FLOOD FREQUENCY FOR UNGAGED CATCHMENTS IN PUERTO RICO

Rafael Segarra-García, Ph.D.
Department of Civil Engineering
University of Puerto Rico

Project No. 4
Grant Agreement No. 14-08-0001-G2043

**FINAL TECHNICAL REPORT
TO
U.S. DEPARTMENT OF THE INTERIOR
GEOLOGICAL SURVEY**

The work on which this report is based were financed in part by the Department of the Interior, U.S. Geological Survey, through the Puerto Rico Water Resources Research Institute.

August 1995

ABSTRACT

The project estimates flood flows for ungaged catchments in Puerto Rico. Flood frequency curves are obtained by using linear regression to estimate mean annual flows for ungaged basins. These are used to estimate flood magnitudes from normalized flood frequency curves derived from the application of discriminant analysis procedures to ungaged basins. Discriminant analysis allows the estimation of ungaged basin parameters from clusters of gaged basins. With these clusters, quantiles are computed for different recurrence intervals and flood frequency curves are derived from this procedure. This method yields smaller standard errors than those obtained utilizing regression techniques in previous efforts in Puerto Rico .

RESUMEN

El proyecto estima las descargas de inundación para cuencas sin aforar en Puerto Rico. Las curvas de frecuencia de inundaciones se obtienen al utilizar regresión lineal para estimar flujos anuales promedios en cuencas sin aforar. Estas descargas se utilizan para estimar magnitudes de inundaciones de curvas de frecuencia de inundaciones derivadas de la aplicación de análisis discriminatorio a cuencas sin aforar. El análisis discriminatorio permite la estimación de parámetros de cuencas sin aforar de conglomerados de cuencas aforadas. Con estos conglomerados, los cuantiles son calculados para los diferentes intérvalos de recurrencia y las curvas de frecuencia de inundaciones se derivaron de este procedimiento. Este método produce menores errores que otros esfuerzos previos que utilizaban técnicas de regresión en Puerto Rico.

TABLE OF CONTENTS

Table of Contents (Cont.)

LIST OF TABLES

List of Tables (Cont.)

LIST OF FIGURES

List of Figures (Cont.)

List of Figures (Cont.)

CHAPTER I

INTRODUCTION

Estimation of flood frequency is one of the most important aspects of catchment studies. It is required for all infrastructure development in areas subject to flooding, for insurance policy formulation, and environmental studies. Due to the high flooding potential of large parts of the developed areas in Puerto Rico, it is necessary to conduct hydrologic studies for proposed development projects within areas susceptible to flooding. These studies require accurate estimates of peak flows to assess the hydraulic effect of projects on flood levels.

Estimation of peak flows related to the regulation frequency, such as the 100-year flow for bridges, is highly uncertain because the available peak flow record in Puerto Rico is short, around 30 years in the best of cases. In many catchments, the record is even shorter, or nonexistent. Traditional approaches to flood frequency estimation, such as the customary approach of the well-known Soil Conservation Service's Curve Number hydrologic procedure, and the US Geological Survey's (USGS) regression approach from López et al. (1979), introduce large errors in peak flood estimation. The Curve Number procedure assumes that the T-year rainfall produces the T-year peak flow in the resulting runoff hydrograph. This is not generally true, since the actual soil moisture dynamics in a basin are never accounted for in a realistic fashion with event-based rainfall/runoff models. The USGS regression approach yielded estimation errors between -38 to +61%, which may be unacceptably large for some studies. A methodology is presented that incorporates recent flood data with modern regionalization techniques to obtain reliable estimates of peak flows for catchments with little or no recorded flood history. The results should be of great value to anyone involved in hazard mitigation, infrastructure development, and water resources planning in flood-prone areas.

This research uses linear regression analysis to estimate mean annual flows for ungaged basins. These are used to estimate flood magnitudes from normalized flood frequency curves derived from the application of discriminant analysis procedures to ungaged basins. Discriminant analysis allows the estimation of ungaged basin parameters in terms of regionalized parameters from clusters of gaged basins. Through the use of these groups or clusters, quantiles are computed for different recurrence intervals and flood frequency curves are derived from this procedure. This research provides a method for estimating peak flows in catchments with little or no recorded flood history. It also helps in the estimation of peak flows in gaged areas, at locations of interest, such as a bridge site, when this position does not correspond to the location of the flow gaging station.

The project will benefit those involved in water resources planning and infrastructure development and ultimately the Puertorrican society in general. Technically, it will use modern hydrologic regionalization procedures to make a more efficient and meaningful use of the available flood flow information in Puerto Rico for the assessment of flood flow levels.

The objectives of this study were:

1) To improve the procedures used to estimate peak flows of a given frequency for use in hydrologic studies of catchments away from the streamflow gaging sites or ungaged tributaries in major river networks.

2) To establish the worth of applying modern regionalization techniques to develop improved procedures for incorporating available flood and catchment attributes in flood frequency estimation.

CHAPTER II

LITERATURE REVIEW

The only relevant major scientific studies on regional flood flow frequency in Puerto Rico are the US Geological Survey (USGS) flood study report by López et al. (1979), and the flood frequency study of Segarra (1991).

In the USGS report, regionalized estimates of the T-year flood are estimated through regression analysis of available flood data. The regression obtained was of the type

$$Q_t = KA^{x}(\text{Ann } P)^{y} \tag{2-1}$$

where Q_t is the T-year flood, K, x and y are regression parameters, A is the catchment area, and Ann P is the average annual precipitation. The standard errors of prediction of these equations range from -38 to +61 percent.

The use of Equation (2-1) has almost become standardized for flood frequency estimation on the island. In this case, a major limitation of the regression approach is the large standard errors obtained, due in part to the short records available. Also, regressions forcefully correlate available data with parameters that may possess considerable internal estimation uncertainty that manifests itself in noisy regional estimates and low correlation.

The study by Segarra (1991) overcame many shortcomings related to the use of regression techniques for regionalizing flood data. The study used the Generalized Extreme Value Distribution, Probability-Weighted Moments (GEV/PWM) technique to regionalize flood frequency information in the development of generalized flood frequency curves. The effort was successful, as the standard errors of estimate were reduced considerably when compared with those obtained from the study of López et al. For 100-yr flood estimation, the GEV/PWM technique yielded a 28.7% standard error as the largest error obtained, compared to 61% from the Geological Survey study.

When a GEV is used as a regional model, the consistency of the flood data for the site modeled can be tested using several goodness-of-fit tests. Chowdhury et al. (1990) used the Kolmogorov-Smirnoff test, the probability plot correlation test, and the sample L moment ratio test, and demonstrated their usefulness.

Segarra (1991) recommended a procedure for estimating peak flows for ungaged basins. For verification purposes, the procedure was tested with basins for which limited data was available. The results exhibited large estimation errors when using mean annual flows obtained from Figure 2-1. The linear regression equation was

$$Q_{mean} = 1495 \ (Area)^{.592} \tag{2-2}$$

where the Area is in mi^2 and Q_{mean} in cfs. The regression analysis used in this study showed a low coefficient of determination and presented problems when predicting the mean annual flows of the ungaged basins. When the actual catchment flow means were used, much lower standard errors were obtained. This pointed to the necessity of correlating mean flood flows with other meaningful catchment parameters. This is one of the motivations for this study.

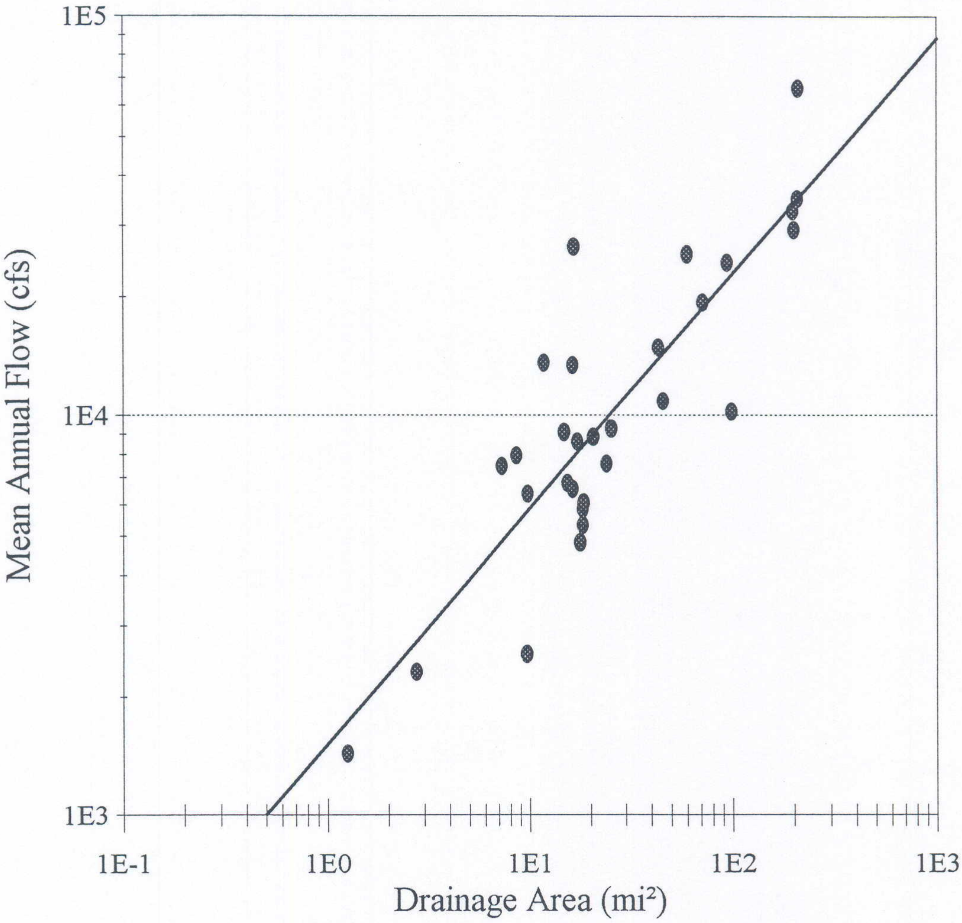

Figure 2-1. Relationship between mean annual flow and drainage area for various catchments in P.R. (Adapted from Segarra, 1991)

Attempts at regionalization are not new (Solomon, 1976; Benson and Matalas, 1967), but past efforts have suffered from limitations when applied to short data bases. The GEV/PWM method [Greenwood et al. (1979), Landwehr et al. (1979), Hosking et al. (1985), and Segarra (1991)] is an excellent tool for estimating flood frequency in areas with short data bases. Another approach using a Pearson type III distribution (Ribeiro-Correa and Rouselle, 1993) also yielded reliable results for catchments with short data bases. The GEV/PWM procedure produces consistently less variable estimates than other procedures commonly utilized. The flood frequency estimates obtained through this procedure can be incorporated into a regionalization scheme to obtain flood frequency estimates for ungaged catchments.

The GEV/PWM algorithm has also been coupled to a multivariate catchment classification scheme (Segarra, 1991) using discriminant analysis. The use of discriminant analysis in the water resources field is relatively recent; one of the most representative applications being the classification of watersheds in Great Britain (Wiltshire, 1986a and 1986b). Discriminant analysis is related to the problem of identification, a branch of decision theory. It has found convenient applications in the medical, anthropological, and biological fields. Basically, it deals with the problem of deciding between a number of alternative hypotheses. For example, an individual may be assigned to one of a number of groups in which he may belong based on a set of observed characteristics. The problem consists of identifying the particular group to which the individual belongs with the highest probability.

As mentioned earlier, the GEV/PWM algorithm has also been coupled to a multivariate catchment classification scheme. This procedure can produce regionalized flood flow frequency curves for homogeneous hydrologic regions. These in turn can be used to estimate quantities for ungaged basins within the similar regions. The test for homogeneity is based on a hypothesis test on a statistic of homogeneity, and a parameter set of catchment characteristic values. Wiltshire (1985) classified the groups on the basis of basin area, average rainfall and urban fraction. For each homogeneous region an improved flood frequency distribution curve was obtained.

Panu and Smith (1988) used regional frequency analysis for estimation of flood flows at ungaged watersheds on the island of Newfoundland in Canada. The island was divided into two homogeneous regions. It was delineated into a North region and a South region. Once watersheds were assigned into one of the two homogeneous regions, the significant watershed characteristics were identified and abstracted for the development of regional relationships between the estimated flood flows of various return periods and the watershed characteristics for the estimation of flood flows at ungaged sites. The important characteristics were watershed area (DA), watershed shape factor (SHAPE), percentage of watershed area controlled by lakes and swamps (ACLS), percentage of barren area (BAREA), mean annual runoff (MAR) for the watershed, and latitude (LAT) of the watershed centroid. MAR was the second most important characteristic after DA in all regional frequency relationships. The general form of the regression equation used in the analysis was

$$Q_T = K + \sum_{i=1}^{n} a_i * p_i \tag{2-3}$$

where Q_T is the estimated annual maximum instantaneous flow (m^3/s) with a T-year return period, K is a regression constant, a_i is the i^{th} regression coefficient, p_i is the i^{th} regression parameter, and n is the number of parameters. The regression equations for all return periods included the most significant watershed characteristics, namely, DA, MAR, ACLS, and SHAPE. An example of a final regression equation is

$$\text{Log } Q_T = K + a \text{ Log DA} + b \text{ Log MAR} + c \text{ Log ACLS} + d \text{ Log SHAPE} \tag{2-4}$$

where, Q_T, K, and a are the same as defined earlier and b, c, and d are regression coefficients. Also, DA (km^2), MAR (mm), ACLS (%), SHAPE (dimensionless), and LAT (degrees) are the watershed characteristics. The regional flood frequency equations were adequate for reliable flood flow estimates in ungaged watersheds on the island of Newfoundland. The climatic environment of this island is a function of several interrelated influences, such as general atmospheric circulation at mid-latitude in the Northern Hemisphere, the location of the island in relation to the North-American mainland, and the presence of a cold oceanic surface caused by the Labrador current around the island. In the central and southern watersheds, flood flows can occur due to rain on melting snow. These conditions are not encountered in Puerto Rico; therefore their parameters cannot be used to classify local streams.

Baldwin and Potter (1987) used a different approach to estimate flood quantiles at ungaged sites. Their study used data from the Kickapoo and Pecatonica rivers located in Wisconsin. They did not use regional regression equations because it was found that sometimes the predicted quantiles had high standard errors. Often, these equations do not account for important physical factors. The study focused on the use of time-area histograms. It was found that the histograms gave reasonable results, but it was noted that further study is necessary to reach definite conclusions about its potential.

In a study made in Greece (Mimikou, 1987), it was observed that when peak discharges were plotted against drainage areas, the regression points were scattered. It was later found that a single envelope curve for the area under study was obtained when the drainage area was replaced by a morphoclimatic index, which is the product of the expected storm duration, the maximum observed average storm intensity for this duration, and the area of the drainage basin. The developed envelope curve can predict peak discharges for ungaged basins in the area under study. This morphoclimatic index procedure could be applied in Puerto Rico once the climatological data analysis necessary for the definition of the index is carried out.

Sherwood (1994) developed multiple-regression equations to estimate maximum flood volumes of d-hour duration and T-year recurrence interval for ungaged streams. The significant explanatory variables in the resulting volume-duration-frequency equations were drainage area, average annual precipitation, and basin-development factor (BDF). The BDF is a measure of channel and basin development that accounts for channel improvements, impervious channel linings, storm sewers, and curb-and-gutter streets. The step-forward and step-backward regression techniques were used to decide which of the explanatory variables should be included in the regression equations. The volume-duration-frequency data sets can be identified by abbreviations in the form dV_T, where V is total volume in millions of cubic feet, d is duration in hours, and T is recurrence interval in years. In Ohio, (Sherwood, 1994) for a one hour duration and 100 year recurrence interval the volume can be calculated as

$$1\,V_{100} = 1.28\,(A)^{.77}(P - 30)^{.51}(13 - BDF)^{-.36} \tag{2-5}$$

where A is the drainage area in mi², BDF is the basin development factor (on a scale of 0 to 12), and P is the average annual precipitation in inches. The volume-duration-frequency equations for the desired recurrence interval can be used to develop a relation between inflow volume and duration for an ungaged site. In Puerto Rico, this procedure cannot be applied directly because the multiple regression equations developed are applicable only to small urban streams in Ohio, whose basin characteristics are similar to the basin characteristics of the sites used in the regression analysis.

In recent studies (Gingras et al., 1994), nonparametric frequency analyses were shown to improve the regional estimates. The method was used in Ontario and Quebec. It revealed unimodal and multimodal annual maximum flood probability density shapes in the area of study. Improvements in regional estimates are possible in terms of single station flood frequency analysis, in terms of homogeneous region delineation, and in terms of regional relationship development. Linear regression was employed in order to assess the need to further investigate nonparametric regression, and to check whether the division of the entire data set into smaller regions led to improved regional relationships. The model used was

$$\text{Log } Q_T = a + b \text{ Log DA} \qquad (2\text{-}6)$$

where, Q_T is the flood of return period T as estimated from the sample using nonparametric frequency, DA is the drainage area, and a and b are regression coefficients. The results showed that lower standard errors were obtained when the nonparametric analyses were employed. This study used 183 natural flow stations from Ontario and Quebec with a record length of at least 20 years. In Puerto Rico there are only 30 flow stations with reasonably acceptable record length. This type of regression could be used in Puerto Rico, but the available flow data on the island is not comparable with that of this study. A larger data base would greatly affect the regression results based solely on the basin area. Hence, it is difficult to apply the same procedure on the island and obtain excellent results.

All these methods used different approaches to determine flood flow levels in ungaged catchments. Most used regional frequency analysis and obtained reliable results. However, it is questionable whether the parameter sets employed in previous studies can be used for regional analysis in Puerto Rico. The regional analysis with locally defined parameter sets should yield more reliable results than those obtained from traditional regression or synthetic generation methods.

CHAPTER III

LINEAR REGRESSION ANALYSIS FOR MEAN FLOWS

A better regression model is needed to predict the mean annual flow in ungaged catchments. Segarra (1991) noted that a better estimate of mean flows could significantly reduce standard errors in flood frequency estimation.

Regression analysis is one of the most widely used statistical techniques for analyzing multifactor data. To develop a model that could give the mean flow for most catchments in Puerto Rico a multiple linear regression analysis was used. The term 'linear' is used because the unknown parameters are in a linear form. In general, the response y may be treated to k regressor variables. The model

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \epsilon \qquad (3\text{-}1)$$

is called a multiple linear regression model with k regressors (Montgomery and Peck, 1992). The parameters β_j, j=0,1, ..., k are called the regression coefficients. This model describes a hyperplane in the k-dimensional space of the regressor variables X_j. The parameters β_j represent the expected change in the response y per unit change X_j when all remaining regressor variables X_i ($i \neq j$) are held constant (Montgomery and Peck, 1992). To estimate the regression coefficients the method of least squares was utilized.

One of the most important aspects in developing a regression model is the selection and number of parameters to be used in the equation. A number of characteristics were considered to be included in the regression analysis. After reviewing the available data it was decided that 11 characteristics were going to be measured for all catchments. The important characteristics were watershed area (AREA in mi²), slope stream (SL in %), mean annual evapotranspiration (ET in inches), watershed shape factor (SH in length/width), percentage of basin area covered by lakes (LAKE in %), stream frequency (SF in number of segments/ area), mean annual precipitation (ANP in inches), mean monthly precipitation in the month of September (MOP in inches), centroid elevation of the basin (CE in meters), the 5 year return period 24 hour rainfall (X5 in inches), and the 25 year return period 24 hour rainfall (X25 in inches). The gaged rivers and their characteristics are listed in Table 3-1. Figure 3-1 illustrates these streams. It must be noted that each topographic characteristic was measured using a 1:120,000 map. This is an important factor because, for example, the value of the stream frequency would be different if the scale of the map changes.

The first model considered all the characteristics obtained for each basin and was of the form

$$Q_{mean} = \beta_0 + \beta_1 AREA + \beta_2 SL + \beta_3 ET + \beta_4 SH + \beta_5 LAKE + \beta_6 SF + \beta_7 ANP + \beta_8 MOP + \beta_9 CE + \beta_{10} X5 + \beta_{11} X25 \tag{3-2}$$

where, Q_{mean} is the mean annual flow in ft³/sec. The model included all the data but was not practical. This equation had a low correlation coefficient and yielded large residuals. Through the use of a backward elimination procedure the model was reduced to eight

Table 3-1. Gaged Basin Characteristics

Basin	Qmean (cfs)	AREA (mi^2)	SL (%)	ET (in)
G. de Manatí	32542	197.00	12.458	55
Cibuco	10198	99.10	20.833	55
De la Plata	34882	208.00	9.968	50
Bayamón	19199	71.90	14.388	55
G. de Loíza	66054	209.00	8.772	55
Herrera	2301	2.75	38.580	50
Grande	7437	7.31	85.979	45
Mameyes	13583	11.80	40.541	55
Fajardo	9062	14.90	49.020	50
Icacos	1426	1.26	5.787	45
Humacao	8605	17.30	13.123	65
Coamo	10820	46.00	31.172	55
Jacaguas	14782	43.50	35.328	55
Inabón	2545	9.70	54.945	55
Bucaná	9287	25.60	49.500	55
Portugues	6041	18.60	40.527	55
Tallaboa	7548	24.20	47.729	55
Guayanilla	8844	20.80	62.500	55
Rosario	6529	16.40	37.634	50
G. de Añasco	24119	94.30	20.375	45
Tanamá	5835	18.40	44.910	50
Valenciano	13355	16.40	22.059	45
Gurabo	25359	60.20	6.061	45
Canóvanas	6351	9.84	51.282	45
Espiritu Santo	7917	8.62	59.524	40
Cerrillos	4810	17.80	111.434	55
Yahuecas	6769	15.40	61.728	45
G. de Patillas	5319	18.30	50.891	60

Table 3-1 (Cont.). Gaged Basin Characteristics

Basin	SH	LAKE (%)	SF (seg. / area)	ANP (in)
G. de Manatí	1.20	1.0	0.467	70
Cibuco	1.75	0.0	0.333	65
De la Plata	3.57	3.0	0.760	68
Bayamón	2.13	0.5	0.654	70
G. de Loíza	1.21	3.0	0.474	89
Herrera	2.75	0.0	0.364	85
Grande	4.50	0.0	0.547	93
Mameyes	3.00	0.0	0.932	80
Fajardo	4.00	0.0	1.007	78
Icacos	2.00	0.0	0.794	120
Humacao	1.75	0.0	0.694	85
Coamo	2.75	0.0	0.761	35
Jacaguas	3.00	1.0	0.529	40
Inabón	4.20	0.0	0.412	37
Bucaná	2.40	0.0	0.586	35
Portugues	5.33	0.0	0.269	35
Tallaboa	4.00	0.0	0.620	45
Guayanilla	2.89	0.0	0.865	40
Rosario	4.50	0.0	0.793	75
G. de Añasco	3.80	1.0	0.594	100
Tanamá	3.00	0.0	0.326	90
Valenciano	2.20	0.0	0.546	70
Gurabo	2.25	0.0	0.664	75
Canóvanas	4.00	0.0	0.915	100
Espiritu Santo	1.56	0.0	0.928	95
Cerrillos	2.33	0.0	0.562	40
Yahuecas	2.67	0.0	0.195	85
G. de Patillas	2.14	0.0	0.546	80

Table 3-1 (Cont.). Gaged Basin Characteristics

Basin	MOP (in)	CE (m)	X5 (in)	X25 (in)
G. de Manatí	7.0	250	6.3	9.0
Cibuco	7.0	100	6.1	9.0
De la Plata	8.0	250	6.0	8.5
Bayamón	7.0	50	6.5	8.6
G. de Loíza	12.0	50	7.5	9.9
Herrera	8.0	50	7.7	10.5
Grande	10.0	200	8.0	11.0
Mameyes	7.0	100	8.0	11.0
Fajardo	10.0	25	8.0	10.9
Icacos	14.0	650	8.8	11.8
Humacao	12.5	100	8.2	12.0
Coamo	10.0	51	6.4	9.5
Jacaguas	10.5	100	6.1	9.0
Inabón	10.0	50	6.3	9.2
Bucaná	11.0	250	6.0	8.9
Portugues	11.0	200	6.2	9.1
Tallaboa	12.0	250	7.5	11.0
Guayanilla	8.0	50	7.6	11.5
Rosario	11.0	50	7.0	9.8
G. de Añasco	13.0	150	6.0	7.7
Tanamá	8.0	100	6.5	8.5
Valenciano	10.0	50	7.9	11.0
Gurabo	11.5	50	7.0	10.0
Canóvanas	12.5	50	7.5	10.0
Espiritu Santo	8.0	150	8.0	10.9
Cerrillos	11.0	150	6.5	9.5
Yahuecas	10.0	100	9.0	13.0
G. de Patillas	9.0	200	7.0	10.3

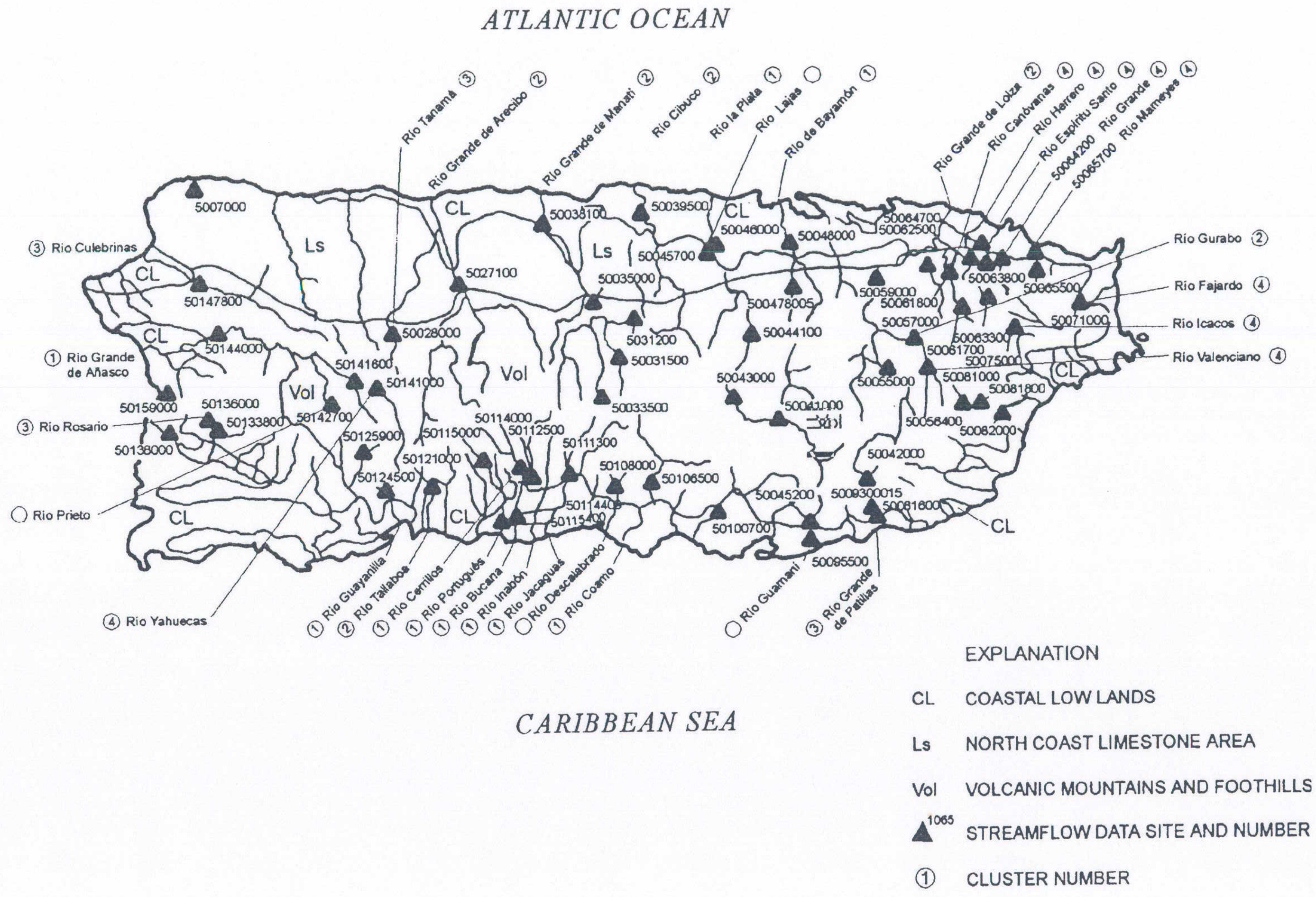

Figure 3-1. Geographical cluster distribution (from Segarra, 1991)

variables. The model still needed other adjustments to obtain a higher correlation coefficient and to be certain that it would predict acceptably. By examining Figure 3-2 we can see that the mean flow is directly proportional to the area. This indicated that one of the most important characteristics in the model was the area of the catchment, but there still was some scatter that could be improved through the use of a transformation.

