REGIONALIZED FLOOD FREQUENCY ANALYSIS FOR BASINS IN PUERTO RICO

by

Natalia Almodóvar Rosario

A thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

CIVIL ENGINEERING
(Environmental and Water Resources Engineering)

UNIVERSITY OF PUERTO RICO
MAYAGÜEZ CAMPUS
2012

Approved by:

____________________ Nazario Ramírez, PhD Member, Graduate Committee	____________ Date
____________________ Jorge Rivera Santos, PhD Member, Graduate Committee	____________ Date
____________________ Rafael Segarra García, PhD President, Graduate Committee	____________ Date
____________________ Ismael Pagán Trinidad, MSCE Chairperson of the Department	____________ Date
____________________ Dana Collins, PhD Representative of Graduate Studies	____________ Date

ABSTRACT

The main goal of this research is to provide a flood regionalization study for Puerto Rico based on recent advances in watershed classification and parameter estimation techniques. The regional flood frequency was achieved using the "index-flood" procedure in combination with the regional *L*-moment algorithm. The main assumption of the "index flood" procedure is that sites form a homogeneous region, and for this matter the heterogeneity measure based on *L*-statistics were used to determine the homogeineity of a region. The hybrid cluster analysis was used as a regionalization procedure to group inland streams into four clusters using physical and climatological characteristics of the basins. Relative root mean square error values lower than 1% demonstrated the high accuracy in the procedure to estimate the quantile curves for the regions. The procedure recommended to estimate quantiles for ungaged basins couples the relationship between the derived regional growth curves and the estimated mean peak streamflow at site.

RESUMEN

El objetivo principal de esta investigación es proveer un análisis de regionalización de frecuencia de inundaciones para Puerto Rico basado en los avances recientes en clasificación de cuencas y nuevas técnicas de estimados de parámetros. El análisis regional fue ejecutado utilizando el procedimiento "índice de flujo" en combinación con el algoritmo regional de momentos *L*. La hipótesis principal del procedimiento "índice de flujo" es que las cuencas forman una región homogénea. El algoritmo de conglomerado híbrido es utilizado para el procedimiento de regionalización, agrupando las cuencas en cuatro regiones según las características físicas y climatológicas. Valores de la raíz cuadrada del error cuadrático medio menores a 1% fueron obtenidos, demostrando la alta precisión en el proceso para estimar los cuantiles regionales. El proceso recomendado para estimar los cuantiles para cuencas sin aforar relaciona las curvas de crecimiento derivadas mediante el análisis regional y el promedio del caudal máximo anual estimado para el lugar.

To God, my parents, friends, and my fiancé.

ACKNOWLEDGEMENT

I am gratefull for the guidance of my advisor, Dr. Rafael Segarra García, who was a key on the completion of this research. Dr. Segarra shared with me his valuable knowledge in Water Resources and probabilities, and for that I am very grateful. I would like to express my very great appreciation to my graduate committee, Dr. Nazario Ramirez and Dr. Jorge Rivera Santos, for their support and guidance in my thesis.

My special thanks are extended to the administrative staff of the Civil Engineering Department, especially to Ismael Pagán Trinidad, Department Director, for believing in my abilities to perform research by allowing me to participate in research internships with the US Army Corps of Engineers. Also I would like to give thanks to Myriam Hernández for her help in courses enrollment in my master degree. The advices and opportunities provided by Dra. Ingrid Padilla are greatly appreciated. Thanks to the graduate scholarship by the US Department of Energy for providing me funding during the development of this study.

I would like to give special thanks to my parents that were always besides me. They have supported all my decisions, especially my goal to pursue graduate studies. I am very thankful for the support, help and love provided by my fiancé Eduardo Alvarez.

But above all, I would like to thank God for allowing me to continue expanding my knowledge and providing me health to help me achieve my goals.

TABLE OF CONTENT

List of Tables

LIST OF FIGURES

List of Symbols

A	Area
AMC	Antecedent Moisture Condition
CDA	Contributing Drainage Area
CL	Channel Length
CN	Curve Number
D_i	Discordancy Measure
DR	Depth-to-rock
GEV	General Extreme Value Distribution
GLO	Generalized Logistic Distribution
GNO	Generalized Normal Distribution
GPA	Generalized Pareto Distribution
GS	Ground Slope
H	Heterogeneity measure
l_1	Sample L-location
l_2	Sample L-scale
λ_1	L-location or mean of distribution
λ_2	L-scale
L-CV	Coefficient of L-variation
MAR	Mean Annual Rainfall
NRCS	Natural Resources Conservation Service
PE3	Pearson Type 3
RMSE	Root Mean Square Error
RI-T	Average T-year 24 hour Rainfall Intensity
SCS	Soil Conservation Service
SP	Average Soil Permeability
t	Sample coefficient of L-variation
t_3	Sample L-skewness
t_4	Sample L-kurtosis

τ	Coefficient of L-variation
$\tau 3$	L-skewness
$\tau 4$	L-kurtosis
TSVQ	Tree-Structure Vector Quantization
USGS	United States Geological Survey
VC	Vegetative Cover
Z_i^{DIST}	Goodness-of-fit measure

1. INTRODUCTION

1.1 Justification

An important public issue regarding natural disasters has been the adequate prediction of catastrophic flooding in populated areas. The process of climate change has increased awareness of the potential for more frequent major flooding. Indeed, the occurrence of unprecedented flooding all across the world prompted in the United States a workshop on Hydrologic Frequency Analysis (Olsen, Kiang, & Waskom, 2010). Here, federal agencies discussed the next step for developing best practices for hydrologic frequency analysis for changing conditions. Although, there was agreement that changes occur in hydrology, there was no consensus on which methods could be used to model either past or potential future hydrological changes.

The island of Puerto Rico is constantly affected by floods and formulating adequate flood management policies is a constant challenge. An integral part of such efforts is the ability to estimate and predict the magnitude and frequency of major flood events. Appropriate technologies for assessing the magnitude and frequency of expected floods are urgently needed in the policy making process. In Puerto Rico, a large percentage of streams are not monitored and there is no reliable way to assess the flooding potential of such systems without recurring to the use of regionalized methods. For flood frequency analyses, the use of traditional rainfall/runoff procedures such as the Natural Resources Conservation Service (NRCS) method (NRCS, 1986) are not reliable since such methods are based on foreign parameters and fixed hydrological responses, in the form of the universal synthetic unit hydrograph. Additional uncertainty in this procedure is introduced by using the T-year rainfall to generate an assumed T-year runoff peak, based on the assumption that both return periods are exactly similar.

A regionalized flood frequency analysis is based on actual streamgage data and results in a more efficient use of the available historical discharge time series for a site. In this type of analysis, the available gage data at monitored streams serves as the basis for estimating the corresponding parameters related to ungaged sites. The United States Geological Survey (USGS) has published three reports addressing the estimation of magnitude and frequency of floods in Puerto Rico based on streamgage records. The procedure developed by the USGS is based on a regional regression analysis approach. The three reports – López & Fields (1970), López, Colón-Diepa, & Cobb (1979), and Ramos-Gines (1999) – all follow the same procedure, each updating the earlier one. The fundamental methodology of these studies consists of estimating the parameters of the Log-Pearson extreme value distribution, with subsequent regionalization of the parameter through regressions with catchment and climatic attributes. The regression-based approach of these studies yielded relatively high standard errors and require the determination of several parameters (such as, contributing drainage area, depth-to-rock, mean annual rainfall, average permeability of soils, vegetative cover, main channel slope, main channel length, average of the ground slope and the average i-year 24-hour rainfall intensity), some of which are difficult to estimate in a practical context.

1.2 Objectives

The objective is to perform a new flood regionalization study for Puerto Rico based on recent advances in watershed classification and parameter estimation techniques of flood frequency distributions.

2. LITERATURE REVIEW

2.1 Flood Frequency Analyses in Puerto Rico

The USGS has published three reports on the estimation of the magnitude and frequency of river floods in Puerto Rico. Their procedure is based on regionalizing flood estimates using regressions with basin physical and climatologic attributes. The reports by López & Fields (1970) and López, Colón-Diepa, & Cobb (1979) presented regressions for estimating peak flows for the 5-, 10-, 25-, 50-, and 100-year return periods. In the latest report, Ramos-Ginés (1999) presented updated regressions for up to a 500-year return period. The lowest standard errors achieved with the USGS procedure are in the range of 22.8 to 29.5 percent.

In the USGS reports, the flood frequency curves for gaged streams were constructed following the procedures in Bulletin 17B from the Interagency Advisory Committee on Water Data (1982), which prescribes the Log-Pearson Type III distribution for estimating flood-frequency parameters. Flood frequency relations for ungaged sites are constructed by regressing the peak flows with catchment attributes such as drainage area and mean annual rainfall. Ramos-Gines (1999) added the depth-to-rock as a significant parameter. Depth-to-rock is not easily estimated at ungaged locations, and alternate expressions are provided that do not include this parameter, albeit with higher standard estimation errors.

Using a different approach to regionalization, Segarra (1998) developed generalized flood frequency relations for ungaged basins by grouping catchments into homogeneous regions. The use of the Log-Pearson distribution was abandoned in favor of the Generalized Extreme Value (GEV) distribution for determining parameters for gaged sites. The study grouped basins into four homogenous regions, as opposed to the single homogenous region employed by López, Colón-Diepa, & Cobb (1979) and Ramos-Ginés (1999). However, Ramos-Ginés argued that the standard errors related to the regression-based approach were lower than those resulting from the approach of Segarra (1998). In part, this result was explained by the fact that the study gleaned the physical catchment data from large-scale printed topographic quadrangles, and this procedure may have introduced scaling errors in the measures. The use of geographical information systems-based tools should result in much more accurate physical parameter estimates that could improve the estimation errors arising from this particular aspect.

2.2 Other Flood Frequency Analyses

Four additional previous studies on flood frequency analysis were evaluated. The studies treat the use of stochastic rainfall-runoff model, kinematic wave peak discharge, and regional frequency analysis using *L*-moments.

The first study was performed on six basins in Sicily by Aronica & Candela (2007). They proposed deriving frequency distributions of peak flows using a semi-distributed stochastic rainfall-runoff model. The authors argue that the benefit of this approach are that the rainfall-runoff model is not complex, has a limited number of parameters, and does not require a major calibration effort resulting in a robust tool for those catchments which are partially or poorly gauged. The authors selected a Monte Carlo simulation approach to determine a derived flood distribution for its practical applicability and flexibility. For ungaged catchments, or catchments where the recording period is not sufficiently long, the authors used previous works where regionalization had been employed by means of cluster analysis. Two modules were defined for the Monte Carlo simulation, (1) a stochastic rainfall generator module, and (2) a catchment response module. For the rainfall module, the Two Components Extreme Value distribution was used and simple relations between the rainfall duration were used to estimate the parameters. The Rational Method was used to define the basin response module; therefore, the effective depth of rainfall was needed to determine the peak discharge. For a simpler procedure the Soil Conservation Service - Curve Number (SCS-CN) method was used to transform the storm depth into effective rainfall (h_e).

The results of the study by Aronica and Candela (2007) demonstrated that for areas greater than 200 km^2 derived distributions are not entirely reliable. The method provides a good estimation for smaller basins and does not require extensive recorded data. The method is less reliable for areas larger than 200 km^2.

In the second study, Cadavid, Obeysekera, & Shen, (1991) derived flood-frequency distributions applicable to small watersheds in which overland flow is considered an important runoff component, such as urban catchments. The cited work is based on the use of the kinematic wave to determine peak discharge and time to peak, as a function of precipitation and watershed characteristics. Even though the model included variables representing catchment geometry and dynamics, the results were not comparable with recorded peak discharges and times to peak data.

Saf (2009) performed a flood frequency analysis for a region of Turkey based on *L*-moments probabilistic method for parameter estimation. The methodology used in the study follows the steps proposed by Hosking & Wallis (1997), and encompassed 36 stations with more than 10 years of data. To define homogenous regions, the author used the *L*-moment diagram and the Z_i^{DIST} statistics criteria. In the study, a Monte Carlo simulation was used to address the accuracy of the estimates selected. These simulated results were compared to the observed estimates, and the accuracies of the procedure to estimate quantiles were determined. The greater relative root-mean-square error (RMSE) was 0.192 percent for a recurrence of 100 years, which can be considered a relatively low level.

