Low Flow Discharges Regional Analysis using Wakeby Distribution in an ungauged basin in Colombia.

Ana C. Arbeláez[1],
Civil Engineer, MSc Water Resources Management

Lina M. Castro[2],
Civil Engineer, Master Student. Assistant Professor in Universidad del Valle – Colombia.

Abstract. In order to estimate low flow return periods for an ungaged watershed in Colombia, a regional analysis of low flow discharges was performed using the Wakeby probability distribution function. This distribution is more flexible than other distributions and separates the distribution's extreme tails; these are advantages for modeling flow discharges. A procedure for estimating the parameters of the distribution is also presented. The frequency analysis results of 18 to 36 long time series is shown. The goodness of fit test was evaluated with the Graphic Correlation Coefficient (GCC), the Standard Error of Fitting (SEF) and the Graphic Test; they were useful tools for testing the best fit with Wakeby distribution, which was better than LogNormal II and Gumbel distribution. The Index Flood Method for Regional Frequency Analysis was applied with Wakeby distribution and the discharge was standardized with the area. Discussions and conclusions are presented.

1 Introduction

Low flow estimations are underlying to: a) water quality management, b) reservoir storage design, water – supply planning and design, water for irrigation and c) aquatic ecosystems impact, wildlife conservation and recreation (Smakhtin 2001). The social and economic lost caused by droughts generally are greater than happened with floods (Kim et al., 2003).

Low flows are normally derived from groundwater discharge or surface discharge from lakes, marshes, and it is closely related with physiographic factors, climatic and geologic characteristics and other ones that affect infiltration and evaporation processes.

[1] Área Metropolitana del Valle de Aburra. This research was developed while the author was professor at Universidad del Cauca. Colombia
Medellín, Calle 75sur No 52-101. Medellín. Colombia
Tel: (574) 3093878
e-mail: acarbela@yahoo.com
[2] Research Group in Water Resources and Soil Development
Engineering Department
Universidad del Valle
Santiago de Cali, Calle 13 No. 100 – 00. Building 344. Office 2009. Colombia
Tel: (572) 3212153 ext 122
e-mail: linacahe@yahoo.com

There are diverse methodologies for low flow forecasting and its application depends of the data available. Low-flow frequency analysis is used to estimate values for different return periods, when the site of interest has gauging stations. Recession curve showed the depletion of streamflow discharge during dry weather periods, and it depends primarily on catchment's geology (e.g. transmissitive, storativity of the aquifers), and the distance from stream channels to basin boundaries (Chow 1988; Smakhtin, 2001). A flow duration curve is a relationship for any given discharge value and the percentage of time that this discharge is equaled or exceeded. With Rainfall – runoff modeling is possible to obtain flood quantiles driven by historical rainfall and evaporation records.

Ungauged catchments has a different problem. Some approaches for low – flow estimation in ungauged catchments may be classifies into: those which using techniques of hydrological regionalization, those which allow any low-flow characteristic to be estimated from synthetic flow time series and those which implies the application of simulation methods which aim at the generation of a continuous flow time series.

The purpose of this paper is exploring the application of Wakeby distribution in order to regionalize low flow discharges in the Cauca Region (Colombia). First we find the best distribution (LogNormal II, Gumbel and Wakeby) to model the records in the region. Later the index flood method was employed to regionalize the Wakeby distribution (Kuczera, 1982). The discharge to different return periods were estimated based in frequency analysis and standardized records.

2 Methods

2.1 Basin Information

Colombia is a country located in the northwestern corner of South America, it has two oceans and it is plenty of rivers that drain to the Atlantic sea, the Pacific Sea and the Amazon River. The Andes, which is the longest and highest mountain ranges in the word, extends from northern Colombia to southern Chile, through the whole South America.

Cauca State is located in the southwestern of Colombia, occupying an area of 31627 km^2. In this region the main river sources of the country are located like Magdalena River and Cauca River (Figure 1).

The whole state has five main watersheds: Patia, Pacific, Caqueta, Magdalena and Cauca basin with an area of 7394 km^2, which is the region of interest of the present paper. This region is located in the Andes mountain range, and has a mean annual rainfall of 2500 mm; the land use is, almost in the entire basin, agriculture. In some places, the water is needed for irrigation, and human use. For that reason a correct estimative of low flow discharge is necessary.