The next model considered was of the form:

$$\log Q_{mean} = 3.17 + 0.593 \log Area \qquad (3\text{-}3)$$

where Q_{mean} and Area are the same as described earlier. This regression model gave lower mean square errors, but had a low coefficient of determination (R^2) equal to 73.5%. The plot of residuals presented on Figure 3-3 shows no obvious model defects. The scatter plot in Figure 3-4 shows an improvement when compared to that without the transformation. The next step was to include more variables in the regression and compare the results. The stepwise procedure was used to select the best variables for the model . The Statistical Analysis Software system, SAS/STAT release 6.03, for the VAX/VMS computer (SAS, 1988) was used to obtain the desired equation. The model produced by the analysis was

$$\log Q_{mean} = 2.02 + 0.76 \log Area + 0.318\, X5 - 0.135\, X25 \qquad (3\text{-}4)$$

where, Q_{mean}, Area, X5 and X25 are the same as described earlier. The coefficient of determination and the correlation coefficient were equal to 89% and 93% respectively. This indicates that 89 percent of the variability in the mean flow Q_{mean} has been explained

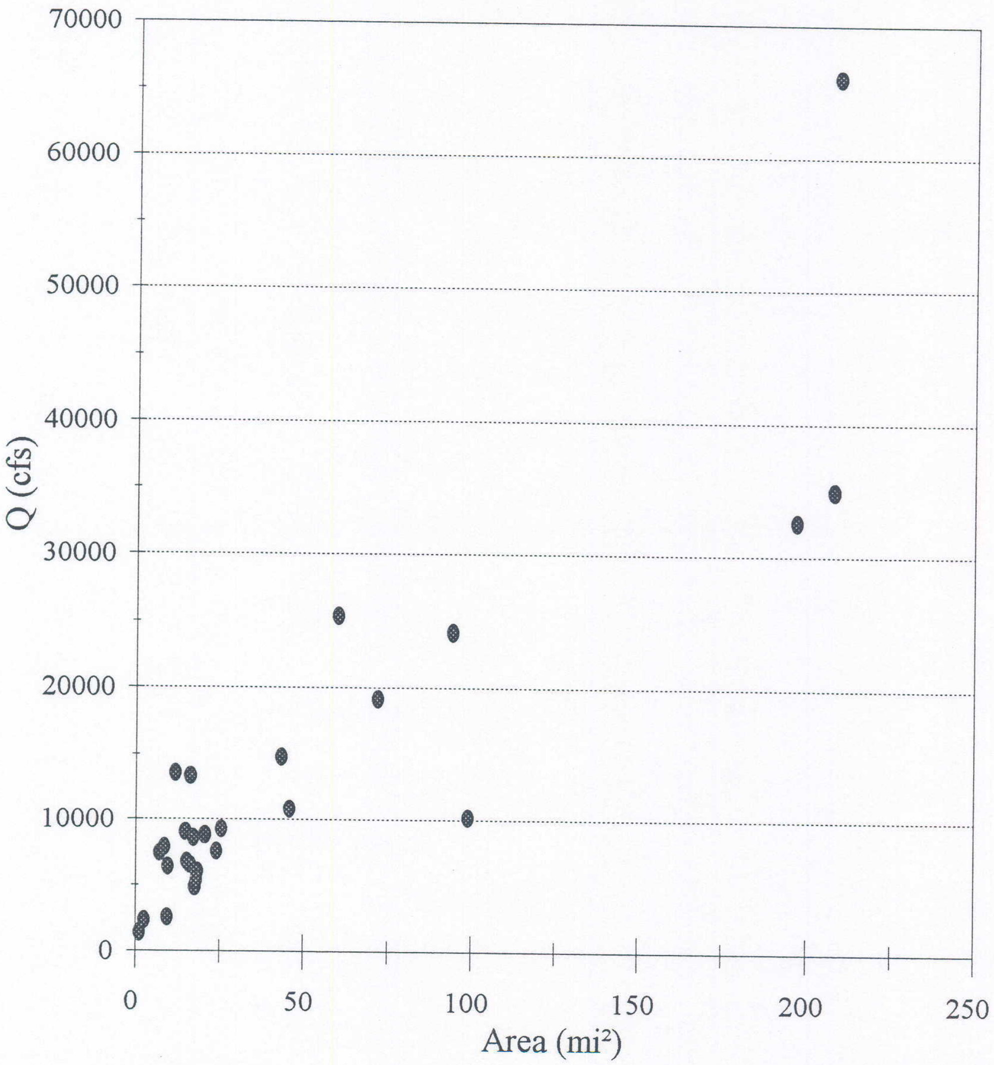

Figure 3-2. Scatter plot of the mean flow (cfs) versus area (mi^2).

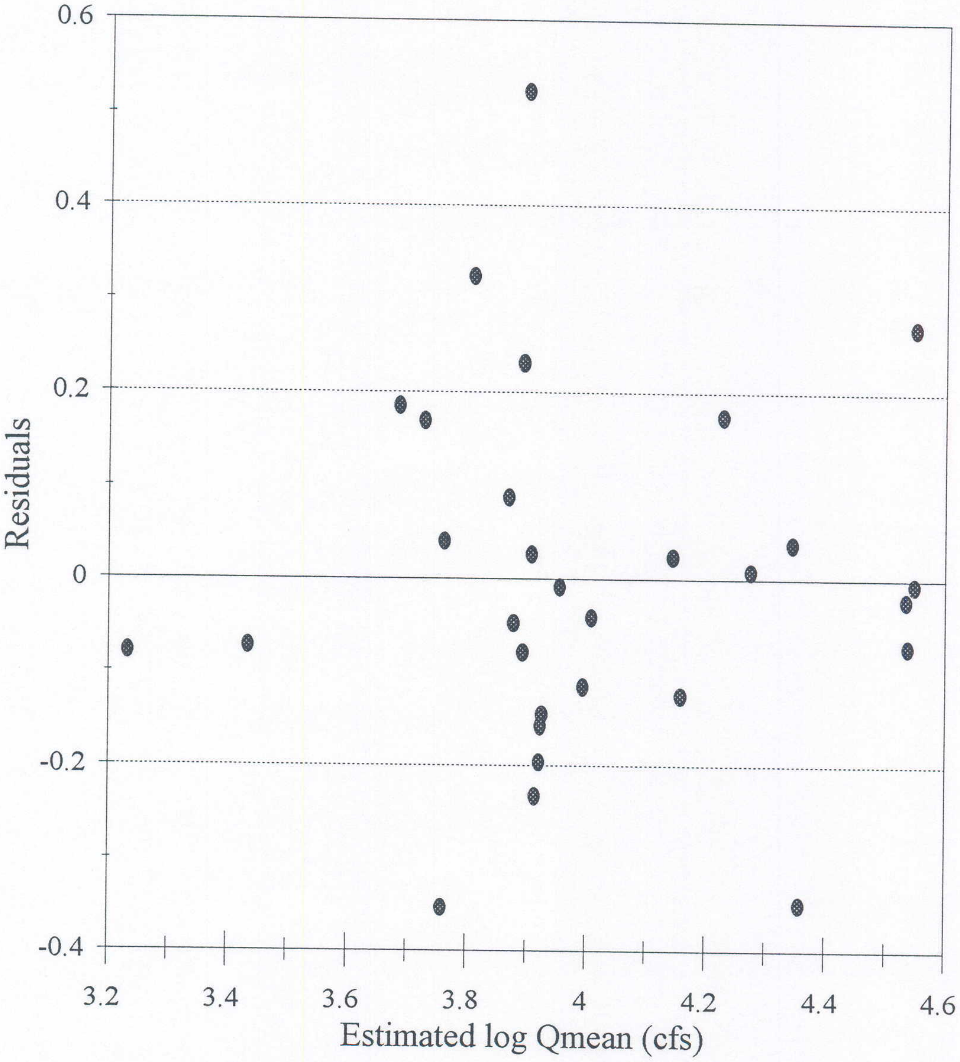

Figure 3-3. Plot of residuals for the transformed data.

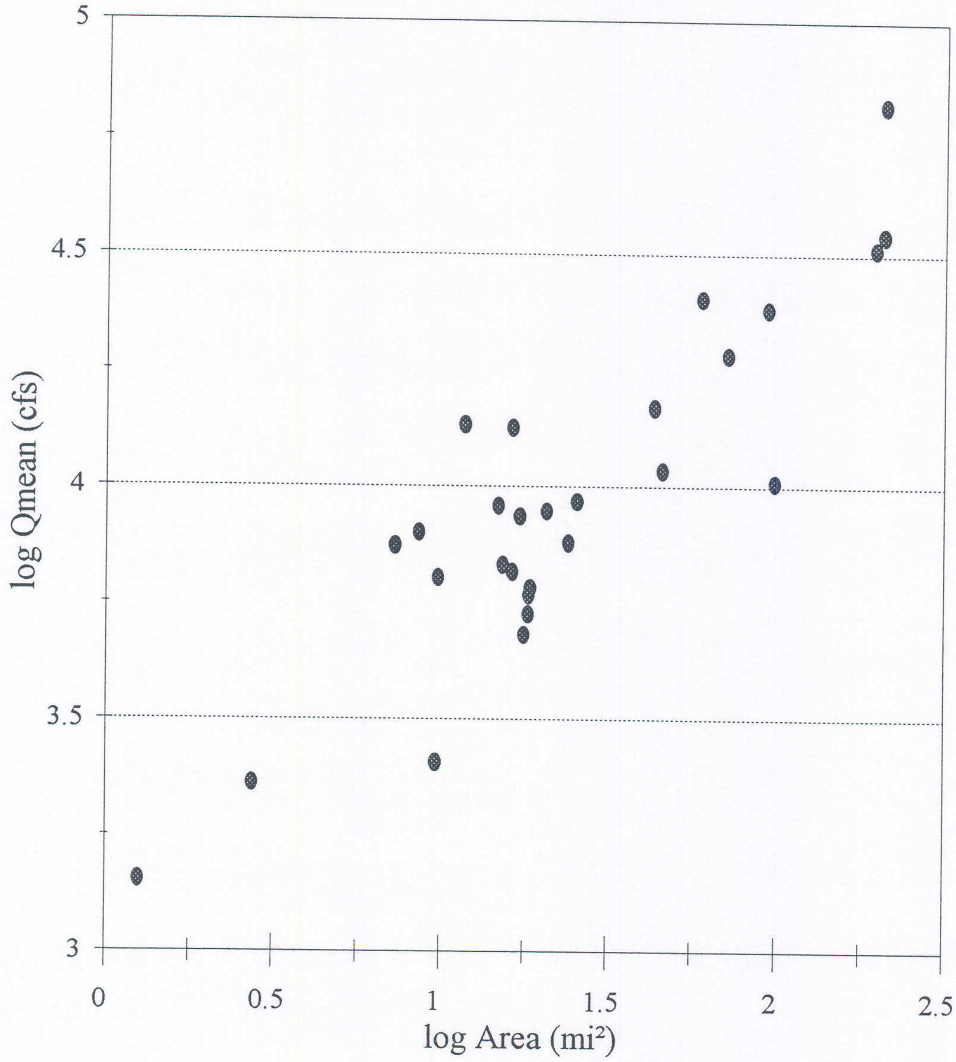

Figure 3-4. Scatter plot for the transformed data.

by the model. It can be seen in Figure 3-5 that the points lie approximately along a straight line, which indicates that the distribution of the residuals is normal. This is important because one of the assumptions of a regression analysis is that the residuals are normally distributed. The expected normal value was determined as

$$\Phi^{-1} = [(i - .5) / n] \tag{3-5}$$

where Φ^{-1} denotes the standard normal cumulative distribution. The residual plot presented in Figure 3-6 shows no obvious defects in the model. The residuals are scattered and do not form a shape that could give the impression that more transformations are needed.

It is also important to examine the difference between the R^2 and the R^2 adjusted. The coefficient of determination adjusted is equal to 87.7%. This small difference indicates that there is no problem of overspecification in the model.

Another tool used is the variance inflation factor (VIF). The VIF for the j^{th} regression coefficient can be written as

$$VIF_j = \frac{1}{1 - R_j^2} \tag{3-6}$$

where R^2_j is the coefficient of multiple determination obtained from regressing x_j on the other regressor variables (Montgomery and Peck, 1992). The VIF was checked to identify any problem of multicollinearity. Variance inflation factors larger than 10 imply serious problems with multicollinearity. The resulting value of 8.5 demonstrates no problem of multicollinearity. Also, the model was checked to determine influential values and no

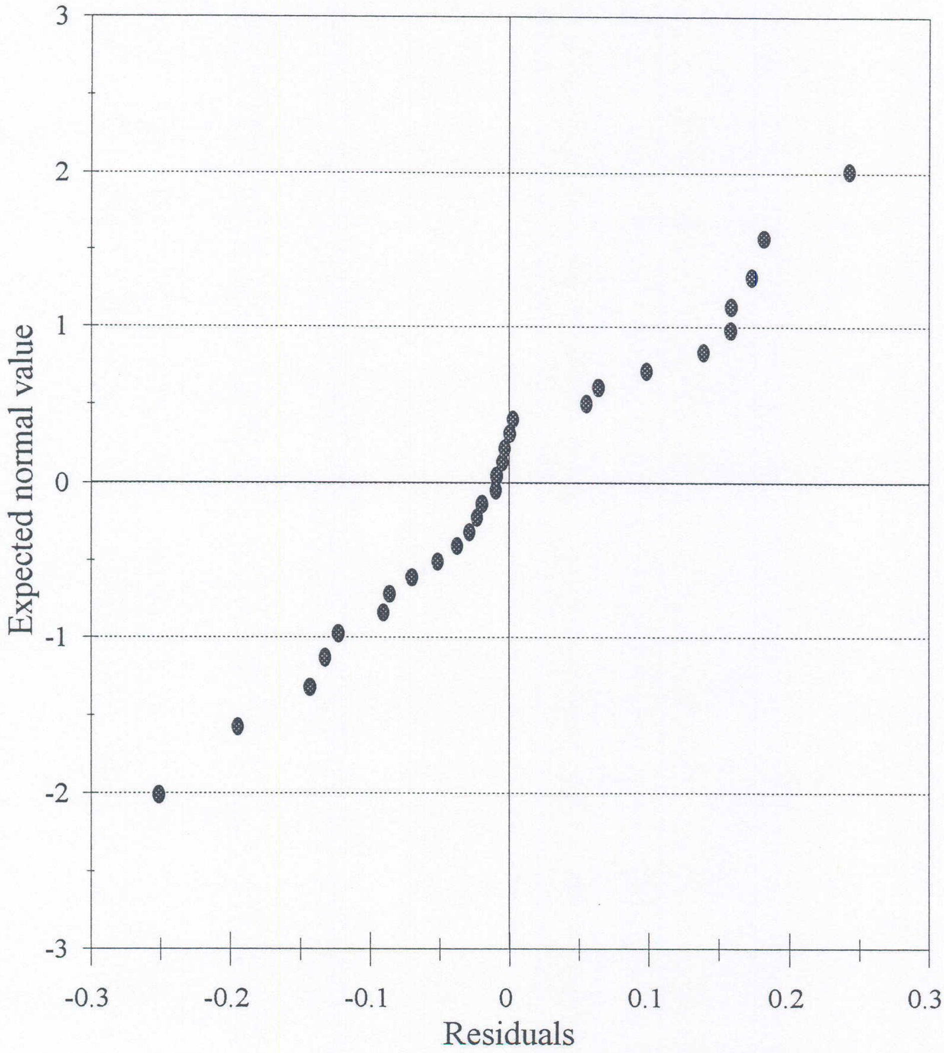

Figure 3-5. Normal probability plot of the residuals, final model.

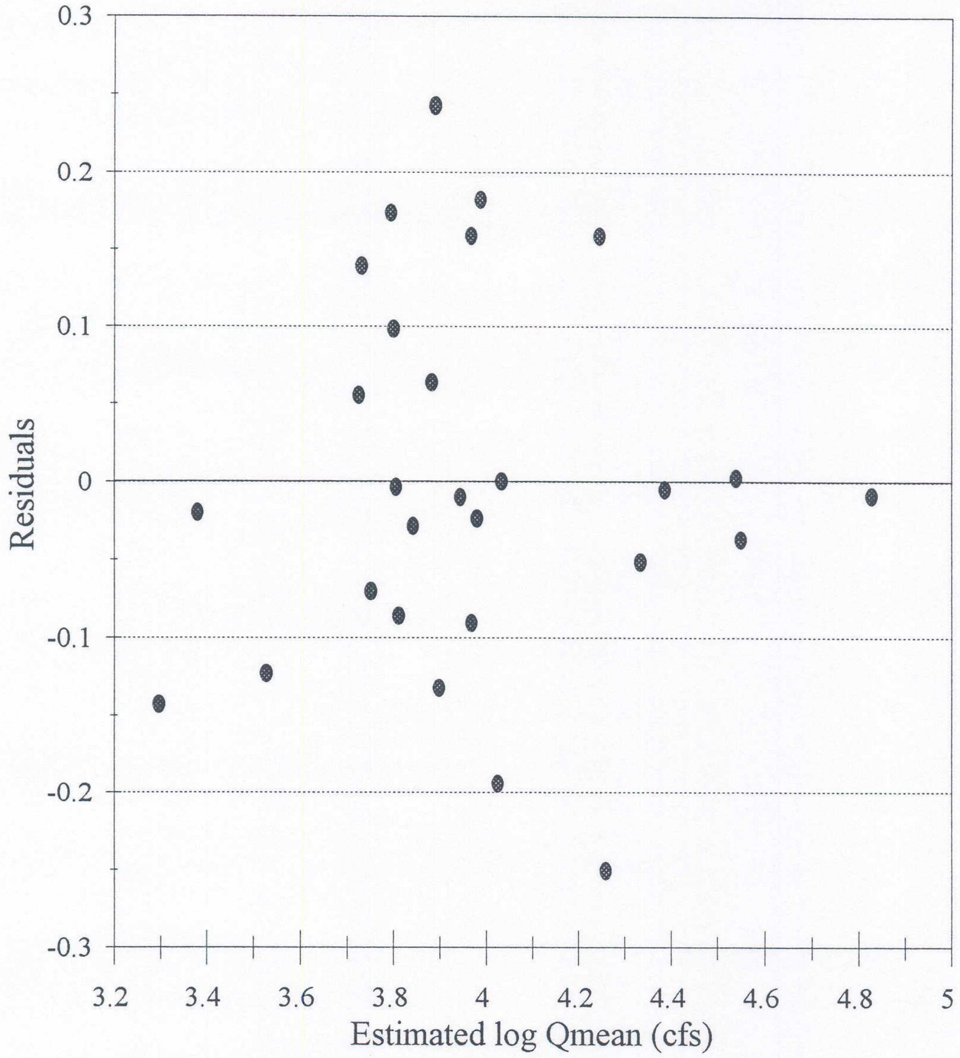

Figure 3-6. Plot of residuals of final regression model.

problem was found. An example of the subroutine used with the Statistical Analysis Software system, SAS/STAT release 6.03, for the VAX/VMS computer (SAS, 1988) is shown in Section A.1 of Appendix A. A complete output from SAS/STAT (SAS, 1988) is presented in Section A.2.

One of the great advantages of this model is that the X5, X25 and the Area are characteristics easily measured from the technical paper no. 42 (US Weather Service, 1961) and topographical maps, which are available to the general public. The resulting model can be used to estimate the mean flow (Q_{mean}) for almost any ungaged catchment in Puerto Rico and can be applied with minimal difficulty and small errors.

CHAPTER IV

DISCRIMINANT ANALYSIS AND APPLICATION

4.1 Discriminant Analysis

Discriminant analysis is related to the problem of identification. Within the present context, it will be used to determine the probability that an ungaged catchment has of belonging to a particular cluster, or group of basins for which generalized flood frequency curves have been defined. From an earlier study (Segarra, 1991), gaged basins in Puerto Rico were grouped into four clusters from which regionalized flood frequency curves were developed. The discriminant procedure will obtain the parameters of a flood frequency curve for an ungaged basin from the parameters of the regionalized flood curves from the four clusters, based on the cluster membership probability (Manly, 1986). The attributes employed in discriminant classification consist of climatological and geomorphological catchment descriptors.

The data for a discriminant function analysis do not need to be standardized to have zero means and unit variances prior to the start of the analysis, as is usual with principal component and factor analysis. This is because the outcome of a discriminant function analysis is not affected in any important way by the scaling of individual variables. A general description of the basic theory of discriminant analysis, following Rao (1973), is presented next.

Let x denote the measurements on an individual and S the sample space of possible values of x. On the basis of observed x, a decision must be reached about the membership of an individual in one of k specified populations. The situation is equivalent to that of choosing one among a given set of alternate hypotheses appropriate to an observed event. A decision rule is specified which allows assigning an individual with measurements x to a given population. For the process, a loss function r_{ij} is defined which determines the loss in assigning a member of i^{th} population to the j^{th}. Letting $P_1(x), \ldots, P_k(x)$ be the probability densities at x with respect to a measure υ in the k populations, the expected loss in applying a given rule, when in fact the individuals come from the i^{th} population, is

$$L_i = \int_{w_1} r_{i1} P_i(x)\, dv + \ldots + \int_{w_k} r_{ik} P_i(x)\, dv \tag{4-1}$$

where $w_1, \ldots, w_k$ represent the mutually exclusive regions into which the sample space S is divided into. As a rule, an individual with measurements x is assigned to the i^{th} population if $x \in w_i$. The loss vector $(L_1, \ldots, L_k)$, corresponding to the k alternative hypotheses is known as the operating characteristic of the decision rule.

Let $\pi_1, \ldots, \pi_k$ be a prior probability of the k populations. The expected loss then reduces to the quantity

$$L = \pi_1 L_1 + \cdots + \pi_k L_k \tag{4-2}$$

Using the expression in Equation (4-1), the expected loss is given by

$$L = \sum_{i=1}^{k} \int_{w_i} (\pi_1 r_{1i} P_1 + \cdots + \pi_k r_{ki} P_k)\, dv \tag{4-3}$$

or

$$L = \int_{w_1} -S_1\, dv + \cdots + \int_{w_k} -S_k\, dv \tag{4-4}$$

where S_i is called the i^{th} discriminant score of an individual (for the i^{th} population). It can be shown that, if w_1^*, ..., w_k^* are mutually exclusive regions covering the whole sample space, such that

$$x \in w_i^* \Rightarrow S_i(x) \geq S_j(x) \qquad \text{for all } j,\ i=1,\ldots,k \tag{4-5}$$

then for such a choice of w_i, the expected loss of Equation (4-4) is a minimum.

For solving the problem, the following must be known:

1) The probability densities, $P_1(U)$, ..., $P_k(U)$, for a given set of measurements U on an individual in the k alternative populations.

2) Prior probabilities π_1, ..., π_k for the populations, which are relative frequencies of individuals of the k populations in the composite population from which an individual to be identified has been observed.

3) The assignment of a loss function, that is, the specification of values r_{ij} representing the loss in identifying an individual of the i^{th} population as a member of the j^{th} population.

Thus, given an individual with measurements U, his discriminant score for the i^{th} population is computed as

$$S_i = -[\pi_1 P_1(U) r_{1i} + \dots + \pi_k P_k(U) r_{ki}] \quad i=1,\dots,k \tag{4-6}$$

In many practical applications it is difficult to assess the losses due to wrong identification, in which case the criterion of minimizing the frequency of wrong identification is adequate. The optimum rule is to assign the individual with measurements U to that population for which the posterior probability has the highest value. The discriminant score for the i^{th} population is then given by

$$S_i = \pi_i P_i(U) \tag{4-7}$$

which is in the form of a posterior distribution.

It is assumed that the distribution of U is p-variate normal in each of the populations. This allows the evaluation of the discriminant score as

$$S_i = (B_i^T C^{-1}) U - \frac{1}{2} B_i^T C^{-1} B_i + \log \pi_i \tag{4-8}$$

in which B_i is the vector of measurement means, and C is the covariance matrix of the measurements. Redefining measurements in terms of basin characteristics, it has been possible to employ discriminatory analysis for the classification of basins into groups or clusters. The underlying assumption is that the set of basin characteristics is considered normally distributed.

The characteristics used for the discriminant analysis were the same as those in the regression procedure. A stepwise procedure was used to determine the optimum number of basin characteristics to be used in the discriminant analysis. This procedure was performed using subroutine DISCRIMINANT of the Statistical Package for Social Sciences, SPSS release 4.0, for the VAX/VMS computer (Nie et al. 1975). In this procedure variables are added one by one until it is found that adding extra variables does not give significantly better discrimination. There are many different criteria that can be used to decide which variables to include and which to leave out. The criterion selected maximizes the smallest F ratio. The F ratio is defined as

$$F = \frac{M_B}{M_W} \tag{4-9}$$

where M_B is the mean square variation between groups and M_W the mean square variation within groups. The criterion selects the variable that maximizes the smallest F ratio between pairs of groups. The default value of the F to enter is 1. This corresponds to a significance level of about 0.5 for large sample sizes (SPSS Inc., 1990). The ungaged catchments and the characteristics used in the discriminant analysis are listed in Table 4-1.

The discriminant analysis of these basins was also performed using subroutine DISCRIMINANT of SPSS release 4.0, for the VAX/VMS computer (Nie et al. 1975). The

Table 4-1. Ungaged Basin Characteristics

Basin	AREA (mi^2)	ET (in)	LAKE (%)
Arroyata	17.41	44.0	0.00
Blanco (Este)	25.92	55.0	0.00
Blanco (Oeste)	124.84	43.0	0.00
Cagüitas	14.10	54.0	0.00
Canovanillas	19.33	56.0	0.00
Caonillas	40.40	54.0	0.82
Cayaguas	10.20	45.0	0.00
Cialitos	17.00	53.0	0.00
Daguao	2.26	60.0	0.00
Grande de Jayuya	44.15	49.0	0.00
Guamaní	12.30	45.0	0.00
Guanajibo	120.00	56.0	0.00
Guayanés (Este)	34.00	65.0	0.00
Guayanés	12.70	55.0	0.00
Guaynabo	12.06	54.0	0.00
Jueyes	11.85	57.0	0.00
Lapa	9.97	55.0	0.00
Majada	16.70	57.0	0.00
Maunabo	12.70	67.0	0.00
Mavilla	9.51	48.0	0.00
Orocovis	10.10	45.0	0.00
Piedras	15.40	57.0	0.00
Santiago	4.99	58.0	0.00
Toro Negro	34.63	50.0	0.17
Turabo	7.40	50.0	0.00
Unibón	5.29	50.0	0.00
Usabón	9.15	42.0	0.00
Viví	5.66	51.0	0.00
Yagüez	6.70	55.0	0.00
Yauco	45.50	57.0	0.31
Yunés	34.54	55.0	0.00

Table 4-1 (Cont.). Ungaged Basin Characteristics

Basin	ANP (in)	MOP (in)	X5 (in)	X25 (in)
Arroyata	60	8.0	6.5	9.7
Blanco (Este)	130	12.0	8.8	12.0
Blanco (Oeste)	84	11.0	7.1	10.0
Cagüitas	76	7.9	6.8	9.5
Canovanillas	100	11.0	8.0	11.5
Caonillas	70	10.3	7.4	10.0
Cayaguas	90	9.0	7.5	10.8
Cialitos	80	9.0	6.8	9.8
Daguao	80	9.0	8.5	11.5
Grande de Jayuya	70	10.0	9.0	13.0
Guamaní	90	10.0	6.8	9.5
Guanajibo	80	11.0	8.0	11.3
Guayanés (Este)	95	11.0	7.3	10.5
Guayanés	88	9.0	8.5	12.0
Guaynabo	67	7.3	6.8	9.0
Jueyes	40	7.0	6.4	9.3
Lapa	74	9.5	6.6	9.8
Majada	65	7.0	6.7	9.5
Maunabo	90	10.5	7.4	10.5
Mavilla	70	8.0	6.9	9.7
Orocovis	80	10.0	7.4	10.8
Piedras	83	7.1	7.0	9.0
Santiago	100	10.0	8.7	12.0
Toro Negro	105	11.5	7.4	10.0
Turabo	90	11.0	7.4	10.0
Unibón	87	10.6	6.8	10.0
Usabón	54	8.0	6.5	10.3
Viví	80	11.0	9.0	12.0
Yagüez	103	13.0	7.2	10.0
Yauco	82	9.0	7.8	11.5
Yunés	70	10.0	6.5	8.8

program computes the covariance matrix, the vectors of the mean of each basin characteristic for each cluster or group, and the prior probability, to supply as output the discriminant scores for each cluster by means of Equation (4-8). The program also gives a "performance matrix" in which the basins are notally allocated to the cluster which yields the highest discriminant score. This notional allocation, based solely on basin characteristic data, is then compared with previous allocation of basins to the cluster derived from flood statistics shown in Table 4-2 (Segarra, 1991).

Utilizing the discriminant scores, the subroutine computes the probability of a new ungaged basin being in each of the four clusters. The probabilities are obtained from

$$P_i = \frac{\exp\,(S_i)}{\sum_{i=1}^{M} \exp\,(S_i)} \tag{4-10}$$

where S_i was previously defined as the cluster discriminant score. The use of Equation (4-10) implies fractional membership, and provides an attractive alternative to unique allocation to a single cluster, since the consequences of allocating the ungaged catchment to the wrong cluster are alleviated. These probabilities are shown in Table 4-3. Table 4-4 shows the cluster to which the ungaged basins had the highest probability of belonging. An example of the data and subroutine used with the subroutine DISCRIMINANT of SPSS release 4.0, for the VAX/VMS computer (Nie et al. 1975), are shown in Section B.1 and B.2 of Appendix B. A complete output given by the SPSS computer package is also presented in Section B.3.