Federal agencies in the United States commonly use Bulletin 17B or a derivative to determine flood quantiles. Lim & Voeller (2009) have compared this method to the *L*-moments approach. The authors performed a frequency analysis using the *L*-moments flood index method on the North basin of the Red River, which flows within the Canadian province of Manitoba, and through three states of the United States: North Dakota, Minnesota, and South Dakota. The authors conclude that the *L*-Moments based index flood method (LMFI) offers several advantages over the older Bulletin 17B approach but suggest further studies to adapt the method for the consideration of the estimation of index floods, particularly for ungaged basins.

The regional flood frequency analysis using *L*-moments is now considered a standard technique in hydrologic studies, as evidenced by its use in the preparation of Atlas 14 (Bonnin, et al., 2006), the new rainfall frequency atlas for Puerto Rico. Therefore, this analysis could provide better flood quantile estimations in Puerto Rico when compared to the results obtained from the Bulletin 17B on the USGS report. The conclusions of Lim & Voeller (2009) and Saf (2009) demonstrate that the procedure can provide low quantile estimation errors.

3. *L*-MOMENTS

The most recent approach developed for the estimation of flood distribution parameters uses *L*-moments. The "*L*" on the *L*-moments stands for the construction of moments from linear functions of order statistics. It has been applied at many locations, including the United States (Lim & Voeller, 2009), Turkey (Saf, 2009), India (Kumar & Chatterjee, 2005), and China (Yang, Hao, & Sun, 2009), and has become a standard technique for flood quantile parameter estimation.

3.1 Method of *L*-moments

L-moments were introduced by Hosking & Wallis (1990), and have several theoretical advantages over other moment estimation procedures. Compared with other moment estimation procedures, *L*-moments are more robust with respect to the outliers, are less subject to bias in estimation, and can approximate their asymptotic normal distribution more closely for finite examples than other methods (Yang, Hao, & Sun, 2009). For a finite sample $X_{i:n}$ denote the *i*th smallest observation, and let $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ be an ascending ordered sample. To determine the distributional parameters of a population of size *n*, the L-moments are defined by,

$$\lambda_r = r^{-1} \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} E\left(X_{r-j:r}\right), \qquad r = 1,2,3, \ldots \qquad (3.1)$$

where,

$$E(X_{r:n}) = \frac{n!}{(r-1)! - (n-r)!}\int_0^1 x(u)u^{r-1}(1-u)^{n-r}\,du. \tag{3.2}$$

Equation 3.1 defines the measurements of the shape parameter of the distribution: λ_1 represent the location parameter of the distribution, λ_2 is the measure of the scale parameter of the distribution, λ_3 is the measure of skewness parameter of the distribution, and λ_4 is the measure of the kurtosis of the distribution. Each of the parameters are described by the following expressions:

$$\lambda_1 = E(X_{1:1}) \tag{3.3}$$

$$\lambda_2 = \frac{1}{2}E(X_{2:2} - X_{1:2}) \tag{3.4}$$

$$\lambda_3 = \frac{1}{3}E(X_{3:3} - 2X_{2:3} + X_{1:3}) \tag{3.5}$$

$$\lambda_4 = \frac{1}{4}E(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}). \tag{3.6}$$

Of this group of *L*-moments, λ_1 (mean) and λ_2 (*L*-scale) are mostly used to summarize the probability distribution. *L*-moment ratios are also useful in the definition of probability distribution. *L*-moment ratios are dimensionless *L*-moments defined by:

$$\tau_r = \lambda_r/\lambda_2, \quad r = 3, 4, 5 \ldots. \tag{3.7}$$

When λ_1 is used as the divisor of λ_2 the coefficient of L-variation (τ) is obtained. In the case were the measure of skewness is divided by the measure of scale, the L-skewness (τ_3) is obtained. Similarly, when the measure of kurtosis is divided by the measure of scale, the *L*-kurtosis (τ_4) is obtained.

Sample *L*-moments are defined in general as:

$$l_{\mathrm{r}} = \frac{1}{\mathrm{r}}\binom{n}{r}^{-1}\sum_{i=1}^{n}\left[\sum_{j=0}^{r-1}(-1)^{j}\binom{r-1}{j}\binom{i-1}{r-1-j}\binom{n-i}{j}x_{i:n}\right]. \qquad (3.8)$$

The estimates of the *L*-moment ratios obtained from the observed data are defined then by:

$$t_r = \frac{l_r}{l_2}, \quad r = 3,4,5 \ldots. \qquad (3.9)$$

The coefficient of L-variation for the sample (t) is obtained then by:

$$t = \frac{l_2}{l_1}. \qquad (3.10)$$

The values of l_r, t_r and t are the natural estimators of λ_r, τ_r, and τ, respectively.

4. REGIONAL FLOOD FREQUENCY ANALYSIS BASED ON *L*-MOMENTS

The main idea of the Regional Flood Frequency Analysis (RFFA) is to use the data from a grouping of statistically similar regions to estimate site quantiles. The aim of regional frequency analysis is not to fit a distribution to a particular data set, but to obtain quantile estimates of the distribution from which future data values will arise (Hosking & Wallis, 1997).

A screening of the data used on a RFFA is recommended to eliminate sites with erroneous data values, trends and outliers that bias in the *L*- moments of a sample. The discordancy measure (D_i), in terms of the *L*-moments is suggested by Hosking and Wallis (1997). A discordant site is a point in a three dimensional space defined by the *L*-CV, *L*-skewness and *L*-kurtosis that lies outside a concentric ellipse determined by the covariance matrix of the sites *L*-moments ratios. For site *i*, the measure indicates the discordancy between the site *L*-moment ratio and the regional average *L*-moment ratio. To define the discordancy measure let N represent the number of sites on a region, $\mathbf{u_i} = \begin{bmatrix} t^{(i)} & t_3^{(i)} & t_4^{(i)} \end{bmatrix}$, and T be the transpose of a matrix (Hosking & Wallis, 1997):

$$D_i = \frac{1}{3} N(\mathbf{u_i} - \bar{\mathbf{u}})^T \boldsymbol{A}^{-1}(\mathbf{u_i} - \bar{\mathbf{u}}) \tag{4.1}$$

$$\boldsymbol{A} = \sum_{i=1}^{N} (\mathbf{u_i} - \bar{\mathbf{u}})(\mathbf{u_i} - \bar{\mathbf{u}})^T \tag{4.2}$$

$$\bar{\mathbf{u}} = N^{-1}\sum_{i=1}^{N}\mathbf{u_i}\,. \tag{4.3}$$

Table 1 presents the corresponding critical value for regions with a certain number of sites. Regions with N sites where their D_i value is greater than D_{crit} can be eliminated from the regionalization procedure for a regional flood frequency analysis.

Table 1. Critical values for discordancy measures.
(From Hosking & Wallis, 1997)

Number of sites in Region	**Critical value (D_{crit})**	**Number of sites in Region**	**Critical value (D_{crit})**
5	1.333	11	2.632
6	1.648	12	2.757
7	1.917	13	2.869
8	2.140	14	2.971
9	2.329	≥ 15	3
10	2.491		

A site with D_i values greater than 3 is always assumed as a discordant site. Only Regions with a number of sites greater than 5 are presented on Table 1 because for regions with fewer sites the D_i statistic is not functional. Hosking and Wallis (1997) showed that $D_i \leq (N\text{-}1)/3$; therefore, for $N \leq 3$ the matrix A given by Equation 4.2 is singular and D_i cannot be calculated. For regions with $N \geq 7$ the rate of change on D_{crit} when adding a site is lower than for $5 \leq N < 7$.

4.1 Heterogeneity Measure

Several modern hydrologic regionalization procedures can be used to cluster inland streams into homogenous groupings. Once regions are formed, a homogeneity test is performed in order to compute heterogeneity measures and assure that homogeneous regions are obtained. One of the advantages of using *L*-moment based methods for testing homogeneity of a particular region is that they avoid assumptions about the form of the underlying probability distribution of the observed data (Rao & Srinivas, 2008). The homogeneity test evaluates the dispersion of the sample *L*-moment ratios between sites. The measures of dispersion are computed as follows:

a. weighted standard deviation of L-CVs (*V*):

$$V = \left\{ \sum_{i=1}^{N} n_i (t^{(i)} - t^R)^2 \middle/ \sum_{i=1}^{N} n_i \right\}^{1/2} \tag{4.4}$$

b. weighted average distance of L-CV/ L-skewness distances (V_2):

$$V_2 = \sum_{i=1}^{N} n_i \left\{ \left(t^{(i)} - t^R\right)^2 + \left(t_3^{(i)} - t_3^R\right)^2 \right\}^{1/2} \middle/ \sum_{i=1}^{N} n_i \tag{4.5}$$

c. weighted average distance of L-skewness/L-kurtosis (V_3)

$$V_3 = \sum_{i=1}^{N} n_i \left\{ \left(t_3^{(i)} - t_3^R\right)^2 + \left(t_4^{(i)} - t_4^R\right)^2 \right\}^{1/2} \middle/ \sum_{i=1}^{N} n_i \tag{4.6}$$

where *N* is the number of sites in a region, n_i is the record length of site *i*, and $t^{(i)}$, $t_3^{(i)}$ *and* $t_4^{(i)}$ are the *L*-moment ratios of site *i*. The values of $t^R,\ t_3^R,\ and\ t_4^R$ are a regional weighted average of the site's *L*-moment ratios.

Three heterogeneity measures H_1, H_2, and H_3 are obtained by substituting each dispersion measure into Equation 4.7 in which μ_{V_i} and σ_{V_i} are the mean and the standard deviation of the N_{sim} values of the measure of dispersion V_i.

$$H_i = \frac{\left(V_i - \mu_{V_i}\right)}{\sigma_{V_i}} \tag{4.7}$$

The values of the heterogeneity measure are then used to determine whether a region is homogeneous or heterogeneous. The following values of H_i where suggested by Hosking and Wallis (1997) regarding each range: $H_i < 1$ as "acceptably homogenous", $1 \leq H_i < 2$ as "possibly heterogeneous", and $H_i \geq 2$ as "definitely heterogeneous". The accuracy of the quantiles estimate is improved by a lower H.

4.2 Goodness-of-fit Measure

An appropriate regional frequency distribution must be selected once the homogenous regions are defined. The aim is to find a distribution that will yield accurate quantile estimates for each site (Hosking & Wallis, 1997). The goodness-of-fit test is necessary to test whether the flow data follows a GEV or a GPA distribution, or any other distribution used in hydrology. It is intended to make an overall comparison of the observed and hypothetical frequencies and determine if it is the best fit (Kottegoda & Rosso, 1997). Hosking & Wallis (1997) used the L-kurtosis of the fitted distribution (τ_4^{Dist}) and L-kurtosis of the region (t_4^R) to define the goodness-of-fit measure for each distribution:

$$Z^{DIST} = (\tau_4^{DIST} - t_4^R + B_4)/\sigma_4 \tag{4.8}$$

where B_4 is bias of τ_4,

$$B_4 = N_{sim}^{-1} \sum_{m=1}^{Nsim} (t_4^{[m]} - t_4^R), \tag{4.9}$$

And, σ_4 is the standard deviation of L-kurtosis values from the simulation,

$$\sigma_4 = \left[\frac{\sum_{m=1}^{Nsim} \left(t_4^{[m]} - t_4^R \right)^2 - N_{sim} B_4^2}{N_{sim} - 1} \right]^{\frac{1}{2}}. \tag{4.10}$$

On the equations, N_{sim} refers to the number of simulated regional data sets generated using a kappa distribution in a similar manner to the heterogeneity statistics. For more details on the use of kappa distribution on the simulation see section 5.2.3 of Hosking and Wallis (1997). Values of Z^{DIST} lower than 1.64 are considered acceptable, at a confidence level of 90%.

