramount interest in planning and design of water works is that related with low flow frequency analysis. Due to the characteristic that design values have, given that they are linked to a return period or to an exceedance probability, the use of mathematical models known as probability distribution functions is a must. Among the most widely used probability distribution functions for hydrological analysis, related with low flow frequency analysis, are (Kite [1]; Salas and Smith [2]; Rao and Hamed [3]; Raynal-Villasenor [4]):

- a) Three parameters Log-Normal (LN3)
- b) Pearson type III (PIII)
- c) Extreme Value Type III (EVIII)
- d) General Extreme Value for the minima (GEVM)

The first three probability distribution functions have been applied to low flow frequency analysis by Matalas [5]. Gumbel [6] developed the theoretical grounds and hydrological applications for the extreme value type III distribution for the minima (EVIIM), the well-known Weibull distribution. This distribution has been applied since the first third of the XX Century to the analysis of dynamic breaking strength of materials, Weibull [7] and [8]. Kite [1] provided with a computer program to estimate the parameters of the EVIIM distribution using the methods of moments (MOM) and maximum likelihood (ML). More recently, Lee and Kim [9] used the two-parameter Weibull distribution with bayesian Markov chain Monte Carlo and maximum likelihood estimates to assess the uncertainty of low frequency analysis. The estimation of ML parameters of EVIIM distribution has some difficulties when using the Newton-Raphson method as have been pointed out by Offinger [10]. Durrans and Tomic [11] compared five methods of estimation of parameters for the Log-Normal distribution in fitting the lower tail of such distribution. Smakhtin [12] made a review of 20 years of research results with regard to low flow hydrology. Yue and Wang [13] studied the scaling of Canadian rivers to regionalize the low flows. Ouarda Taha et al. [14] presented a brief review of statistical models that are commonly used in the estimation of low flows both at sites with a reliable stream flow record and sites remote from data sources. Hao and Singh [15] applied the maximum entropy method to the Burr III distribution and compared the results with the MOM, ML and probability weighted moments (PWM); they found no differences on the quantiles for small return period, the differences increased for large period returns. Iacobellis [16] studied the evaluation of a flow duration curve with assigned a T-year return period with beta and complementary beta distributions.

The use of the general extreme value distribution for the minima (GEVM) with PWM estimators for the parameters and quantiles are proposed in the paper. A complete example of application of the proposed methodology is contained in the paper, through the application of common spreadsheets framework provided by Excel<sup>®</sup> (Excel is a registered trademark of Microsoft Corporation, Inc.).

## 2. PROBABILITY DISTRIBUTION AND DENSITY FUNCTIONS OF THE GEVM

The probability distribution function of the GEVM distribution for the minima is, Raynal [17]:

$$\Pi(x) = \exp \left\{ - \left[ 1 - \beta(\omega - x) / \alpha \right]_{\beta}^1 \right\} \quad (1)$$

where  $\omega$ ,  $\alpha$  and  $\beta$  are the location, scale and shape parameters, respectively.  $\Pi(x)$  is the probability distribution function of the random variable  $x$  and for the case of low flow frequency analysis is equal to the exceedance probability,  $Pr(X > x)$ . The scale parameter must meet the condition that  $\alpha > 0$ . The domain of variable  $x$  in GEVM distribution is as follows:

1) For  $\beta < 0$ :

$$-\infty < x \leq \omega - \alpha / \beta \quad (2)$$

2) For  $\beta > 0$ :

$$\omega - \alpha / \beta \leq x < \infty \quad (3)$$

The probability density function for the GEVM distribution is, Raynal [17]:

$$\pi(x) = \frac{1}{\alpha} \exp \left\{ - \left[ 1 - \beta(\omega - x) / \alpha \right]^{1/\beta} \right\} \left[ 1 - \beta(\omega - x) / \alpha \right]^{1/\beta - 1} \quad (4)$$

where  $\pi(x)$  is the probability density distribution of random variable  $x$ .