Table 4-2. Cluster Arrangement for Gaged Basins (from Segarra, 1991)

Basin	Cluster
G. de Manatí	2
Cibuco	2
De la Plata	1
Bayamón	1
G. de Loíza	2
Herrera	4
Grande	4
Mameyes	4
Fajardo	4
Icacos	4
Humacao	1
Coamo	1
Jacaguas	1
Inabón	1
Bucaná	1
Portugues	1
Tallaboa	2
Guayanilla	1
Rosario	3
G. de Añasco	1
Culebrinas	3
Tanamá	3
G. de Arecibo	1
Valenciano	4
Gurabo	2
Canóvanas	4
Espiritu Santo	4
Cerrillos	1
Yahuecas	4
G. de Patillas	3

Table 4-3. Ungaged Basin Probabilities of Belonging to Clusters

Basin	Cluster Number			
	1	2	3	4
Arroyata	0.029	0.000	0.609	0.363
Blanco (Este)	0.000	0.000	0.036	0.964
Blanco (Oeste)	0.000	1.000	0.000	0.000
Cagüitas	0.004	0.000	0.992	0.003
Canovanillas	0.008	0.001	0.738	0.253
Caonillas	0.443	0.001	0.511	0.044
Cayaguas	0.000	0.000	0.029	0.971
Cialitos	0.010	0.000	0.987	0.003
Daguao	0.003	0.000	0.169	0.828
Grande de Jayuya	0.000	0.025	0.000	0.975
Guamaní	0.000	0.000	0.874	0.125
Guanajibo	0.000	1.000	0.000	0.000
Guayanés (Este)	0.211	0.004	0.785	0.000
Guayanés	0.000	0.000	0.010	0.990
Guaynabo	0.010	0.000	0.983	0.007
Jueyes	0.663	0.000	0.337	0.000
Lapa	0.031	0.000	0.969	0.000
Majada	0.023	0.000	0.976	0.000
Maunabo	0.046	0.000	0.954	0.000
Mavilla	0.004	0.000	0.733	0.263
Orocovis	0.001	0.000	0.053	0.947
Piedras	0.001	0.000	0.997	0.002
Santiago	0.000	0.000	0.063	0.937
Toro Negro	0.006	0.001	0.860	0.133
Turabo	0.005	0.000	0.805	0.190
Unibón	0.003	0.000	0.992	0.004
Usabón	0.015	0.000	0.304	0.680
Viví	0.000	0.000	0.000	1.000
Yagüez	0.008	0.000	0.991	0.000
Yauco	0.129	0.014	0.554	0.303
Yunés	0.678	0.012	0.310	0.000

Table 4-4. Highest Probability Clusters for Ungaged Basins

Basin	Cluster
Arroyata	3
Blanco (Este)	4
Blanco (Oeste)	2
Cagüitas	3
Canovanillas	3
Caonillas	3
Cayaguas	4
Cialitos	3
Daguao	4
Grande de Jayuya	4
Guamaní	3
Guanajibo	2
Guayanés (Este)	3
Guayanés	4
Guaynabo	3
Jueyes	1
Lapa	3
Majada	3
Maunabo	3
Mavilla	3
Orocovis	4
Piedras	3
Santiago	4
Toro Negro	3
Turabo	3
Unibón	3
Usabón	4
Viví	4
Yagüez	3
Yauco	3
Yunés	1

After the characteristic group is obtained, discriminant analysis will assign the catchments to each of the four groups, or clusters of river basins defined for Puerto Rico, with an estimate of the probability of belonging to each group. Then, the estimate of the T-year flood X_T for a catchment is given by

$$X_T = \frac{\sum_{i=1}^{M} P_i T_i}{\sum_{i=1}^{M} P_i} \qquad (4\text{-}11)$$

where T_i is the i^{th} cluster quantile estimate of the T-year flood, and P_i is the posterior probability of the new ungaged basin being in each of the M clusters. The quantile estimates for each cluster are listed in Tables 4-5 to 4-8. Equation (4-11) is used to construct a dimensionless frequency curve for a new ungaged basin considering several return periods. The flood frequency curves for the ungaged basins are presented in Appendix C.

4.2 Application

In this chapter and in the preceding the individual components of the research were shown. The final product incorporates the regression and the discriminant analysis to obtain flood frequency curves for ungaged catchments. To better illustrate the use of each analysis an example will be presented.

Table 4-5. Quantiles for Cluster 1 (from Segarra, 1991)

Return Period (years)	Quantile Xt
2	0.615
3	0.885
4	1.100
5	1.279
10	1.959
15	2.463
20	2.880
25	3.244
30	3.569
35	3.867
40	4.142
45	4.399
50	4.642
55	4.872
60	5.090
65	5.300
70	5.500
75	5.694
80	5.880
85	6.060
90	6.235
95	6.405
100	6.570
200	9.233
300	11.241
400	12.916
500	14.380

Table 4-6. Quantiles for Cluster 2 (from Segarra, 1991)

Return Period (years)	Quantile Xt
2	0.771
3	1.103
4	1.334
5	1.514
10	2.096
15	2.461
20	2.734
25	2.955
30	3.141
35	3.303
40	3.447
45	3.576
50	3.695
55	3.803
60	3.904
65	3.998
70	4.087
75	4.170
80	4.249
85	4.323
90	4.395
95	4.463
100	4.528
200	5.470
300	6.078
400	6.537
500	6.910

Table 4-7. Quantiles for Cluster 3 (from Segarra, 1991)

Return Period (years)	Quantile Xt
2	0.893
3	1.096
4	1.231
5	1.334
10	1.651
15	1.839
20	1.975
25	2.082
30	2.171
35	2.247
40	2.313
45	2.372
50	2.426
55	2.474
60	2.519
65	2.561
70	2.599
75	2.635
80	2.669
85	2.702
90	2.732
95	2.761
100	2.788
200	3.172
300	3.407
400	3.579
500	3.716

Table 4-8. Quantiles for Cluster 4 (from Segarra, 1991)

Return Period (years)	Quantile Xt
2	0.889
3	1.131
4	1.289
5	1.407
10	1.765
15	1.972
20	2.119
25	2.234
30	2.328
35	2.408
40	2.478
45	2.539
50	2.595
55	2.645
60	2.691
65	2.733
70	2.773
75	2.810
80	2.844
85	2.877
90	2.907
95	2.936
100	2.964
200	3.343
300	3.570
400	3.734
500	3.862

As shown in the foregoing and current chapter, it is necessary to measure basin climatological and geomorphological characteristics to develop flood frequency curves for ungaged catchments. For the purpose of this example, the 100-year peak flow of Río Guadiana basin is computed. The characteristics for this watershed are presented in Table 4-9. First, to calculate the mean annual flow we must use Equation 3-4.

Example 4-1: Mean annual flow of Río Guadiana basin.

The mean annual flow of the 5.54 mi^2 Río Guadiana basin is computed by means of Equation (3-4) as

$$Log\ Q_{mean} = 2.02 + .76 * Log\ 5.54 + .318 * 7 - .135 * 10$$

$$Q_{mean} = 2891\ cfs$$

After determining the mean annual flow, the attempt is made to group the basin into one of the four clusters defined for the Island. The discriminant analysis described in Chapter 4 is used to determine the membership probability of the catchment in each of the four clusters. Using subroutine DISCRIMINANT of SPSS release 4.0, for the VAX/VMS computer (Nie et al. 1975), the probabilities shown in Table 4-10 are obtained. From the table, it can be seen that for all practical purposes (with a probability of 99.5%), the Río Guadiana basin can be grouped with the catchments in Cluster 3. Therefore, the flood frequency parameters for this basin will be largely determined from the flood frequency parameters from Cluster 3 basins.

Table 4-9. Basin Characteristics for Examples

Río Guadiana Basin			
AREA (mi^2)	ET (in)	LAKE (%)	ANP (in)
5.54	55	0	80.00

Río Guadiana Basin (Cont.)		
MOP (in)	X5 (in)	X25 (in)
8	7	10

Table 4-10. Basin Probabilities of Belonging to Clusters

	Cluster Number			
Basin	1	2	3	4
Guadiana	0.001	0.000	0.995	0.004

The next step involves the estimation of the T-year quantile. Using Equation 4-11 we can obtain the quantile estimate for a 100-year recurrence interval.

Example 4-2: 100-year quantile for Río Guadiana basin.

The quantile of the 100-year flood for Río Guadiana basin, from Equation (4-11) is given by :

$$X_{100} = 6.57 * .0015 + 0 + 2.788 * .99465 + 2.964 * .0043$$

$$X_{100} = 2.8$$

To obtain the T-year maximum flow for the basin, the T-year quantile, X_T, is multiplied by the mean annual flow.

Example 4-3: 100-year maximum flow for Río Guadiana basin.

The 100-year flow is obtained as the product of the quantile X_{100} and the mean annual flow Q_{mean}:

$$Q_{100} = X_{100} * Q_{mean} = 8095 \quad cfs$$

The same procedure was followed for the 31 ungaged catchments shown in Table 4-1. The recurrence intervals ranged from 2 to 500 years. The flood frequency curves developed for each basin can be found in Appendix C.

CHAPTER V

DISCUSSION OF RESULTS

The study derived flood frequency distributions for 31 ungaged streams in Puerto Rico by using discriminant analysis procedures. The distributions obtained will be of use to water resource specialists for all projects related to flood flow management. Excellent results were obtained when the methodology was tested with streams for which flood flow records were available, but had not been included in the original regionalized flood frequency estimation study.

The study demonstrated that seven geomorphologic characteristics are necessary to obtain the flood frequency curves by means of the proposed procedure. In future studies other characteristics such as stream frequency and drainage density could be included in the procedure. These parameters depend significantly on the precision with which geomorphological attributes are measured. The implementation of geographical information systems (GIS) and remote sensing procedures would be of immense aid in the accurate measurement of geomorphological characteristics. Refined measurements could, in principle, allow inclusion of particular geomorphological attributes into the augmented discriminating parameter set used for the analysis.

Regarding the flood frequency results obtained, the curves for the Río Yunés, Río Jueyes, Río Guayanés Este and Río Caonillas basins are rather steep curves. Wiltshire and Beran (1987) noted in their study that a basin with a steep frequency curve is representative of small, wet, steep and impermeable catchments. They also demonstrated that such a basin could be expected to produce consistently large floods with a relatively small coefficient of variation. On the other hand, the large, wet and flat watersheds could be expected to produce floods of small specific runoff and small coefficient of variation.

To test the accuracy of the flood frequency procedure, two gaged rivers not used in the present work or in the previous flood frequency study were analyzed as if they were ungaged basins. These basins were the Río Grande de Manatí at Ciales and the Río De la Plata at Proyecto la Plata. The watershed characteristics necessary for the proposed method are shown in Table 5-1. All procedures presented in the previous chapter were used to estimate the flood flows. The computed flood flows were then compared with the historical data. The measure of performance of the proposed procedure is the standard error. To compute the error, Log-Pearson and Probability Weighted Moments fitting techniques were used to obtain the T-year flood from the historical record at both catchments. This was then compared to the T-year flood estimated from the ungaged catchment procedure. The relative estimation error of the estimated flows for the two catchments was computed as

$$REE\ (\%) = (100)\left|\frac{Q_R - Q_E}{Q_R}\right| \tag{5-1}$$

Here Q_R is the real T-year maximum flow from the Generalized Extreme Value Probability Weighted Moments (GEV/PWM) or Log Pearson Type III distribution of the available data, and Q_E is the T-year maximum flow estimated using the procedure developed in this

Table 5-1. Basin Characteristics

Basin	AREA (mi^2)	ET (in)	LAKE (%)	ANP (in)
G. Manatí at Ciales	136	55	1	70
De la Plata at Proyecto la Plata	54.8	51	0	70

Table 5-1. Basin Characteristics (Cont.)

Basin	MOP (in)	X5 (in)	X25 (in)
G. Manatí at Ciales	7.0	6.3	9.0
De la Plata at Proyecto la Plata	7.5	6.6	8.5

study. Figures 5-1 and 5-2 show the curves for each technique; it is seen that the three flood frequency curves behave in a similar manner.

Tables 5-2 to 5-5 present the standard errors for the two, presumed ungaged, basins. It can be seen that the errors are relatively low, ranging between 0.35% and 17.45%. Two distributions were used to compare the results and test the precision of the procedure. The GEV/PWM and the Log Pearson Type III distributions gave similar flood frequency estimates as those calculated by the ungaged catchment procedure presented. Greis and Wood (1981) also used regional regression estimates and had an average standard error of 59.1% in their study of 16 basins in Arizona. Aaron and Kibler (1979) used the Log Pearson Type III distribution to test their regional regression estimates and reported a range of standard errors for flood events of 31.5% to 71.7%, which are well above those presented in this study. Also, in the study by López et al. (1979) standard errors ranged from -38 to 61%. This contrasts with the standard errors obtained in the present study, which are considerably smaller. It also proves that the proposed procedure gives reliable results of flood frequency estimation for ungaged catchments in Puerto Rico.

Segarra (1991) noted that with better estimates of the mean annual flow lower standard errors could be observed. This also indicates that the regression model used for estimating the mean flows predicts in the desired manner. Certain care must be taken when using the multiple linear regression equation. Extrapolation should be avoided whenever possible. It is important not to predict outside the range of the original observations. It is

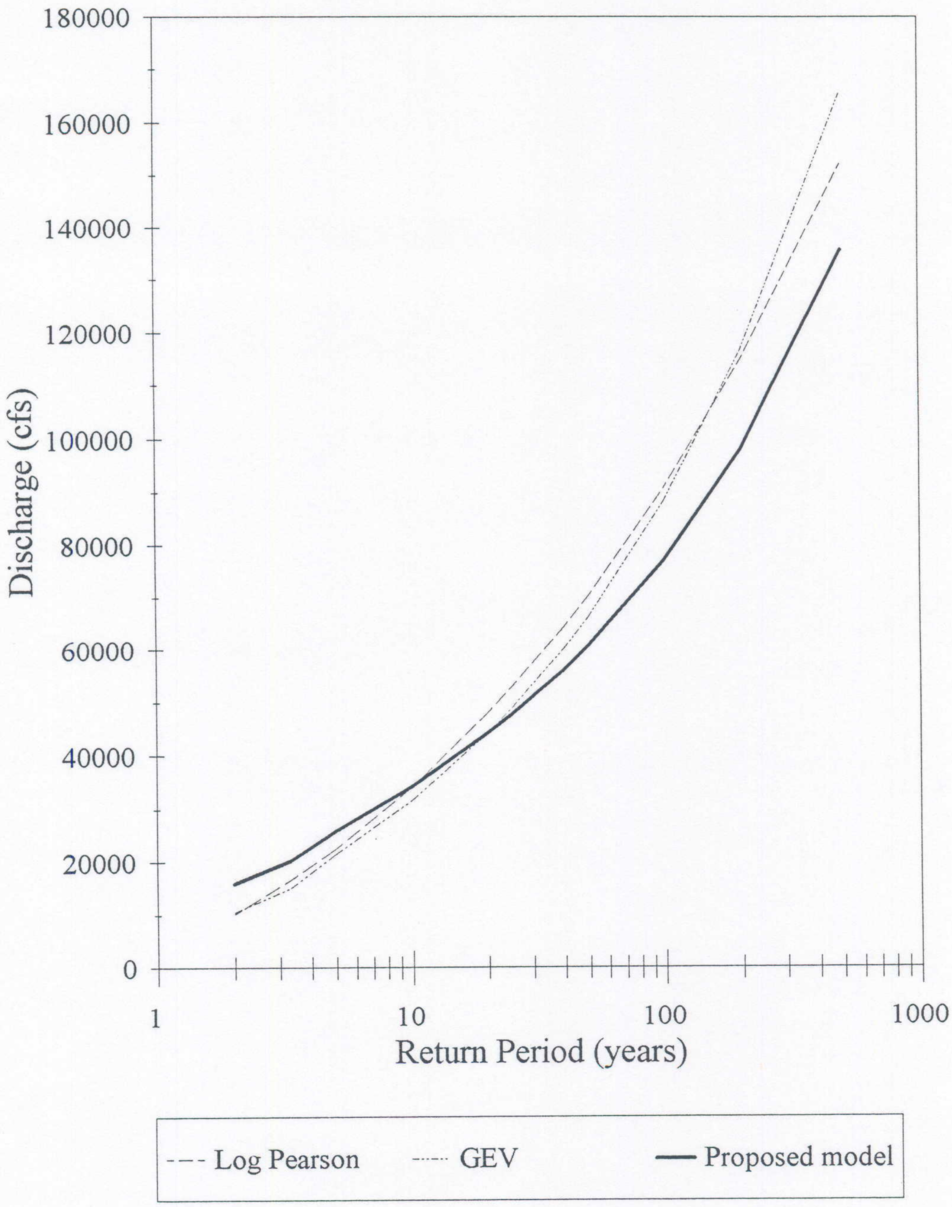

Figure 5-1. Comparison of flood frequency distributions for Río De la Plata at Proyecto la Plata

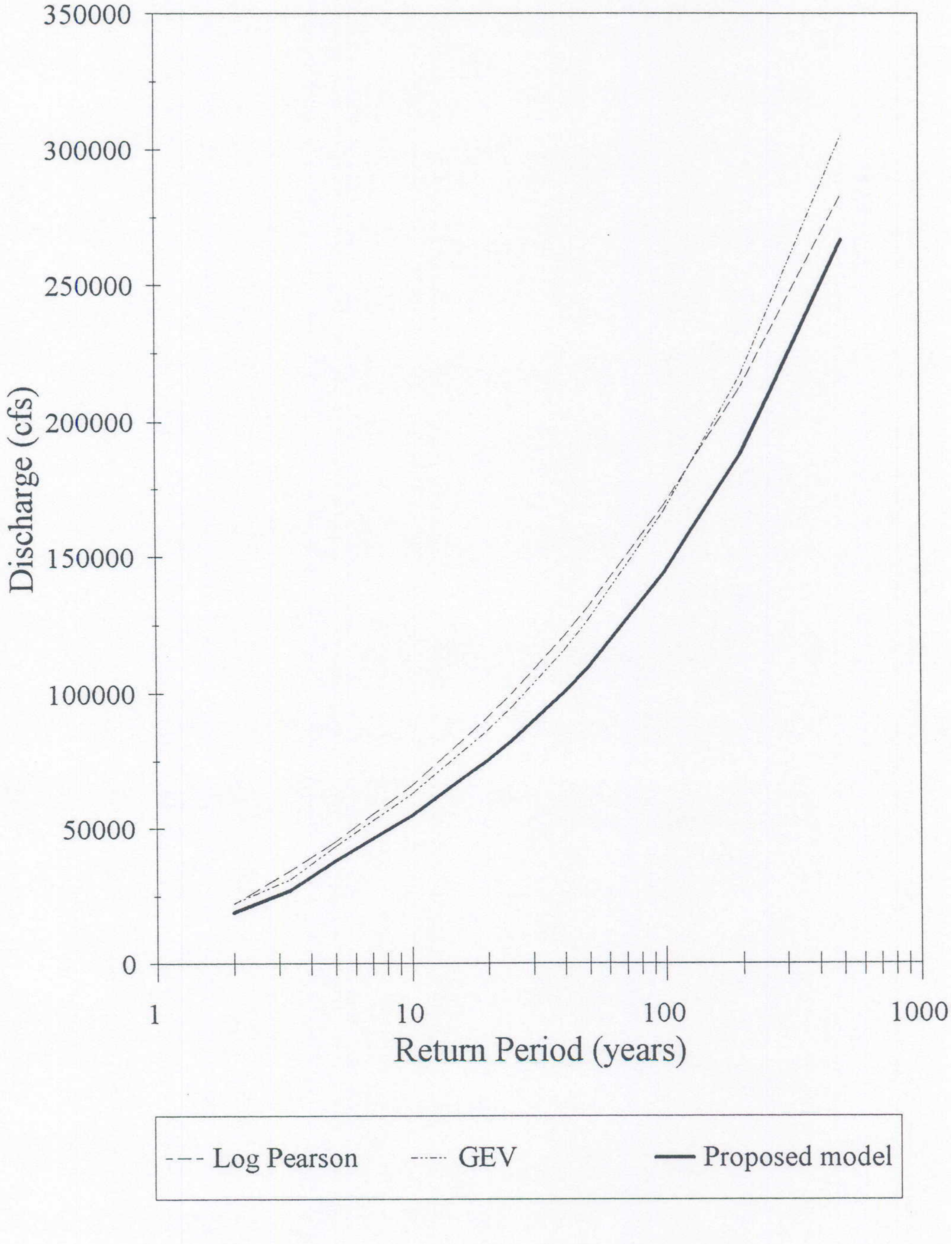

Figure 5-2. Comparison of flood frequency distributions for Río Grande de Manatí at Ciales

Table 5-2. Standard Errors Using Log Pearson Type III Distribution

Grande de Manatí at Ciales Basin			
Return Period	Qr *	Qe **	% error
10	66021	54877	16.88
25	100063	82604	17.45
50	131619	109659	16.68
100	169047	143972	14.80

* Qr is the T-year maximum flow from the distribution.

** Qe is the T-year maximum flow estimated using the proposed procedure.

Table 5-3. Standard Errors Using Generalized Extreme Value Distribution

Grande de Manatí at Ciales Basin			
Return Period	Qr *	Qe **	% error
10	62799	54877	12.61
25	95134	82604	13.17
50	126865	109659	13.56
100	166867	143972	13.72

* Qr is the T-year maximum flow from the distribution.

** Qe is the T-year maximum flow estimated using the proposed procedure.

Table 5-4. Standard Errors Using Log Pearson Type III Distribution

La Plata at Proyecto La Plata Basin			
Return Period	Qr *	Qe **	% error
10	34093	34211	0.35
25	48921	47697	9.80
50	66163	60652	13.66
100	90713	76975	15.14

* Qr is the T-year maximum flow from the distribution.

** Qe is the T-year maximum flow estimated using the proposed procedure.

Table 5-5. Standard Errors Using Generalized Extreme Value Distribution

La Plata at Proyecto La Plata Basin			
Return Period	Qr *	Qe **	% error
10	31594	34211	8.28
25	48921	47697	2.50
50	66163	60652	8.33
100	88160	76975	12.69

* Qr is the T-year maximum flow from the distribution.

** Qe is the T-year maximum flow estimated using the proposed procedure.

very possible that a model that fits well in the region of the original data will perform poorly outside that region. The areas of the watersheds ranged from 1.26 to 209 square miles.

CHAPTER VI

CONCLUSION AND RECOMMENDATIONS

With the purpose of estimating better flood flows for water resources projects, a linear regression model coupled with a discriminant analysis was used to develop flood frequency curves for ungaged catchments in Puerto Rico. The following conclusions were obtained from the research:

1. The regression model can be used to estimate the mean annual flow for almost any ungaged catchment in Puerto Rico, using known values of the catchment area and the 5 and 25-year 24 hour storm. The use of this model is limited to the range of areas used; these were between 1.26 and 209 square miles.

2. Islandwide, 31 flood frequency curves for ungaged basins were developed using the discriminant procedure.

3. For the discriminant analysis procedure, it was determined that the optimal parameter set for ungaged basin classification consisted of six attributes: catchment area, evapotranspiration, fraction of area covered by lake, annual precipitation, September precipitation, and the 5-year, 24 hour rainfall.

4. The largest estimation error obtained, when comparing results with gaged streams, was in the order of 17.4%, for the 25-year peak flow. This error compares favorably with errors from other related procedures, showing that the flood frequency

estimates for Puerto Rico are significantly improved with the present study, thus establishing the worth of the proposed procedure.

The application of the procedure yielded reliable results with the available data. However, other developments can be considered to improve the procedures as part of the dynamics of the theoretical evolution of these methodologies. With these objectives in mind, the following recommendations are proposed:

1. With the implementation of GIS algorithms and remote sensing procedures, other geomorphological characteristics could be included in the discriminant data set.

2. Further study is needed to interpret the extreme flow response of the catchments, based on the particular shape of the derived flood frequency curve.

3. No confidences limit assessment has been formulated for discriminant analysis results. Statistical studies considering the effects of the regionalization parameter estimation procedure are needed to develop a distribution of errors related to peak flow estimation.

BIBLIOGRAPHY

Aaron, G., and D. F. Kibler, Critique-Ungaged site flood estimation test for guidelines, discussion paper presented at Annual Meeting, AGU, San Francisco, California, 1979.

Baldwin, E. P., and K. W. Potter, Improving flood quantile estimation on ungaged watersheds, Proceedings of the International Symposium on Flood Frequency and Risk Analyses, LSU, Baton Rouge, May 1987.

Benson, M. A., and N. C. Matalas, Synthetic hydrology based on regional statistical parameters, Water Resources Research, 3(4), 931-935, 1967.

Chowdhury, J. U., J. R. Stedinger and L. H. Lu, Goodness-of-Fit tests for regional generalized extreme value flood distributions, Water Resources Research, 27(7), 1765-1776, 1990.

Gingras, D., K. Adamowski and P. J. Pilon, Regional flood equations for the provinces of Ontario and Quebec, Water Resources Bulletin, 30(1), 55-67, 1994.

Greenwood, J. A., J. M. Landwehr, M.C. Matalas, and J. R. Wallis, Probability weighted moments: Definition and relation to parameters of several distributions expressed in inverse form, Water Resources Research, 15(6), 1049-1054, 1979.

Greis, N. P., and E. F. Wood, Regional flood frequency estimation and network design, Water Resources Research, 17(4), 1167-1177, 1981.

Hosking, J. R. M., J. R. Wallis, and E. F. Wood, Estimation of the generalized extreme-value distribution by the method of probability-weighted moments, Technometrics, 27(3), 251- 261, 1985.

Landwehr, J. M., N. C. Matalas, and J. R. Wallis, Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles, Water Resources Research , 15 (5), 1055-1064, 1979.

López, M. A., E. Colón-Dieppa, and E. D. Cobb, Floods in Puerto Rico, magnitude and frequency, U.S.G.S. Water Resources Investigations WRI 78-141, June 1979.

Manly, Bryan F., Multivariate Statistical Methods: A Primer, Chapman and Hall, New York, 1986.

Mimikou, M., Regional treatment of flood data, Proceedings of the International Symposium on Flood Frequency and Risk Analyses, LSU, Baton Rouge, May 1987.

Montgomery D. C., and E. A. Peck, Introduction to Linear Regression Analysis, 2nd ed., John Wiley & Sons, New York, 1992.

Nie, N. H., Hadlai, C., Jenkings, J. G., Steinbrenner, K., Bent, D. H., Statistical Package for Social Sciences, Mc Graw-Hill Inc., New York, N. Y., 1975.

Panu, U. S., and D. A. Smith, Estimating flood flows at ungaged sites in Newfoundland, Proceedings of the VIth IWRA World Congress on Water Resources, June 1988.

Rao, C. R., Linear Statistical Inference and Its Applications, 2nd ed., John Wiley & Sons, New York, 1973.

Ribeiro-Correa, J., and J. Rousselle, A hierarchical and empirical Bayes approach for the regional Pearson type III distribution, Water Resources Research, 29(2), 435-444, 1993.

SAS Institute Inc., SAS/STAT User's Guide Release 6.03 Edition, SAS Institute Inc., Cary, N. C., 1988.

Segarra, R., Improved regionalized flood frequency estimates for Puerto Rico, Final Technical Report, U.S. Department of the Interior, Geological Survey, Puerto Rico Water Resources Research Institute, July 1991.

Sherwood, J. M., Estimation of volume-duration-frequency relations of ungaged small urban streams in Ohio, Water Resources Bulletin, 30(2), 261-269, 1994.

Solomon, S. I., Parameter regionalization and network design, in Stochastic Approaches to Water Resources, Vol. 1, edited by H. W. Shen, Water Resources Publications, Fort Collins, Colorado, 1976.

SPSS Inc., SPSS Reference Guide Release 4.0, SPSS Inc., Chicago, Illinois, 1990.

U. S. Weather Bureau, Generalized Estimates of Probable Maximum Precipitation and Rainfall-Frequency Data for Puerto Rico and Virgin Islands, Technical Paper no. 42, Washington D. C., 1961.

Wiltshire, S. E., Grouping basics for regional flood frequency analysis, Hydrological Sciences-Journal, 30(1, 3), 151-159, 1985.

Wiltshire, S. E., Regional flood frequency analysis I: Homogeneity Statistics, Hydrological Sciences-Journal, 31(3,9), 321-333, 1986a.

Wiltshire, S. E., Regional flood frequency analysis II: Multivariate classification of drainage basins in Britain, Hydrological Sciences-Journal, 31(3, 9), 335-346, 1986b.

Wiltshire, S. E. and M. Beran, Multivariate techniques for the identification of homogeneous flood frequency regions, Proceedings of the International Symposium on Flood Frequency and Risk Analyses, LSU, Baton Rouge, May 1987.