4.3 Index Flood

The basic assumption of the index flood method is that the flood distribution of a homogeneous region is the same except for a site specific scale or an index factor. Therefore, a regionalized frequency analysis with application of the index flood procedure can be used to define the quantiles of particular sites in a region. At an homogeneous region with *N* sites, the quantile function at site *i* is defined as follows:

$$Q_i(F) = \mu_i q(F), \qquad i = 1, \cdots, N \tag{4.11}$$

where *q(F)* is the dimensionless T-year flow value estimated from the region (regional growth curve), and μ_i is the index factor. Hosking & Wallis (1997) recommended the mean of the observed annual peak streamflow data at site *i* as the index flood. They also recommended a method in which the parameters are estimated separately at each site, where *L*-moment parameters are denoted by $\hat{\theta}_k^{(i)}$. The weighted regional average of the site estimates is defined as follows:

$$\hat{\theta}_k^R = \sum_{i=1}^{N} n_i\, \hat{\theta}_k^{(i)} \Big/ \sum_{i=1}^{N} n_i \tag{4.12}$$

Substituting these estimates into *q(F)* results in the estimated regional growth curve $\hat{q}(F) = q\left(F; \hat{\theta}_1^R, \dots, \hat{\theta}_p^R\right)$ where p refers to the undetermined parameters. The site-*i* quantile estimates are obtained from the estimates of μ_i and *q(F):*

$$\hat{Q}_i(F) = \hat{\mu}_i \hat{q}(F) \tag{4.13}$$

For ungaged basins, an index flood ($\hat{\mu}_i$) relates these basins to clusters where data is available. The parameters obtained from the *L*-moments are then used to estimate quantiles *q(F)* for each recurrence interval. Depending on the distribution fitted to each region, the parameters estimated with the *L*-moments are used to define the regional growth curve equation. Substitution the regional parameters into the distribution growth curve yields *q(F).* For the set of distributions considered in this study the quantile functions are defined as follows:

- Generalized Extreme Value (GEV):

$$x(F) = \xi - \frac{\alpha}{k}\left(1 - (-\log F)^k\right) \tag{4.14}$$

Where the location parameter (ξ), scale parameter (α), and shape parameter (*k*) functions are defined as:

$$k \approx 7.8590c + 2.9554c^2 \tag{4.15}$$

$$c = \frac{2}{3+\tau_3} - \frac{\log 2}{\log 3} \qquad (4.16)$$

$$\alpha = \frac{\lambda_2 k}{(1-2^{-k})\Gamma(1+k)} \qquad (4.17)$$

$$\xi = \lambda_1 - \alpha\{1 - \Gamma(1+k)\}/k \qquad (4.18)$$

- Pearson Type 3 (Pe3):

  The distribution's parameters are the first three (ordinary) moment ratios: the mean (μ) as location parameter, the standard deviation (σ) as scale parameter, and the skewness (γ) as shape parameter. If $\gamma \neq 0$, it follows that $\alpha = 4/\gamma^2$, $\beta = \frac{1}{2}\sigma|\gamma|$, and $\xi = \mu - 2\sigma/\gamma$. The probability density function for this distribution is as follows:

$$f(x) = \frac{|x-\xi|^{\alpha-1} exp(-|x-\xi|/\beta)}{\beta^{\alpha}\Gamma(\alpha)}. \qquad (4.19)$$

  To obtain the values of the parameters the value of α is first determined, depending on the values possessed by the skewness (τ_3). If $0 < |\tau_3| < 1/3$, and $z = 3\pi\tau_3^2$, then

$$\alpha \approx \frac{1 + 0.2906z}{z + 0.1882z^2 + 0.0442z^3}. \qquad (4.20)$$

  On the contrary, if $1/3 \leq |\tau_3| < 1$, $z = 1 - |\tau_3|$ then

$$\alpha \approx \frac{0.36067z - 0.59567z^2 + 0.25361z^3}{1 - 2.78861z + 2.56096z^2 - 0.77045z^3} \qquad (4.21)$$

  Consequently, the parameters for the Pearson Type 3 distribution are defined as follows,

$$\gamma = 2\alpha^{-1/2} sign(\tau_3) \qquad (4.22)$$

$$\sigma = \lambda_2 \pi^{1/2} \alpha^{1/2} \Gamma(\alpha) / \Gamma\left(\alpha + \frac{1}{2}\right) \tag{4.23}$$

$$\mu = \lambda_1 \tag{4.24}$$

where λ_1, $\lambda_2, and\ \tau_3$ refers to the mean, L-scale, and L-skewness as defined in Section 3.1.

- Wakeby:

$$x(F) = \xi + \frac{\alpha}{\beta}\left(1 - (1 - F)^{\beta}\right) - \frac{\gamma}{\delta}\left(1 - (1 - F)^{\delta}\right) \tag{4.25}$$

The Wakeby distribution is considered a more robust distribution because it has five parameters: ξ, *α*, *β*, *γ*, and *δ*, allowing for a wider range of distribution shapes. Hosking and Wallis (1997) defined the expression to calculate the parameters values in terms of the L-moments. In the case where ξ is unknow:

$$(N_2C_3 - N_3C_2)z^2 + \ (N_1C_3 - N_3C_1)z + \ (N_1C_2 - N_2C_1) = 0 \tag{4.26}$$

where,

$$\begin{aligned} N_1 &= 3\lambda_2 - 25\lambda_3 + 32\lambda_4, \quad C_1 = 37\lambda_2 - 85\lambda_3 + 203\lambda_4 - 125\lambda_5, \\ N_2 &= -3\lambda_2 + 5\lambda_3 + 8\lambda_4, \quad C_2 = -7\lambda_2 + 25\lambda_3 + 7\lambda_4 - 25\lambda_5, \\ N_3 &= 3\lambda_2 + 5\lambda_3 + 82, \quad C_3 = 7\lambda_2 + 5\lambda_3 - 7\lambda_4 - 5\lambda_5. \end{aligned} \tag{4.27}$$

The parameters *β* and –*δ* are the roots of equation 4.26, with *β* being the largest root value. With this information the rest of the parameters are obtained from:

$$\alpha = \frac{(1+\beta)(2+\beta)(3+\beta)\{(1+\delta)\lambda_2 - (3-\delta)\lambda_3\}}{\{4(\beta+\delta)\}}, \tag{4.28}$$

$$\gamma = -\frac{(1-\delta)(2-\delta)(3-\delta)\{(1-\beta)\lambda_2 - (3+\beta)\lambda_3\}}{4(\beta+\delta)}, \tag{4.29}$$

$$\xi = \lambda_1 - \frac{\alpha}{1+\beta} - \frac{\gamma}{1-\beta}. \tag{4.30}$$

The GEV and Pe3 distributions were considered for this study due to their extensive use in hydrology. The GEV has demonstrated accepted standard errors when fitted to a specific site in Puerto Rico as presented by Sánchez (1995). Similarly, the log-Pearson Type 3 distribution was used on the last report by the USGS as part of the procedure stated in Bulletin 17B by U.S. Interagency Advisory Committee on Water Data (1982).

In this section it has been shown the general steps to compute a flood frequency curve for a defined region. To relate the flood quantiles of a region to a site within the region, the index flood procedure is used. Before all these procedures are applied, sites must be grouped into regions. For this reason, next hydrological regionalization procedures taken under consideration for the analysis are presented.

5. HYDROLOGICAL REGIONALIZATION

Estimating the flood frequency relation for an ungaged site requires performing a regional flood frequency analysis. The procedure is based on the similarity of gaged basins to ungaged ones. "The process of identifying similar watersheds for pooling peak flow information is known as regionalization" (Rao & Srinivas, 2008). Several methods have been developed for this purpose, such as methods of residual and cluster analysis.

The method of residuals, a procedure used frequently by the United States Geological Survey (USGS), relies on the formation of homogeneous regions based on the positive and negative signs of residuals pertaining to a regional regression model that relates flood quantiles at each gauged site to watershed characteristics. Under this method, the delineation of homogenous regions on occasions are arranged to coincide with geographic and/or hydrologic boundaries. Bulletin 17B by U.S. Interagency Advisory Committee on Water Data (1982) describes the procedure for performing the regional analysis. It involves the use of residuals from a regression equation relating the t-year flood level Q_t, at each gauged site to the physical and climatic characteristics of the watershed. Consequently, the t-year flood levels are estimated from a log-Pearson Type III probability distribution as fitted to the log-transformed maximum annual flood series at each gauged site using a regionalized coefficient of skewness.

A more appropriate technique for regionalization to be considered in this study is cluster analysis. Clustering is a process by which a set of feature vectors is divided into clusters or groups such that the feature vectors within a cluster are as similar as possible. Examples of feature vectors are the following: drainage area, average basin slope, main stream slope, stream length, runoff coefficient, geographical location attributes, mean annual rainfall, and precipitation intensities.

Cluster analysis is considered the most practical method to group sites in homogenous regions (Hosking & Wallis, 1997). Clustering algorithms are classified as hard clustering and fuzzy clustering. Hard clustering is structured in that a feature vector has a membership to a cluster of one. On the contrary in fuzzy clustering a feature vector can belong to more than one cluster. Hard clustering can be subdivided as hierarchical and partitional clustering, as seen on Figure 1. Under the hierarchical classification exists the following categories: single linkage, complete linkage, average linkage, and Ward's linkage. These clustering methods differ on how the nearest neighbor to a chosen cluster is defined. For single linkage the distance between clusters is the smallest distance between a pair of feature vectors, for the complete linkage it is the greatest distance between a pair of feature vectors, and for the average linkage it is the average distance between all pairs of feature vectors. In Ward's cluster method the objective function minimizes the sum of squares of deviations of the feature vectors from the centroid of their respective clusters. From these hierarchical algorithms, the average-link clustering, and Wards method tend to form clusters of roughly the same size, producing good results according to Hosking & Wallis (1997).

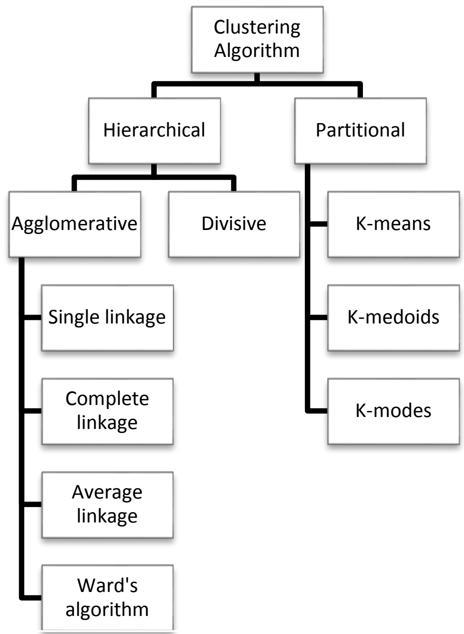


Figure 1. Hard clustering algorithms

(Adapted from Rao & Srinivas, 2008)

The partitional clustering is categorized in three algorithms: K-means, K-medoids and K-modes. In the K-means, the mean of feature vectors within the cluster is the representative centroid, while for the K-medoids it is the median of feature vectors. The K-modes is the result of a modified K-means algorithm. The mode of feature vectors is the representative centroid.

5.1 Hybrid Clustering

Rao & Srinivas (2008) describe other types of cluster analysis: hybrid cluster analysis, fuzzy cluster analysis and artificial neural networks. The hybrid clustering analysis mentioned by Rao & Srinivas (2008) is a combination between the hierarchical and partitional clustering. Hierarchical clustering has a disadvantage in that the feature vector cannot move from one cluster to another minimizing the objective function, an advantage that exists in the partitional clustering. In an investigation performed by Rao & Srinivas (2006), three hybrid-clustering algorithms where investigated: single linkage, complete linkage, and Ward's linkage combined with the partitional cluster, K-means. The main objective of the K-means analysis is to minimize an objective function F, defined by:

$$F = \sum_{k=1}^{K} \sum_{j=1}^{n} \sum_{i=1}^{N_k} d^2\left(x_{ij}^{k} - x_{\cdot j}^{k}\right) \tag{5.1}$$

where d denotes distance between clusters, K denotes the number of clusters, N_k represents the number of feature vectors in cluster k; x^k_{ij} denotes the rescaled value of attribute j in the feature vector i assigned to cluster k; and $x^k_{\cdot j}$ is the mean value of attribute j for cluster k. It was found that the hybrid cluster between Ward's linkage and K-means provides the lowest F values. Even though the procedure eventually did not provide fully homogeneous areas, the results were considered acceptable. For a regional flood frequency analysis, all regions must be homogenous to attain the best result.

