Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 198 (2022) 560–565

Procedia Computer Science

www.elsevier.com/locate/procedia

The 2nd International Workshop of Innovation and Technologies (IWIT 2021)

November 1-4, 2021, Leuven, Belgium

A preliminary analysis for selecting the best hydrological probability density functions of annual peak flows associated to various return periods in some rivers of Colombia

Óscar E. Coronado-Hernández[a,*], Julio Ramos-Urzola[a], Luis F. Julio-Amigo[a], Jairo R. Coronado-Hernández[b], Gustavo Gatica[c], Nohora Mercado-Caruso[b]

[a]*Universidad Tecnológica de Bolívar, Cartagena 131001, Colombia; ocoronado@utb.edu.co*
[b]*Universidad de la Costa, Barranquilla 080002, Colombia; {jcoronad18, nmercad1}@cuc.edu.co*
[c]*Universidad Andres Bello, Santiago de Chile 7500971, Chile; ggatica@unab.cl*

Abstract

Analysis of extreme events of annual flow peaks are used for sizing hydraulic structures for specified return period. Cumulative distribution functions are applied to annual flow peak records in order to obtain extreme values with different return periods. In Colombia, when performing a frequency analysis, hydrological planners often do not know a priori the best cumulative distribution function for making analysis. In the present research, annual flow peak records from 49 hydrometric stations located in important rivers were collected, with the objective of determining the most representative cumulative distribution function. The best results were achieved using the generalized extreme value (GEV) cumulative distribution function with the maximum likelihood method.

Keywords: annual flow peaks; Colombia; cumulative distribution function; hydrometric stations

* Corresponding author. Tel.: +57-3013715398
E-mail address: ocoronado@utb.edu.co

1877-0509
10.1016/j.procs.2021.12.286

1. Introduction

Annual flow peaks records are used to estimate extreme values for different return periods [1]. These values are used to plan hydraulic projects such as urban drainage networks, flood protection, overflows [2], and for basin erosion analysis [3]. In recent years, due to anthropogenic and climate changes, the trends of records can change probability occurrence of these events. Then, they are likely to increase or decrease in frequency [4], making it necessary to determine the cumulative distribution functions that best simulate actual trends of annual flow peaks [5,6].

Typical analyzed cumulative distribution functions to know the trend in the recorded annual flow peaks are: Generalized Extreme Value (GEV) [7], Gumbel [1], Log-Pearson Type III [8] and Pearson Type III [9]. The parameters of the above distributions are determined by applying the maximum likelihood method and method of moments [10]. To select the probability distribution function that best fits the trend of the data, different goodness of fit tests are usually used, such as the Chi Square test [11,12].

The extension of Colombia presents variety in its topography, which is mainly composed of the Colombian Andean Mountain, an independent mountain system, costal zones, and the inter-Andean valleys. The meteorology is affecting by the inter-tropical convergence zone, high pressure systems, the East Caribbean Winds, the inter-tropical front, the Mesoscale meteorology, the mid-latitude trough, the South Atlantic convergence zone, and the tropical trough of the upper troposphere. These variations produce different hydrological behaviors in the watershed areas of rivers in Colombia [13].

In this study, 49 hydrometric stations with annual flow peaks records distributed in some rivers of Colombia were analyzed no only to detect the most representative hydrological distribution in the analyzed stations but also to elaborate curves of runoff per unit watershed areas that can be used to have reference values of annual flow peaks associated to different return periods. Results show that the GEV, with the maximum likelihood method, presented the best fit of annual flow peaks series using the Chi-Square test. At the end, a practical application of the curves of runoff per unit watershed area versus watershed areas is presented in order to see the use of the current research

Nomenclature

f	hydrological probability density function
L	likelihood function
n	sample
M_i	modelled value
Me	mean
Max	Maximum
Min	Minimum
Sd	Standard deviation
q	water flow
Rec_i	Recorded value
X^2	Chi-square test
β	location parameter
α	scale parameter
μ	mean of annual flow peaks series
κ	shape parameter
ϕ	centroid of a HPDF
γ	gamma function.

2. Material and Methods

This section shows the used methodology to know the best hydrological probability density function (HPDF), which can be used for engineers and designers to estimate a priori reference values of annual flow peaks when a frequency analysis is performed.