APPENDIX A

SUBROUTINE AND OUTPUT FOR REGRESSION ANALYSIS USING SAS/STAT COMPUTER PACKAGE

A.1 Subroutine for Regression Analysis Using SAS/STAT

```
data ejemplo;
infile 'reg2123.data';
option linesize=80 pagesize=60;
input y a s e sh l sf ap mp ce x5 x25;
a1=LOG(a);
y1=LOG(y);
cards;
proc reg;model y1= a1 x5 x25/r;
```

A.2 Output from Regression Analysis Using SAS/STAT

The SAS System

Model: MODEL1
Dependent Variable: Y1

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	16.85047	5.61682	65.272	0.0001
Error	24	2.06527	0.08605		
C Total	27	18.91573			

Root MSE	0.29335	R-square	0.8908
Dep Mean	9.14416	Adj R-sq	0.8772
C.V.	3.20803		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	4.653941	0.69967934	6.652	0.0001
A1	1	0.760226	0.05869839	12.951	0.0001
X5	1	0.731392	0.18467076	3.961	0.0006
X25	1	-0.311309	0.12859466	-2.421	0.0234

Obs	Dep Var Y1	Predict Value	Std Err Predict	Residual	Std Err Residual	Student Residual
1	10.3903	10.4764	0.114	-0.0861	0.270	-0.319
2	9.2299	9.8078	0.100	-0.5778	0.276	-2.094
3	10.4597	10.4539	0.117	0.00582	0.269	0.022
4	9.8626	9.9809	0.111	-0.1183	0.272	-0.435
5	11.0982	11.1188	0.177	-0.0206	0.234	-0.088

Obs	Dep Var Y1	Predict Value	Std Err Predict	Residual	Std Err Residual	Student Residual
6	7.7411	7.7860	0.124	-0.0449	0.266	-0.169
7	8.9142	8.5930	0.085	0.3213	0.281	1.145
8	9.5166	8.9570	0.081	0.5596	0.282	1.984
9	9.1118	9.1655	0.087	-0.0536	0.280	-0.191
10	7.2626	7.5924	0.159	-0.3298	0.246	-1.338
11	9.0601	9.0828	0.118	-0.0227	0.268	-0.085
12	9.2892	9.2881	0.089	0.0011	0.280	0.004
13	9.6012	9.1818	0.094	0.4194	0.278	1.509
14	7.8419	8.1250	0.123	-0.2831	0.266	-1.063
15	9.1364	8.7367	0.111	0.3996	0.272	1.471
16	8.7063	8.5779	0.105	0.1284	0.274	0.469
17	8.9290	9.1373	0.088	-0.2083	0.280	-0.744
18	9.0875	8.9397	0.127	0.1478	0.265	0.559
19	8.7840	8.8494	0.063	-0.0654	0.286	-0.228
20	10.0908	10.1016	0.146	-0.0108	0.254	-0.043
21	8.6716	8.9759	0.124	-0.3043	0.266	-1.144
22	9.4996	9.1341	0.075	0.3655	0.284	1.289
23	10.1409	9.7758	0.076	0.3651	0.283	1.289
24	8.7564	8.7645	0.096	-0.00815	0.277	-0.029
25	8.9768	8.7494	0.087	0.2274	0.280	0.812
26	8.4785	8.6394	0.090	-0.1609	0.279	-0.576
27	8.8201	9.2682	0.160	-0.4481	0.246	-1.822
28	8.5790	8.7771	0.081	-0.1981	0.282	-0.703

```
                              Cook's
Obs    -2  -1-0  1  2           D

  1    |         |      |     0.005
  2    |     ****|      |     0.143
  3    |         |      |     0.000
  4    |         |      |     0.008
  5    |         |      |     0.001
  6    |         |      |     0.002
  7    |         |**    |     0.030
  8    |         |***   |     0.081
  9    |         |      |     0.001
 10    |       **|      |     0.187
 11    |         |      |     0.000
 12    |         |      |     0.000
```

```
                          Cook's
Obs     -2 -1 -0 1 2        D

 13     |       |***  |   0.065
 14     |     **|     |   0.060
 15     |       |**   |   0.090
 16     |       |     |   0.008
 17     |      *|     |   0.014
 18     |       |*    |   0.018
 19     |       |     |   0.001
 20     |       |     |   0.000
 21     |     **|     |   0.070
 22     |       |**   |   0.029
 23     |       |**   |   0.030
 24     |       |     |   0.000
 25     |       |*    |   0.016
 26     |      *|     |   0.009
 27     |    ***|     |   0.351
 28     |      *|     |   0.010

Sum of Residuals                    0
Sum of Squared Residuals       2.0653
Predicted Resid SS (Press)     2.8278
```

APPENDIX B

DATA, SUBROUTINE AND OUTPUT FOR DISCRIMINANT ANALYSIS

USING SPSS COMPUTER PACKAGE

B.1 Discriminant Analysis data for SPSS computer package

Obs	Cl	Area	SL	ET	SH	Lake	SF	ANP	MOP	CE	X5	X25
01	0	01741	00255	44	250	000	126	060	080	350	65	097
02	0	02592	00432	55	182	000	069	130	120	300	88	120
03	0	12484	00236	43	204	000	051	084	110	350	71	100
04	0	01410	00194	54	208	000	064	076	079	100	68	095
05	0	01933	00293	56	296	000	078	100	110	150	80	115
06	0	04040	00313	54	316	082	025	070	103	300	74	100
07	0	01020	00182	45	185	000	078	090	090	200	75	108
08	0	01700	00301	53	404	000	018	080	090	250	68	098
09	0	00226	00150	60	137	000	133	080	090	150	85	115
10	0	04415	00323	49	180	000	102	070	100	400	90	130
11	0	01230	00446	45	102	000	122	090	100	400	68	095
12	0	12000	00129	56	194	000	065	080	110	300	80	113
13	0	03400	00193	65	281	000	059	095	110	250	73	105
14	0	01270	00832	55	192	000	039	088	090	600	85	120
15	0	01206	00258	54	178	000	091	067	073	100	68	090
16	0	01185	00328	57	225	000	076	040	070	100	64	093
17	0	00997	00703	55	191	000	100	074	095	250	66	098
18	0	01670	00224	57	162	000	174	065	070	200	67	095
19	0	01270	00328	67	283	000	150	090	105	025	74	105
20	0	00951	00275	48	656	000	116	070	080	250	69	097
21	0	01010	00438	45	150	000	307	080	100	700	74	108
22	0	01540	00115	57	237	000	123	083	071	025	70	090
23	0	00499	00797	58	270	000	060	100	100	050	87	120
24	0	03463	00326	50	282	017	095	105	115	500	74	100
25	0	00740	00343	50	173	000	189	090	110	200	74	100
26	0	00529	00324	50	362	000	057	087	106	100	68	100
27	0	00915	00286	42	090	000	361	054	080	300	65	103
28	0	00566	00412	51	325	000	124	080	110	450	90	120
29	0	00670	00215	55	289	000	164	103	130	100	72	100
30	0	04550	00271	57	249	031	079	082	090	300	78	115
31	0	03454	00330	55	145	000	119	070	100	400	65	088
32	0	00554	00434	55	258	000	253	080	080	360	70	100
33	2	19700	01246	55	120	100	047	070	070	250	63	090
34	2	09910	02083	55	175	000	033	065	070	100	61	090
35	1	20800	00997	50	357	300	076	068	080	250	60	085
36	1	07190	01439	55	213	050	065	070	070	050	65	086
37	2	20900	00877	55	121	300	047	089	120	050	75	099
38	4	00275	03858	50	275	000	036	085	080	050	77	105
39	4	00731	08598	45	450	000	055	093	100	200	80	110

Obs	Cl	Area	SL	ET	SH	Lake	SF	ANP	MOP	CE	X5	X25
40	4	01180	04054	55	300	000	093	080	070	100	80	110
41	4	01490	04902	50	400	000	101	078	100	025	80	109
42	4	00126	00579	45	200	000	079	120	140	650	88	118
43	1	01730	01312	65	175	000	069	085	125	100	82	120
44	1	04600	03117	55	275	000	076	035	100	051	64	095
45	1	04350	03533	55	300	100	053	040	105	100	61	090
46	1	00970	05495	55	420	000	041	037	100	050	63	092
47	1	02560	04950	55	240	000	059	035	110	250	60	089
48	1	01860	04053	55	533	000	027	035	110	200	62	091
49	2	02420	04773	55	400	000	062	045	120	250	75	110
50	1	02080	06250	55	289	000	087	040	080	050	76	115
51	3	01640	03763	50	450	000	079	075	110	050	70	098
52	1	09430	02038	45	380	100	059	100	130	150	60	077
53	3	01670	01327	45	400	000	228	090	100	050	60	075
54	3	01840	04491	50	300	000	033	090	080	100	65	085
55	1	20000	01437	55	380	200	055	075	060	150	60	075
56	4	01640	02206	45	220	000	055	070	100	050	79	110
57	2	06020	00606	45	225	000	066	075	115	050	70	100
58	4	00984	05128	45	400	000	092	100	125	050	75	100
59	4	00862	05952	40	156	000	093	095	080	150	80	109
60	1	01780	11143	55	233	000	056	040	110	150	65	095
61	4	01540	06173	45	267	000	020	085	100	100	90	130
62	3	01830	05089	60	214	000	055	080	090	200	70	103

B.2 Subroutine for Discriminant Analysis Using SPSS computer package

```
set printback = yes
file handle datos/name='riverdat.'
data list file=datos/grupo 5 a 9-13(z,2) b 17-21(z,2) c 26-27(z,0)
                          d 31-33(z,2) e 38-40(z,2) f 45-47(z,2)
                          g 52-54(z,0) ar 59-61(z,1) sl 65-67 x5 69-70(z,1)
                          t 72-74(z,1)
discriminant groups = grupo(1,4)
      /variables=a b c d e f g ar sl x5 t
      /method=maxminf
      /priors=equal/save= probs=prb class=prdclas
      /classify=unclassified
      /statistics=table coeff
      /plot=cases
save outfile=probab.out
```

B.3 Output of Discriminant Analysis Using SPSS computer package

<FF>12-Feb-95 SPSS RELEASE 4.1 FOR VAX/VMS

VAX U.P.R. MAYAGUEZ CAMPUS License Number 19824
This software is functional through June 30, 1995.

Try the new SPSS Release 4.1/4.0 features:

* LOGISTIC REGRESSION procedure
* EXAMINE procedure to explore data
* FLIP to transpose data files
* MATRIX Transformations Language
* ALL-IN-1 Interface To SPSS
* CATEGORIES Option:
* conjoint analysis
* correspondence analysis
* GRAPH interface to SPSS Graph
* LISREL7/PRELIS procedure

See the new SPSS documentation for more information on these new features.

```
 1 0   set printback = yes
 2 0   file handle datos/name='riverdat.'
 3 0   data list file=datos/grupo 5 a 9-13(z,2) b 17-21(z,2) c 26-27(z,0)
 4 0             d 31-33(z,2) e 38-40(z,2) f 45-47(z,2)
 5 0             g 52-54(z,0) ar 59-61(z,1) sl 65-67 x5 69-70(z,1)
 6 0             t 72-74(z,1)
```

This command will read 1 records from 1DIA4:[841878395]RIVERDAT.;

Variable	Rec	Start	End	Format
GRUPO	1	5	5	F1.0
A	1	9	13	Z5.2
B	1	17	21	Z5.2
C	1	26	27	Z2.0
D	1	31	33	Z3.2
E	1	38	40	Z3.2
F	1	45	47	Z3.2
G	1	52	54	Z3.0
AR	1	59	61	Z3.1
SL	1	65	67	F3.0
X5	1	69	70	Z2.1
T	1	72	74	Z3.1

```
 7 0   discriminant groups = grupo(1,4)
 8 0      /variables=a b c d e f g ar sl x5 t
 9 0      /method=maxminf
10 0      /priors=equal/save= probs=prb class=prdclas
11 0      /classify=unclassified
12 0      /statistics=table coeff
13 0      /plot=cases
```

There are 13,771,008 bytes of memory available.

SINCE ANALYSIS= WAS OMITTED FOR THE FIRST ANALYSIS ALL VARIABLES
ON THE VARIABLES= LIST WILL BE ENTERED AT LEVEL 1.

```
<FF>12-Feb-95  SPSS RELEASE 4.1 FOR VAX/VMS

FOLLOWING VARIABLES HAVE BEEN CREATED:

 NAME       LABEL

--------    ---------------------------------------

PRDCLAS     --- PREDICTED GROUP FOR ANALYSIS   1

PRB1        --- PROBABILITY   1 FOR ANALYSIS   1

PRB2        --- PROBABILITY   2 FOR ANALYSIS   1

PRB3        --- PROBABILITY   3 FOR ANALYSIS   1

PRB4        --- PROBABILITY   4 FOR ANALYSIS   1



-------------------- D I S C R I M I N A N T   A N A L Y S I S ---------------

ON GROUPS DEFINED BY GRUPO


     62 (UNWEIGHTED) CASES WERE PROCESSED.
     32 OF THESE WERE EXCLUDED FROM THE ANALYSIS.
     32 HAD MISSING OR OUT-OF-RANGE GROUP CODES.
     30 (UNWEIGHTED) CASES WILL BE USED IN THE ANALYSIS.


NUMBER OF CASES BY GROUP

               NUMBER OF CASES
 GRUPO   UNWEIGHTED    WEIGHTED  LABEL

    1          12          12.0
    2           5           5.0
    3           4           4.0
    4           9           9.0

   TOTAL       30          30.0



-------------------- D I S C R I M I N A N T   A N A L Y S I S ---------------

ON GROUPS DEFINED BY GRUPO


ANALYSIS NUMBER      1
```

<FF>12-Feb-95 SPSS RELEASE 4.1 FOR VAX/VMS

VAX U.P.R. MAYAGUEZ CAMPUS License Number 19824

STEPWISE VARIABLE SELECTION

SELECTION RULE: MAXIMIZE MINIMUM F BETWEEN GROUPS
MAXIMUM NUMBER OF STEPS 22
MINIMUM TOLERANCE LEVEL.................. 0.00100
MINIMUM F TO ENTER 1.0000
MAXIMUM F TO REMOVE 1.0000

CANONICAL DISCRIMINANT FUNCTIONS

MAXIMUM NUMBER OF FUNCTIONS 3
MINIMUM CUMULATIVE PERCENT OF VARIANCE... 100.00
MAXIMUM SIGNIFICANCE OF WILKS' LAMBDA 1.0000

PRIOR PROBABILITY FOR EACH GROUP IS 0.25000

--------------- VARIABLES NOT IN THE ANALYSIS AFTER STEP 0 ---------------

VARIABLE	TOLERANCE	MINIMUM TOLERANCE	F TO ENTER	MINIMUM F	BETWEEN	GROUPS
A	1.0000000	1.0000000	4.7495	0.5197818E-01	3	4
B	1.0000000	1.0000000	1.3011	0.1067879E-01	1	3
C	1.0000000	1.0000000	5.0914	0.3095238	2	3
D	1.0000000	1.0000000	1.5871	0.1696962	1	3
E	1.0000000	1.0000000	1.7367	0.0000000E+00	3	4
F	1.0000000	1.0000000	1.5220	0.2336165	1	2
G	1.0000000	1.0000000	6.7023	0.2744611	3	4
AR	1.0000000	1.0000000	0.40517E-01			
SL	1.0000000	1.0000000	0.16929			
X5	1.0000000	1.0000000	13.066	0.1611810	1	3
T	1.0000000	1.0000000	5.7164	0.1204583	1	3

****

AT STEP 1, C WAS INCLUDED IN THE ANALYSIS.

		DEGREES OF FREEDOM			SIGNIF.	BETWEEN GROUPS	
WILKS' LAMBDA	0.62994	1	3	26.0			
EQUIVALENT F	5.09135		3	26.0	0.0066		
MINIMUM F	0.309524		1	26.0	0.5827	2	3

****

AT STEP 1, C WAS INCLUDED IN THE ANALYSIS.

		DEGREES OF FREEDOM			SIGNIF.	BETWEEN GROUPS	
WILKS' LAMBDA	0.62994	1	3	26.0			
EQUIVALENT F	5.09135		3	26.0	0.0066		
MINIMUM F	0.309524		1	26.0	0.5827	2	3

<FF>12-Feb-95 SPSS RELEASE 4.1 FOR VAX/VMS

--------------- VARIABLES IN THE ANALYSIS AFTER STEP 1 ---------------

VARIABLE	TOLERANCE	F TO REMOVE	MINIMUM F	BETWEEN GROUPS	
C	1.0000000	5.0914			

--------------- VARIABLES NOT IN THE ANALYSIS AFTER STEP 1 ---------------

VARIABLE	TOLERANCE	MINIMUM TOLERANCE	F TO ENTER	MINIMUM F	BETWEEN	GROUPS
A	0.9831576	0.9831576	4.3520	1.366283	3	4
B	0.9889848	0.9889848	1.3403	0.7291888	1	3
D	0.9499463	0.9499463	1.7741	0.7382890	1	3
E	0.9806041	0.9806041	1.7423	0.2418008	1	2
F	0.9367357	0.9367357	1.3733	0.4056002	1	2
G	0.9403522	0.9403522	3.0630	0.7370470	2	3
AR	0.9667064	0.9667064	0.11546			
SL	0.9984052	0.9984052	0.12509			
X5	0.8678058	0.8678058	13.969	0.2463872	2	3
T	0.8040885	0.8040885	8.7696	0.4906170	2	3

****

AT STEP 2, A WAS INCLUDED IN THE ANALYSIS.

		DEGREES OF	FREEDOM		SIGNIF.	BETWEEN	GROUPS
WILKS' LAMBDA	0.41382	2	3	26.0			
EQUIVALENT F	4.62089		6	50.0	0.0008		
MINIMUM F	1.36628		2	25.0	0.2734	3	4

--------------- VARIABLES IN THE ANALYSIS AFTER STEP 2 ---------------

VARIABLE	TOLERANCE	F TO REMOVE	MINIMUM F	BETWEEN GROUPS	
A	0.9831576	4.3520			
C	0.9831576	4.6752			

--------------- VARIABLES NOT IN THE ANALYSIS AFTER STEP 2 ---------------

VARIABLE	TOLERANCE	MINIMUM TOLERANCE	F TO ENTER	MINIMUM F	BETWEEN	GROUPS
B	0.8014877	0.7967652	0.29706			
D	0.9293126	0.9255619	1.4428	1.290647	1	3
E	0.1891648	0.1891648	2.3588	0.9200129	3	4
F	0.9359342	0.9201901	1.1953	1.163274	1	2
G	0.7389849	0.7389849	5.6613	0.9184941	3	4
AR	0.7922390	0.7922390	1.0857	0.8823112	3	4
SL	0.9937566	0.9785800	0.11985			
X5	0.8018144	0.8018144	11.456	1.358256	1	3
T	0.6303774	0.6303774	8.2461	1.401276	1	3

<FF>12-Feb-95 SPSS RELEASE 4.1 FOR VAX/VMS

* *

AT STEP 3, T WAS INCLUDED IN THE ANALYSIS.

		DEGREES OF	FREEDOM		SIGNIF.	BETWEEN	GROUPS
WILKS' LAMBDA	0.20378	3	3	26.0			
APPROXIMATE F	6.00230		9	58.6	0.0000		
MINIMUM F	1.40128		3	24.0	0.2667	1	3

--------------- VARIABLES IN THE ANALYSIS AFTER STEP 3 ---------------

VARIABLE	TOLERANCE	F TO REMOVE	MINIMUM F	BETWEEN GROUPS
A	0.7707613	4.0498		
C	0.7961593	7.4183		
T	0.6303774	8.2461		

--------------- VARIABLES NOT IN THE ANALYSIS AFTER STEP 3 ---------------

VARIABLE	TOLERANCE	MINIMUM TOLERANCE	F TO ENTER	MINIMUM F	BETWEEN	GROUPS
B	0.7999333	0.6291549	0.27746			
D	0.8691068	0.5895382	0.42873			
E	0.1777752	0.1559123	2.8310	1.007763	1	3
F	0.9195351	0.6193321	0.60595			
G	0.7256378	0.6020015	4.6267	2.017452	1	2
AR	0.7352811	0.5850565	1.5727	1.324161	1	3
SL	0.9732794	0.6173880	0.36294E-01			
X5	0.1332426	0.1047538	4.9294	2.009756	1	2

* *

AT STEP 4, G WAS INCLUDED IN THE ANALYSIS.

		DEGREES OF	FREEDOM		SIGNIF.	BETWEEN	GROUPS
WILKS' LAMBDA	0.12708	4	3	26.0			
APPROXIMATE F	6.01669		12	61.1	0.0000		
MINIMUM F	2.01745		4	23.0	0.1255	1	2

--------------- VARIABLES IN THE ANALYSIS AFTER STEP 4 ---------------

VARIABLE	TOLERANCE	F TO REMOVE	MINIMUM F	BETWEEN GROUPS
A	0.6020015	6.1526		
C	0.7481041	3.2207		
G	0.7256378	4.6267		
T	0.6189919	6.9527		

<FF>12-Feb-95 SPSS RELEASE 4.1 FOR VAX/VMS

--------------- VARIABLES NOT IN THE ANALYSIS AFTER STEP 4 ---------------

VARIABLE	TOLERANCE	MINIMUM TOLERANCE	F TO ENTER	MINIMUM F	BETWEEN	GROUPS
B	0.6609376	0.5826438	0.57520			
D	0.8210151	0.5863841	0.58801			
E	0.1776505	0.1490511	2.6744	3.163515	1	3
F	0.9089867	0.5921753	0.30255			
AR	0.6107751	0.4800418	3.2332	1.543852	1	2
SL	0.9218246	0.6019781	0.25657			
X5	0.0664263	0.0615681	2.1042	1.575336	1	2

****

AT STEP 5, E WAS INCLUDED IN THE ANALYSIS.

		DEGREES OF		FREEDOM	SIGNIF.	BETWEEN	GROUPS
WILKS' LAMBDA	0.09312	5	3	26.0			
APPROXIMATE F	5.55479		15	61.1	0.0000		
MINIMUM F	3.16352		5	22.0	0.0266	1	3

--------------- VARIABLES IN THE ANALYSIS AFTER STEP 5 ---------------

VARIABLE	TOLERANCE	F TO REMOVE	MINIMUM F	BETWEEN GROUPS
A	0.1490511	6.4241		
C	0.7321993	3.2655		
E	0.1776505	2.6744		
G	0.7251288	4.3862		
T	0.5833052	7.4558		

--------------- VARIABLES NOT IN THE ANALYSIS AFTER STEP 5 ---------------

VARIABLE	TOLERANCE	MINIMUM TOLERANCE	F TO ENTER	MINIMUM F	BETWEEN	GROUPS
B	0.6602578	0.1469845	0.45881			
D	0.7866357	0.1407163	0.21089			
F	0.9063065	0.1472253	0.27004			
AR	0.4700729	0.0944386	5.0014	3.407383	1	2
SL	0.9210451	0.1489315	0.25093			
X5	0.0593758	0.0580952	2.0847	2.520255	1	3

****

AT STEP 6, AR WAS INCLUDED IN THE ANALYSIS.

		DEGREES OF		FREEDOM	SIGNIF.	BETWEEN	GROUPS
WILKS' LAMBDA	0.05432	6	3	26.0			
APPROXIMATE F	5.99065		18	59.9	0.0000		
MINIMUM F	3.40738		6	21.0	0.0166	1	2

<FF>12-Feb-95 SPSS RELEASE 4.1 FOR VAX/VMS

--------------- VARIABLES IN THE ANALYSIS AFTER STEP 6 ---------------

VARIABLE	TOLERANCE	F TO REMOVE	MINIMUM F	BETWEEN GROUPS
A	0.0944386	11.126		
C	0.6933587	3.6746		
E	0.1367258	4.3667		
G	0.5801122	6.9204		
AR	0.4700729	5.0014		
T	0.5726607	7.2811		

--------------- VARIABLES NOT IN THE ANALYSIS AFTER STEP 6 ---------------

VARIABLE	TOLERANCE	MINIMUM TOLERANCE	F TO ENTER	MINIMUM F	BETWEEN	GROUPS
B	0.6554316	0.0924497	0.13409			
D	0.7460583	0.0941246	0.54544			
F	0.9058337	0.0939126	0.26025			
SL	0.8512112	0.0909159	0.20469			
X5	0.0591893	0.0575608	1.6194	2.864146	1	2

****

AT STEP 7, X5 WAS INCLUDED IN THE ANALYSIS.

		DEGREES OF		FREEDOM	SIGNIF.	BETWEEN	GROUPS
WILKS' LAMBDA	0.04370	7	3	26.0			
APPROXIMATE F	5.45225		21	58.0	0.0000		
MINIMUM F	2.86415		7	20.0	0.0305	1	2

--------------- VARIABLES IN THE ANALYSIS AFTER STEP 7 ---------------

VARIABLE	TOLERANCE	F TO REMOVE	MINIMUM F	BETWEEN	GROUPS
A	0.0888985	9.4803			
C	0.6917099	2.9181			
E	0.1233189	3.7713			
G	0.2998137	2.5611			
AR	0.4685967	4.2796			
X5	0.0591893	1.6194			
T	0.0575608	0.40497	3.507929	1	2

--------------- VARIABLES NOT IN THE ANALYSIS AFTER STEP 7 ---------------

VARIABLE	TOLERANCE	MINIMUM TOLERANCE	F TO ENTER	MINIMUM F	BETWEEN GROUPS
B	0.6511762	0.0568525	0.15713		
D	0.7320148	0.0545739	0.47961		
F	0.8576535	0.0534771	0.47276		
SL	0.8203894	0.0551043	0.29191		

<FF>12-Feb-95 SPSS RELEASE 4.1 FOR VAX/VMS

* *

AT STEP 8, T WAS REMOVED FROM THE ANALYSIS.

		DEGREES OF	FREEDOM		SIGNIF.	BETWEEN	GROUPS
WILKS' LAMBDA	0.04636	6	3	26.0			
APPROXIMATE F	6.52763		18	59.9	0.0000		
MINIMUM F	3.50793		6	21.0	0.0146	1	2

---------------- VARIABLES IN THE ANALYSIS AFTER STEP 8 ----------------

VARIABLE	TOLERANCE	F TO REMOVE	MINIMUM F	BETWEEN GROUPS
A	0.0909243	10.998		
C	0.6996926	3.7648		
E	0.1317997	4.6379		
G	0.5167914	5.5202		
AR	0.4729472	4.9991		
X5	0.5888622	9.7336		

---------------- VARIABLES NOT IN THE ANALYSIS AFTER STEP 8 ----------------

VARIABLE	TOLERANCE	MINIMUM TOLERANCE	F TO ENTER	MINIMUM F	BETWEEN GROUPS
B	0.6511762	0.0568525	0.15713		
B	0.6592887	0.0890156	0.13002		
D	0.7720797	0.0906293	0.56016		
F	0.9231477	0.0906300	0.31861		
SL	0.8569612	0.0881554	0.20391		
T	0.0575608	0.0575608	0.40497		

SUMMARY TABLE

STEP	ACTION ENTERED	ACTION REMOVED	VARS IN	WILKS' LAMBDA	SIG.	MINIMUM F	SIG.	BETWEEN	GROUPS	LABEL
1	C		1	.62994	.0066	.30952	.5827	2	3	
2	A		2	.41382	.0008	1.36628	.2734	3	4	
3	T		3	.20378	.0000	1.40128	.2667	1	3	
4	G		4	.12708	.0000	2.01745	.1255	1	2	
5	E		5	.09312	.0000	3.16352	.0266	1	3	
6	AR		6	.05432	.0000	3.40738	.0166	1	2	
7	X5		7	.04370	.0000	2.86415	.0305	1	2	
8		T	6	.04636	.0000	3.50793	.0146	1	2	

CLASSIFICATION FUNCTION COEFFICIENTS
(FISHER'S LINEAR DISCRIMINANT FUNCTIONS)

GRUPO =	1	2	3	4
A	0.4394749	0.5860006	0.3250350	0.3602794
C	2.455648	2.273153	2.333832	1.704320
E	-19.65037	-27.27453	-16.45773	-17.90372
G	-0.8168025E-01	-0.1377795	0.1278106	0.6871054E-01
AR	5.165937	5.879350	3.848164	3.348077

<FF>12-Feb-95 SPSS RELEASE 4.1 FOR VAX/VMS

GRUPO =	1	2	3	4
X5	12.41028	15.74977	11.57603	18.77521
(CONSTANT)	-139.8112	-163.8022	-126.0031	-138.6844

CANONICAL DISCRIMINANT FUNCTIONS

FUNCTION	EIGEN VALUE	PERCENT OF VARIANCE	CUMULATIVE PERCENT	CANONICAL CORRELATION	:	AFTER FUNCTION	WILK'S LAMBDA	CHI-SQUARED	D.F.	SIGNIF
					:	0	0.0463552	73.714	18	0.0000
1*	5.54238	74.91	74.91	0.9204078	:	1	0.3032737	28.635	10	0.0014
2*	1.57620	21.30	96.22	0.7821969	:	2	0.7812950	5.9233	4	0.2050
3*	0.27993	3.78	100.00	0.4676591	:					

* MARKS THE 3 CANONICAL DISCRIMINANT FUNCTIONS REMAINING IN THE ANALYSIS.