The hydrologic data used in this paper are concerned to instantaneous low flow discharge; the gauging stations are shown in the Figure 2. The fundamental basin morphometric information and the main statistical parameters of the hydrologic data are presented in the Table 1. The Enso phenomena, and specially the cold phase The Niño, has influence in the Colombian hydrologic cycle and exactly in the minimum discharges. The length of the data is not enough to separate the Niño event years to do the frequency analysis. Nevertheless, this influence was already included because the hydrological year was considered from June to May to the next year, when The Niño phenomena has the maximum activity.

Figure 1 Basin Location.

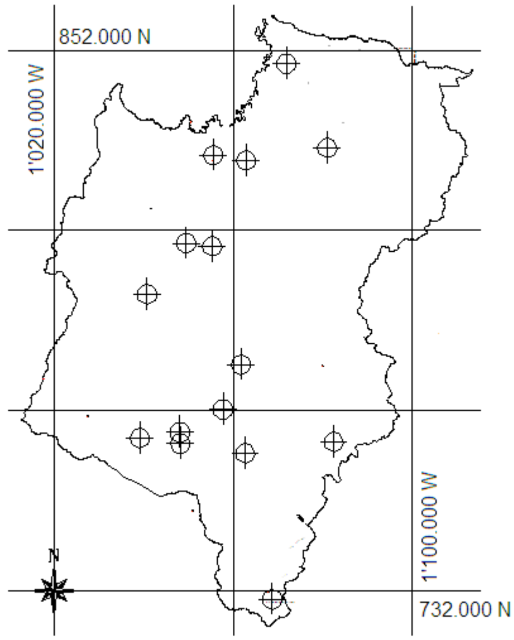

Figure 2. Gauging station location

Table 1. Morphometric basin characteristics and main gauging records statistics

River and Gauging Station	Coordinate		Area km^2	Gauging Height	Record years	Main statistics		
	N	E				μ	σ	γ
Cauca (Lomitas)	727,358.55	1,073,897.64	26.4	2,900	20	0.81	0.3	-0.278
Cauca (Pan de Azúcar)	791,852.81	1,042,358.09	3071.3	1,151	28	26.43	6.48	-0.018
Hondo (Puente Carretera)	765,804.15	1,039,183.75	216.6	1,740	20	1.89	0.6	0.977
La Teta (Lomitas)	828,718.35	1,055,315.66	128.4	1,022	26	0.76	0.34	0.886
Ovejas (Abajo)	808,444.87	1,055,324.75	527.3	1,263	36	5.5	1.52	0.257
Ovejas (Los Cámbulos)	806598.85	1,047,912.83	585.5	1,143	18	7.68	1.73	0.553
Palacé (Malvasá)	764,869.76	1,082,469.54	89.6	2,850	33	0.88	0.26	0.237
Palacé (Puente Carretera)	772,230.17	1,057,749.34	255.4	1,758	24	2.55	0.88	-0.899
Palo (Puerto Tejada)	849,001.30	1,071,979.03	1524.2	966	35	8.97	4.07	0.699
Piedras (Puente Carretera)	762,433.36	1,062,696.66	59.2	2,000	24	2.55	0.88	-0.899
Quinamayó (Puente Fe/rril)	827,524.18	1,063,036.57	165.5	1,000	29	0.59	0.49	1.451
Cr. Saté (Puente Carretera)	767,219.78	1,048,142.90	19.6	1,470	24	0.12	0.09	1.67

Where μ is the mean, σ is the standard deviation, and γ is the skewnes coefficient of the minimum discharges sample.

2.1.1 Minimum Discharge Frequency Analysis

The objective of a frequency analysis is to relate the magnitude of extreme events to their frequency using probability distributions, such as Gumbel, Log Normal II and Wakeby. It is assumed that the variables used in the analysis are independent and identically distributed (Chow, 1988).