At the beginning, the HPDFs of Gumbel, GEV, Pearson III, Log-Pearson III, and Normal were applied to annual peak flow records in order to compute values for various return periods. These HPDF were used due to their utilization in recent researches [12,14]. The annual series of maximum flow values for the different stream gauge stations was fit using the Hyfran programme [15], which is a practical tool for computing a HDPF. The used HPDF are presented in Table 1.

Table 1. Hydrological probability density function

HPDF	Equation	Equation Number	Reference
Gumbel (Gum)	$f(q) = \frac{1}{\alpha} e^{\left[-\frac{q-\mu}{\alpha} - e^{-\frac{q-\mu}{\alpha}}\right]}$	(1)	[16]
GEV	$f(q) = \frac{1}{\alpha}\left[1 - \frac{k}{\alpha}(q-\mu)\right]^{\frac{1}{k}-1} e^{-\left[1-\frac{k}{\alpha}(q-\mu)\right]^{\frac{1}{k}}}$	(2)	[7]
Pearson III (Pea)	$f(q) = \frac{1}{\lvert\alpha\rvert\gamma(k)}\left(\frac{q-\beta}{\alpha}\right)^{k-1} e^{\left[-\left(\frac{q-\beta}{\alpha}\right)\right]}$	(3)	[9]
Normal	$f(q) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]}$	(4)	[11]

where:
f(q) = hydrological probability distribution, q = water flow, μ = mean of annual flow peaks series, α = scale parameter, κ = shape parameter, β = location parameter, and γ = gamma function.

To fit the parameters of each HPDF, the Maximum Likelihood and Method of Moments were used to adjust the HPDF of annual flow peaks records for having the estimation of the parameters. Table 2 shows a description of the used methods.

Table 2. Methods for adjusting a HPDF

Method	Description	Equation	Equation Number	Referen ce
Maximum Likelihood (ML)	This method adjusts the best estimation of a parameter of a HPDF maximazing the joint probability ocurrence of an observed sample.	$L = \prod_{i=1}^{n} f(q_i)$	(5)	[17]
Method of Moments (MM)	This method finds a good fit of parameters of a HPDF considering these functions around the origin are the same to the corresponding moments of a sample data.	$\varphi = \int_{-\propto}^{\propto} q f(q) dq$	(6)	[18]

where:
L= likelihood function, n= sample, ϕ = centroid of a HPDF.

The Chi-squared goodness of fit test was used to evaluate whether the HPDF fit the trend of the data appropriately. The formulation of the Chi-squares (X^2) is:

$$X^2 = \frac{\sum_i^n (Rec_i - M_i)^2}{M_i} \tag{1}$$

where, M_i and Rec_i are modeled and recorded values, respectively.

Considering the total hydrometric stations used in the current research, the chi-square values of mean, standard deviation, and maximum and minimum values were analyzed to determine the best HPDF.

3. Results and discussion

3.1. Case study and data

The case study corresponds to the watershed areas of the following rivers in Colombia: Aracataca, Ariguaní, Badillo, Креaré, Cauca, Ceibas, Cesar, Chivor, Frío, Garupal, Guavio, Lengupá, Magdalena, Nare, Nechí, Negro, Páez, Ranchería, Rucio, San Jorge, Sinú, Sogamoso, Sumapaz, Sutatausa, Ullucos, Une, Upía, and Vetas. To select the hydrometric stations, the availability of information was taken into account, as well as how evenly they were distributed throughout Colombia. Figure 1 presents the location of the used hydrometric stations for this analysis. In the compiled data of annual flow peaks records in Colombia, the following aspects were considered for each hydrometric station: (i) outliers data were eliminated of series; and (ii) registered periods of hydrometric stations were used considering at least thirty two (32) years. The annual peak flow records from 49 hydrometric stations distributed in Colombia were collected, which have watershed area varying from 35 to 161292 km^2. The annual peak flow records were extracted from the Institute of Hydrology, Meteorology and Environmental Studies (IDEAM in Spanish). Appendix 1 presents a list of the analyzed hydrometric stations. These kinds of analyses are needed in many places [19, 20].