STANDARDIZED CANONICAL DISCRIMINANT FUNCTION COEFFICIENTS

	FUNC 1	FUNC 2	FUNC 3
A	1.55056	2.69942	1.01548
C	0.63148	-0.49215	0.25139
E	-0.75973	-1.84324	-1.44392
G	-0.68877	-0.67860	0.88038
AR	0.93204	0.48123	0.12295
X5	-0.59713	1.02643	-0.43376

STRUCTURE MATRIX:

POOLED WITHIN-GROUPS CORRELATIONS BETWEEN DISCRIMINATING VARIABLES
AND CANONICAL DISCRIMINANT FUNCTIONS
(VARIABLES ORDERED BY SIZE OF CORRELATION WITHIN FUNCTION)

	FUNC 1	FUNC 2	FUNC 3
X5	-0.46940*	0.40883	-0.28678
C	0.31711*	-0.13465	-0.07446
D	0.18978*	-0.14043	-0.16738
E	0.17852*	0.11241	-0.11706
F	-0.09621*	-0.04219	0.07817
A	0.27275	0.28280*	0.18563
G	-0.34301	0.09212	0.62074*
T	-0.33018	0.36798	-0.40682*
B	0.00020	-0.09381	-0.36251*
SL	0.04957	0.13759	0.17098*
AR	-0.00438	0.04238	-0.07880*

CANONICAL DISCRIMINANT FUNCTIONS EVALUATED AT GROUP MEANS (GROUP CENTROIDS)

GROUP	FUNC 1	FUNC 2	FUNC 3
1	1.74615	-0.55304	-0.39445
2	1.98149	1.97805	0.56554
3	-0.78330	-1.96224	0.92850
4	-3.08090	0.51058	-0.20093

<FF>12-Feb-95 SPSS RELEASE 4.1 FOR VAX/VMS

CASE SEQNUM	MIS VAL	SEL	ACTUAL GROUP	HIGHEST PROBABILITY GROUP	P(D/G)	P(G/D)	2ND HIGHEST GROUP	P(G/D)	DISCRIMINANT SCORES...		
1			UNGRPD	3	0.2757	0.6088	4	0.3626	-1.3928	-0.8703	-0.5906
2			UNGRPD	4	0.1504	0.9636	3	0.0363	-2.8265	0.5567	2.0891
3			UNGRPD	2	0.0033	1.0000	1	0.0000	1.2590	5.2209	2.2009
4			UNGRPD	3	0.9440	0.9923	1	0.0042	-1.0716	-2.1871	0.4305
5			UNGRPD	3	0.3790	0.7383	4	0.2532	-1.4004	-0.3275	1.1019
6			UNGRPD	3	0.1567	0.5114	1	0.4435	-0.4389	-1.0056	-1.1164
7			UNGRPD	4	0.7583	0.9714	3	0.0286	-3.1286	-0.5280	0.1106
8			UNGRPD	3	0.9947	0.9871	1	0.0096	-0.8048	-1.8462	0.6827
9			UNGRPD	4	0.5189	0.8284	3	0.1686	-1.9328	-0.4393	-0.4169
10			UNGRPD	4	0.0015	0.9746	2	0.0253	-1.9371	4.1711	-1.0198
11			UNGRPD	3	0.6213	0.8737	4	0.1254	-1.9588	-1.3812	0.7020
12			UNGRPD	2	0.0048	1.0000	1	0.0000	2.1453	5.2809	1.9800
13			UNGRPD	3	0.1666	0.7852	1	0.2108	1.0901	-1.5537	2.1096
14			UNGRPD	4	0.9654	0.9903	3	0.0096	-2.6147	0.2964	-0.1129
15			UNGRPD	3	0.7811	0.9828	1	0.0099	-1.0487	-2.0871	-0.0701
16			UNGRPD	1	0.2477	0.6626	3	0.3372	0.6205	-2.1565	-0.9351
17			UNGRPD	3	0.8472	0.9689	1	0.0309	-0.0976	-2.4011	0.5456
18			UNGRPD	3	0.8327	0.9764	1	0.0228	-0.4715	-2.3379	0.1338
19			UNGRPD	3	0.4404	0.9542	1	0.0458	0.6426	-2.5529	1.4912
20			UNGRPD	3	0.3855	0.7333	4	0.2628	-1.8381	-1.3687	-0.3265
21			UNGRPD	4	0.7750	0.9469	3	0.0525	-2.2327	-0.1121	-0.2412
22			UNGRPD	3	0.8550	0.9970	4	0.0023	-1.4303	-2.5373	0.7622
23			UNGRPD	4	0.7607	0.9372	3	0.0627	-2.6411	-0.2766	0.3949
24			UNGRPD	3	0.4006	0.8601	4	0.1333	-1.3279	-0.4220	1.4512
25			UNGRPD	3	0.5921	0.8046	4	0.1902	-1.5804	-0.9148	0.5115
26			UNGRPD	3	0.9855	0.9924	4	0.0042	-1.1118	-2.0025	0.7331
27			UNGRPD	4	0.1958	0.6803	3	0.3044	-1.6677	-0.8394	-1.1347
28			UNGRPD	4	0.3909	0.9999	3	0.0000	-2.6838	1.9510	-1.0795
29			UNGRPD	3	0.8731	0.9913	1	0.0083	-0.3617	-1.8463	1.6421
30			UNGRPD	3	0.2228	0.5538	4	0.3034	-0.8137	0.0200	0.2541
31			UNGRPD	1	0.4766	0.6777	3	0.3105	1.0463	-1.1223	0.9012
32			UNGRPD	3	0.7860	0.9946	4	0.0043	-1.4769	-2.4956	0.3829

CLASSIFICATION RESULTS -

ACTUAL GROUP	NO. OF CASES	PREDICTED GROUP MEMBERSHIP 1	2	3	4
--------------------	------	--------	--------	--------	--------
GROUP 1	0	0 0.0%	0 0.0%	0 0.0%	0 0.0%
GROUP 2	0	0 0.0%	0 0.0%	0 0.0%	0 0.0%
GROUP 3	0	0 0.0%	0 0.0%	0 0.0%	0 0.0%
GROUP 4	0	0 0.0%	0 0.0%	0 0.0%	0 0.0%
UNGROUPED CASES	32	2 6.3%	2 6.3%	19 59.4%	9 28.1%

<FF>12-Feb-95 SPSS RELEASE 4.1 FOR VAX/VMS

PERCENT OF "GROUPED" CASES CORRECTLY CLASSIFIED: 0.00%

CLASSIFICATION PROCESSING SUMMARY

62 CASES WERE PROCESSED.
0 CASES HAD AT LEAST ONE MISSING DISCRIMINATING VARIABLE.
32 CASES WERE USED FOR PRINTED OUTPUT.
62 CASES WERE WRITTEN INTO THE ACTIVE FILE.

CLASSIFICATION RESULTS -

ACTUAL GROUP	NO. OF CASES	PREDICTED GROUP MEMBERSHIP 1	2	3	4
GROUP 1	0	0	0	0	0
		0.0%	0.0%	0.0%	0.0%
GROUP 2	0	0	0	0	0
		0.0%	0.0%	0.0%	0.0%
GROUP 3	0	0	0	0	0
		0.0%	0.0%	0.0%	0.0%
GROUP 4	0	0	0	0	0
		0.0%	0.0%	0.0%	0.0%
UNGROUPED CASES	32	2	2	19	9
		6.3%	6.3%	59.4%	28.1%

PERCENT OF "GROUPED" CASES CORRECTLY CLASSIFIED: 0.00%

CLASSIFICATION PROCESSING SUMMARY

62 CASES WERE PROCESSED.
0 CASES HAD AT LEAST ONE MISSING DISCRIMINATING VARIABLE.
32 CASES WERE USED FOR PRINTED OUTPUT.
62 CASES WERE WRITTEN INTO THE ACTIVE FILE.