2.1.2 Return Period (TR)

The minimum discharge return period is the mean time interval in which the discharge may be exceeded at least one time in this period; the exceedance probability $P(X \geq x_{T_R})$ is related to the return period in this way (Kim, T., Valdes, J. y Yoo, Ch., 2003):

$$P(X \geq x_{T_R}) = \frac{1}{T_R} \tag{1}$$

2.1.3 Wakeby Distribution

This is a more flexible distribution than other common distribution used in hydrology, because is a five parameter distribution. Probability Weight Moment Method (PWM) is used to obtain the Wakeby parameter distribution. It has a particular advantage because the original data are not with high exponent, like they are in the traditional Moment Method. This distribution is defined in an inverse way as follows:

$$Q_{\min} = -a*(1-F)^{b} + c*(1-F)^{-d} + e \tag{2}$$

Where Q_{min} is the minimum discharge associated with a specific return period, F is the exceedance probability, *a*, *b*, *c*, *d*, and *e* are the parameter distribution. (Houghton, 1978).

2.1.4 Plotting Position

It represents the empirical cumulative probability distribution of the sample, and is represented by the general equation:

$$F_{i,emp} = \frac{i-a}{n+1-2a} \tag{3}$$

In this paper the empirical probabilities given by Landwher, Blom, Weibull, Cunnane and Gringorten are used, they are referred in the Tabla 2:

Table 2. Empirical probabilities

Author	**Expression**	**Author**	**Expression**	**Author**	**Expression**
Weibull	$\frac{i}{n+1}$	Gringorten	$\frac{i-0.44}{n+0.12}$	Cunnane	$\frac{i-0.4}{n+0.2}$
Blom	$\frac{i-3/8}{n+1/4}$	Landwehr	$\frac{i-0.35}{n}$		

2.1.5 Goodness of Fit Test

To determine which is the best distribution that represents the original data, the typical Goodness of Fit Tests Kolmogorov – Smirnov and χ^2 (Chow, 1988) are used, but the cumulative theoretical probability distribution must be known. It is not possible with the Wakeby distribution, other Goodness of Fit Tests were implemented: Graphic Correlation Coefficient (GCC), Standard Error of Fitting (SEF) and Graphic Test.

- Graphic Correlation Coefficient (GCC)

$$r = \frac{\sum (Q_i - \widehat{\mu})(Q_{calculated\ i} - \overline{Q_{calculated}})}{\left[\sum (Q_i - \widehat{\mu})^2 \sum (Q_{calculated\ i} - \overline{Q_{calculated}})^2\right]^{0.5}} \tag{4}$$

If this value is near 1 means the theoretical distribution represents correctly the sample data.

- Standard Error of Fitting (SEF)

$$SEF = \left[\frac{\sum_{i=1}^{n} (Q_i - Q_{calculated\ i})^2}{(n - np)}\right]^{1/2} \tag{5}$$

Where n is the record length, and np is the number of parameters of the distribution.

- Graphic Test is the representation graphic of the sample and the theoretical distribution. When both of them are closer each other it is considered a good fit.

3 Regional Analysis

This technique is used in an ungauged basin; a correct use requires an homogeneous region, with similar climatologic characteristics, and of course, gauging basin closer to the study area.

The most typical way is correlating the mean μ and the standard deviations σ with the main morphometric characteristics, and applying the frequency factor presented by Chow (1988).

$$Q = \widehat{\mu} + K_T \widehat{\sigma} \tag{6}$$

Where K_T is the frequency factor of the probability distribution that has the best fit in the study region. The most common distribution used in hydrologic studies are Gumbel and LogNormal

This technique has the disadvantage that it is required to know the frequency factor of the probability distribution. In the case of Wakeby, it is not known. For that reason, another technique should be used, like Index Flood.

The Index Flood assumes the discharge probability distribution is the same in a region and all the records could be used in a single long series, but it is necessary the used of a scale parameter that reflects the effect of different areas, rainfall, land cover and land use. In this method, the discharges are standardized with a morphometric or a hydrologic parameter like drainage area or bankfull discharge and all the standardized records are used in a single long series to do again the frequency analysis (Maidment, 1993). In the present paper, the parameter used to the standardized process was the drainage area (A), because there is a strong correlation between the mean discharge and the area, as shows the following expression.

$$\mu_{\min} = \lambda A^{\theta} \qquad (7)$$

In the case of minimum discharges θ is close to 1 (Poveda et al., 2002). Similar behaviour was obtained with Cauca basin regional data; this result is shown in Figure 3. Some specific stations do not reflect the same behaviour like Cauca (Lomitas) and Piedras (Puente Carretera), with the highest rates of (Q/A) and Quinamayó with the lowest rate, those stations were excluded for the next analysis.