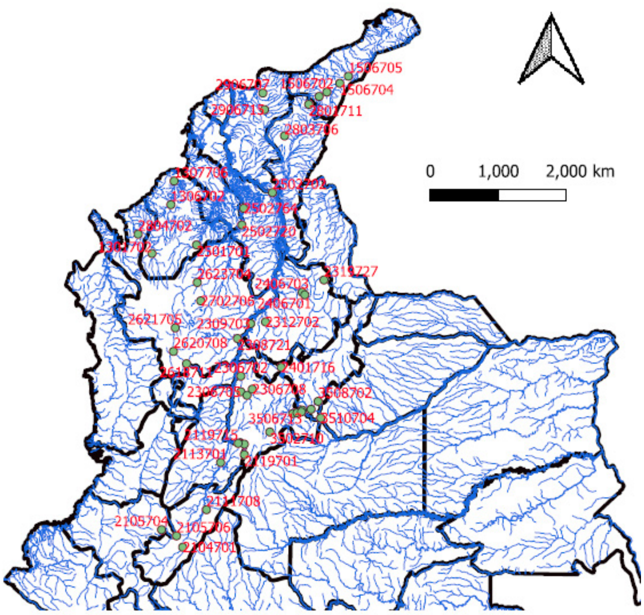

Fig. 1. Distribution of hydrometric stations used.

3.2. HPDF modeling and Chi-square test analysis

This section shows the results obtained from the methodology presented above, where the best HPDF for each hydrometric station to compute annual peak flow for different return periods is analyzed. Table 3 presents the compilation of the Chi-square values obtained for all hydrometric stations.

Table 3. Values of Chi-square test for the Adjustments of the HPDF for all hydrometric stations

CFD	Me	Max	Min	Sd
Gev-MV	5.93	16.95	1.18	4.04
Gev-MM	6.76	31.60	1.26	5.25
Gum-MV	7.96	44.44	1.00	7.47
Gum-MM	8.27	35.20	0.74	6.58
Normal	9.71	25.60	1.00	6.67
Pea -MV	38.63	408.00	1.26	94.67
Pea -MM	6.88	41.80	0.00	6.90

where,
Me (Mean); Max (Maximum); Min (Minimum); and Sd (Standard deviation)

According to results presented in Table 3, the best fits were obtained using the Gev-MV distribution function, with an average Chi-square value of 5.93 in the 49 hydrometric stations used. Also, the lowest standard deviation of the Chi-square test was found in the Gev-MV with a value of 4.04. The worst-fitting distribution function according to the Chi-square test was the Pea-MV (with an average value of 38.63). The Gev-MM, Gum-MV, Gum-MM, and Pea-MM distributions also presented adequate fits in the analyzed rivers of Colombia, with respect to the Gev-MV, with Chi-square values ranging from 6.76 to 8.27. The maximum value of the Chi-square test was found using the Pea-MV with a value of 408.00, while the minimum value was reached with the Pea-MM with a value of 0.0.

4. Conclusions

Based on the current regional analysis of annual peaks flow data in Colombia, it was found that the Gev-MV hydrological probability distribution best fits the trend of the data in all analyzed hydrometric stations, since the best fit using the Chi-squared was obtained. In the current analysis the following probability distributions were evaluated: Gev-MM, Gum-MV, Gum-MM, Normal, Pea-MV and Pea-MM.

Future studies should analyze the influence of other mixed probability functions for Colombia, as well as the incidence of increasing or decreasing trends due to the occurrence of climate and anthropogenic changes in some regions. Also, more hydrometric stations should be collected in the regions of Colombia. It is recommended to analyze the effect caused by the El Niño and La Niña phenomena.

Appendix A. Codes of Used Hydrometric Stations

3510704	2306702	2406703	2113701
3508702	2306705	2406701	2309703
3508701	2306706	2104701	1301702
3506713	2306708	2105704	2502702
3506704	2308721	2105706	2906715
3506703	2309703	2111708	2803706
3506701	2319727	2119701	2804702

2702706	2312702	1506702	2502764
2623704	2401716	1506704	2906707
2621705	1307706	1506705	1306702
2620708	2801711	2119703	
2618711	2801708	2119715	
2501701	3502710	2502720	

References

[1] Chow, V. T., D.R. Maidment, and L. W. Mays. (1994) "Applied Hydrology". Bogotá: McGraw-Hill, Colombia.