APPENDIX C

FLOOD FREQUENCY CURVES FOR THE 31 UNGAGED BASINS

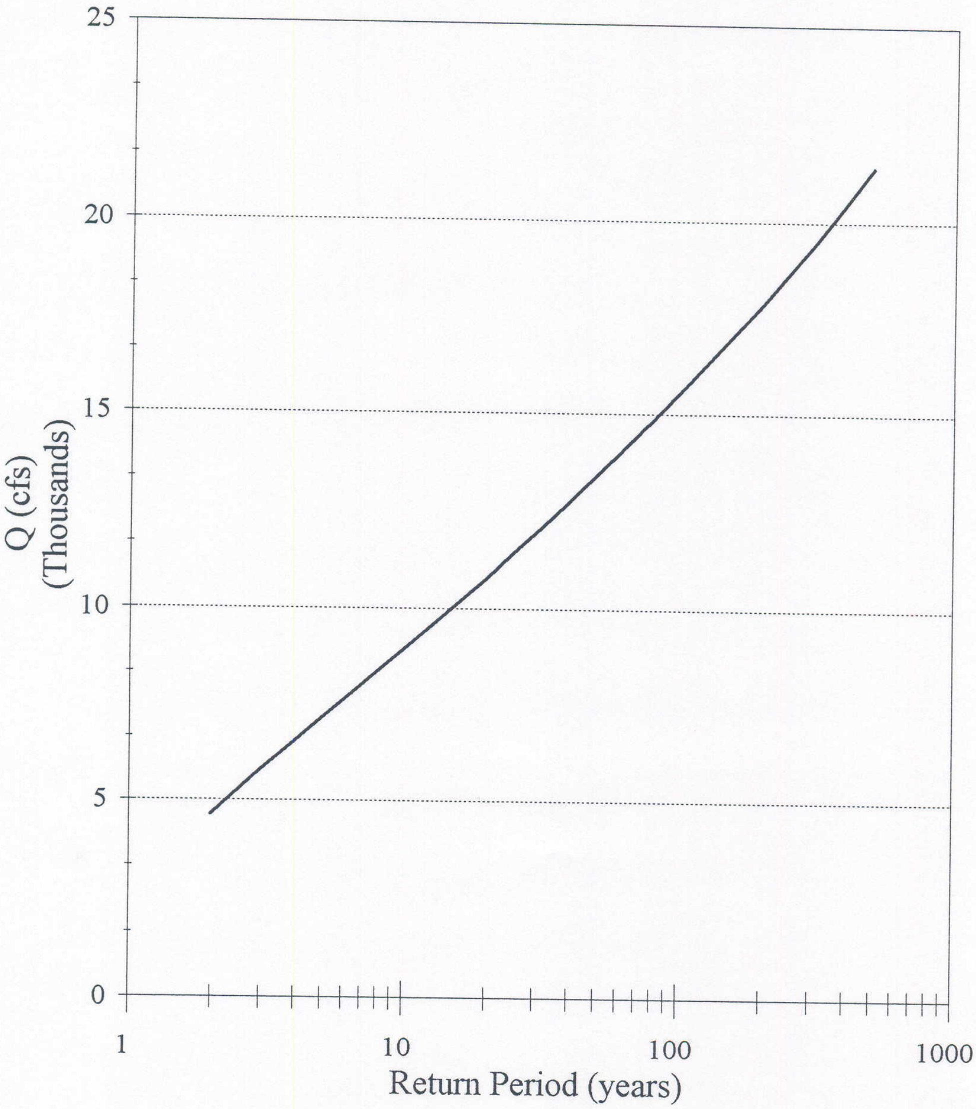

Figure C-1. Flood Frequency Curve for Río Arroyata basin

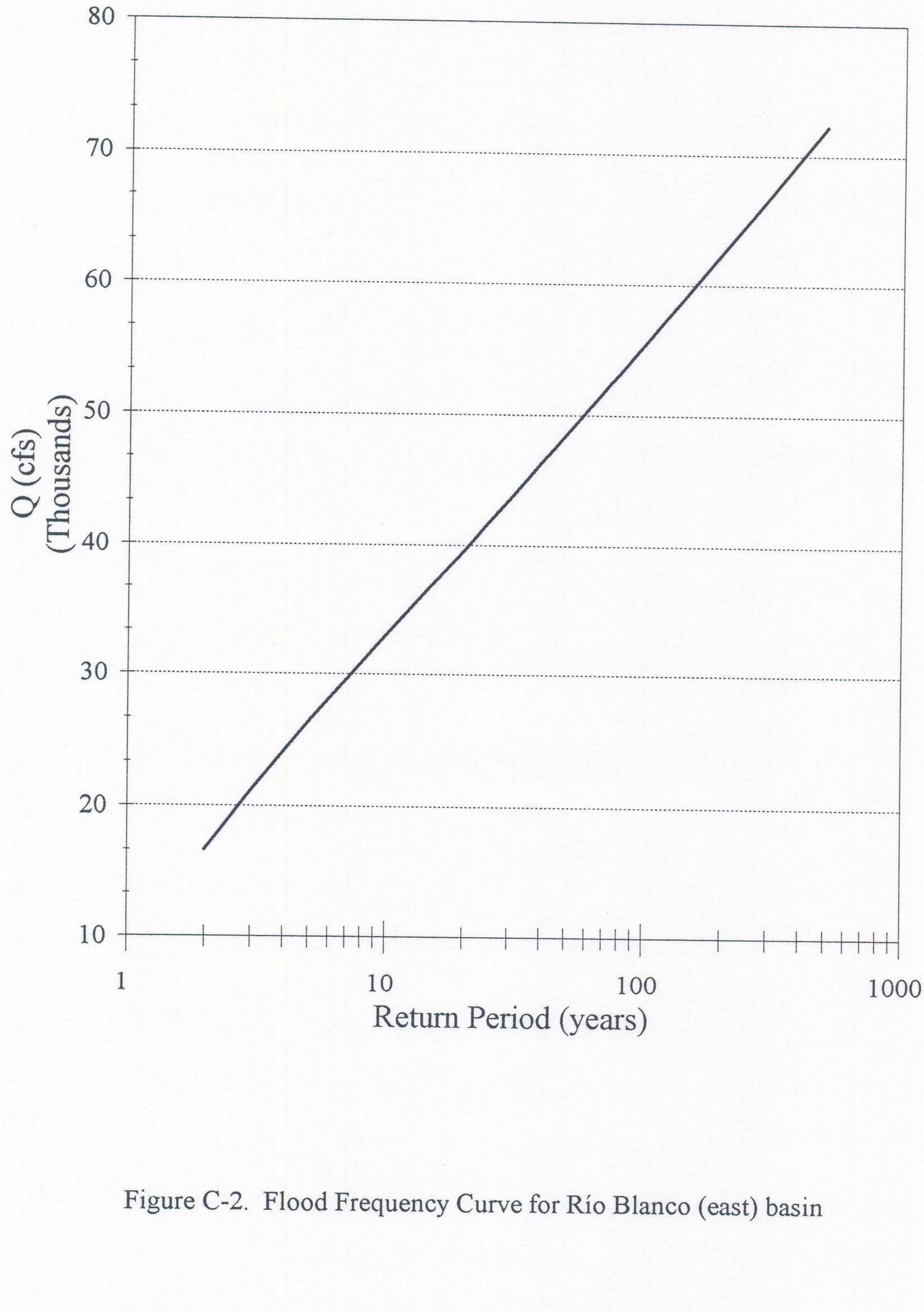

Figure C-2. Flood Frequency Curve for Río Blanco (east) basin

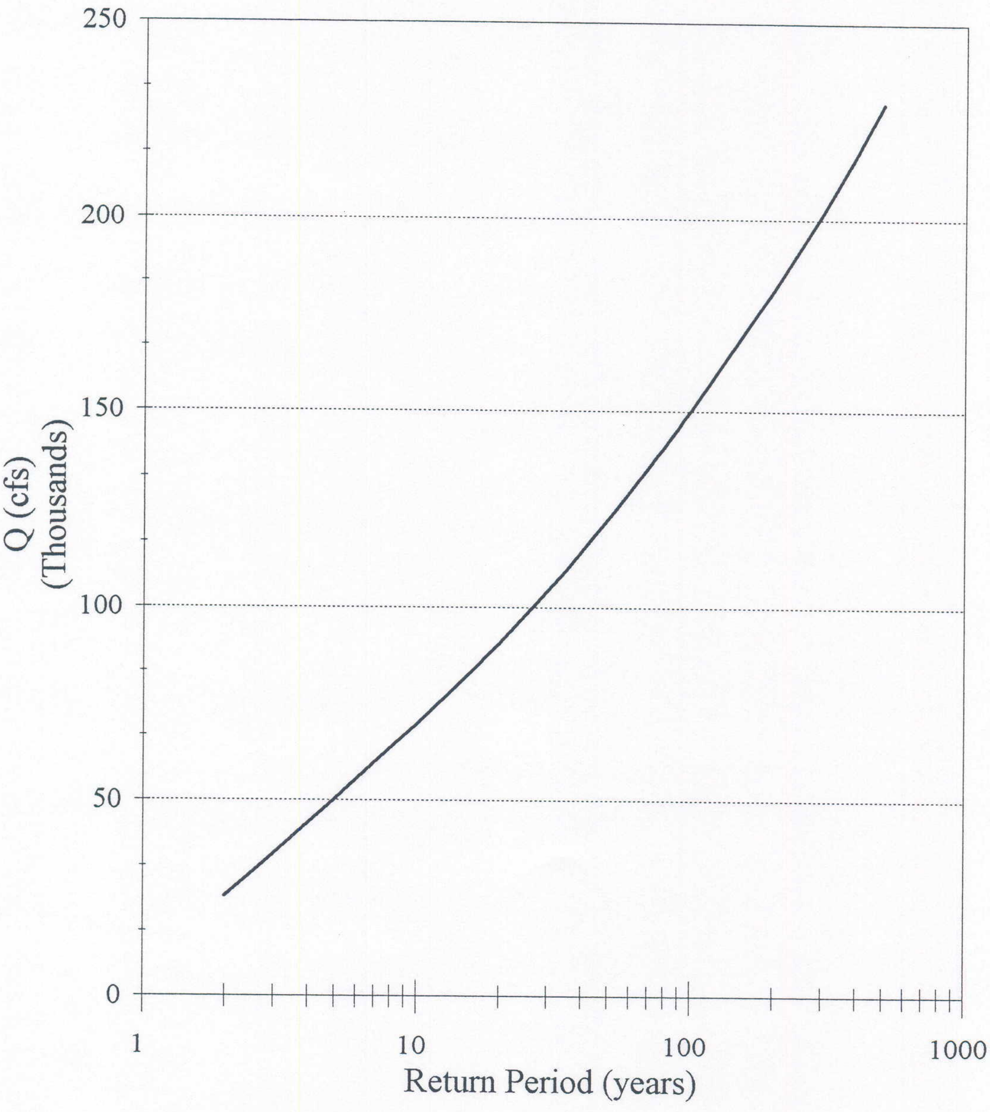

Figure C-3. Flood Frequency Curve for Río Blanco (west) basin

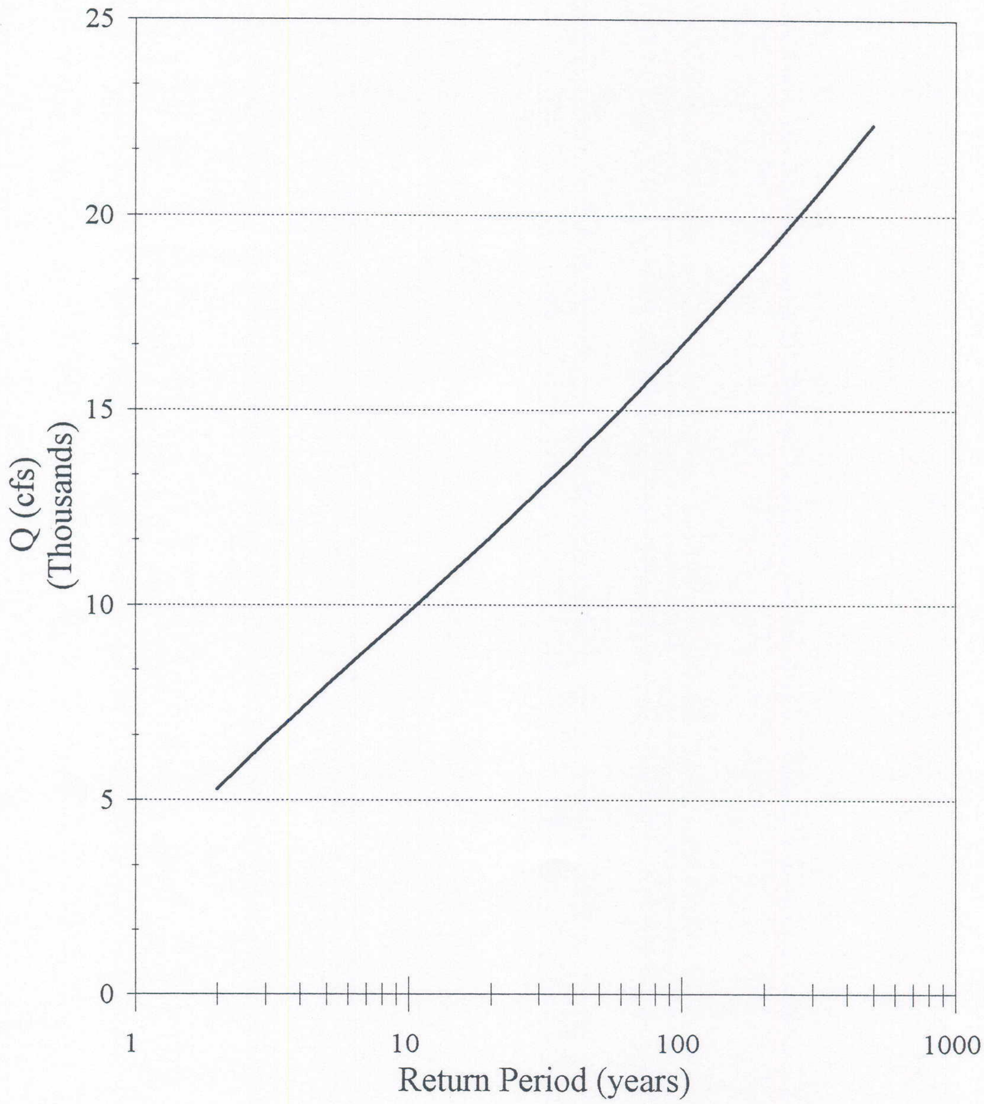

Figure C-4. Flood Frequency Curve for Río Cagüitas basin

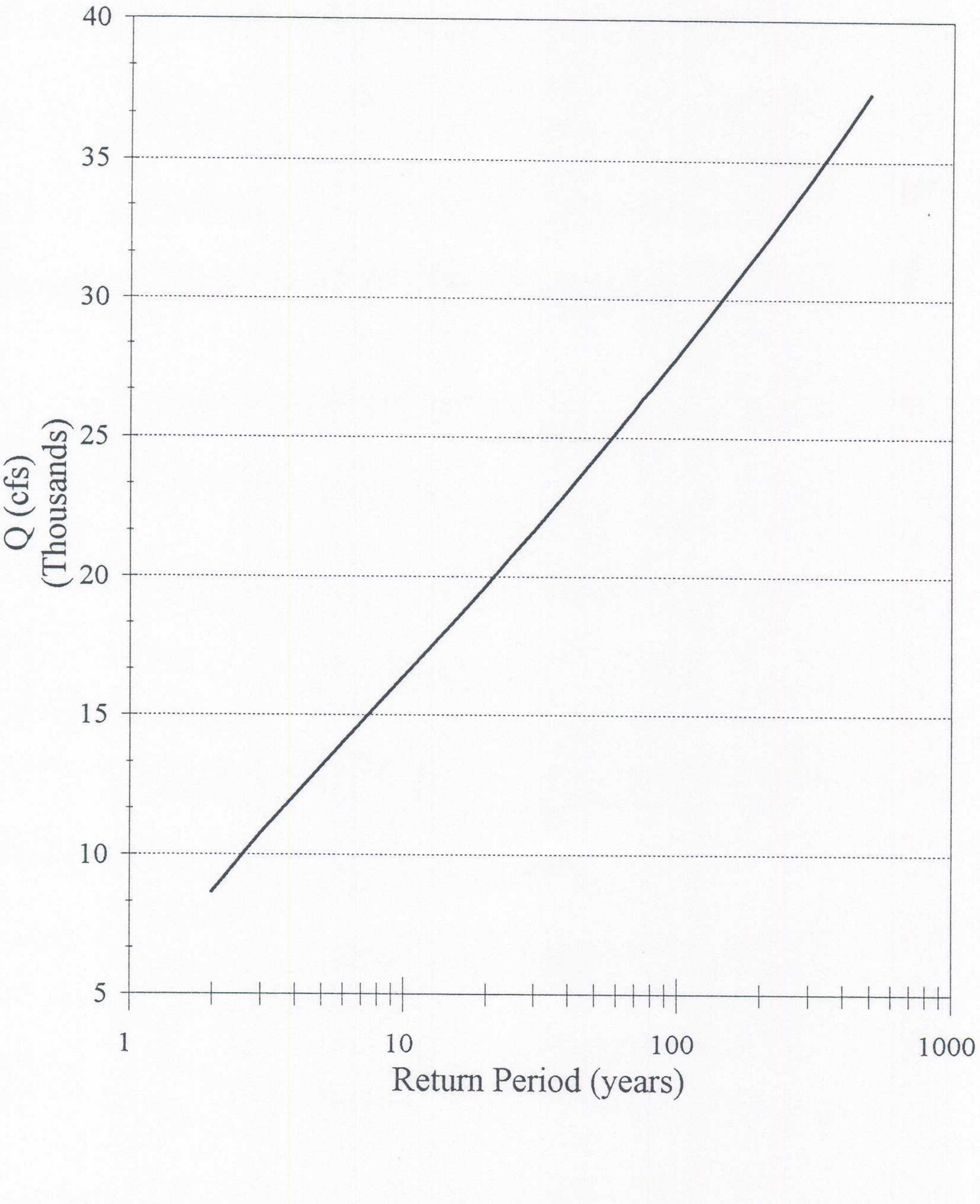

Figure C-5. Flood Frequency Curve for Río Canovanillas basin

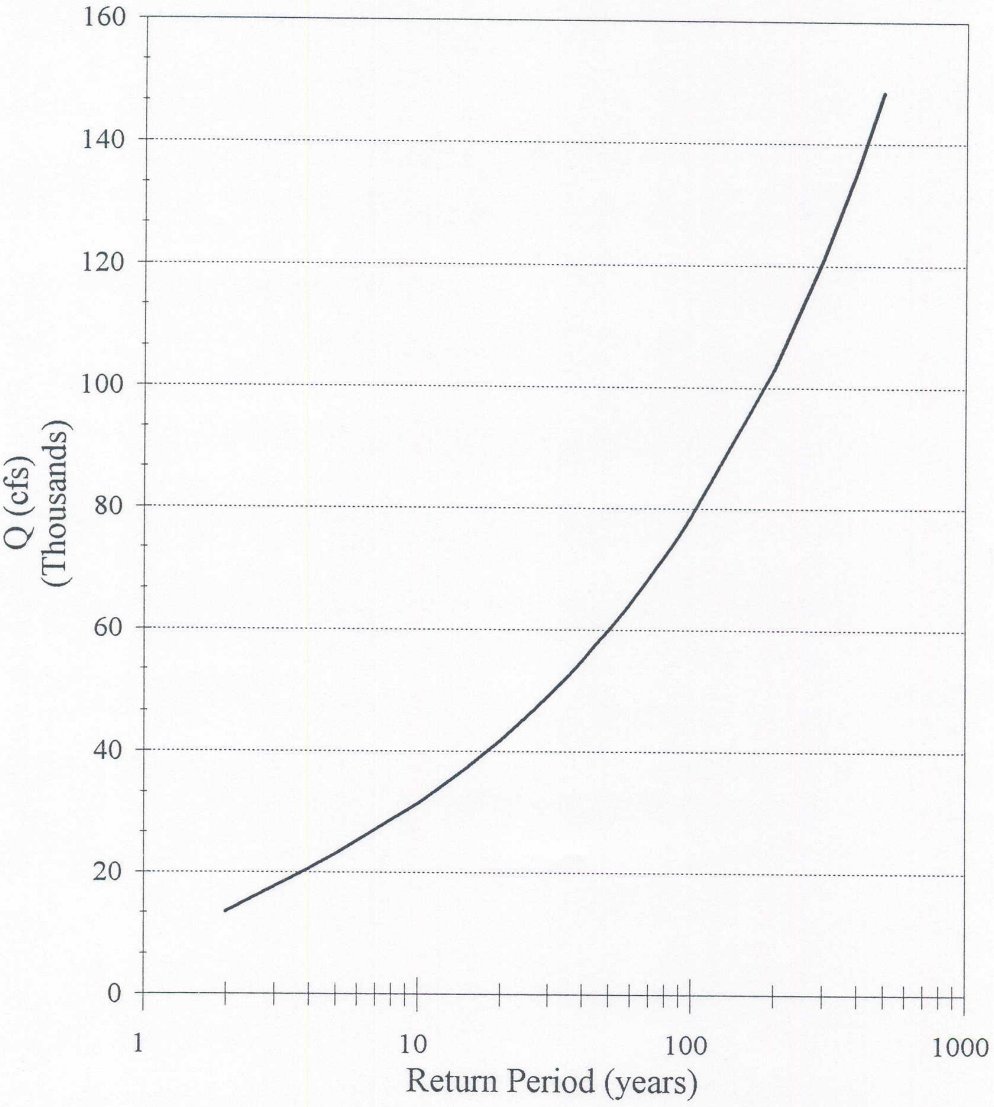

Figure C-6. Flood Frequency Curve for Río Caonillas basin

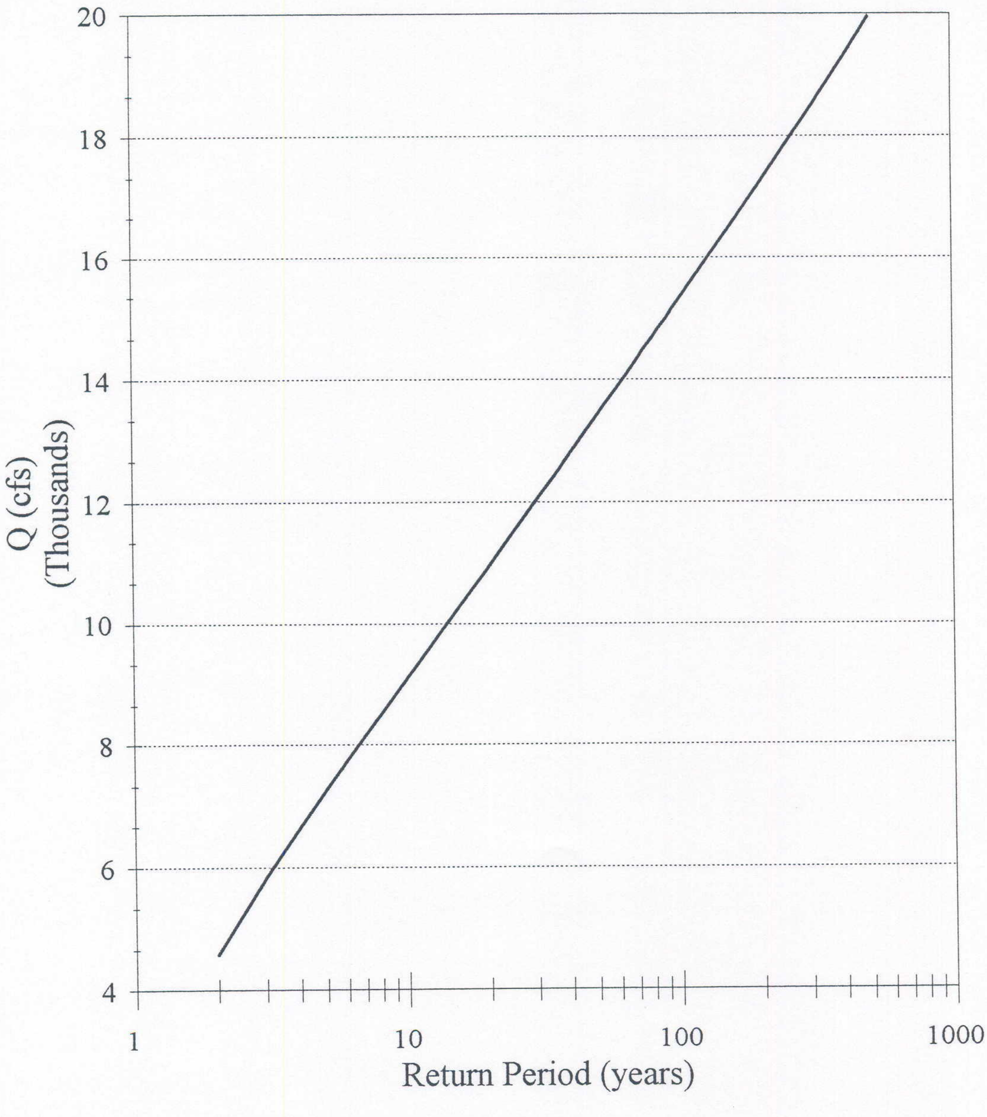

Figure C-7. Flood Frequency Curve for Río Cayaguas basin

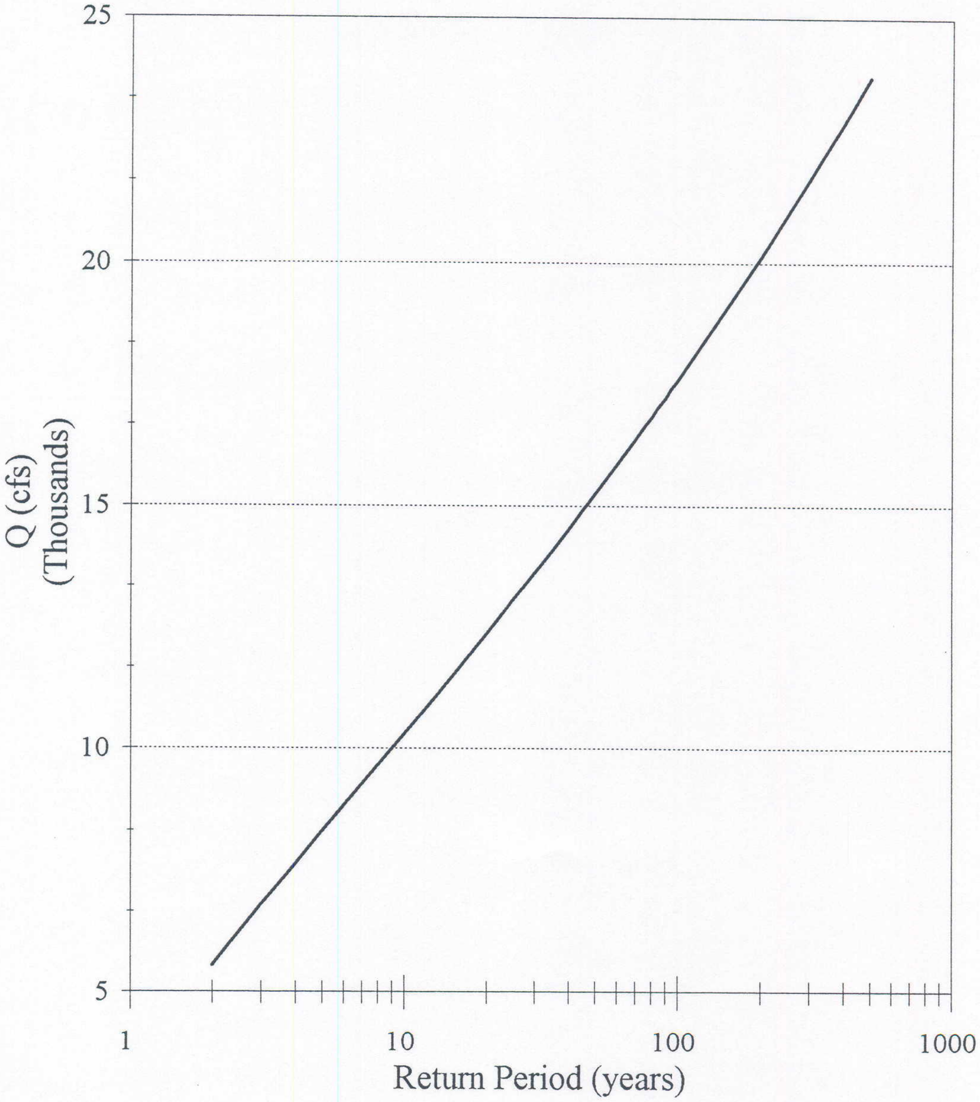

Figure C-8. Flood Frequency Curve for Río Cialitos basin

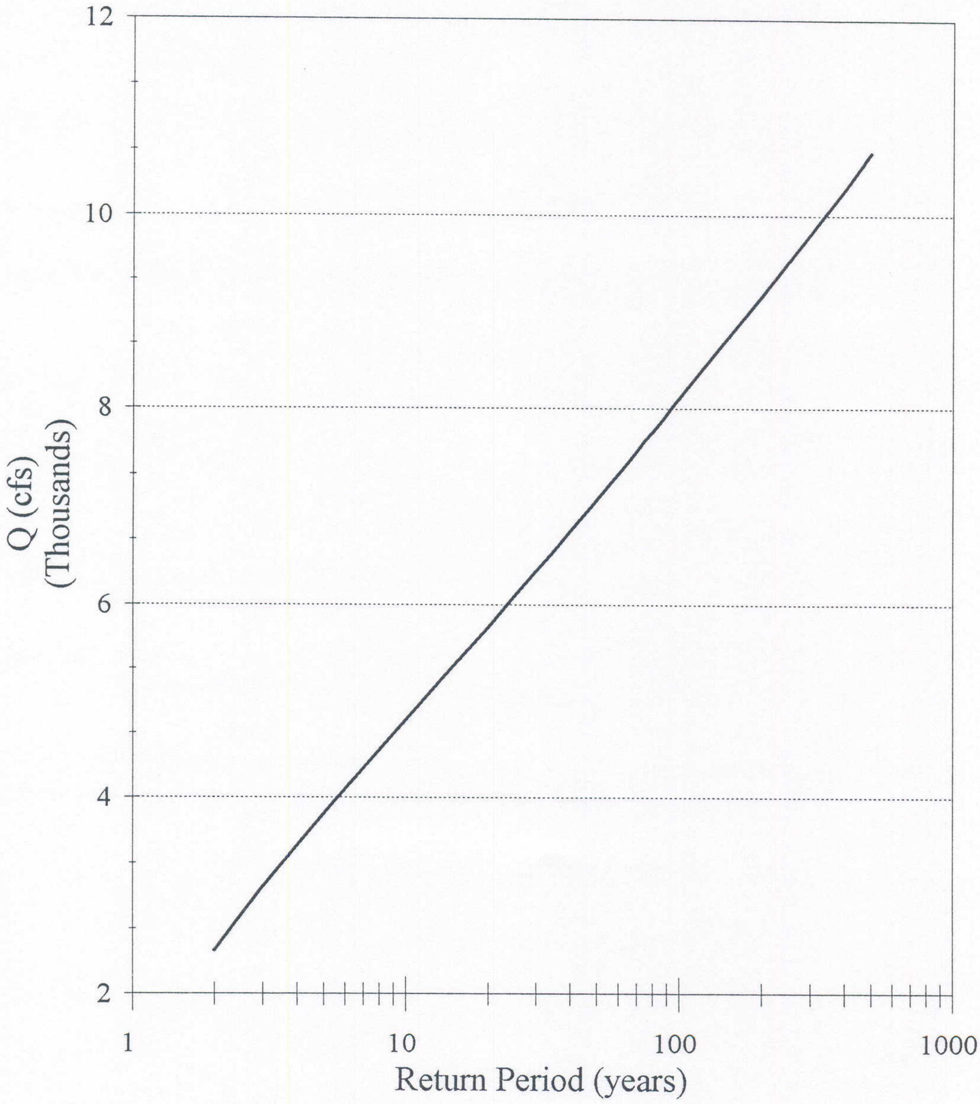

Figure C-9. Flood Frequency Curve for Río Daguao basin

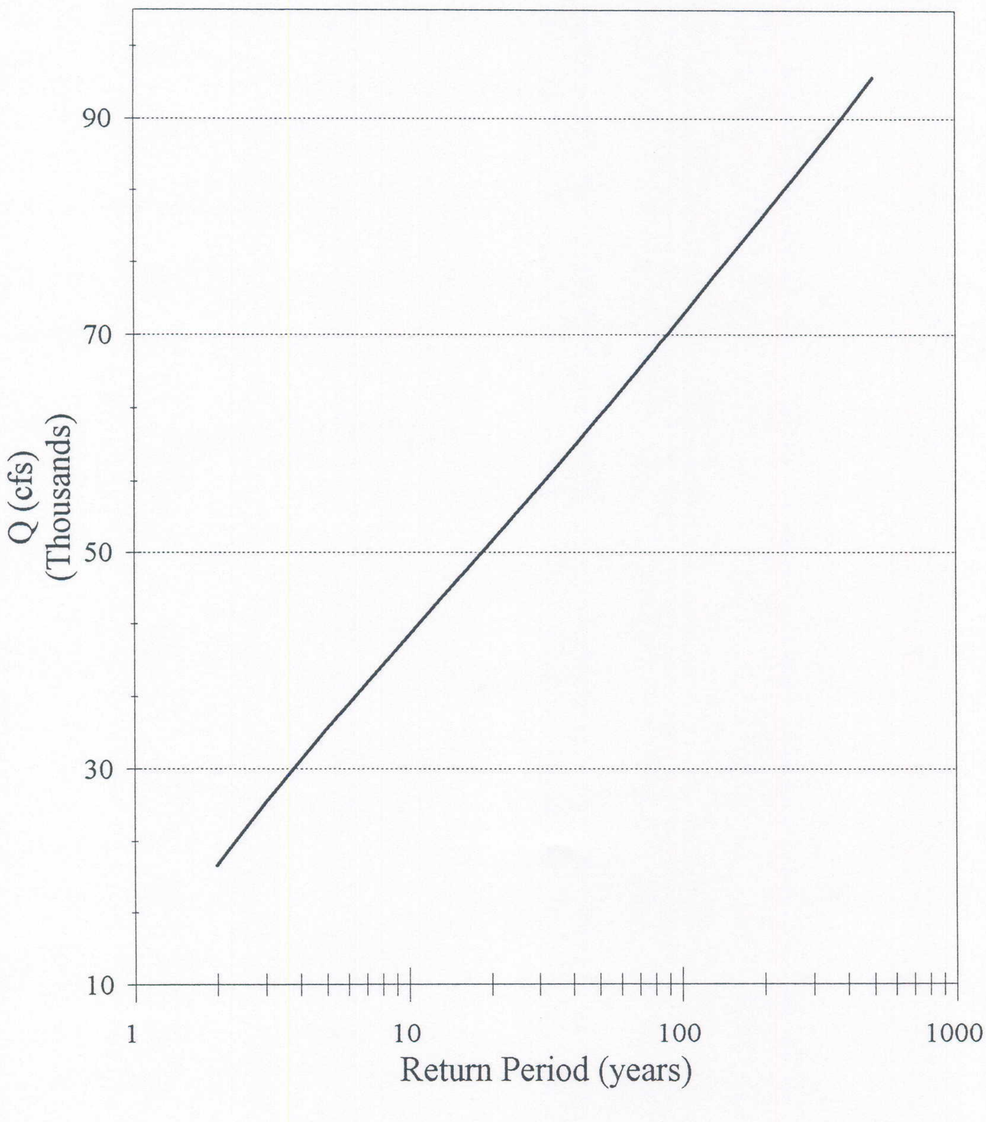

Figure C-10. Flood Frequency Curve for Río Grande de Jayuya basin