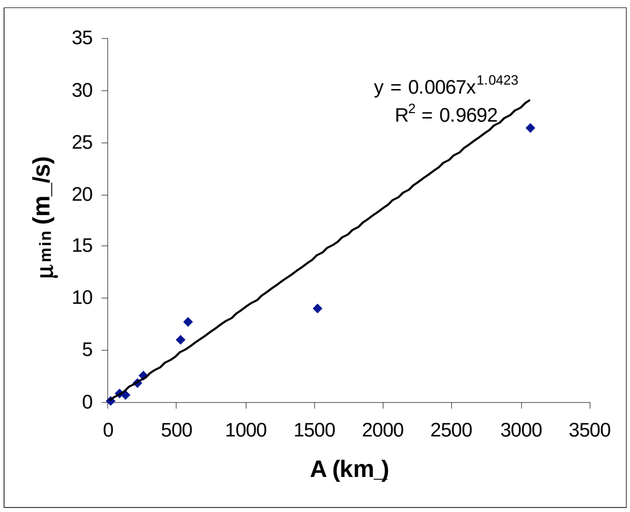

Figure 3 Relation between mean minimum discharges and drainage area.

4 Results

The frequency analysis process was done with the Software HydroStat v. 1.0 that runs in Java (Castro and Hoyos, 2004). It has an interface to introduce the data, and then the statistical parameters, the probability graphic and the Goodness of Fit Tests are calculated. Finally, it estimates different discharges associated with specific return periods. The Software allows to chose between Wakeby, Gumbel or LogNormal distribution.

4.1 Frequency Analysis

Goodness of Fit Tests Graphic Correlation Coefficient (GCC), Standard Error of Fitting (SEF) and Graphic Test were used for testing the best fit with Wakeby distribution which has a behaviour better than LogNormal II and Gumbel distribution. SEF was lower than the mean discharge and GCC, in all the stations, was closer to 1, which means that the theoretical distribution has a good fit.

Table 3 Goodness of Fit Tests for the best fit distribution and Q_{min} (T_R) m³/s

Gauging Station	**Distribution**	**GCC**	**SEF**	**T_R year**				
				5	**10**	**20**	**50**	**100**
Cauca (Pan de Azúcar)	Wakeby, Landwehr	0.9892	1.28	21.0	16.9	14.1	12.1	11.3
Hondo (Puente Carretera)	LnNormal, Blom	0.9709	0.15	1.41	1.23	1.11	0.98	0.9
La Teta (Lomitas)	Wakeby, Landwehr	0.984	0.07	0.44	0.35	0.31	0.28	0.27
Ovejas (Abajo)	Wakeby, Gringorten	0.9947	0.17	4.25	3.45	2.9	2.51	2.37
Ovejas (Los Cámbulos)	Gumbel ML, Weibull	0.9836	0.33	6.13	5.58	5.18	4.77	4.52
Palacé (Malvasá)	Wakeby, Landwehr	0.992	0.04	0.62	0.52	0.46	0.42	0.41
Palacé (Puente Carretera)	Wakeby, Landwehr	0.9804	0.19	1.90	1.28	0.72	0.27	0.09
Palo (Puerto Tejada)	LnNormal, Weibull	0.9814	0.83	5.54	4.53	3.84	3.18	2.81
Creek. Saté (Puente Carretera)	Wakeby, Landwehr	0.9965	0.01	0.05	0.03	0.01	0.01	0.00

Discharges calculated for 100-year return period were always positives (or cero), which is consistent with physical phenomena. The best fit was, in most cases, with Wakeby distribution (Table 3); for that reason, this distribution was used for the regional analysis. Stations Hondo, Palo (Puerto Tejada) and Ovejas (Los Cámbulos) were not excluded to the analysis, even the Wakeby distribution did not fit with this data. Figure 4 shows the Graphic Test for some stations, they have a good fit because the theoretical distribution is near the record data.