Chipman & Tibshirani (2006) propose another use of the hybrid clustering algorithm in a study to cluster gene expression data of breast cancer tumors. The algorithm is a merge between the agglomerative and divisive clustering methods, in addition to taking into consideration the mutual cluster. The mutual clusters are formed by individuals that are closer to each other than to any other individual (Chipman & Tibshirani, 2006). The main idea is that the mutual points should not be separated. By definition the distance (d) between any point on cluster S must be less than any other point not located in *S:*

$$\forall x \in S, y \notin S, d(x, y) > diameter(S) \equiv max_{w \in S, z \in S} d(w, z). \qquad (5.2)$$

The agglomerative or bottom-up method forms clusters by joining individuals with the nearest distance, while the divisive or top-down method forms clusters by successively separating the individuals from a unified group. The algorithm used for the top-down method is the K-means with K = 2 being applied in a recursive way, also known as tree-structure vector quantization (TSVQ). The algorithm proposed by Chipman & Tibsshirani (2006) first defines the mutual clusters, then performs a top-down clustering via K-means, and lastly applies the K-means algorithm within each mutual cluster. The distance metrics used to define the cluster in K-means is the square Euclidean distance because it is suitable for clusters with spherical shapes.

5.2 Fuzzy Clustering

Another analysis proposed by Rao & Srinivas (2008) is fuzzy clustering, which has the flexibility of allowing basins to be grouped into more than one cluster simultaneously. This method is dependent in iterative optimization of a fuzzy objective function. The objective function is defined as

$$Minimize\ J(U,V:X) = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^{\mu} d^2(x_{ik}, v_i) \qquad (5.3)$$

where,

$$d^2(x_{ik}, v_i) = (x_{ik} - v_i)^T A_i (x_{ik} - v_i) \qquad (5.4)$$

and, it is subjected to the following constraints,

$$\textstyle\sum_{i=1}^{c} u_{ik} = 1 \qquad \forall k \in \{1,2,\dots,N\} \qquad (5.5)$$

$$0 < \textstyle\sum_{k=1}^{N} u_{ik} < N \qquad \forall i \in \{1,2,\dots,c\}. \qquad (5.6)$$

On the equations, d represent the distance, x_{ik} denotes the rescaled value of y_{ik}, which in turn represents the value of attribute i in a k-th feature vector, v_i characterizes the centroid of one of the i cluster, μ is a fuzzifier, and u_{ik} denotes the membership of a feature vector in a i-th fuzzy cluster.

To identify optimal allocation of feature vectors on c clusters, the author describes the following validity measures: partition coefficient, partition entropy, fuzziness performance index, normalized classification entropy, extended Xie-Beni, Fukuyama-Sugeno, and Kwon's index. During the process of application of the algorithm the more effective indices must be selected for identification of optimal regions. Rao & Srinivas (2008) found that the fuzzy clustering algorithm generated less effort in adjusting the clusters to find homogenous regions.

Bhaskar & O'Connor (1989) compared the methods mentioned earlier, residuals and cluster analysis, in the state of Kentucky. The method of residuals was performed using statistics of hydrologic variables, discriminant analysis, and a regression analysis relating the t-year log-Pearson Type-III flood quantile to watershed physical characteristics. On the cluster analysis, regions were delineated using the coefficient of variation of the log-transformed maximum annual flood series (LCV), and the specific mean annual flood (QSP). Comparing the results, it was concluded that cluster regions have lower variability within each of the regions with respect to the mean and median of hydrologic variables. The discriminant analysis indicated that the regions formed by the method of residual display different watershed characteristics than those formed by the method of residual discrimination values.

From the methods discussed above, the hybrid cluster analysis by Chipman & Tibshirani (2006) will be used as the procedure to group sites into regions in Puerto Rico. This method is simple and the definition of the mutual clusters is convinient when ungaged basins need to be assigned to a region. Although Rao & Srinivas (2008) identified the fuzzy clustering method as an easier method for adjustment, a site can be grouped in more than one cluster simultaneously. The characteristic of this method can create difficulties if used in a flood frequency analysis since it is expected that a site is only contained by one region.

6. AREA OF STUDY

The study area encompasses a total of 166 USGS stations with peak streamflow data. A screening of the data was performed, eliminating sites affected by hydraulic structures, short records, and broken records. McCuen & Galloway (2010) made a study that provides a graphical approach to assess the expected accuracy in 2-, 10-, and 100-year magnitudes resulting from a log-Pearson Type III analysis. With a 90 % confidence the relative error for stations with n = 10 was approximately 1.6, while for n = 25 it was 1.0. This suggest that in order to minimize the introduction of large errors in the analysis due to short records, only sites with records greater than 25 years should perhaps be selected. In addition, all stations located downstream of a flow regulation structure and channels were not included in the study. Another consideration for the selection of the stations was the number of years without data. Only those stations missing one year where included in the analysis, leaving a total of 30 stations. These are listed in Table 2, and located in Figure 2 on a map of Puerto Rico. Most of the stations conglomerated are in the eastern part of Puerto Rico, while the others are more scattered across the southern and western regions.

Table 2. USGS stations evaluated in this study

id	Site Number	Station Name
1	50028000	RIO TANAMA NR UTUADO
2	50028400	RIO TANAMA AT CHARCO HONDO
3	50031200	RIO GRANDE DE MANATI NR MOROVIS
4	50034000	RIO BAUTA NR OROCOVIS
5	50035000	RIO GRANDE DE MANATI AT CIALES
6	50038100	RIO GRANDE DE MANATI AT HWY 2 NR MANATI
7	50038320	RIO CIBUCO BLW COROZAL
8	50039500	RIO CIBUCO AT VEGA BAJA
9	50043000	RIO DE LA PLATA AT PROYECTO LA PLATA
10	50047850	RIO DE BAYAMON NR BAYAMON
11	50050900	RIO GRANDE DE LOIZA AT QUEBRADA ARENAS
12	50051310	RIO CAYAGUAS AT CERRO GORDO
13	50055000	RIO GRANDE DE LOIZA AT CAGUAS
14	50056400	RIO VALENCIANO NR JUNCOS
15	50057000	RIO GURABO AT GURABO
16	50061800	RIO CANOVANAS NR CAMPO RICO
17	50063800	RIO ESPIRITU SANTO NR RIO GRANDE
18	50064200	RIO GRANDE NR EL VERDE
19	50065500	RIO MAMEYES NR SABANA
20	50067000	RIO SABANA AT SABANA
21	50071000	RIO FAJARDO NR FAJARDO
22	50075000	RIO ICACOS NR NAGUABO
23	50090500	RIO MAUNABO AT LIZAS
24	50092000	RIO GRANDE DE PATILLAS NR PATILLAS
25	50112500	RIO INABON AT REAL ABAJO
26	50115000	RIO PORTUGUES NR PONCE
27	50124200	RIO GUAYANILLA NEAR GUAYANILLA
28	50138000	RIO GUANAJIBO NR HORMIGUEROS
29	50144000	RIO GRANDE DE ANASCO NR SAN SEBASTIAN
30	50147800	RIO CULEBRINAS AT HWY 404 NR MOCA

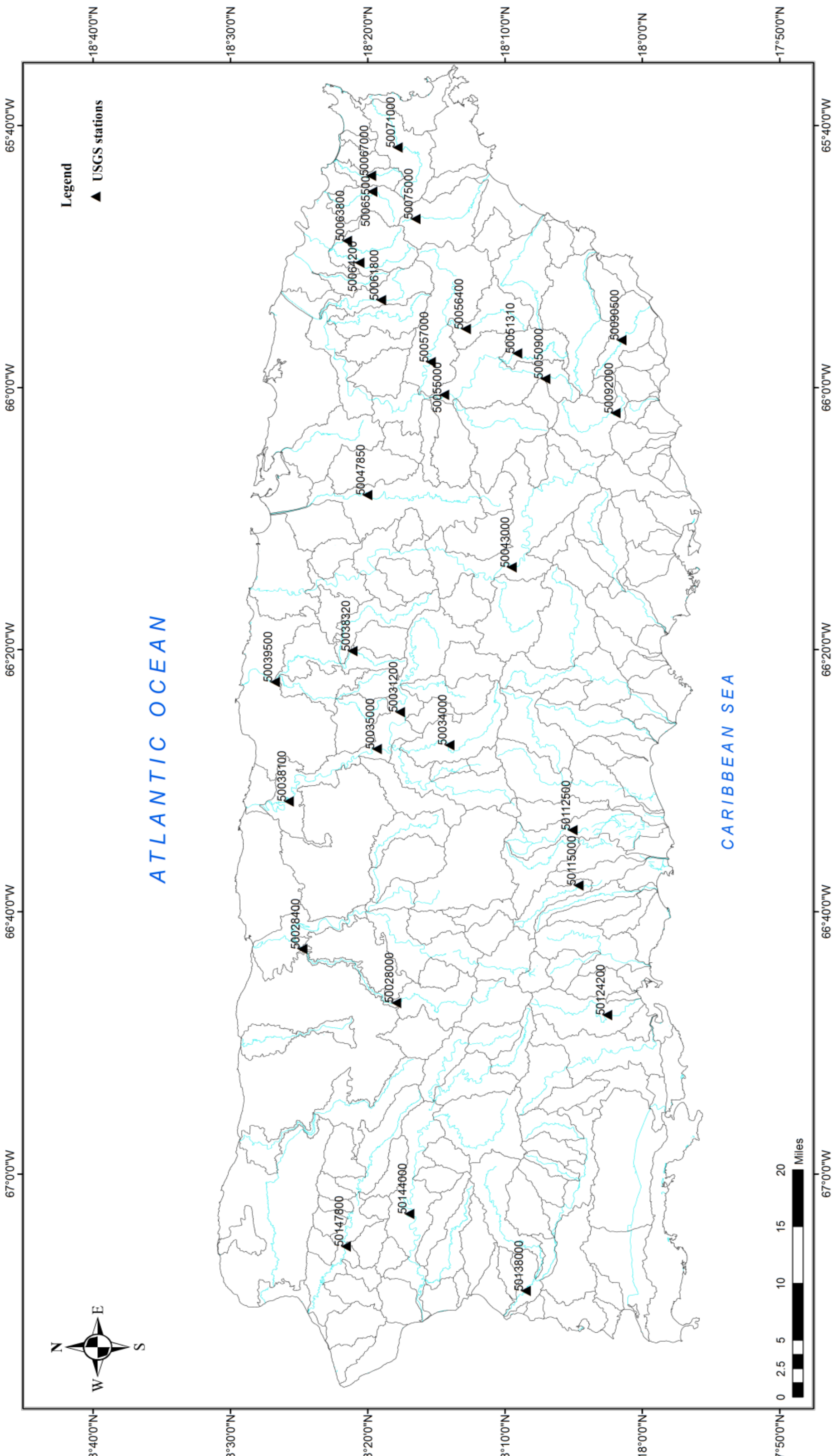


Figure 2. Location of USGS stations selected for the flood frequency analysis.

7. RESULTS AND DISCUSSIONS

7.1 Sample *L*-moments and *L*-moment ratios

The *L*-moments and *L*-moment ratios (l_1, t, t_3, t_4 and t_5) of each USGS station are presented in Table 13 in Appendix A.1, where its *L*-statistics are displayed. These were calculated for the 30 stations in Table 2 using the R source code (Hosking, 2008). The *L*-moment ratio diagram is presented on Figure 3. In the diagram it is observed that the *L*-moment ratios are spread through the values of 0.03 to 0.6 for the *L*-skewness, and from 0.1 to 0.47 for the *L*-kurtosis.