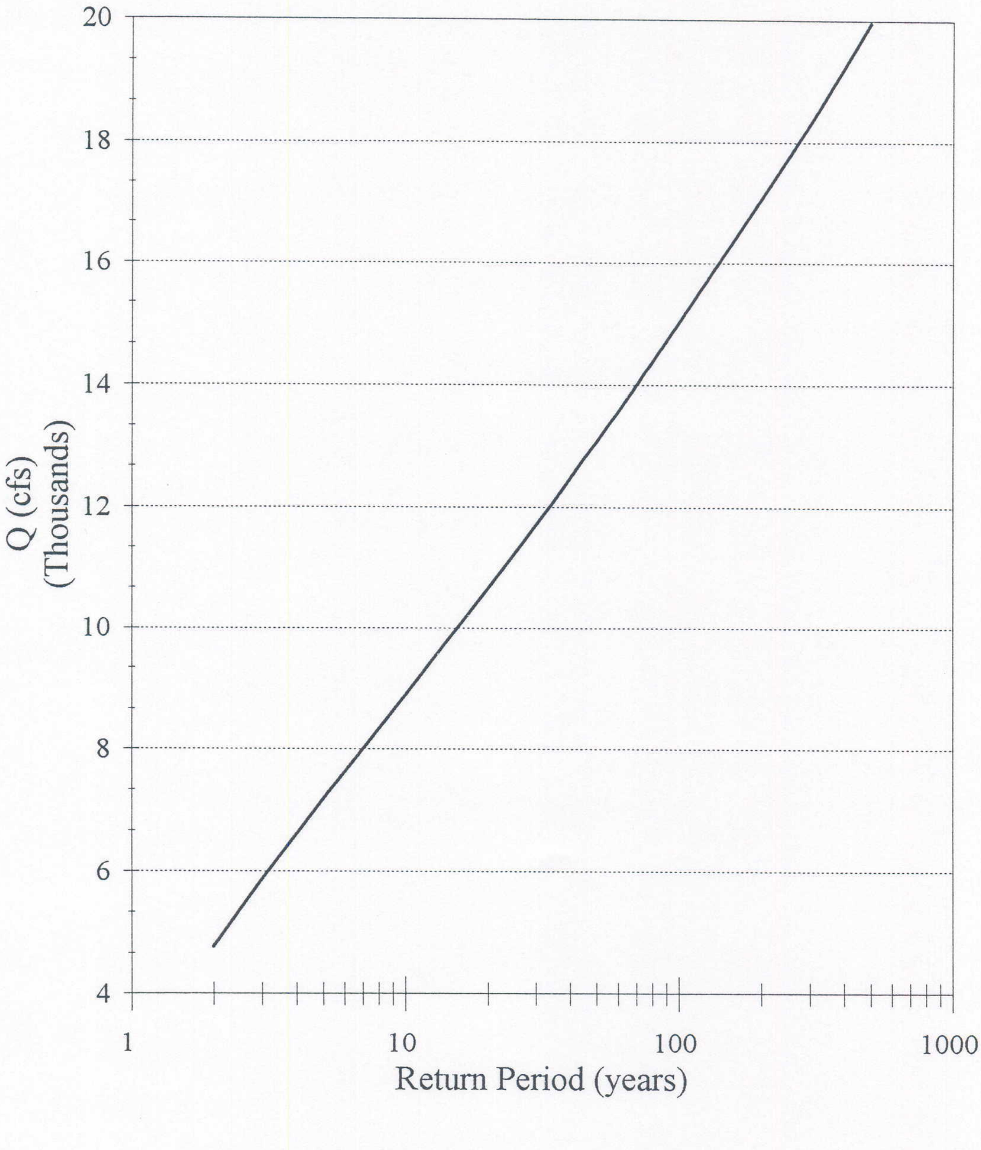

Figure C-11. Flood Frequency Curve for Río Guamaní basin

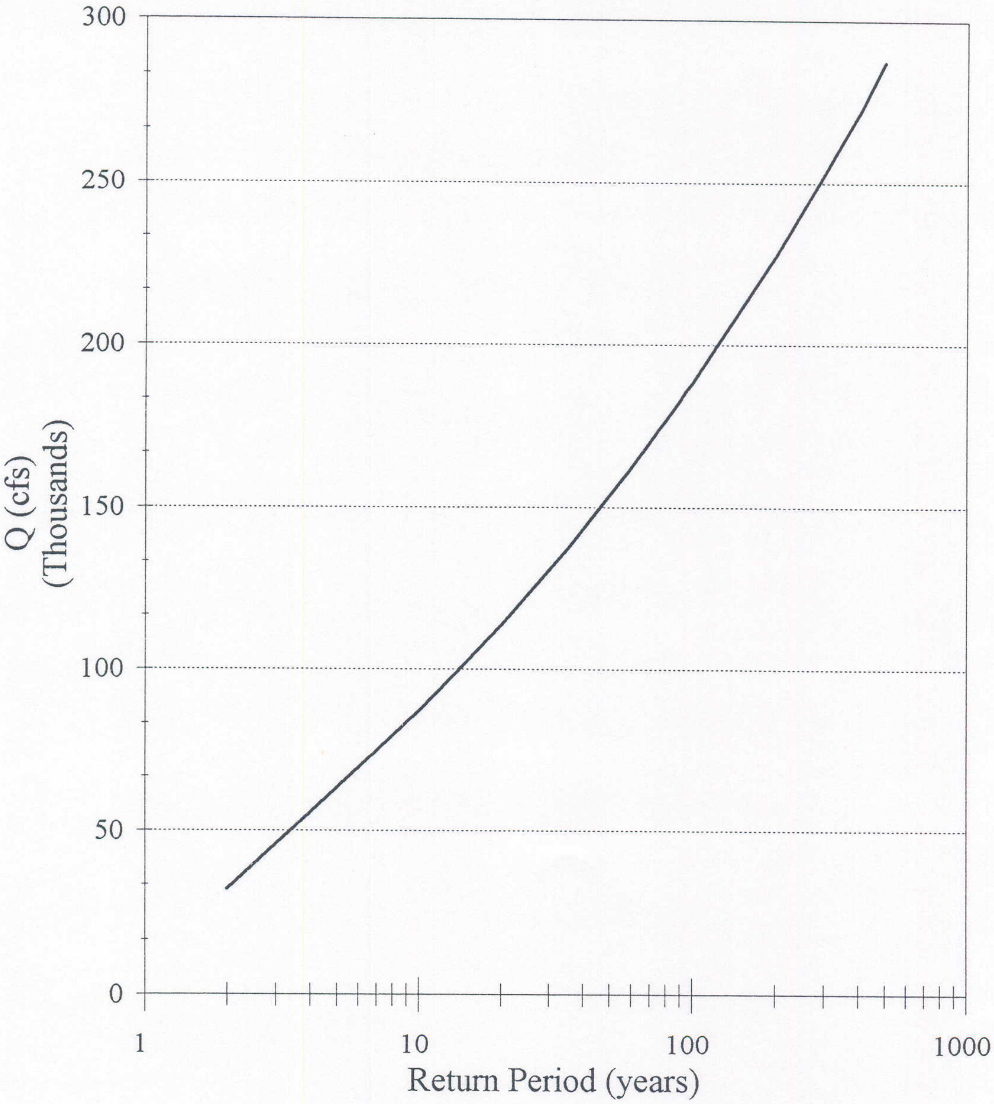

Figure C-12. Flood Frequency Curve for Río Guanajibo basin

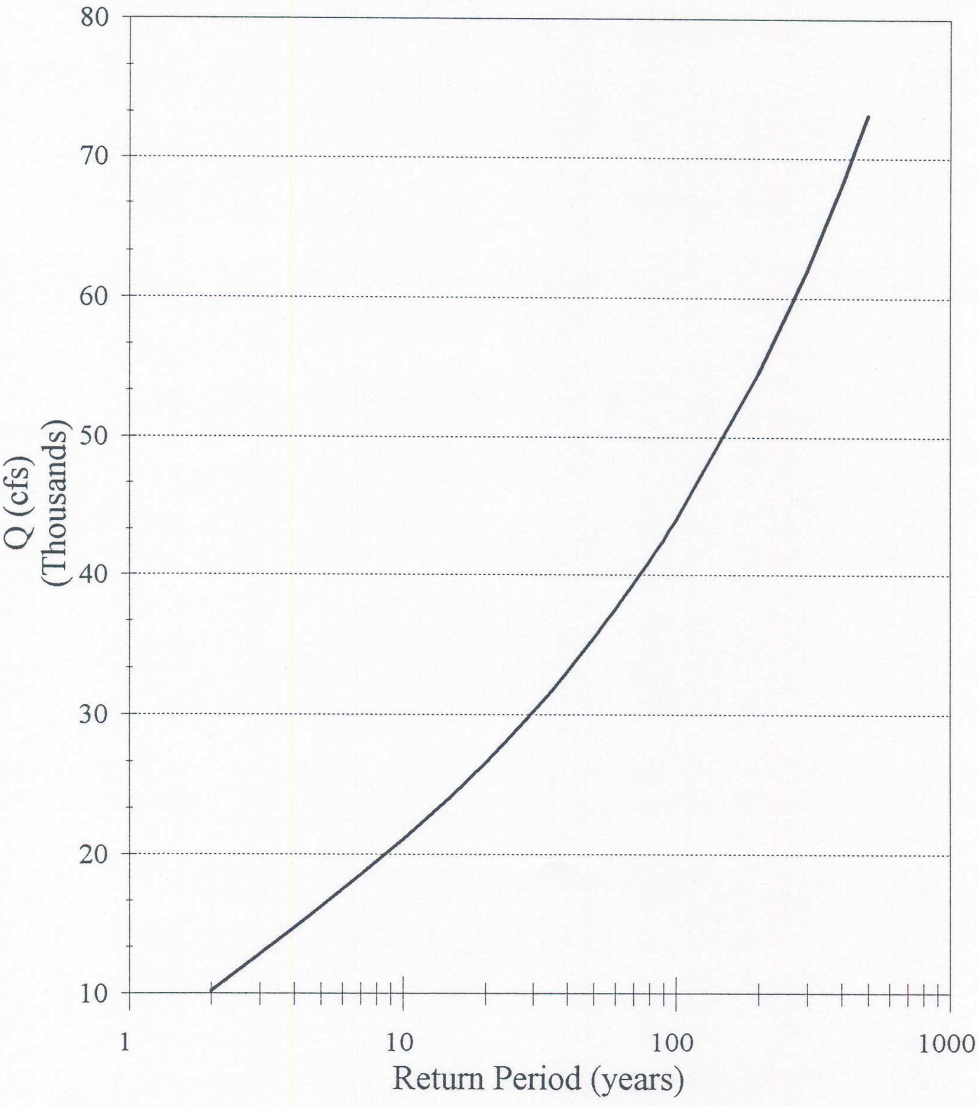

Figure C-13. Flood Frequency Curve for Río Guayanés (east) basin

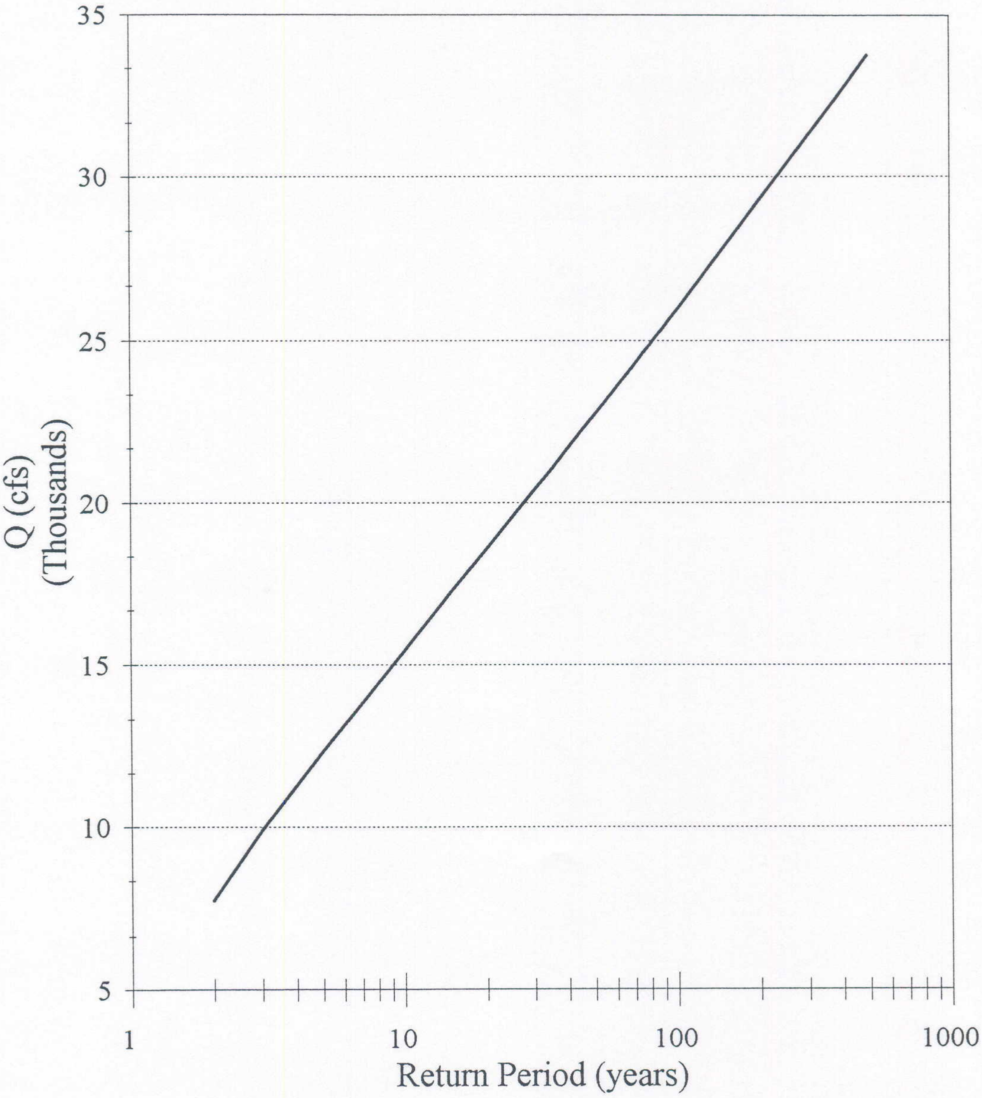

Figure C-14. Flood Frequency Curve for Río Guayanés (west) basin

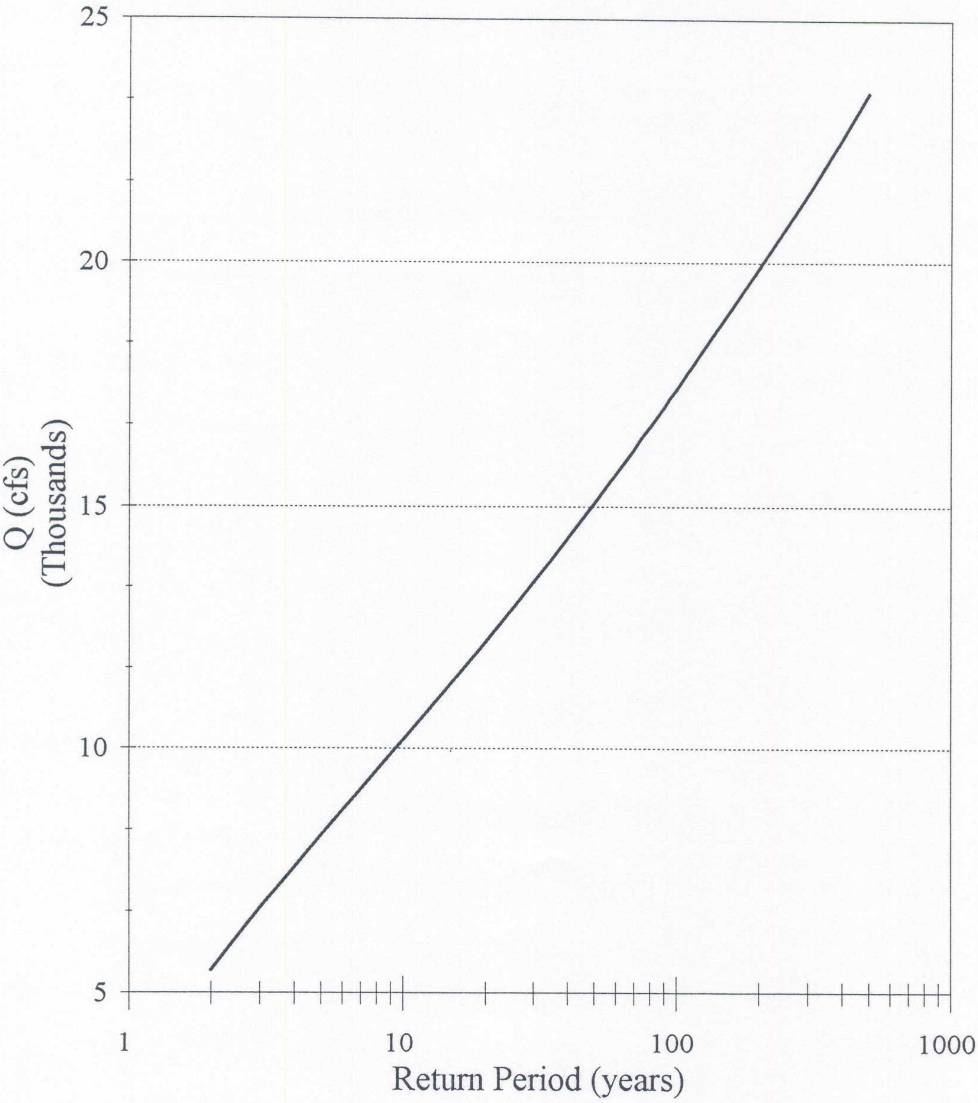

Figure C-15. Flood Frequency Curve for Río Guaynabo basin

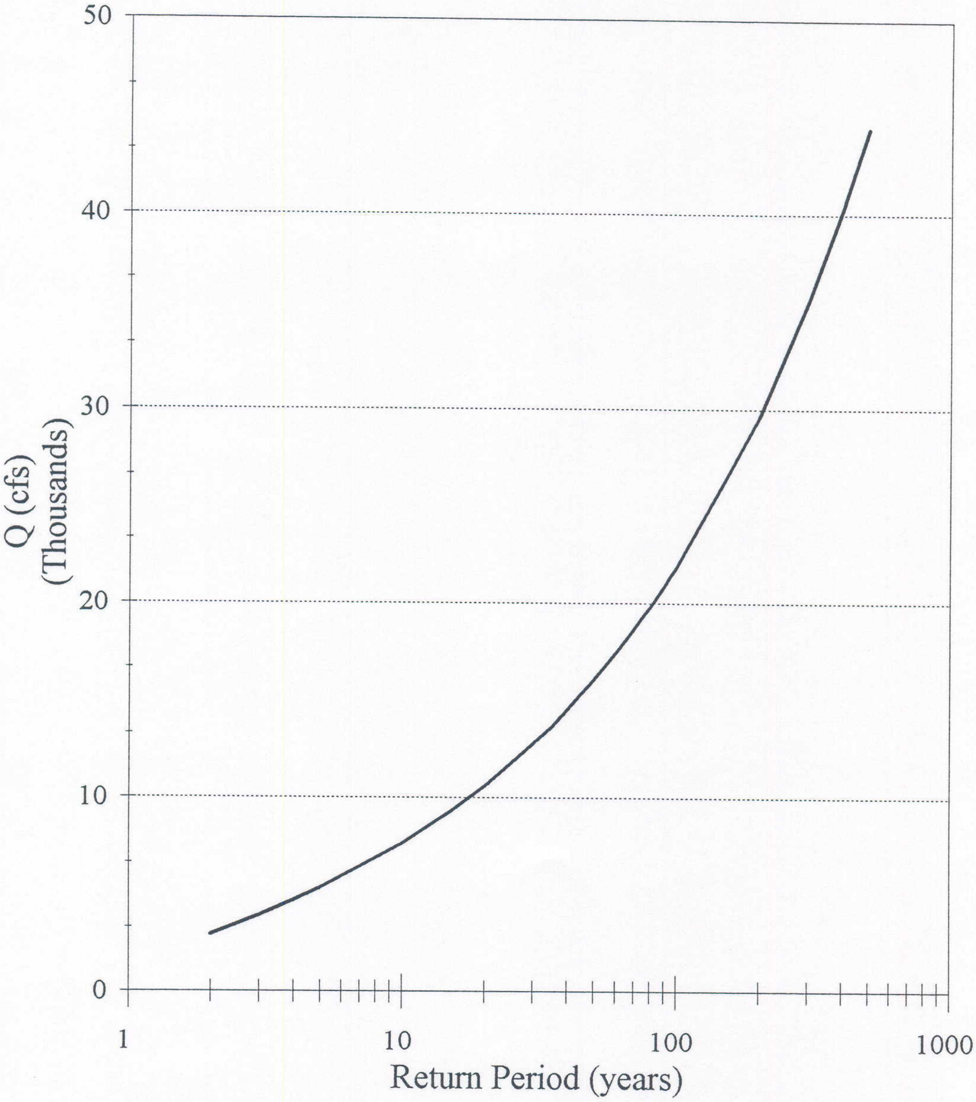

Figure C-16. Flood Frequency Curve for Río Jueyes basin

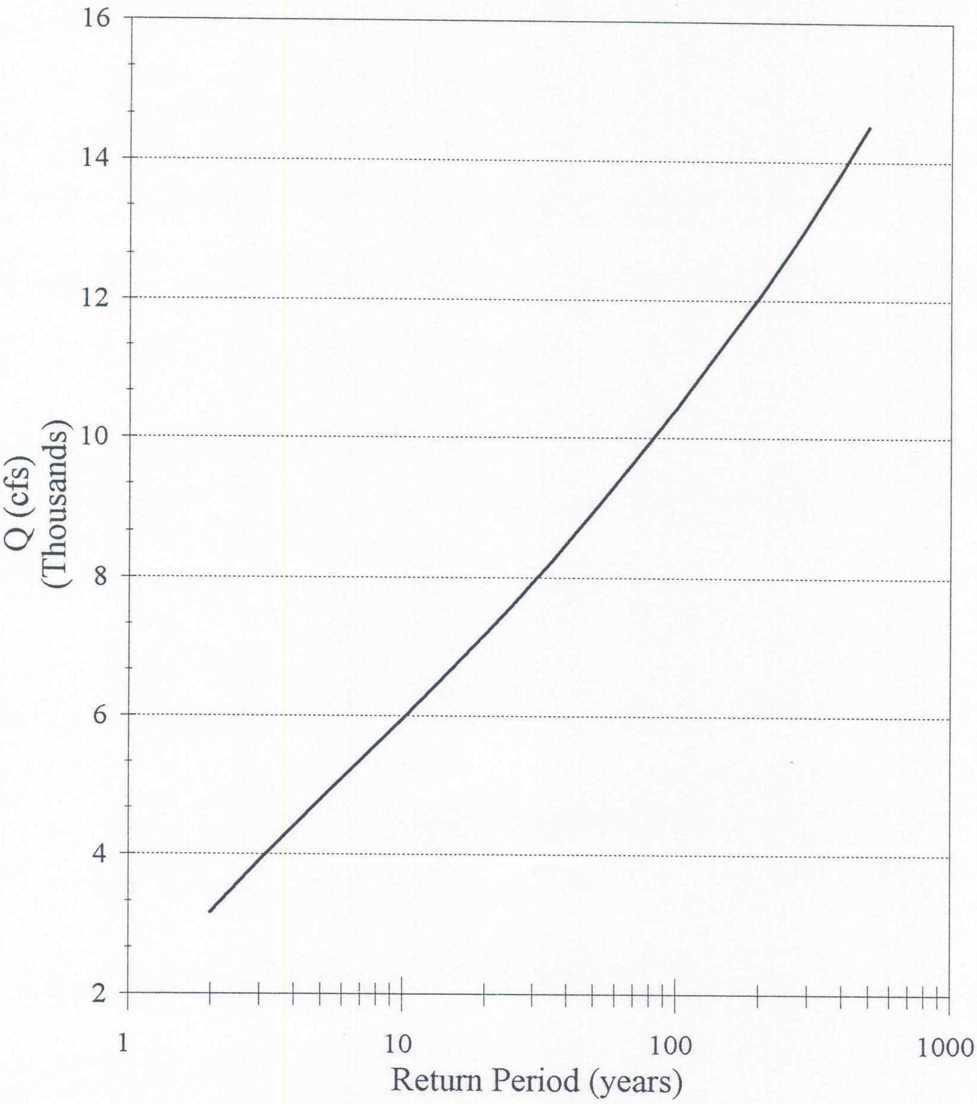

Figure C-17. Flood Frequency Curve for Río Lapa basin

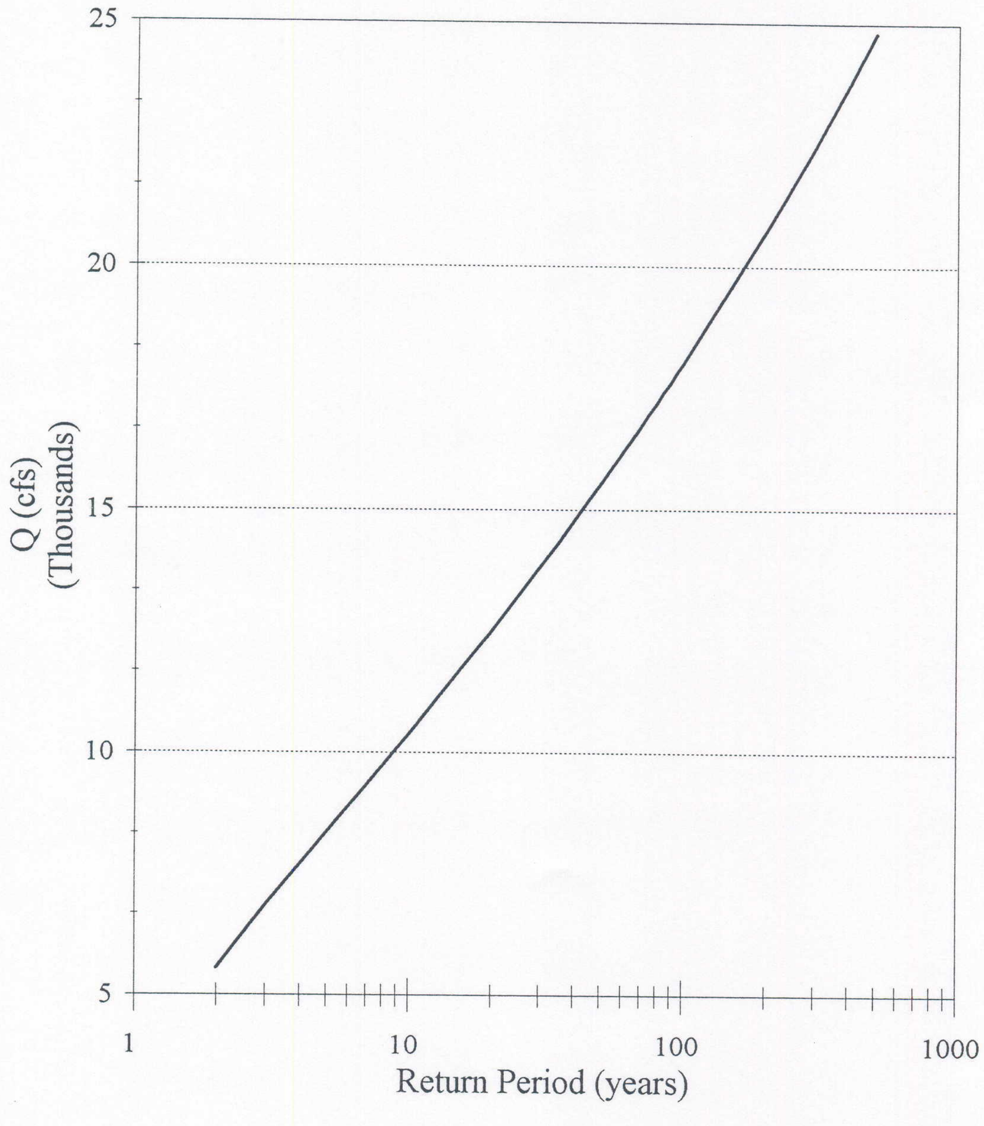

Figure C-18. Flood Frequency Curve for Río Majada basin

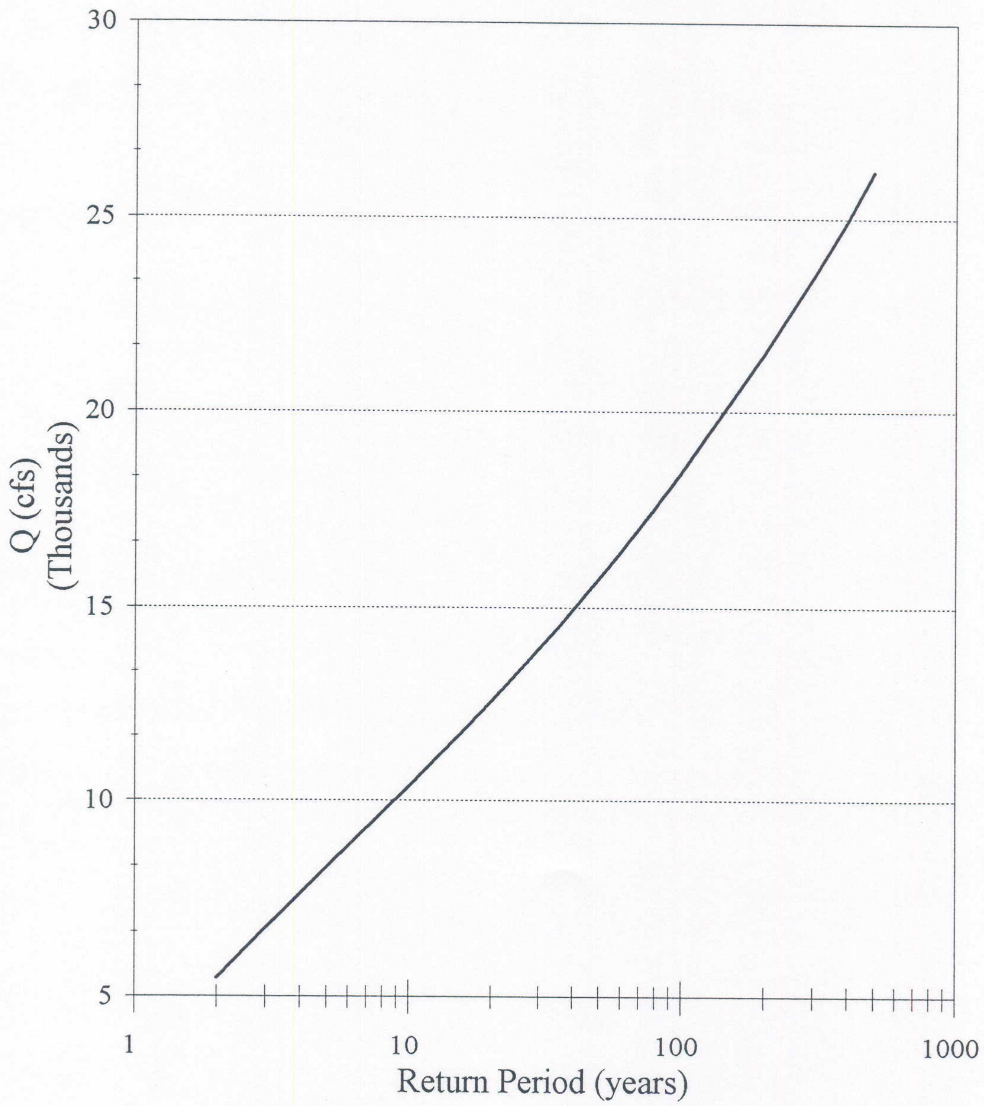

Figure C-19. Flood Frequency Curve for Río Maunabo basin

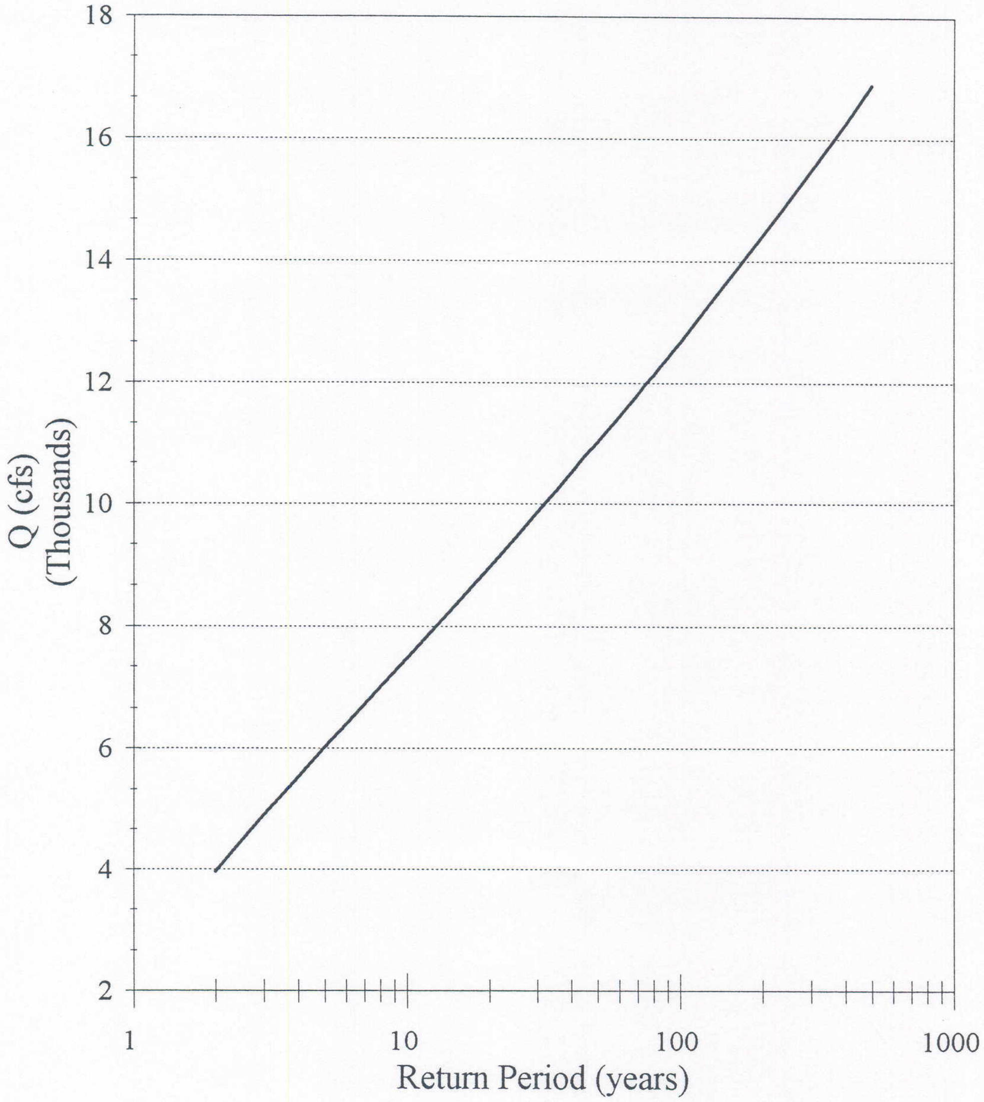

Figure C-20. Flood Frequency Curve for Río Mavilla basin

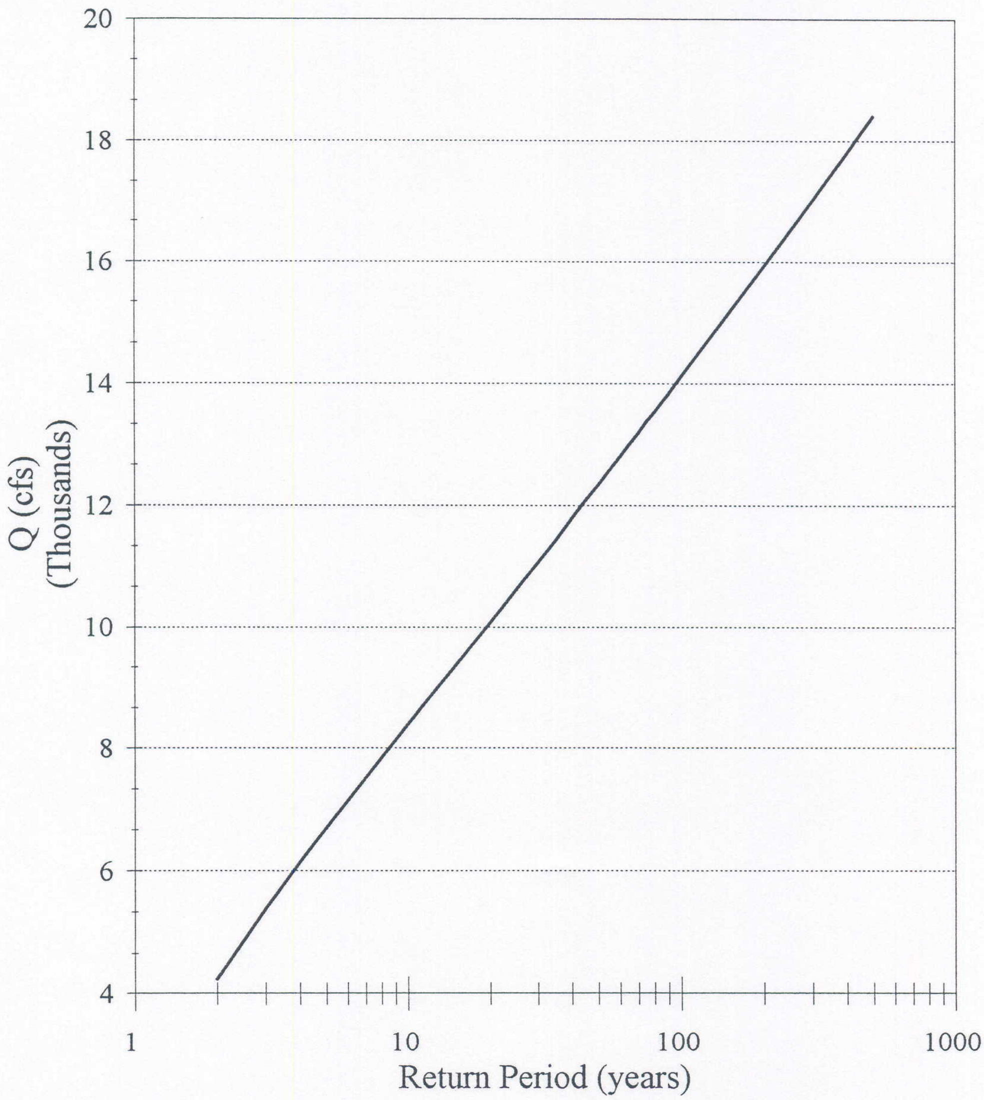