### 3. PWM ESTIMATORS FOR THE PARAMETERS OF THE GEVM DISTRIBUTION

By definition, Greenwood et al [18], a probability distribution function,  $F(x) = \text{Prob}(X \leq x)$ , can be characterized by the probability weighted moments :

$$M_{i,j,k} = \int_0^1 x^i F^j(x) (1 - F(x))^k dF(x) \quad (5)$$

where  $i, j$  and  $k$  are real numbers. Taking the following convention, Greenwood et al [18]:

$$M_{1,0,k} = M_{(k)} \quad (6)$$

an unbiased estimator of  $M_{(k)}$  is, Maciunas Landwehr et al [19]:

$$M_{(k)} = \frac{1}{N} \sum_{i=1}^{N-k} \frac{\binom{N-i}{k}}{\binom{N-1}{k}} \quad (7)$$

and  $k$  is a non-negative integer and the  $x$ 's,  $i = 1, 2, \dots, N$ , have been rank ordered from  $x_1$  to  $x_N$ :

$$x_1 < x_2 < \dots < x_N \quad (8)$$

With the aid of equations (57) and (4), the following expressions had been found for the general form of the probability weighted moments for the GEV distribution for the minima, Raynal-Villasenor [20]:

$$M_{(k)} = \frac{\left(\omega - \frac{\alpha}{\beta}\right)}{(1+k)} + \frac{\alpha \Gamma(1+\beta)}{\beta(1+k)^\beta} \quad (9)$$

where  $\Gamma(\cdot)$  is the complete gamma function. Eq. (9) is valid only if  $\beta > -1$ . Now, using such an equation to set up a system of equations as many parameters as the distribution has, one can obtain the following expressions for the parameters of the GEV distribution for the minima:

$$\hat{\alpha} = \frac{(\hat{M}_{(0)} - 2\hat{M}_{(1)})^2 \hat{\beta}}{\Gamma(1 + \hat{\beta})(\hat{M}_{(0)} + 4\hat{M}_{(3)} - \hat{M}_{(1)})} \tag{10}$$

$$\hat{\omega} = \hat{M}_{(0)} - \left[ \frac{(\hat{M}_{(0)} - 2\hat{M}_{(1)})^2}{(\hat{M}_{(0)} + 4\hat{M}_{(3)} - \hat{M}_{(1)})} \right] \left[ 1 - \frac{1}{\Gamma(1 + \hat{\beta})} \right] \tag{11}$$

$$\hat{\beta} = \frac{\text{Ln} \left[ \frac{(\hat{M}_{(0)} - 2\hat{M}_{(1)})^2}{(\hat{M}_{(0)} + 4\hat{M}_{(3)} - \hat{M}_{(1)})} \right]}{\text{Ln}(2)} \tag{12}$$

where:

$$\hat{M}_{(0)} = \frac{1}{n} \sum_{i=1}^n x_i \tag{13}$$

$$\hat{M}_{(1)} = \frac{1}{n(n-1)} \sum_{i=1}^{n-1} x_i (n - i) \tag{14}$$

$$\hat{M}_{(2)} = \frac{1}{n(n-1)(n-2)} \sum_{i=1}^{n-2} x_i (n - i)(n - i - 1) \tag{15}$$

**4. DESIGN VALUES FOR THE GEVM DISTRIBUTION**

The design values (quantiles) for the GEVM distribution can be obtained by inverting the GEVM distribution function:

$$Q_T = \omega + \frac{\alpha}{\beta} \left\{ \left[ -\text{Ln}\left(1 - \frac{1}{T_r}\right) \right]^\beta - 1 \right\} \tag{16}$$

where  $Q_T$  are the design values and  $T_r$  is the return period associated with such design values.

**5. GOODNESS OF FIT TESTS FOR THE PARAMETERS OF THE GEVM DISTRIBUTION**

The two goodness of fit tests considered in this paper are:

- 1) Standard error of fit, SEF, Kite [1]

$$SEF = \left[ \frac{\sum_{i=1}^N (x_i - y_i)^2}{(N - n_p)} \right]^{1/2} \tag{17}$$

where  $x_i$  are the descending ordered historical values of the sample,  $y_i$  are the values produced by the distribution function corresponding to the same return periods of the historical values,  $N$  is the sample size, and  $n_p$  is the number of parameters of the distribution function, in this case  $n_p = 3$ .