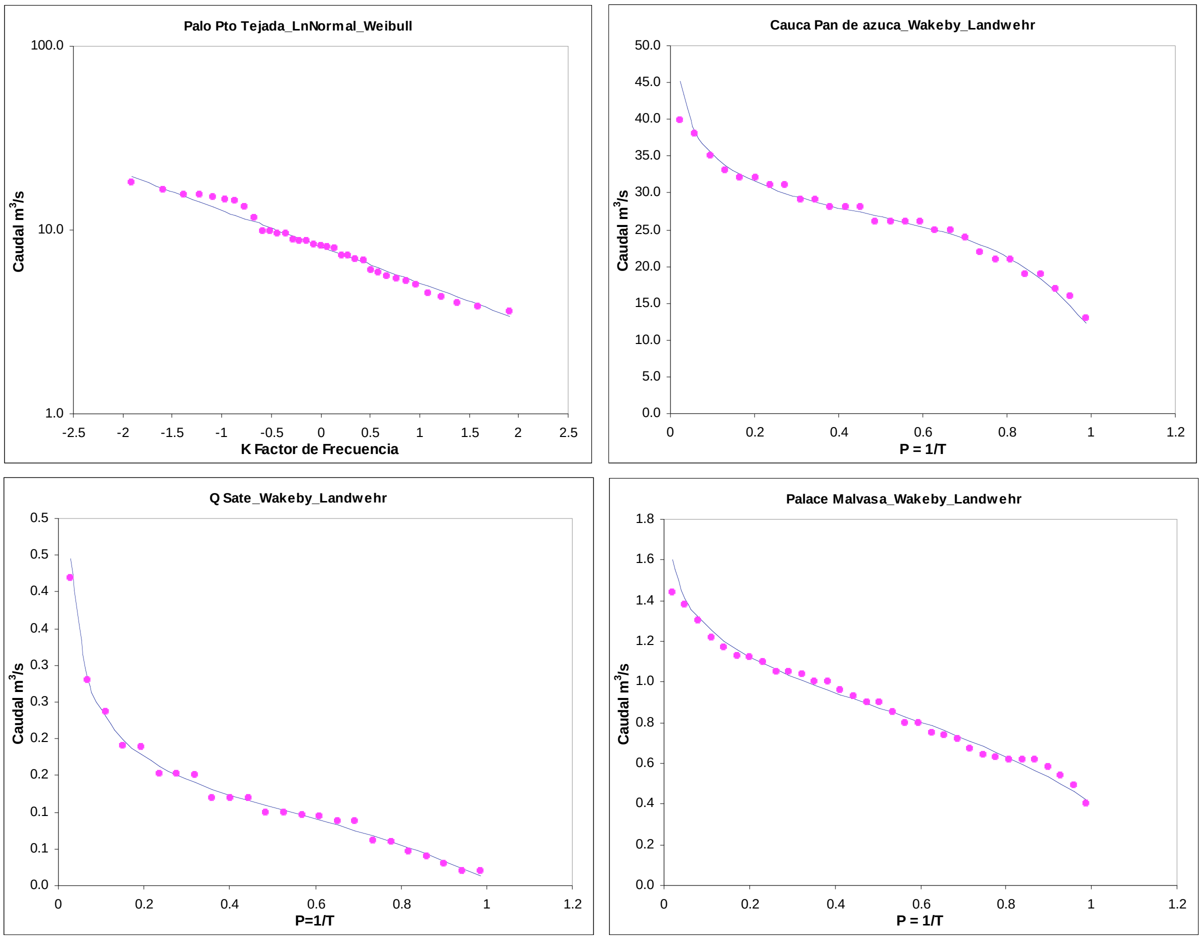

Figure 4 Best Graphic Fit for some stations. P is the exceedance probability, and T is the return period.