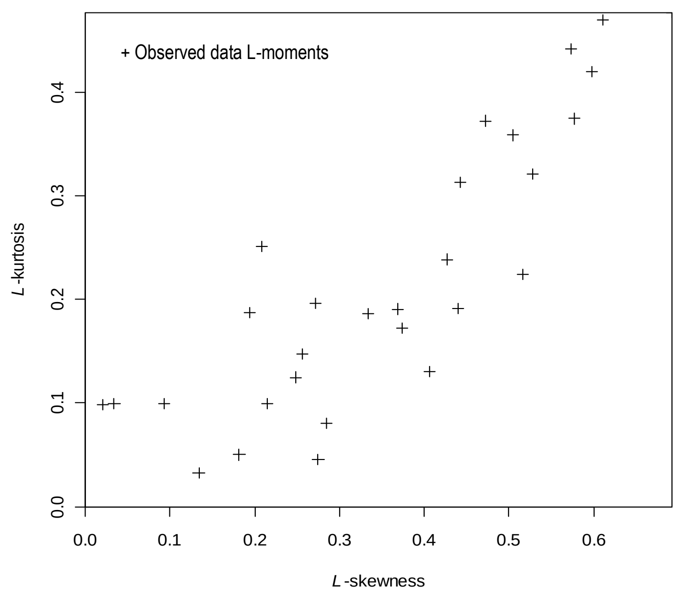


Figure 3. *L*-moment ratio diagram for the 30 USGS sites selected.

Using the *L*-moment statistics results, the discordancy measure and the heterogeneity measures were computed. The discordancy measure (D_i) was calculated using Equation 4.1 to determine which stations have potential errors in the data at the particular site. Considering Puerto Rico as a single region, only one station presented a value greater than 3.0, the critical value for regions with more than 15 stations. Station number 50075000, located on Rio Icacos near Naguabo presented a value of 3.47, and for this reason it was excluded from the study. The exclusion of discordant sites from the analysis, the use of records with potential data errors can be avoided. The heterogeneity measures H_1, H_2, and H_3 were computed from Equation 4.7, yielded values of 9.64, 8.62, and 5.73, respectively. These values demonstrate that the area of study is heterogeneous and must be divided into sub-regions.

7.2 Cluster Analysis

The cluster analysis was performed using only site characteristic data which includes physical and climatological basin characteristics. Hosking & Wallis (1997) strongly recommend the use of site characteristics because at-site statistics are used to determine the homogeneity of the group, and the assessment can be affected when at-site statistics are used on both procedures. Most of the site characteristics for this analysis were obtained from the USGS report by Ramos-Ginés (1999). In this USGS report the following characteristics were described: location, contributing drainage area (CDA), depth-to-rock (DR), mean annual rainfall (MAR), average permeability of soils (SP), vegetative cover (VC), main channel slope (CS), main channel length (CL), average of the ground slope (GS), and the average T-year 24-hour rainfall intensity (RI-T). MAR values were obtained from a mean annual rainfall raster from climate data collected at weather stations from 1990 to 2000. The correlation coefficient between the mean annual flow and the sites characteristics was calculated to determine which characteristics must be selected for further analysis. Results are shown in the following table.

Table 3. Correlation between mean annual flow (MAF) and site characteristics.

Site Characteristics	Correlation	Site Characteristics	Correlation
MAF vs CDA	0.920	MAF vs RI-10	-0.347
MAF vs CS	-0.495	MAF vs RI-25	-0.372
MAF vs MAR	0.0129	MAF vs RI-50	-0.321
MAF vs SP	-0.220	MAF vs RI-2*CDA	0.911
MAF vs VC	-0.095	MAF vs RI-5*CDA	0.912
MAF vs CL	0.881	MAF vs RI-10*CDA	0.911
MAF vs GS	0.014	MAF vs RI-25*CDA	0.904
MAF vs RI-2	-0.395	MAF vs RI-50*CDA	0.903
MAF vs RI-5	-0.415	MAF vs MAR*CDA	0.917

Table 3 shows that the mean annual flow is highly correlated to the contributing drainage area (correlation = 0.920), and also to the channel length (correlation = 0.881). Since the contributing drainage area has a greater effect on the floods, it was selected as one of the independent attributes for the cluster analysis. Another attribute considered in the analysis was the rainfall. The mean annual rainfall (MAR) demonstrates a high correlation value when the volume of rainfall within the area of the basin was evaluated. In Table 3, channel slope presents a low correlation value (correlation = -0.495). Even though the channel slope correlation did not yield an acceptable value, it was nevertheless included in the cluster analysis having then two physical characteristics for the analysis. The final selection of site characteristics as feature vectors for the cluster analysis are presented in Table 4.

Table 4. Selected sites characteristics used in regionalization.

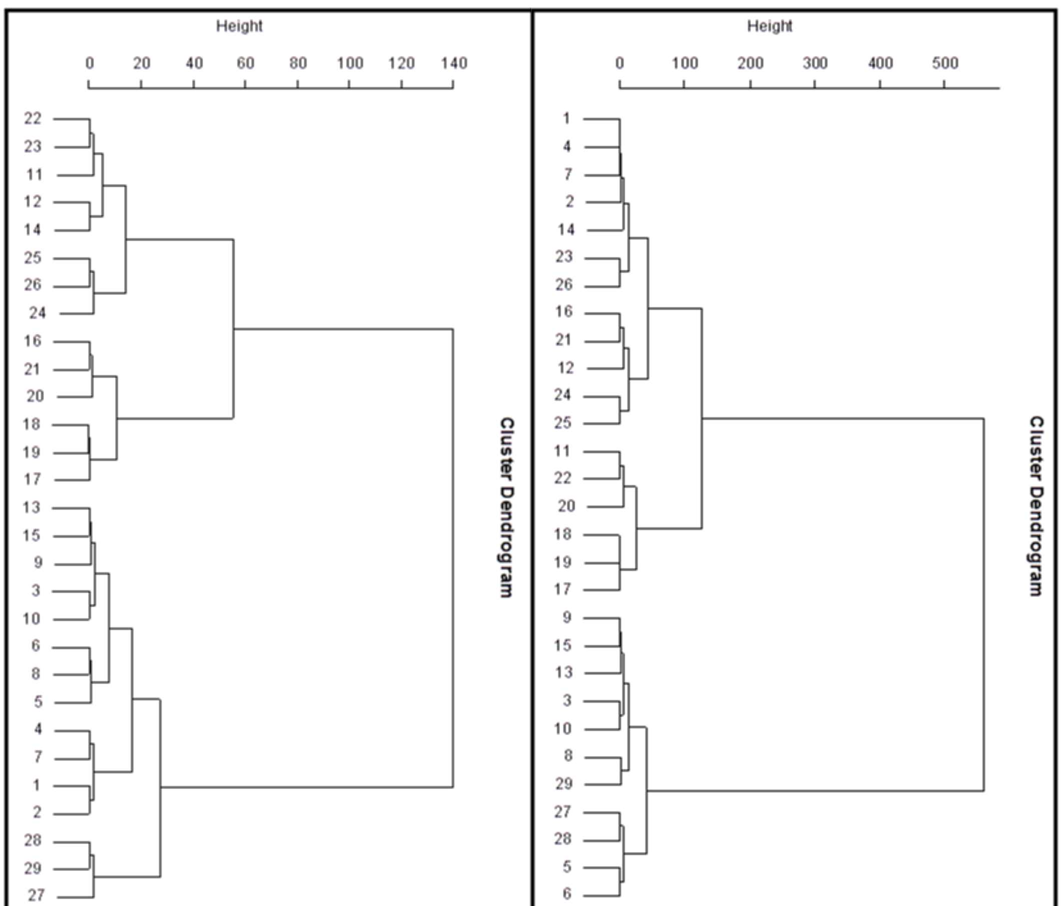
(CDA and CS characteristics adapted from Ramos-Ginés, 1999)

id	Site Number	Latitude	Longitude	CDA (mi^2)	MAR (in)	CS (ft/mi)
1	50028000	18.30056	66.78278	18	72.32	139.7
2	50028400	18.41444	66.71444	22.2	63.11	102.4
3	50031200	18.29583	66.41306	55.2	68.50	86.25
4	50034000	18.23611	66.455	16.7	74.61	89.62
5	50035000	18.32389	66.46	134	68.19	76.76
6	50038100	18.43111	66.52694	165	59.76	54.57
7	50038320	18.35361	66.33528	15.2	62.99	198
8	50039500	18.44806	66.37472	81.6	59.33	73.6
9	50043000	18.16028	66.22889	63.2	59.37	69.37
10	50047850	18.33556	66.13694	41.7	59.33	59.79
11	50050900	18.11944	65.98944	5.99	72.76	334.4
12	50051310	18.1575	65.95806	10.2	77.56	55.37
13	50055000	18.2425	66.00944	89.6	63.23	105
14	50056400	18.21611	65.92611	16.4	69.17	126
15	50057000	18.25833	65.96806	60.1	66.34	171.3
16	50061800	18.31889	65.88917	10.2	66.97	332.4
17	50063800	18.36028	65.81361	8.64	64.80	364.5
18	50064200	18.345	65.84167	7.34	65.59	449.6
19	50065500	18.32944	65.75111	6.8	62.60	492.2
20	50067000	18.33111	65.73111	3.91	60.28	406.1
21	50071000	18.29889	65.695	14.8	52.95	281.2
22	50090500	18.02722	65.94	5.29	66.10	225.9
23	50092000	18.03444	66.03278	18.4	60.94	227
24	50112500	18.08611	66.56278	9.68	60.51	476.2
25	50115000	18.07917	66.63361	8.8	58.15	280
26	50124200	18.04444	66.79806	18.9	50.83	238.6
27	50138000	18.14333	67.14917	120	62.36	75.35
28	50144000	18.28472	67.05139	134	74.84	81.67
29	50147800	18.36167	67.0925	71.3	65.91	53.41

It has been recommended that the variables selected for the cluster analysis be transformed in order to avoid scale differences (Rao & Srinivas, 2006;

Chipman & Tibshirani, 2006). The procedure to assign the transformations can be arbitrary, and characteristics that are deemed more important can be assigned a greater value. The transformations applied to the group of characteristics in this study were chosen following an application presented by Hosking & Wallis (1997) for the formation of regions for Appalachian streamflow data. A nonlinear transformation was achieved by applying a logarithmic transformation to the drainage area; in this manner, sites with outlier values can themselves form a cluster. The variables were standardized by dividing the individual characteristics' values by the standard deviation of the transformed characteristic. Two different approaches (*a* and *b*) were followed in terms of the importance of the characteristics of sites assigned to the clusters. Approach *a* assumes that all characteristics have the same importance. On the contrary, approach *b* took under consideration that the contributing drainage area is highly correlated with the mean annual flow; therefore, the area was multiplied by 4.0 to assign an importance value equal to the other characteristics all together. The R program algorithm used for both approaches to cluster analysis is presented on Appendices B.1 for approach *a,* and B.2 for approach *b*. Let $X_{gh,}$ be the nomenclature to identify the clusters evaluated where X denotes the cluster, g denotes approach and h denotes the number of clusters that the data was divided.

A dendrogram is a visual representation of relationship between sites. The height on the dendrogram represents the difference in distance between sites. The dendrograms illustrated in Figure 4 shows that the differences in approaches had an impact in the manner that the sites were assigned to each cluster. The Euclidean distance between those sites seems to change within both approaches. Other sites, such as sites 16 and 21, demonstrated to be mutual in both methods, demonstrating that the characteristics between these sites are very similar. If the island were divided into only two regions, approach *a* (characteristics equally important) would have a similar number of sites in both regions (14 sites for Cluster A_{a2} and 15 sites for Cluster B_{a2}). Under approach *b*, assuming that area is the most important characteristic, Cluster A_{b2} would be larger, comprising 18 sites, while Cluster B_{b2} would include of 11 sites.



(a) (b)

Figure 4. Cluster dendrograms from hybrid algorithm

(a) Equal importance characteristics

(b) Area important.

Comparing Clusters A and B between both approaches, sites 1, 2, 4, and 7 would belong to Cluster B_{a2} under approach *a,* and to Cluster A_{b2} under approach *b*. Further differences are found when the sites are grouped into four clusters. Due to the limitation in the number of sites available, it is not recommended to use more than four clusters. The next step is to evaluate whether the clusters are homogeneous by applying the heterogeneity measures to each cluster, and then, decide which is the best configuration for grouping the sites.

7.3 Discordancy and heterogeneity measures

Now that the gauging stations have been grouped into regions, the discordancy and heterogeneity measures must be computed to determine if the regions can be classified as homogeneous. As discussed earlier, the aim is to define the most probable homogeneous regions because this represents the similarity of the *L*-moments between the sites. The test for heterogeneity was carried out for both approaches and the results are displayed in Table 5. In approaches *a* and *b* it is not suitable to split the island in two regions because the resulting clusters are "definitely heterogeneous" ($H_i > 2$). When four regions are considered, the heterogeneity values are lowered considerably in Cluster A_{a4} ("acceptably homogeneous"), Cluster B_{a4} ("possibly heterogeneous") and Cluster D_{b4} ("possibly heterogeneous"). Because of the higher values of the heterogeneity measure in the remaining clusters, those regions are classified as "definitely heterogeneous". Considering the lower values of H_i under approach *a*, this approach is selected for further modification until the values of H_i are lowered.