- 2) Mean absolute relative deviation, MARD, Jain and Singh [21]

$$MARD = \frac{100}{N} \sum_{i=1}^N \left| \frac{(x_i - y_i)}{x_i} \right| \tag{18}$$

**6. NUMERICAL EXAMPLE**

The gauging station Villalba is located in the San Pedro River in Northwestern Mexico and has been selected to analyze its sample of annual one-day low flows, using the GEVM distribution with the PWM method of estimation of its parameters and design values.

The geographical location of gauging station Villalba, Mexico is shown in figure 1.

The first step in the computations is to obtain basic statistics of the one-day low flow sample and such statistics have been obtained by the application of common spreadsheets framework provided by Excel<sup>®</sup> (Excel is a registered trademark of Microsoft Corporation, Inc.), they are shown in figure 2.

The parameters, the goodness of fit measures, and design values and its confidence limits obtained through the use of the application of common spreadsheets framework provided by Excel<sup>®</sup> (Excel is a registered trademark of Microsoft Corporation, Inc.), they are shown in figure 3.

The comparison between the histogram of flood data and the theoretical probability density function is shown in figure 4. The figure 5 shows the empirical and theoretical frequency curves for the PWM estimation of parameters for the GEVM distribution to the one-day low flow sample of gauging station Villalba, Mexico. All the figures mentioned before have been obtained through the use of the application of common spreadsheets framework provided by Excel<sup>®</sup> (Excel is a registered trademark of Microsoft Corporation, Inc.)

## 7. DISCUSSION OF RESULTS

The easy use of proposed methodology has been shown by the development of the numerical example. By using the common spreadsheets framework provided by Excel<sup>®</sup> (Excel is a registered trademark of Microsoft Corporation, Inc.), the user has all the time on sight the formulas and results and a possible error could be spotted very easily.

The tables shown in figure 3 contain all the required results for a low flow frequency analysis study for a particular set of low flow data. In these tables are contained the values of the parameters, their goodness of fit measures. Two different measures of goodness of fit are provided to choose among competing models. The design values for several return periods are depicted in figure 4. The information contained in the graphs are produced by the common spreadsheets framework provided by Excel<sup>®</sup> (Excel is a registered trademark of Microsoft Corporation, Inc.), are informative on how good is the adjustment of a particular probability distribution function to a particular set of data, this is given by the graph showing the low flow data and the adjusted model (figure 6). The graph that shows the theoretical probability density function and histogram of low flow data is figure 5.

## 8. CONCLUSIONS

A proposed methodology has been presented for low flow frequency analysis, by using the GEVM distribution coupled with PWM method. The use of the common spreadsheets framework provided by Excel<sup>®</sup> (Excel is a registered trademark of Microsoft Corporation, Inc.) is particularly useful in education and training. The proposed methodology compares well with the existing probability distribution functions when the PWM method is applied. The straightforward application of the proposed methodology to real data, as it has been shown in example contained in the paper, makes it a versatile tool to train students or technical personnel in the field with a personal computer and a printer.

## 9. ACKNOWLEDGEMENTS

The author wish to express their gratitude to the Universidad de las Americas, Puebla for the support provided in the realization of this paper.