4.2 Regional Analysis

To use the Index Flood Method, a common period to the records was used. The resultant expression, using Wakeby distribution, Figure 5, is:

$$\frac{Q_{\min}}{A} = 1.102 + 3.337\left[1-\left(1-\frac{1}{T_R}\right)^{8.73}\right] - (-13.97)*\left[1-\left(1-\frac{1}{T_R}\right)^{-(-0.525)}\right] \qquad (8)$$

Where Q_{min} is in l/s and A is the drainage area in km². Goodness of Fit Tests were accepted with values for Graphic Correlation Coefficient GCC of 0.99 and Standard Error Fit (SEF) of 0.55.

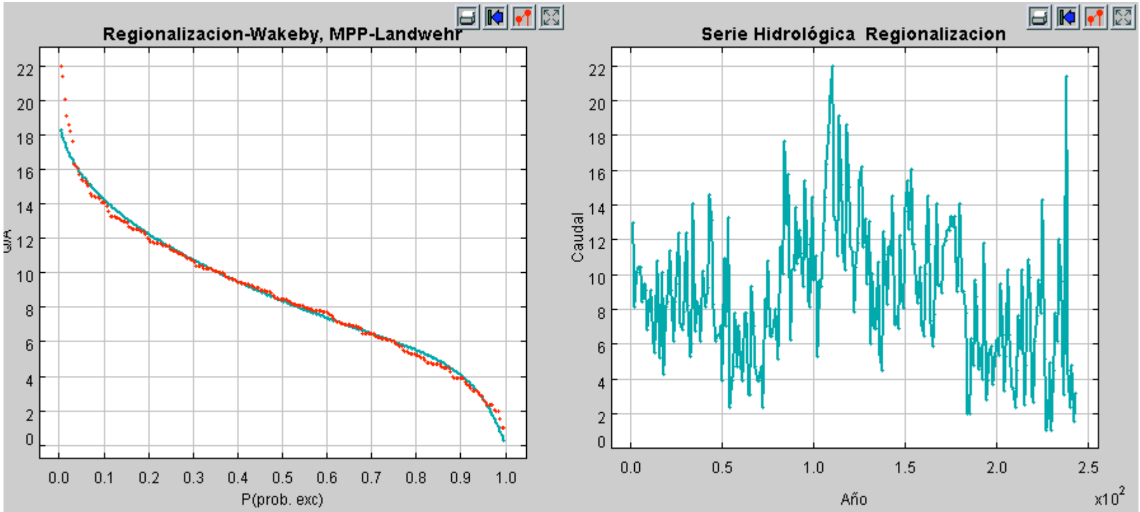

Figure 5 Standardized series (Q_{min}/A) Graph and Graphic Test

A comparison between results from regional analysis and frequency analysis shows a good general trend, with the exception of the Saté Creek that has the highest errors, it could be because the area basin is small, and this regional analysis might not be used in small basins.

5 Conclusions

Goodness of Fit Tests: Graphic Correlation Coefficient (GCC), Standard Error of Fitting (SEF) and Graphic Test were useful tools for testing the best fit distribution. In the study area, Cauca watershed of Cauca State, the majority of the stations were better modelled with Wakeby than with LogNormal II and Gumbel; the best fit was checked with GCC, SEF and Graphic Test.

The mean minimum discharge is lineally correlated with the drainage area. This was the parameter to standardize the originals gauging records. The long regional series was correctly modelled with Wakeby distribution.

Regional analysis is a useful tool for predicting the low discharge in ungauged basins. In the present paper, the results were good in most cases, except for the Saté Creek in which the regional analysis overestimates the minimum discharge. This is a small basin and this regional analysis must not be used to this kind of basins.