Table 5. Heterogeneity measures for approaches *a* (characteristics equally important) and *b* (area important).

Number of Regions	Cluster	H_1	H_2	H_3
Characteristics Equally Important				
2 Regions	Cluster A_{a2}	3.48	5.41	4.27
	Cluster B_{a2}	7.82	6.2	4.41
4 Regions	Cluster A_{a4}	-0.3	1.27	1.17
	Cluster B_{a4}	1.3	1.36	0.92
	Cluster C_{a4}	4.54	3.8	2.45
	Cluster D_{a4}	5.31	4.74	3.12
Area Important				
2 Regions	Cluster A_{b2}	4.32	5.78	4.31
	Cluster B_{b2}	7.19	5.06	4.28
4 Regions	Cluster A_{b4}	3.38	4.29	3
	Cluster B_{b4}	2.88	4.1	3.5
	Cluster C_{b4}	7.08	5.07	3.47
	Cluster D_{b4}	1.82	2.34	2.12

Hosking & Wallis (1997) recommended that regions resulting from clustering analysis be adjusted to obtain lower *H* values for the RFFA. In this study lower *H* values were obtained transferring one or more sites from one region to another, and eliminating one or more sites from the data set. Cluster C_{a4} was divided and various sites were transferred to Cluster $A_{a4.}$ In addition, one mutual cluster from Cluster A_{a4} was transferred to Cluster $B_{a4.}$ Despite the fact that Cluster D_{a4} is "definitely heterogeneous", when sites were transferred to another cluster, the H_i values increased. These sites are located on western Puerto Rico and the number of sites that meets the requirement for sites selection on this study is considerably low, with only 3 sites. The USGS station 50038320 was eliminated from the cluster analysis since its incorporation into any cluster increased the heterogeneity measure. The final structuring of clusters (Cluster 1, Cluster 2, Cluster 3, and Cluster 4), and their *L*-moment ratio diagram is presented in Appendix C.1. The diagrams in this appendix shows the sites *L*-skewness and

L-kurtosis, and the curves for each distribution. The diagrams provide a image representation of which distribution better fits the sites in each specific region. In addition,

Figure 5 illustrates the location of the individual sites and their corresponding cluster. Finally, as show on Table 6, Cluster 1 is considered “acceptably homogeneous”, Cluster 2 and Cluster 3 as “possibly heterogeneous”, and Cluster 4 as “definitely heterogeneous”.

Table 6. Heterogeneity measures for final selection of clusters.

Cluster	**Heterogeneity Measures**		
	H_1	H_2	H_3
Cluster 1	0.66	0.68	-0.11
Cluster 2	1.29	1.25	0.82
Cluster 3	1.31	1.54	1.6
Cluster 4	5.1	4.75	3.15

For Cluster 1 the heterogeneity measure obtained from the measured based on *L*-skewness and *L*-kurtosis (H_3) has a negative value. Hosking and Wallis (1997) explain that negative *H* values indicates that there is less dispersion between the at-site sample *L*-CV values when compared to what is expected from an homogeneous region. They recommend that for values lower than -2, the data must be examined for possible cross-correlation between sites frequency distributions. Consequently, the value of -0.11 is not a major concern. Hereafter, the groups of sites will be referred to as regions instead of clusters.

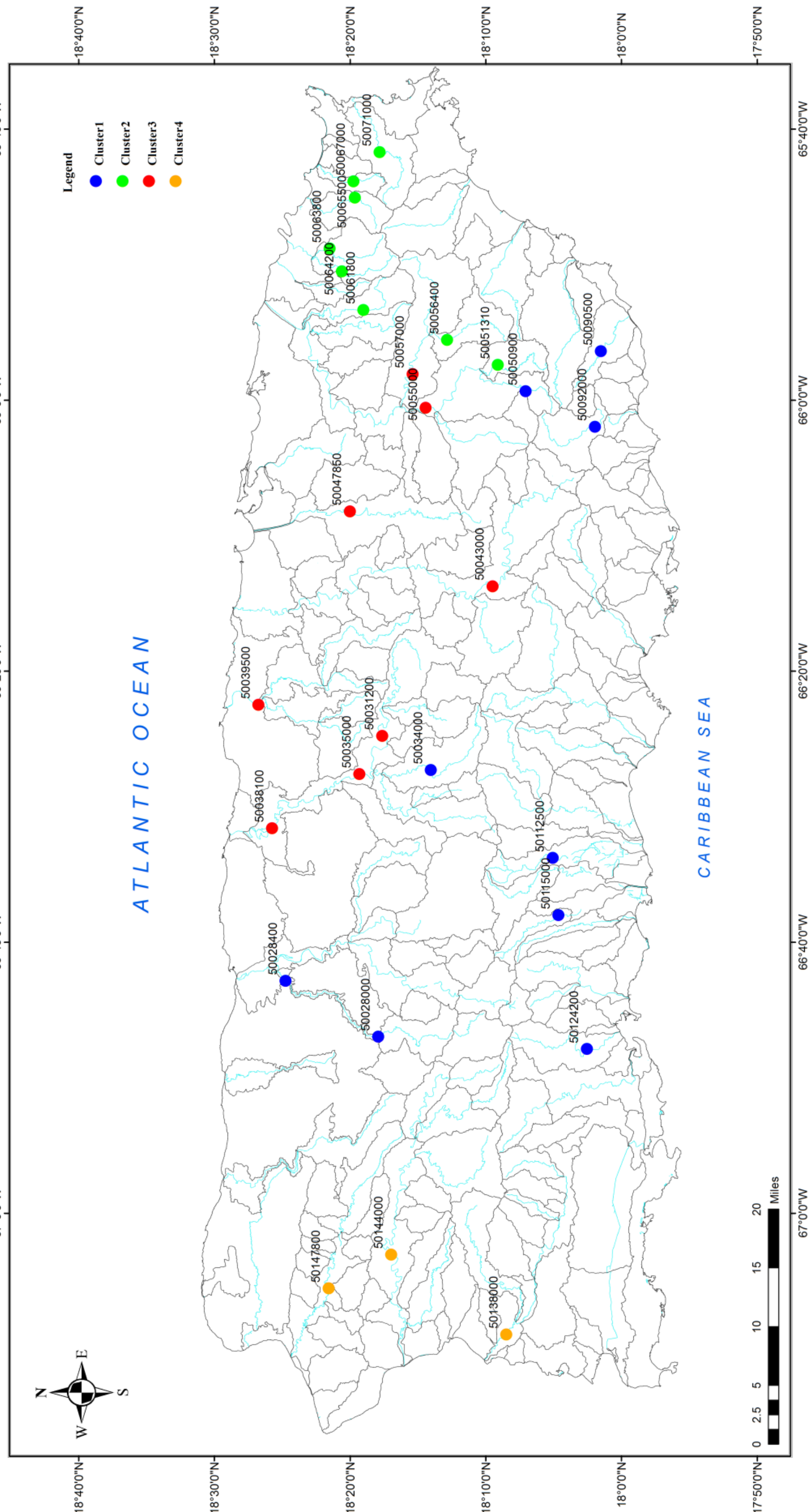


Figure 5. Assignment of sites to clusters

Once the regions were defined, the goodness-of-fit measure was calculated for each. Values of Z^{Dist} were computed using the L-kurtosis of the sample region and each fitted distribution. As shown in Table 7, Regions 1 and 4 possess acceptable Z values for the GEV distribution, while the Pearson Type 3 fits best Regions 2 and 3.

Table 7. Goodness-of-fit-measures

Region	**Distributions *Z* values**			
	GEV	**GNO**	**Pe3**	**GPA**
Region 1	-0.54	-1.4	-2.86	-1.65
Region 2	2.32	1.81	0.84	-1.56
Region 3	3.01	1.96	0.14	0.44
Region 4	-0.9	-1.36	-2.16	-1.63

It was initially desired to use the same distribution for all regions under study, but this could not be achieved since not all regions under same distribution comply with the critical value $Z_{crit} = 1.64$. The GEV has provided a good fit to flood data at many locations but not on a regional level in all cases. Even though the GEV distribution fits Cluster 4, it is nevertheless a heterogeneous region. The Wakeby distribution is mentioned by Hosking & Wallis (1997) to be a more robust frequency distribution, and may well provide a better estimate for a heterogeneous area. Because of this reason the Wakeby distribution was chosen to fit Region 4.

7.4 Regional Quantiles

The index flood was used as the basis for the estimation of the regional growth curves. The dimensionless T-year flood value, *q(F)*, was estimated for each region. The values are presented in Table 8.

Table 8. Estimated quantiles for clustered regions in Puerto Rico for various return periods.

Region	Quantile					Fitted Distribution
	5 year	10 year	25 year	50 year	100 year	
1	1.30	1.91	3.01	4.16	5.68	GEV
2	1.45	1.85	2.33	2.69	3.03	Pe3
3	1.56	2.22	3.12	3.80	4.48	Pe3
4	1.33	1.95	3.03	4.10	5.43	Wakeby

Regions 1 and 4 have similar regional quantile values from 5- to 100- years return period, and also a higher upper tail when compared to the other regions. In a GEV distribution the upper tail behavior is related to the shape parameter k. For values of $k < 0$ the tail weight is heavier for a GEV distribution than for a Pe3 distribution. Hosking and Wallis (1997) define the tail weight as "the rate at which quantile increases as return period is extrapolated beyond the range of the data". For a Wakeby distribution the tail is heavier when $\delta > 0$. In Region 1, $k = -0.41$, and in Region 2, $\delta = 0.33$, thus it is expected that these regions have heavier tails than Regions 2 and 3.

7.5 Quantiles Estimation Errors

In order to assess the overall deviation of the estimated quantiles from the true quantiles, the regional average root mean square error (RMSE) was computed. Hosking & Wallis (1997) assert that the relative RMSE is the criterion to which they give most weight in judging whether one estimation procedure is superior to another. The relative RMSE of the site-i quantile is defined by the following:

$$R_i(F) = \left[M^{-1} \sum_{m=1}^{M} \left\{ \frac{\hat{Q}_i^{[m]}(F) - Q_i(F)}{Q_i(F)} \right\}^2 \right]^{\frac{1}{2}} \times 100\% \tag{7.1}$$

where M is the total of repetitions, $\hat{Q}_i^{[m]}(F)$ is the site-i quantile estimate for nonexceedance probability F, and $Q_i(F)$ is the site-i quantile true value implied by the specified distribution. For an evaluation of the performance of an estimation over a region, the regional average relative RMSE of the estimated quantile is calculated as follows:

$$R^R(F) = N^{-1} \sum_{i=1}^{N} R_i(F) \tag{7.2}$$

The available source code used to determine the relative RMSE is based on a Monte Carlo simulation using the same characteristics from the region under study. On the simulation, the L-moment ratios at the individual simulated sites are in accordance with the heterogeneity measure for the region (Hosking & Wallis, 1997). The observed regional estimates and the simulated values were compared, determining the accuracy of the procedure in estimating quantiles.

Table 9 shows the percent values obtained for the regional average relative RMSE. This values are considerably low for all regions, but it was expected that

Region 4 would yield a higher RMSE value (RMSE=0.79) because of its heterogeneous classification.

Table 9. Regional average relative root mean square error values for various return periods.