[2] Clarke, R. T, R. Dias de Paiva, and C. Bertacchi. (2009) "Comparison of methods for analysis of extremes when records are fragmented:A case study using Amazon basin rainfall data". Journal of Hydrology 26-29.

[3] Arnaez, J., T. Lasanta, P. Ruiz-Flan, and L. Ortigosa. (2007) "Factors affecting runoff and erosion under simulated rainfall in Mediterranean vineyards." Soil & Tillage Research 324-334.

[4] Yang, T., Q. Shao, Z. C.. Hao, X. Chen, Z. Zhang, C-Y Xu, and L. Sun. (2009) "Regional frequency analysis and spatio-temporal pattern characterization of rainfall extremes in the Pearl River Basin, China." Journal of Hydrology, 386-405.

[5] Obeysekera, J., and J.D. Salas. (2014) "Quantifying the uncertainty of design floods under nonstationary conditions." Journal of Hydrological Engineering **19**, 1438–1446.

[6] Obeysekera, J., and J.D. Salas. (2016) "Frequency of recurrent extremes under nonstationarity." Journal of Hydrological Engineering **21**, 04016005.

[7] Grego, J.M., and P. A. Yates. (2010) "Point and standard error estimation for quantiles of mixed flood distribution." Journal of Hydrology **391**, 289–301.

[8] Koutrouvelisa, I.A., and G. C Canavos. (2000) "A comparison of moment-based methods of estimation for the log-Pearson type 3 distribution." Journal of Hydrology 71−81

[9] Xuewu, J., D. Jing, H.W. Shen, J.D. Salas, J.D. (2003) "Plotting positions for Pearson type-III distribution." Journal of Hydrology **74**, 1−29

[10] Makkonen, L. (2006) "Problems in the extreme value analysis." Structural safety, 405-419.

[11] Maidment, D. (1992) "Handbook of Hydrology." United State of America: Mc Graw – Hill.

[12] Coronado-Hernández, Ó.E., E. Merlano-Sabalza, E., Z. Díaz-Vergara, Z., and J.R. Coronado-Hernández, (2020). "Selection of Hydrological Probability Distributions for Extreme Rainfall Events in the Regions of Colombia." Water **12**, 1397..

[13] IDEAM. "Regionalization of Colombia according to the stationarily of the monthly mean precipitation trough an analysis of main components." Available online: http://www.ideam.gov.co/documents/21021/21789/Regionalizaci%25C3%25B3n%2bde%2bla%2blluvia%2ben%2bColombia.pdf/92287f96-840f-4408-8e76-98b668b83664 (accessed on 4 March 2021).

[14] González-Álvarez, Á., O.M. Viloria-Marimón, Ó.E. Coronado-Hernández, A.M. Vélez-Pereira, K. Tesfagiorgis, and J.R. Coronado-Hernández, J.R. (2019) "Isohyetal Maps of Daily Maximum Rainfall for Different Return Periods for the Colombian Caribbean Region." Water **11**, 358.

[15] Chaire en Hydrologie Statistique (CHS). (2002) "Hyfran. Logiciel pour l'analyse fre´quentielle en hydrologie." INRS-Eau, rapport technique.

[16] Gumbel, E.J. (1941) "The return period of flood flows." Ann. Math. Stat. **2**, 163–190.

[17] Aldrich, John. (1997) "R. A. Fisher and the Making of Maximum Likelihood 1912-1922." Statistical Science **12**, 3, 162–176.

[18] Hazelton, M.L. (2011) "Methods of Moments Estimation. In: Lovric M. International Encyclopedia of Statistical Science." Springer, Berlin, Heidelberg,

[19] Li, P., Yu. Z., Jian, P., and Wu, C. (2021). "Spatiotemporal characteristics of regional extreme precipitation in Yangtze River Basin". Journal of Hydrology. In Press.

[20] Muthuvel, D., and Mahesha, A. (2021). "Copula-Based Frequency and Coincidence Risk Analysis of Floods in Tropical-Seasonal Rivers". Journal of Hydrologic Engineering 25, 5.