Figure C-21. Flood Frequency Curve for Río Orocovis basin

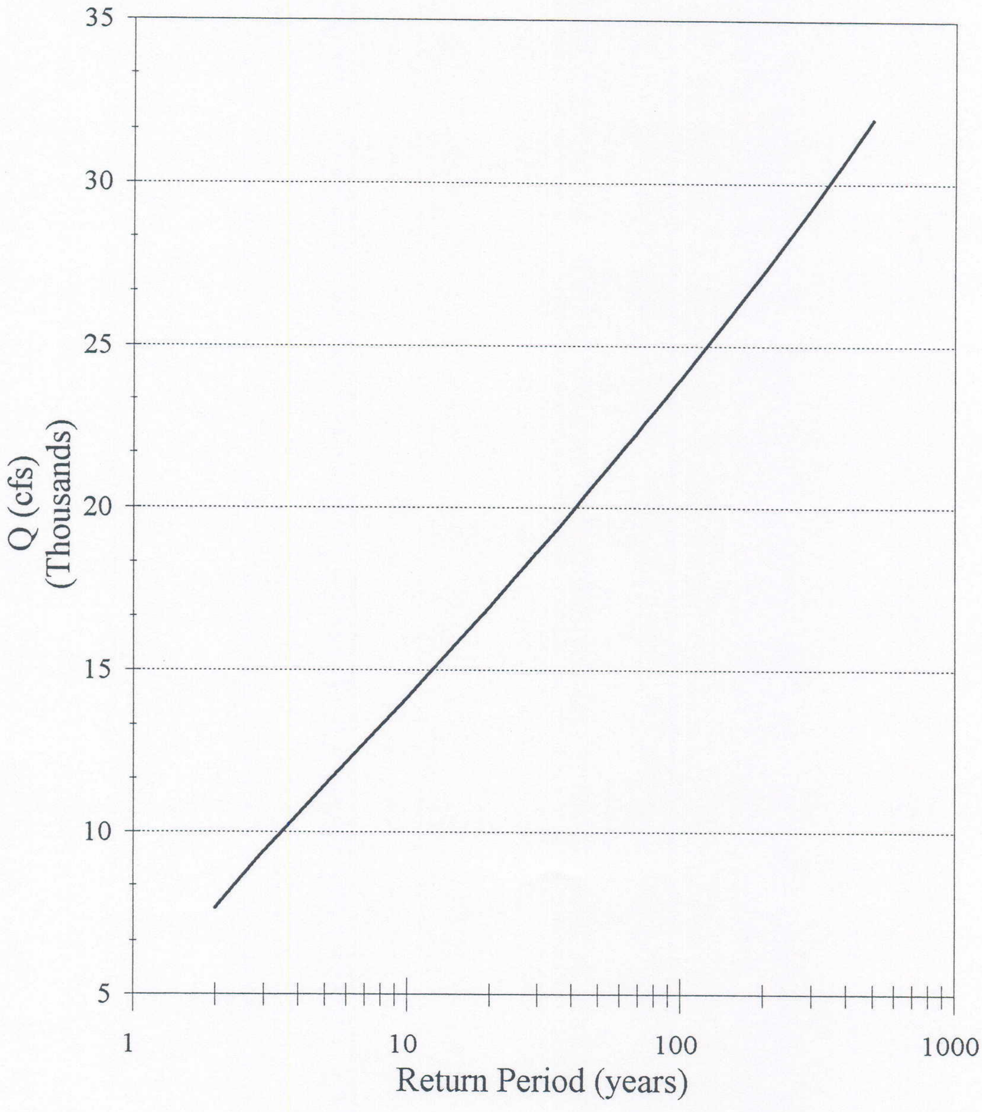

Figure C-22. Flood Frequency Curve for Río Piedras basin

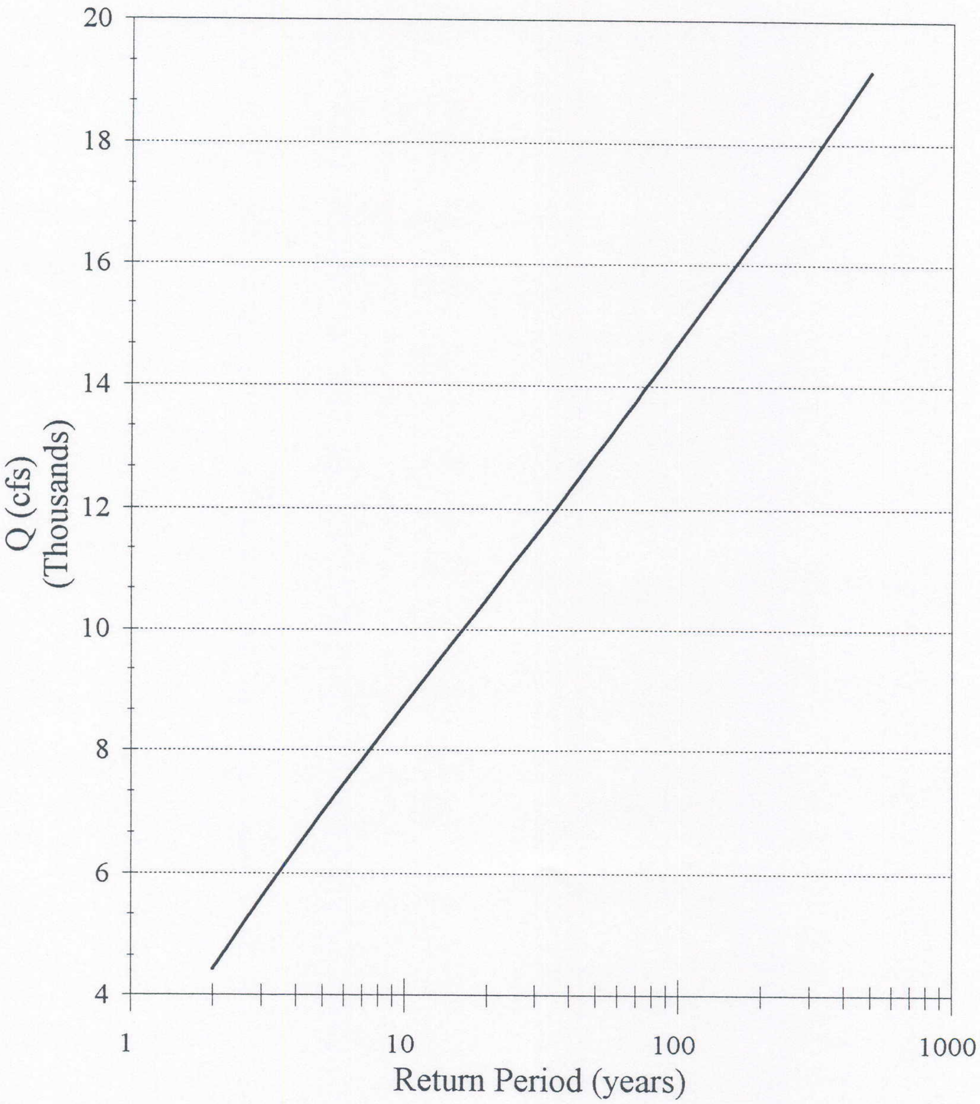

Figure C-23. Flood Frequency Curve for Río Santiago basin

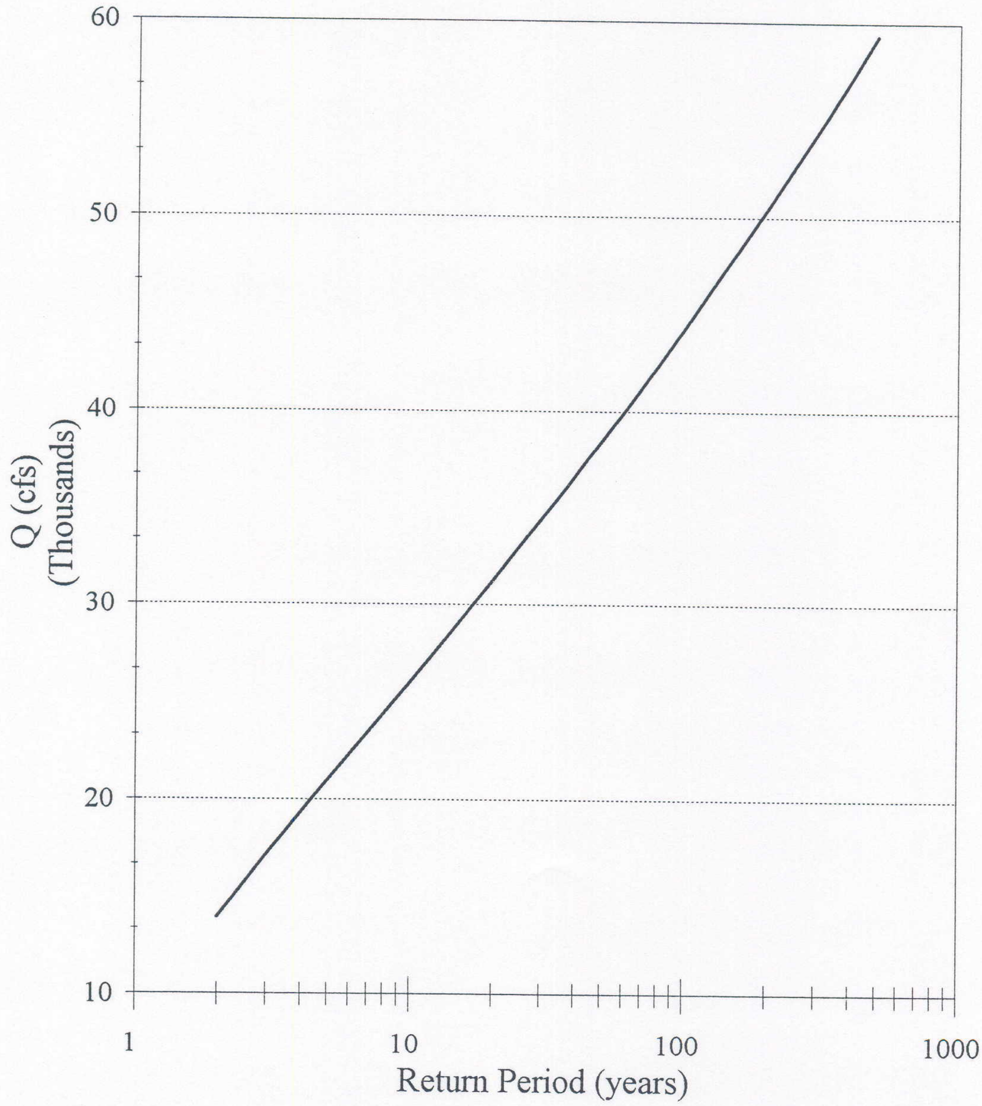

Figure C-24. Flood Frequency Curve for Río Toro Negro basin

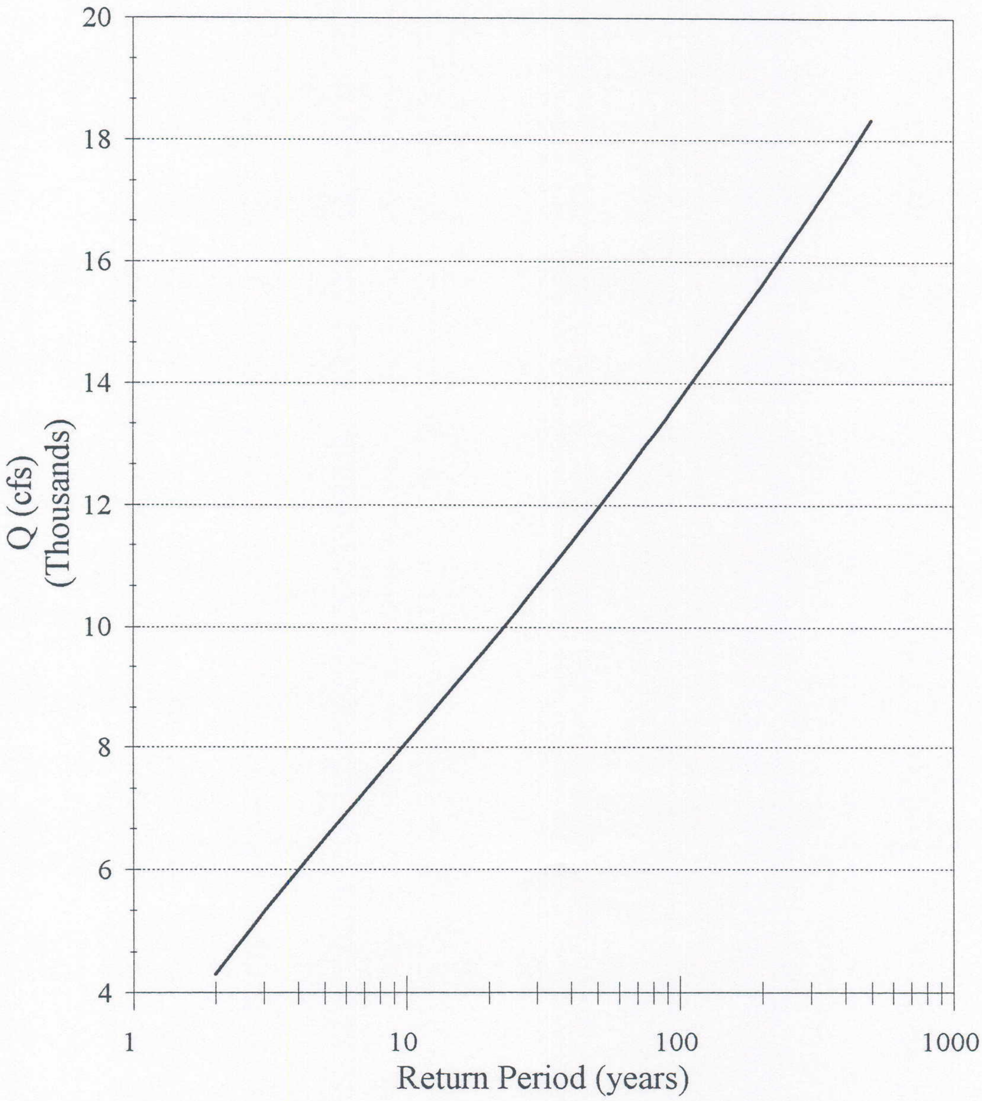

Figure C-25. Flood Frequency Curve for Río Turabo basin

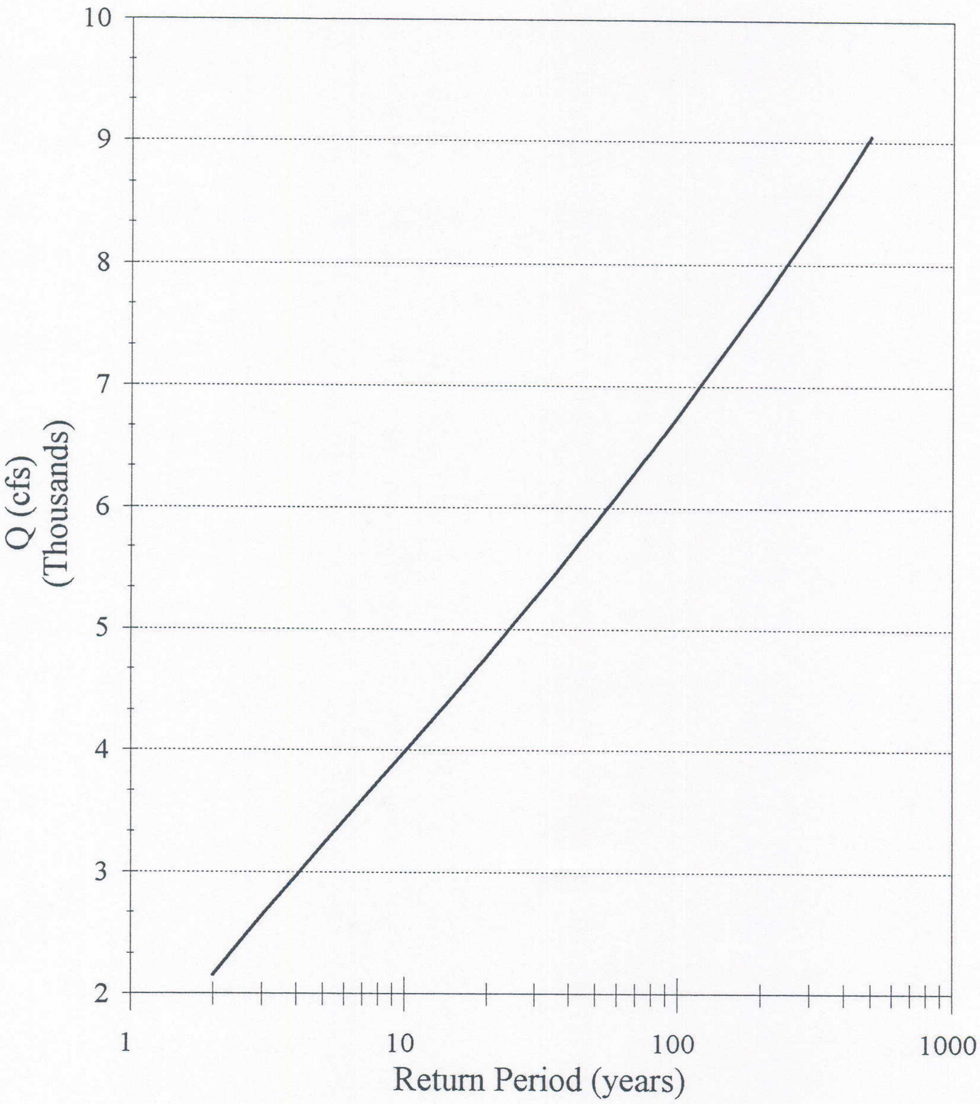

Figure C-26. Flood Frequency Curve for Río Unibón basin

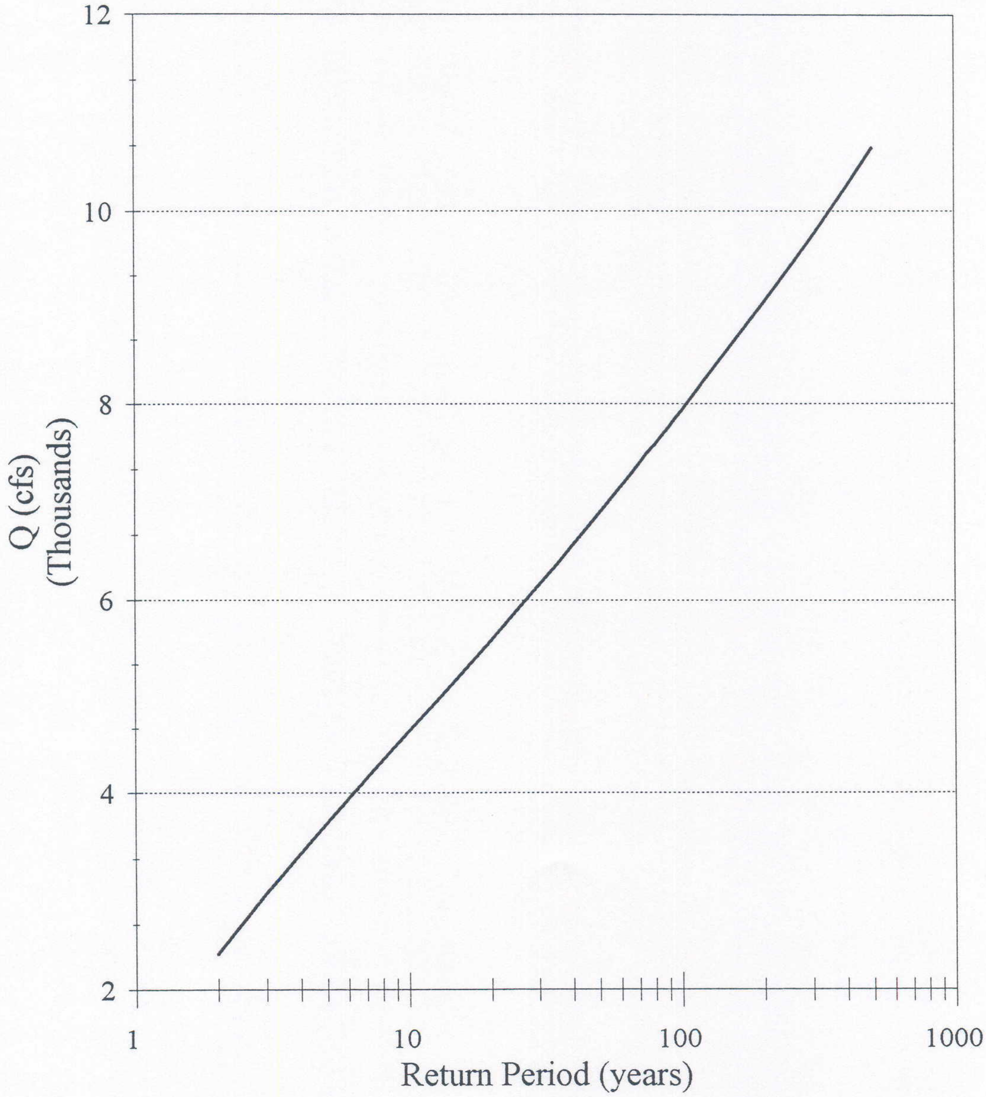

Figure C-27. Flood Frequency Curve for Río Usabón basin

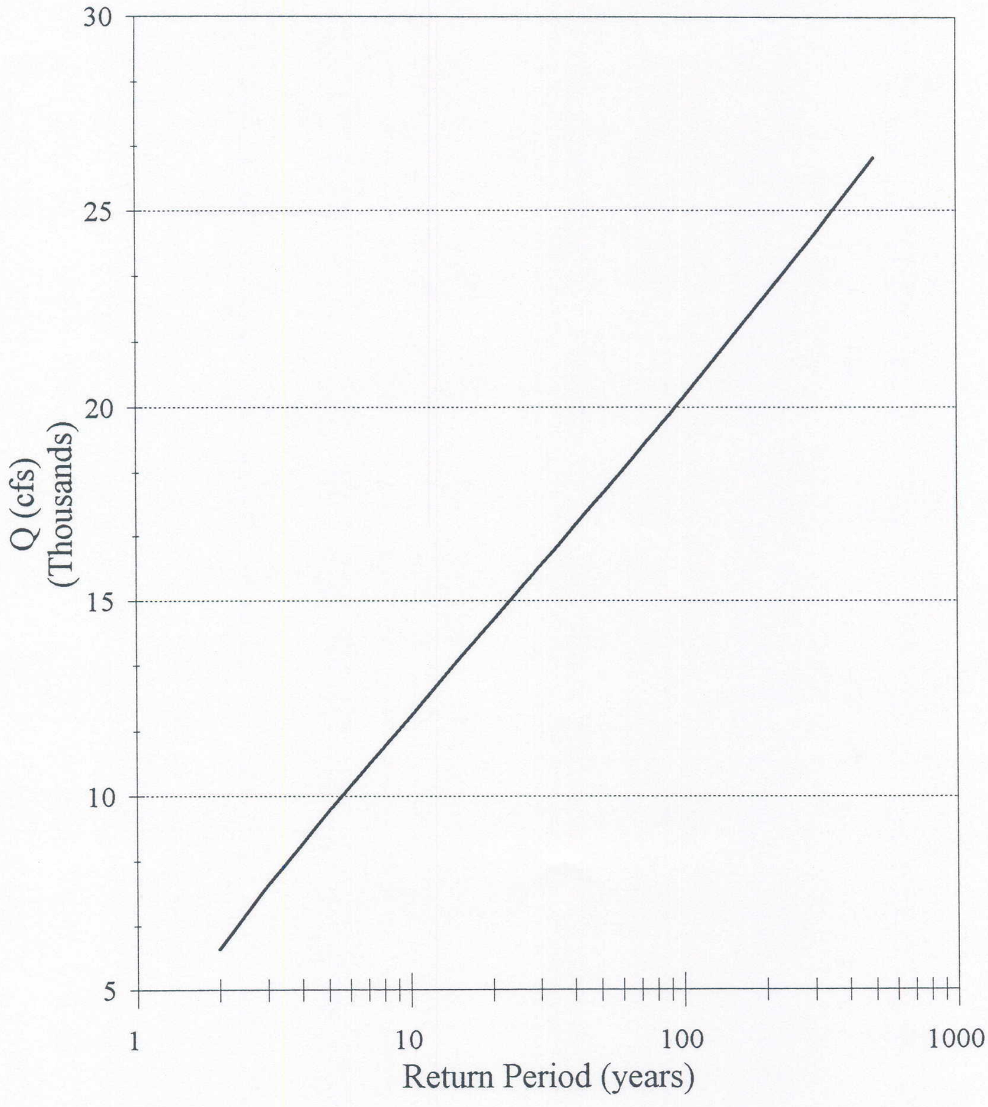

Figure C-28. Flood Frequency Curve for Río Viví basin

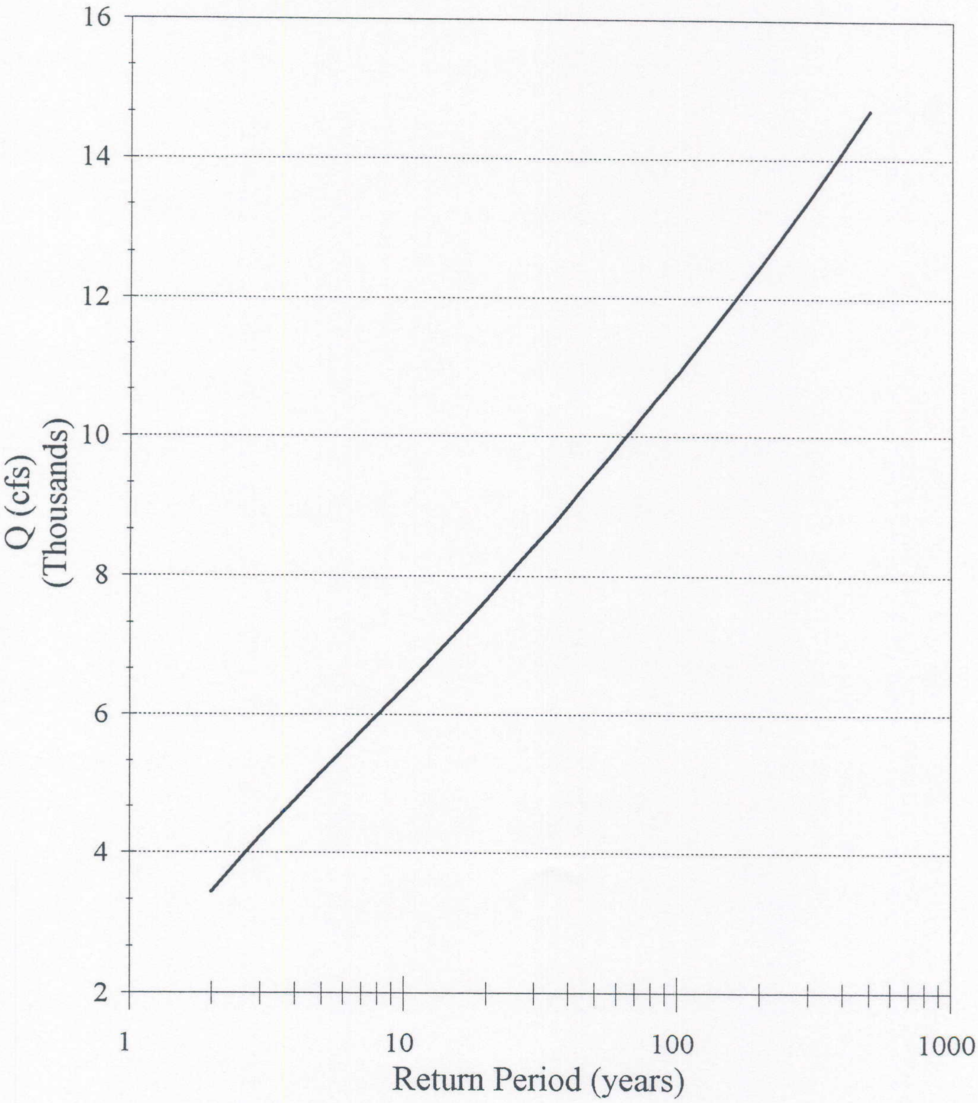

Figure C-29. Flood Frequency Curve for Río Yagüez basin

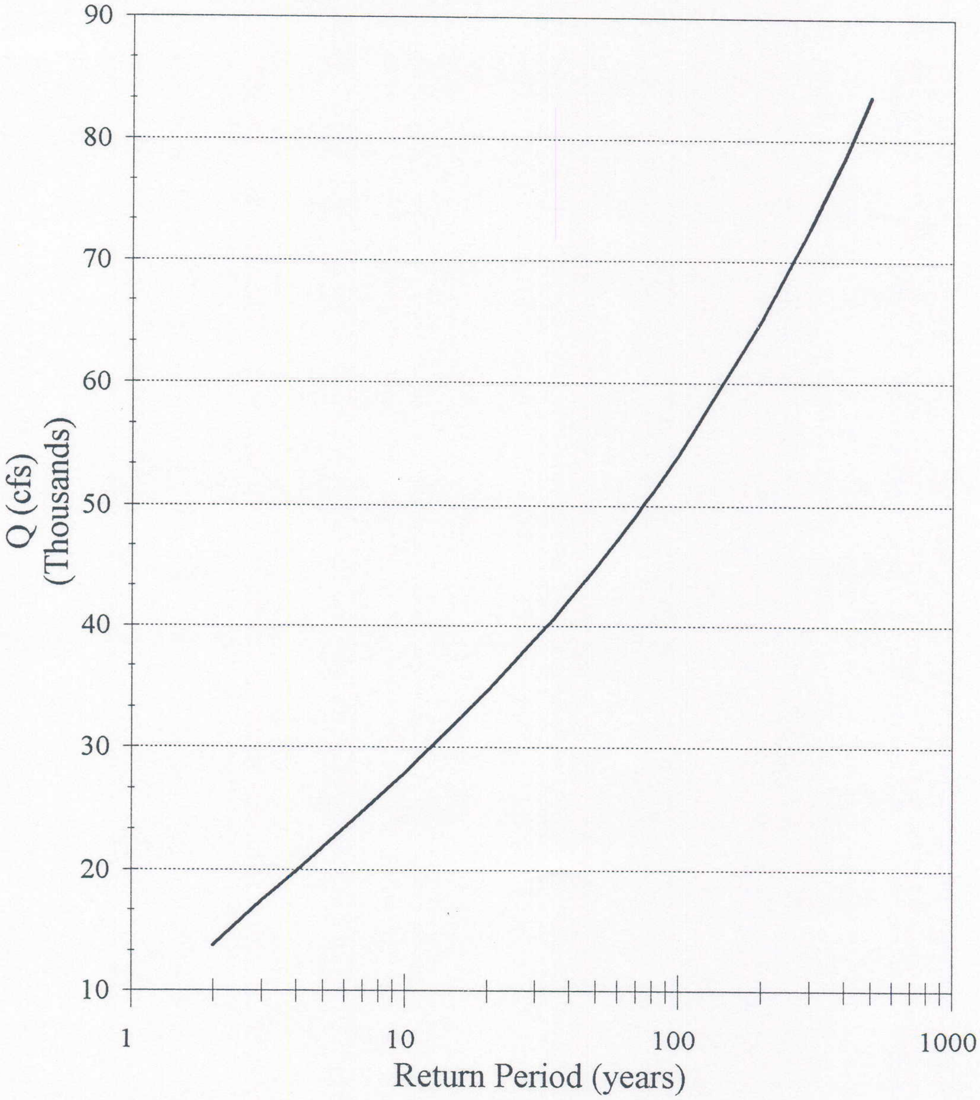

Figure C-30. Flood Frequency Curve for Río Yauco basin

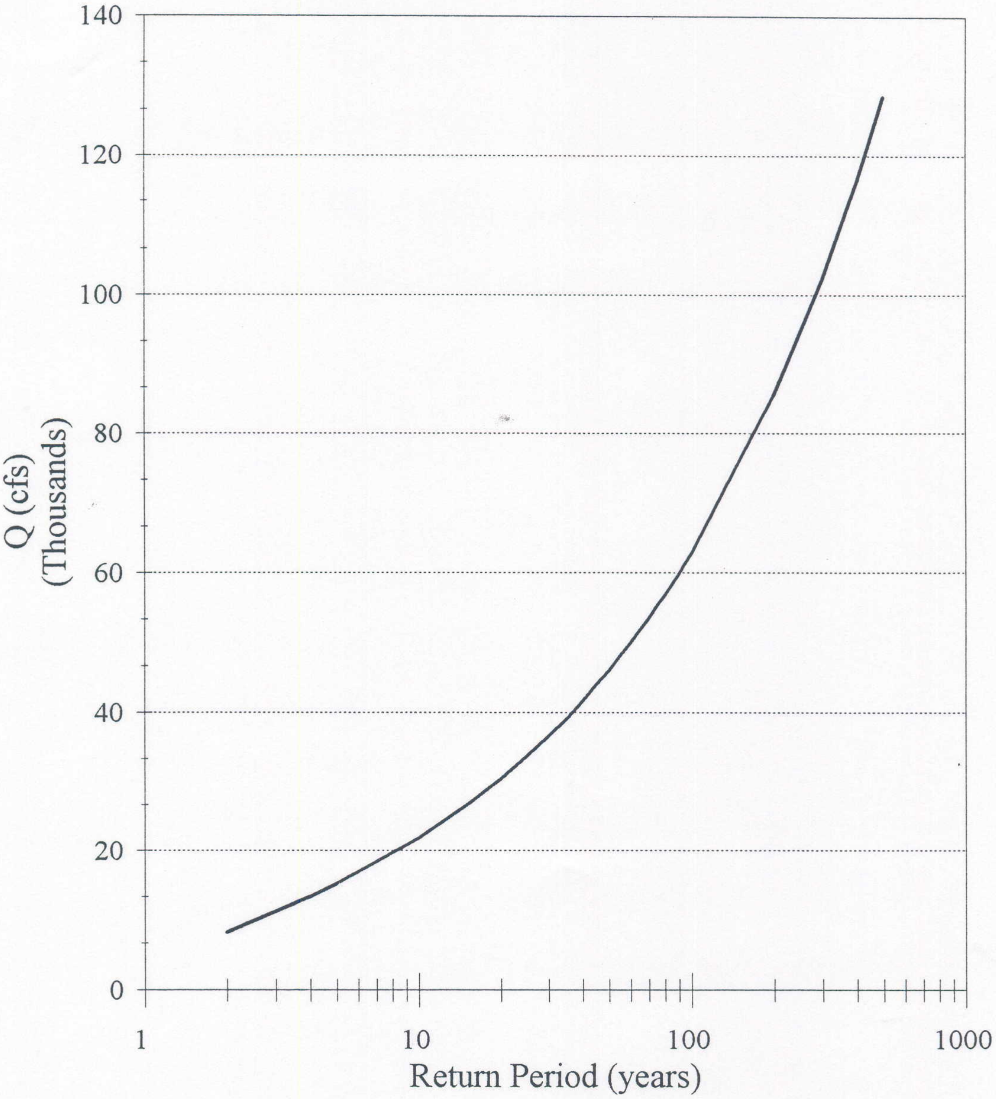

Figure C-31. Flood Frequency Curve for Río Yunés basin