Region	Quantile RMSE %				
	5 years	**10 years**	**25 years**	**50 years**	**100 years**
Region 1	0.048	0.026	0.120	0.301	0.601
Region 2	0.010	0.029	0.061	0.087	0.114
Region 3	0.008	0.037	0.095	0.145	0.198
Region 4	0.084	0.086	0.179	0.398	0.790

These RMSE values are not comparable with the RMSE results from the USGS latest report (Ramos-Ginés, 1999). The approach used on the USGS report is based on the differences between the estimated T-year quantile at site *i* based on at-site streamflow record ($Q_{rec,i}$), and the regional regression estimate at site *i* ($Q^*_{rec,i}$). The USGS used formulas to determine the RMSE were the following:

$$RMSE_{log} = \left[\frac{\sum\left(logQ_{rec,i} - logQ^*_{rec,i}\right)^2}{N_p}\right]^{\frac{1}{2}} \quad (7.3)$$

and in percent,

$$RMSE_{\%} = 100\left(e^{(5.302)\left(RMSE_{log}\right)} - 1\right)^{\frac{1}{2}} \quad (7.4)$$

Thus, the meaning of the RMSE computed by the USGS is different from that used on this study. Ramos-Ginés (1999) claims that the regression equation used to estimate the 100-year flood quantile yielded a RMSE of 43 percent, the smallest value obtained when compared to previous studies. Therefore, the 43 percent refers to the RMSE between two estimated quantiles and not to the accuracy of the individual estimate, unlike the results obtained through the Monte Carlo simulation.

The empirical quantile of the distribution of estimates is obtained by calculating the average on a region of the ratio between estimated and true quantile values, and accumulating a histogram of the values resulting from the ratio. Then, the 90% error bounds for the true quantile lie within the interval,

$$\frac{\hat{Q}(F)}{U_{0.5}(F)} \leq Q(F) \leq \frac{\hat{Q}(F)}{L_{0.5}(F)} \,. \tag{7.5}$$

In Figure 6, the continuous line represents the regional quantile values while the dotted lines indicates the lower and upper error bounds of the regional quantile values for a 90% confidence. The regional growth curves illustrate how the error bounds increase for return periods approximately greater than 20 years. Also, for Region 1 and 4 the upper bounds in the 100-year return period are wider. For Region 1 the GEV parameter k is less than 0, thus when the distribution is fitted to the data, the data will not lie close to the upper bound, as opposed to when $k > 0$ (Hosking and Wallis, 1997). The wider upper bound in these regions is also related to the heavier tails that both regions have.

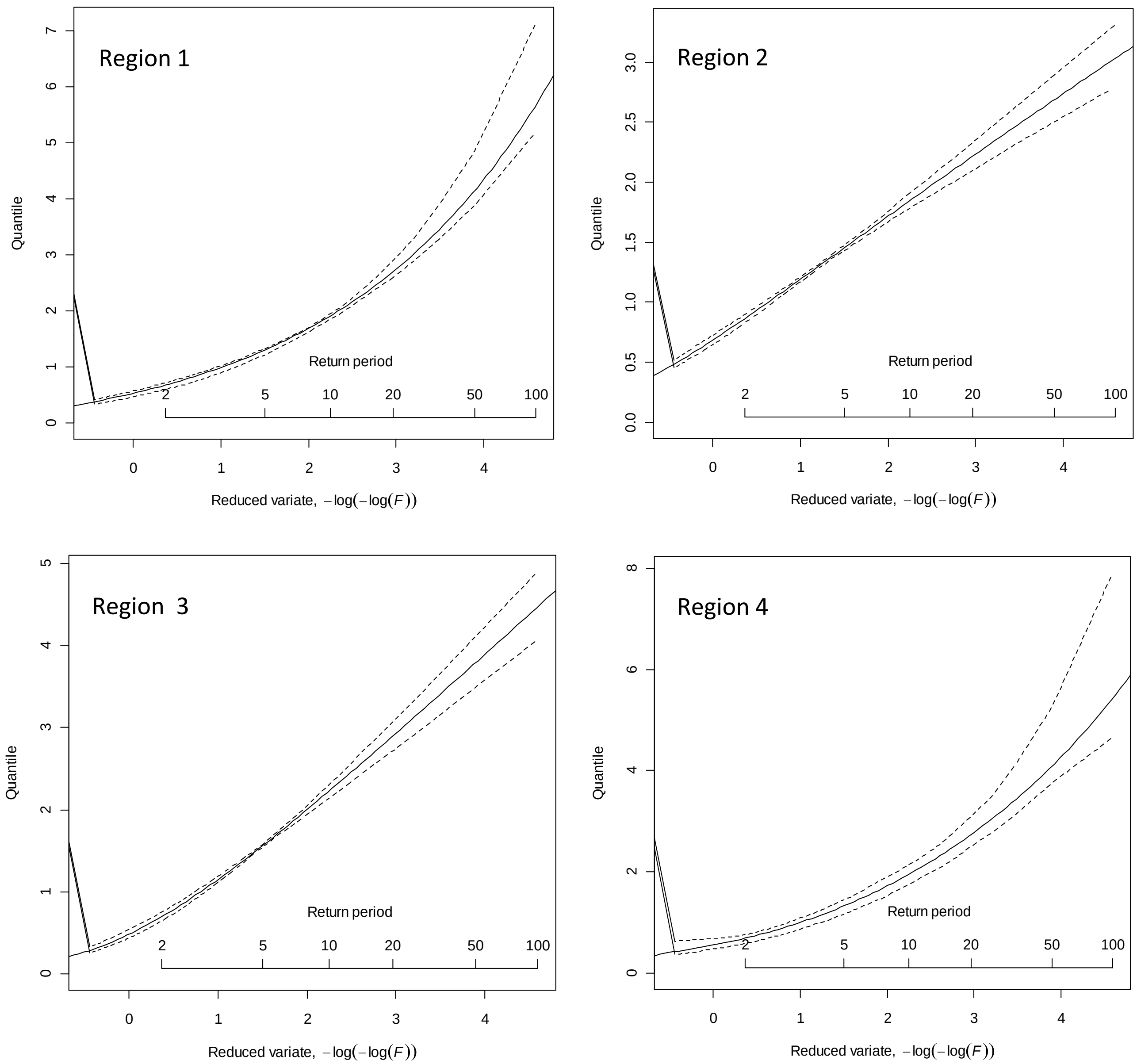


Figure 6. Regional growth curves and error bounds for 90% confidence

7.6 Comparison of regional and at-site estimation

At-site estimates were performed to compare and evaluate the differences with the regional estimates. It was assumed that the GEV distribution was the best fit for each site. The resulting percent errors were calculated and presented in Table 24 on Appendix C.3, page 117. These values are used to compare the magnitude of the difference between the two quantile values. The formula for the percent error, where the at-site quantile ($q_i(F)$) is considered as the real value and the regional site quantile ($\hat{q}_i(F)$) as the estimated value, is given by:

$$\%\ error = \frac{\hat{q}_i(F) - q_i(F)}{q_i(F)}. \tag{7.6}$$

For USGS station 50028000 from Region 1, located in Tanamá River, the percent error for 100-year estimate is equal to 95%. Figure 7 is shows the plot of the site quantile estimated from the regional growth curve (black line) and the at-site quantile (red line) obtained from the GEV distribution. For return periods less than 50-years the at-site quantile fits the data accurately, but for the 100-year return period the regional estimate quantile curve is closer to the data. Even though the percent error is high for this station, the regional estimate is a better estimate for the higher return period at this station, as seen on Figure 7. Other consideration when analyzing extreme data, like annual peak streamflow, is the accuracy of the flow value measured at the river. For most extreme events, the magnitude of the flows is computed via approximate means since these flows are usually beyond the range of the streamflow gage calibration curves for the river.

Contrary to station 50028000 in Region 1, other stations that comprise the region have lower percent errors at 100-year return periods. For example, the USGS station 50092000 in Rio Grande de Patillas, with a percent error of only -4%. Regions 2 and 3 also are composed by USGS stations with percent errors less than 20% between at-site and regional quantiles. In general, for the 5-, 10-, and 25-year return periods, the standard error values are considerably lower for most stations.

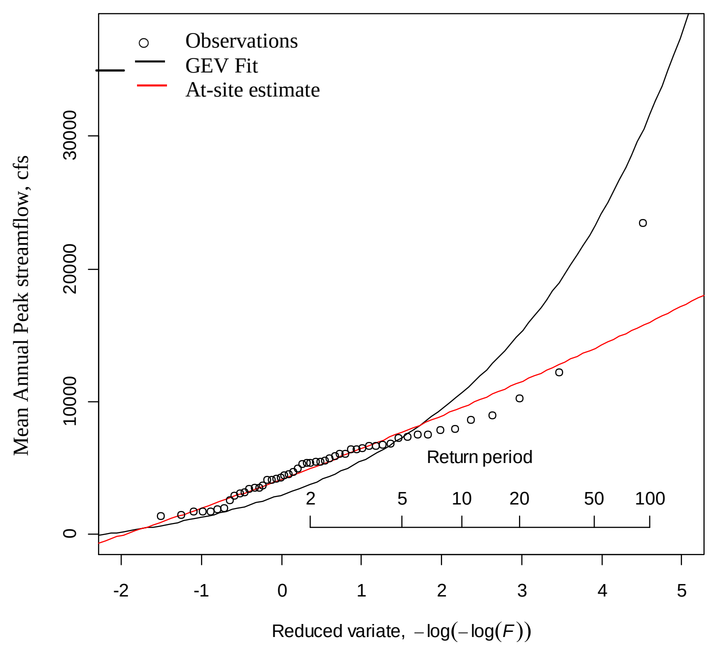


Figure 7. USGS station 50028000 at-site and estimated quantiles.

The percent errors were considerably high for all the three sites that comprise Region 4. The higher value corresponds to USGS station 50147800 at Culebrinas River with an error of 218%, which indicates an overestimated regional quantile. USGS station 50138000 at Guanajibo River also possesses higher percent errors at the 100-year return period with a value of 44%. These results suggest that for Region 4 basins in western Puerto Rico, another parameter estimation may be called for. In addition, a possible reason for the high percent error values in some of the stations may be related to the assumption that the GEV was the appropriate distribution for all the sites.

8. APPLICATION TO UNGAGED BASINS

Following the same procedure as for the index flood method, the quantiles for an ungaged basin can be estimated. Only two requisites would be required to obtain the ungaged site quantile: the regional flood frequency relationship, and the sample mean of the data at site *i*, considered in this study as the scale value or index flood. Next are listed the parameters for the respective distribution of the regions, calculated using the L-moments and L-moments ratios of the specific region.

- Region 1: GEV Parameters
  - $\xi = 0.53016$
  - $\alpha = 0.37084$
  - $k = -0.41599$
- Region 2: Pe3 Parameters
  - $\mu = 1$
  - $\sigma = 0.632164$
  - $\gamma = 1.306465$
- Region 3: Pe3 Parameters
  - $\mu = 1$
  - $\sigma = 0.950069$
  - $\gamma = 2.104329$
- Region 4: Wakeby Parameters
  - $\xi = -0.105$
  - $\alpha = 6.176$
  - $\beta = 14.454$
  - $\gamma = 0.473$
  - $\delta = 0.330$

With the use of these parameters the dimensionless regional quantile function is defined for each distribution. For Region 1, characterized by a GEV distribution, with *F* being the nonexceedance probability, the function desired is as follows:

$$q(F) = 0.530 - \frac{0.371}{-0.416}(1 - (-\log F)^{-0.416}) \tag{8.1}$$

Similarly, the *q(F)* function for Region 4 with a nonexceedance probability *F*, characterized by a Wakeby distribution is:

$$q(F) = -0.105 + \frac{6.176}{14.454}(1 - (1 - F)^{14.454}) - \frac{0.473}{0.330}(1 - (1 - F)^{0.330}). \tag{8.2}$$

An exception is presented at Region 2 and Region 3 where the best fitted distribution, as determined by the goodness-of-fit measure, is the Pearson Type 3. As stated earlier, the probability density function of this distribution does not possess an analytical solution. Therefore, tabulated values of the *q(F)* must be used. Table 18 to Table 21, from Appendix C.2, provide these values with an *F* sequence from 0.2 to 0.99 with an increase of 0.01. For the estimation of the scale factor μ_i, the linear regression model developed by Sanchez (1995) was used. The regression takes into consideration the contributing drainage area (CDA) in mi^2, the 5-year return period 24 hr rainfall (X_5) in inches, and the 25-year return period 24 hr rainfall (X_{25}) in inches. The last two variables values are obtained from NOAA Atlas 14 (Bonnin, et al., 2006). With a correlation coefficient of 93 percent, the formula derived was:

$$\log(Q_{mean}) = 2.02 + 0.76 * \log(CDA) + 0.318x_5 - 0.135x_{25} \tag{8.3}$$

8.1 Accuracy of the index flood method for ungaged basin analysis

To test for accuracy of the regional flood frequency analysis on an ungaged basin, one station that was used on the analysis will be considered as ungaged. The station was extracted from its corresponding region and the procedure to determine the regional quantile was repeated without it. The scale factor μ_i in Equation 4.13 was substituted by Q_{mean} value. The USGS station 50092000 at Rio Grande de Patillas belonging to Region 1 was selected for the test. The contributing drainage area for the station is 18 mi^2, and the 5-year and the 25-year return period 24 hr rainfall are 6.84 and 11.4 inches, respectively. A mean annual flow value of 4075 cfs is obtained by substituting the characteristic values in the regression function. This value is not exactly comparable with the mean annual peak streamflow data obtained from the record data of the USGS station.