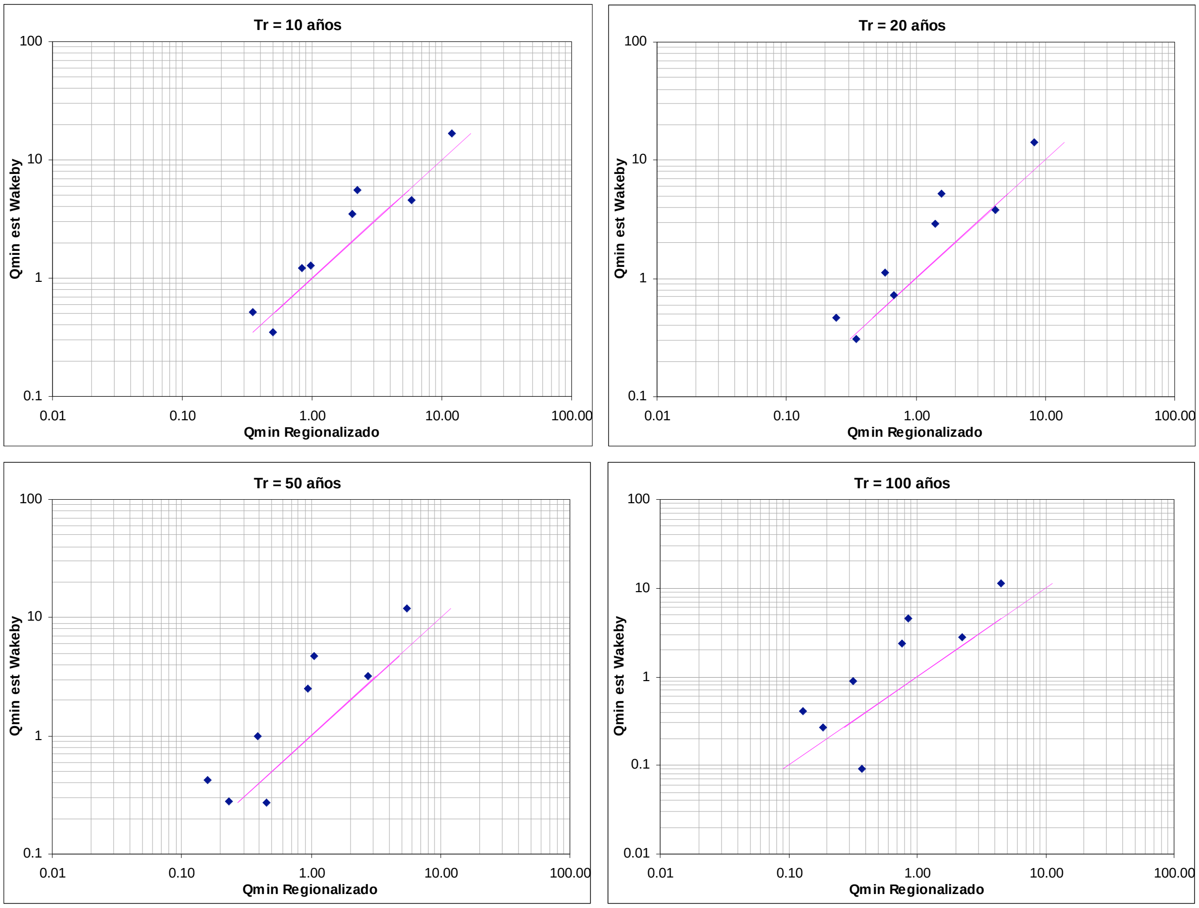

Figure 6 Comparison between the discharge from regional analysis and frequency analysis, for different return periods.

Acknowledgements. This research was supported by the Universidad del Cauca.

6 References

Castro L, y Hoyos, C., 2004: Análisis de Frecuencia de crecientes con diferentes distribuciones de probabilidad. Final Project, Civil Engineer, Universidad del Cauca (Colombia).

Chow, V., 1988: Hidrología Aplicada. McGraw Hill, Inc. New York. Capítulo 11.

Houghton, J. 1978. Birth of a parent: The Wakeby Distribution for modelling flood flows. Water Resources Research, 14(6). 1105 – 1108

Kim, T., Valdes, J., Yoo, CH. 2003: Nonparametric Approach for Estimating Return Periods of Droughts in Arid Regions. Journal of Hydrologic Engineering Sep/Oct.

Landwerh, Matalas., 1979: Estimation of Parameters and Quantiles of Wakeby Distributions. Water Resources Research, 15(6). 1361 – 1379.

Maidment, R., 1993: Handbook of Hydrology. Mc GrawHill.

Poveda G., Velez J., Mesa O., Ceballos L., Zuluaga D., Hoyos C., 2002: Estimación de Caudales Mínimos para Colombia Mediante Regionalización y Aplicación de la Curva de Recesion de Caudales. Meteorología Colombiana. 6, 73-80.