## 10. REFERENCES

- [1] G. W. Kite, Flood and Risk Analyses in Hydrology, Water Resources Publications, Littleton. (1988).
- [2] J. D. Salas and R. Smith, Computer Programs of Distribution Functions in Hydrology, Colorado State University, Fort Collins. (1980).
- [3] R. Rao and K. H. Hamed, Flood Frequency Analysis, CRC Press, Boca Raton. (2000).
- [4] J. A. Raynal-Villasenor Frequency Analysis of Hydrologic Extremes, Lulu.com, <http://www.lulu.com/spotlight/flodro4dot0atgmaildotcom> . (2010).
- [5] N. C. Matalas, Probability Distribution of Low Flows, USGS Professional Paper No. 434-A, US Printing Office, Washington. (1963).
- [6] E. J. Gumbel, Statistics of Extremes, Columbia University Press, New York. (1958).
- [7] W. A. Weibull, Statistical Theory of Strength of Materials, Ingeniörs Vetenskaps Akademien Handlingar, No. 151, Generalstabens Litografiska Anstalts Förlang, Stockholm. (1939a).
- [8] W. A. Weibull, The Phenomenon of Rupture of Solids, Ingeniörs Vetenskaps Akademien Handlingar, No. 153, Generalstabens Litografiska Anstalts Förlang, Stockholm. (1939b).
- [9] K. S. Lee and S. U. Kim, Identification of Uncertainty in Low Flow Frequency Analysis Using Bayesian MCMC Method, Hydrol Proc 22(12), 1949-1964 DOI: 10.1002/hyp.6778. (2008).
- [10] R. Offinger, Schätzer in drei Parametrischen Weibull-Modellen und Untersuchung ihrer Eigenschaften mittels Simulation, Ph. D. diss., Diplomarbeit Institut für Mathematik der Universität Augsburg, Augsburg. (1996).

- [11] S. R. Durrans and S. Tomic, Comparison of Parametric Tail Estimators for Low Flow Frequency Analysis, *J. Am. Wat. Resour. Assoc.*, 37(5), 1203-1214 DOI: 10.1111/j.1752-1688.2001.tb03632.x. (2001).
- [12] V. Smakhtin, Low Flow Hydrology: A Review, *J. Hydrol.*, 240(3-4), 147-186 DOI: 10.1016/S0022-1694(00)00340-1. (2001).
- [13] S. Yue and C. Y. Wang, Scaling of Canadian Low Flows, *Stoch. Environ. Res. Risk Assess.*, 18(5), 291-305 DOI: 10.1007/s00477-004-0176-6. (2004).
- [14] B. M. J. Ouarda Taha, C. Charron and A. St-Hilaire, Statistical Model and the Estimation of Low Flows, *Can. Wat. Resour. J.*, 33(2)SI, 195-205. (2008).
- [15] Z. Hao Z and V. P. Singh, Entropy-Based Parameter Estimation for Extended Three-Parameter Burr III Distribution for Low Flow Frequency Analysis, *Trans ASABE*, 52(4), 1193-1202. (2009).
- [16] V. Iacobellis, Probabilistic Model for the Estimation of T Year Flow Duration Curves, *Wat. Resour. Res.*, 44(2), DOI: 10.1029/2006WR005400. (2008).
- [17] J. Raynal, J., Moment estimators of the GEV distribution for the minima, *Appl. Water Sci. J.*, DOI 10.1007/s13201-012-0052-3. (2012).
- [18] J. A. Greenwood, J. Maciunas Landwehr, N. C. Matalas and J. R. Wallis, Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressible in Inverse Form, *Wat. Resour. Res.*, 15(5), 1049-1054. (1979).
- [19] J. M. Maciunas Landwehr, N. C. Matalas, and J. R. Wallis, Probability Weighted Moments Compared with Some Traditional Techniques in Estimating Gumbel Parameters and Quantiles, *Wat. Resour. Res.*, 15(5), 1055-1064. (1979).
- [20] J. A. Raynal-Villasenor, Computation of probability weighted moments estimators for the parameters of the general extreme value distribution (maxima and minima), *Hydrol. Sci. Tech. J.*, 3(1-3), 47-52. (1987).
- [21] D. Jain and V. P. Singh, Estimating Parameters of EV1 Distribution for Flood Frequency Analysis, *Wat. Resour. Bull.*, 23(1) 59-71. (1987).



Figure 1. Location of gauging station Villalba, Mexico

	A	B
1	<b>LOW-FLOW FREQUENCY ANALYSIS</b>	
2	Gauging Station: Villalba, Mexico (1939-1991)	
3		
4	<i>Descriptive Statistics</i>	
5	Mean	0.3306
6	Standard Error	0.0201
7	Median	0.3280
8	Mode	0.4850
9	Standard Deviation	0.1465
10	Sample Variance	0.0214
11	Kurtosis	0.1906
12	Skewness	0.5853
13	Range	0.6250
14	Minimum	0.0790
15	Maximum	0.7040
16	Sum	17.5214
17	Count	53.0000
18	Largest(1)	0.7040
19	Smallest(1)	0.0790
20	Confidence Level(95.0%)	0.0404

Figure 2. Data statistics for gauging station Villalba, Mexico

	A	B	C	D	E
1	<b>LOW-FLOW FREQUENCY ANALYSIS</b>				
2	Gauging Station: Villalba, Mexico (1939-1991)				
3	General Extreme Value Distribution				
4	Method: Probability Weighted Moments				
5					
6	<b>Probability Weighted Moments</b>				
7	$M_{(0)} =$	0.3306	$m^3/s$		
8	$M_{(1)} =$	0.1240	$m^3/s$		
9	$M_{(3)} =$	0.0469	$m^3/s$		
10					
11	<b>Estimation of Parameters and Goodness of Fit Tests</b>				
12	$1+\beta =$	1.4534			
13	$\Gamma(1+\beta) =$	0.8893			
14					
15	Location Parameter =		0.3687	$m^3/s$	
16	Scale Parameter =		0.1561	$m^3/s$	
17	ShapeParameter =		0.4534		
18					
19	SEF =	0.02		MARD =	3.15

Figure 3. Estimation of parameters (GEVM-PWM) and goodness of fit measures for gauging station Villalba, Mexico

	A	B	C
22	<b>Design Values</b>		
23			<b>Design</b>
24			<b>Values</b>
25	<b>Return Period</b>		$Q_T$
26	<b>Tr (years)</b>	<b>F(x)</b>	<b>(m<sup>3</sup>/s)</b>
27	2	0.5	0.42
28	5	0.8	0.54
29	10	0.9	0.59
30	20	0.95	0.62
31	50	0.98	0.65
32	100	0.99	0.67
33	500	0.998	0.69
34	1000	0.999	0.70

Figure 4. Design values (quantiles) (GEVM-PWM) for gauging station Villalba, Mexico

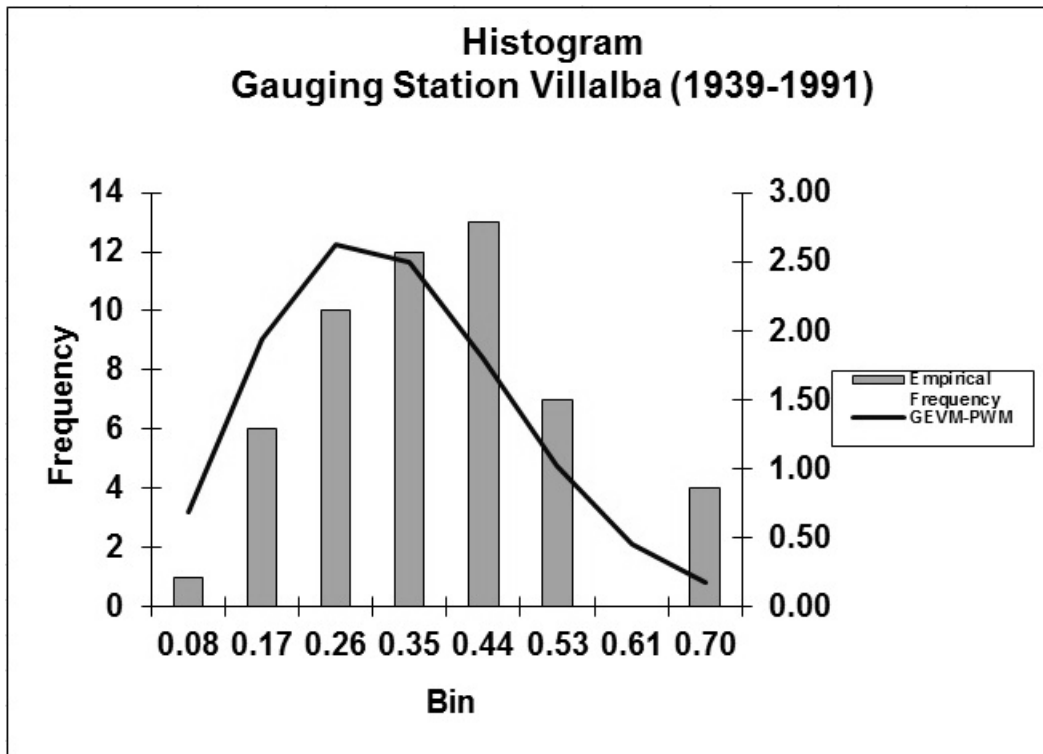


Figure 5. Histogram and theoretical probability density function for gauging station Villalba, Mexico

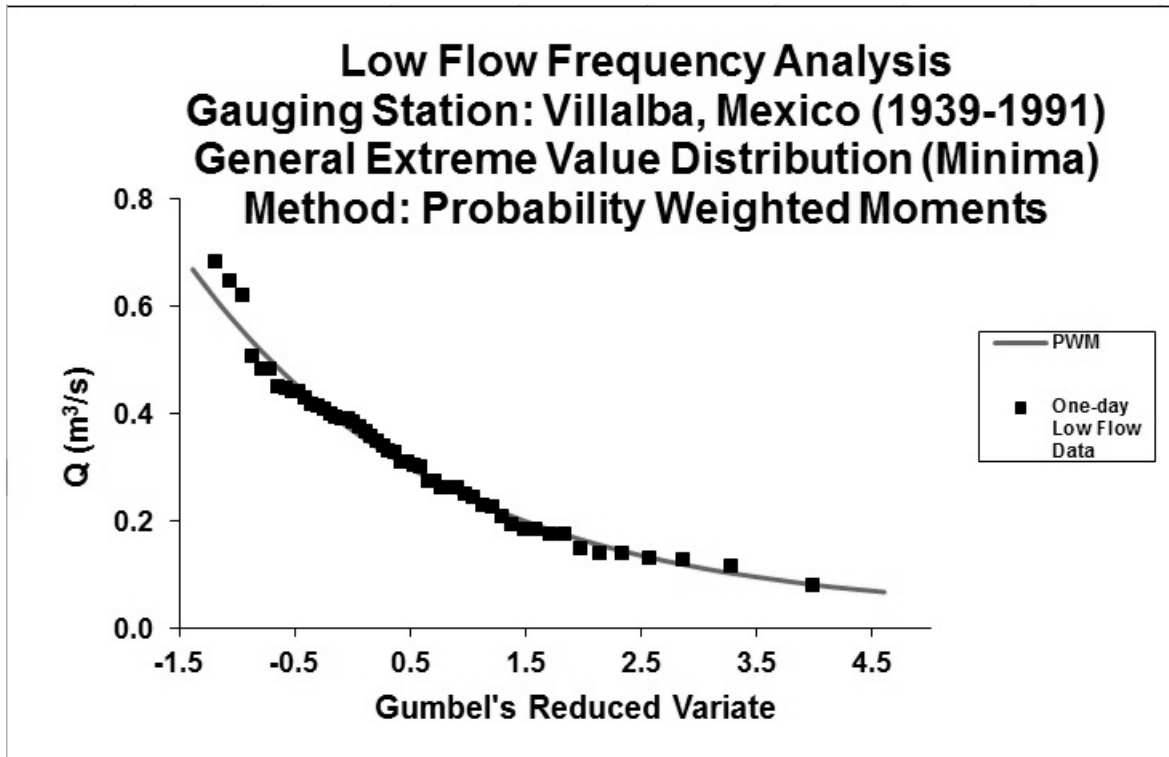


Figure 6. Empirical and theoretical frequency curves for several models applied to one-day low flow data at gauging station Villalba, Mexico