Table 10. Comparison of at-site and regional quantile estimates for site 50092000 for various return periods.

Procedure	Quantile				
	5-year	**10-year**	**25-year**	**50-year**	**100-year**
At- site	8269.7 cfs	12029.03 cfs	18629.61 cfs	25342.15 cfs	34068.24 cfs
Regional Growth Curve	1.298138	1.908236	3.012798	4.169084	5.710218
Q_{mean} equation	4075 cfs				
Site estimate	5289.91 cfs	7776.06 cfs	12277.15 cfs	16989.02 cfs	23269.14 cfs
Percent Error	-36%	-35%	-34%	-33%	-32%
Q_{mean}	6217.364 cfs				
Site Estimate	8071.00 cfs	11864.20 cfs	18731.66 cfs	25920.71 cfs	35502.50 cfs
Percent Error	-2%	-1%	1%	2%	4%

Table 10 shows the at-site quantile, the regional growth curve of Region 1 with station 50092000 excluded from the analysis, the site quantile estimate from the regional analysis, and the percent error. High values of the percent error results from the use of the Q_{mean} estimated from the regression equation of Equation 8.3. The percent error values decrease significantly when the mean annual peak streamflow data from the record data was used as the index flood on Equation 4.13. The larger percent error of 4% is obtained at the 100-year return period. The graph in Figure 8 depicts graphically the similarity between the at-site estimate (red line) and the quantile estimated from the regional growth curve and the mean annual peak streamflow of the record data.

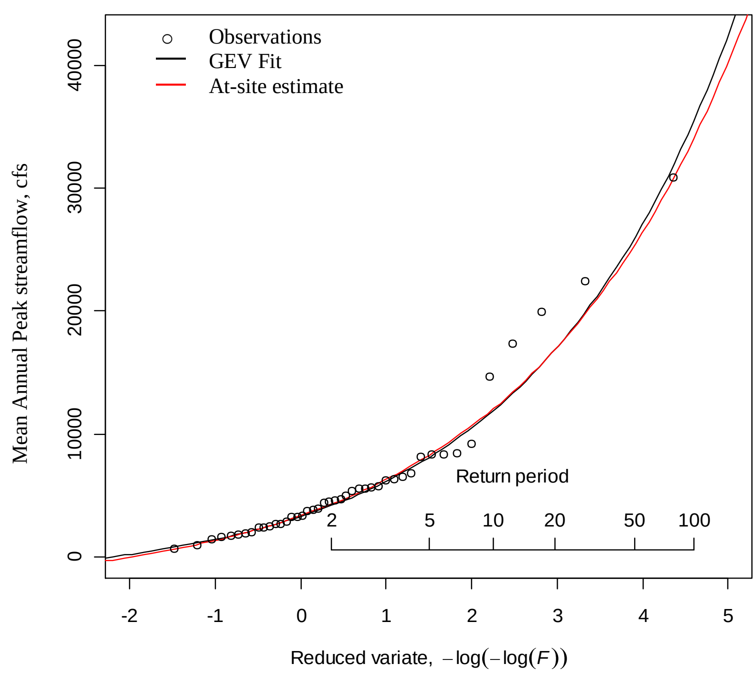


Figure 8. Comparison of USGS station 5009200 at-site and estimated quantiles with index flood equal to record data mean annual peak.

Among those stations that were excluded from this study, USGS station 50073200 was chosen randomly in order to evaluate the accuracy of the proposed methodology on stations with short record. The USGS station 50073200 at Rio Daguao, located at eastern Puerto Rico, has a 15 year annual peak streamflow record. This station was excluded from the analysis due to its short period of record. A GEV distribution was fitted to the data for the at-site quantiles. The contributing drainage area for the station is 2.26 mi^2, and the 5-year return period 24 hr rainfall and the 25-year return period 24 hr rainfall are 8.5 and 11.5 inches, respectively. A mean annual flow value of 2,751.84 cfs is obtained from regression equation.

To determine into which region Station 50073200 can be grouped into, the location of the site was evaluated, and in this case was assigned to Region 2. This information was validated implementing the hybrid cluster analysis. It was found that Station 50073200 forms a mutual cluster with Stations 50051310 and 50056400, therefore it is grouped into Region 2. Using the mean annual flow value of 2,751.84 cfs, and the *q(F)* values from Table 8 for Region 3, the estimated site quantiles for site 50073200 were computed and are presented in Table 11. The percent error between the at-site and regional estimated quantile is considerably high for all the return periods. This can be related to the bias from a short record and the estimated mean annual streamflow. The percent errors for this station are comparable with Site 50065500 from Region 2, included on this analysis in Table 24.

Table 11. Comparison of at-site and regional quantile estimates for site 50073200 for various return periods.

Procedure	Quantile (cfs)				
	5-year	10-year	25-year	50-year	100-year
At- site	4929.57	6972.38	10005.41	12634.81	15615.22
Site estimate	3915.67	4988.85	6429.10	7561.90	8743.63
Percent Error	-21%	-28%	-36%	-40%	-44%

Figure 9 depicts the underestimated quantile (black line). If the Q_{mean} value obtained from the regression equation were nearer average of the annual peak streamflow value of the station, a better estimate would be produced.

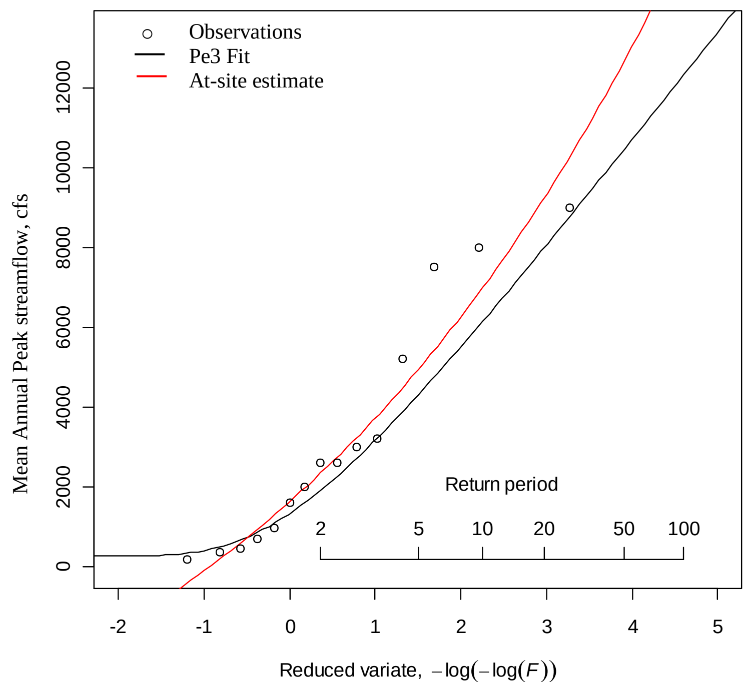


Figure 9. Comparison of USGS station 50073200 at-site and estimated quantiles.

The results obtained from the comparison of quantile estimates in USGS station 50092000 emphasize the importance of the estimate of the index flood in a regional frequency analysis. In addition, it was discussed that for the estimation of quantiles with short records the procedure provides relatively good results despite the error involved in the index flood estimation with the regression equation and the short record.

8.2 Application at an ungaged site

The ungaged river Unibón, located in the Cibuco river watershed at northern Puerto Rico, was selected to apply the regional flood frequency analysis. The river is located inside Region 3 and using the clustering analysis, it was determined that Region 3 is the best option to allocate this river. A mean peak flow of 3,481.91 cfs for this river was estimated with Equation 8.3. With the use of *q(F)* for Region 3, the quantile estimate for 100-year at Unibón river is 15,583 cfs.

Table 12. Regional quantile estimated at Unibón River for various return periods.

Site	Quantile				
	5-year	10-year	25-year	50-year	100-year
Regional	1.561548	2.228793	3.11895	3.79612	4.47551
Unibón	5437.17 cfs	7760.46 cfs	10859.90 cfs	13217.75 cfs	15583.32 cfs

It is evident that the procedure to determine the quantile estimate for an ungaged basin is very simple. Once the site characteristics are found, in this case CDA, MAR, CS, X_5, and X_{25}, the quantiles for nonexceedance values from 0.2 to 0.99 are easily obtained from the information provided in Table 18 to Table 21, from pages 111 to 114. In cases where only the 5-, 10-, 25-, 50- , and 100-year return periods are needed, Table 8 is the best option if only the return periods mentioned are needed.

9. CONCLUSIONS

The objectives of this study were to perform a new flood regionalization study for Puerto Rico based on recent advances in watershed classification and parameter estimation techniques of flood frequency distributions. They were accomplished with the use of the hybrid clustering algorithm and the L-moments. Consequently, a regional flood frequency analysis based on the method of L-moments was carried out using annual peak streamflow data. The following results and conclusions were obtained from this analysis:

- According to the heterogeneity measures applied in the study, Puerto Rico is definitely heterogeneous when considered as a single region.
- The grouping of gage sites in Puerto Rico that provided the more acceptable heterogeneity measures was a four-region grouping, from which only one region is considered as "acceptably homogeneous", two as "possibly heterogeneous", and one as "definitely heterogeneous".
- The general extreme value distribution provided the best fit only to Region 1, while for Regions 2 and 3 the best fit turned out to be the Pearson Type 3 distribution. For Region 4, a Wakeby distribution was applied following an observation from the related literature that this five-parameter distribution could perhaps provide a more robust fitting.
- Monte Carlo simulations demonstrated that, based on the root mean square error (RMSE), that the regionalized flood frequency procedure applied to the available peak streamflow data is accurate. The larger RMSE for a 100-year flood was obtained in Region 4 with a 0.79%. This value is very

low, but when the percent error between at-site and estimated quantile was calculated, errors higher than 40% were obtained between the few sites that comprise the region. Region 4 was a region in which a larger percent of the sites yielded higher errors. This can be explained by the relatively high heterogeneity values obtained for this region.

- The determination of quantiles at ungaged sites is straightforward with the index flood method. The accuracy of the procedure is dependent on the quality of the relation available for the mean annual peak streamflow estimates.
- Presently, the available annual peak flow records are of relatively short durations, mostly not exceeding 50 years. Most of the available stream gage data have one or two very large peaks notably different from the rest of the series. This situation introduces considerable uncertainty in the regionalization process that can only be remedied over time as more annual peaks are incorporated in the data base.

10. RECOMMENDATIONS

The accuracy of the flood frequency study is conditioned on the quality of several fundamental parameters, particularly the procedure for ungaged basins which relies on the availability of a relation of estimating the mean annual peak streamflow. Further effort is required to update and further qualify the relation of the mean annual flood with catchment morphological parameters. This could improve the predictive feature of flood frequency quantiles for ungaged basins.

The particular nature of the annual peak flow series available on the island most surely induces a high degree of uncertainty in flood estimation for the higher return periods. Island peak flow data is characterized by nearly uniform values interspersed with a few inordinate extreme flows. Because of the relatively short periods of record, the presence of a small number of seemingly outlying peaks, in some cases exceeding a tenfold increase in magnitude over the nominal values, will certainly bias the analysis. Only the long-term sampling of annual data will serve to reduce the attendant estimation uncertainty when predicting the high-end flows such as the 100-year flood or above.

Although it is impossible to completely eliminate estimation uncertainty, improvements in the theoretical analysis of extreme value events should provide more accurate procedures. However, these must be accompanied by more accurate measurements of extreme floods. At present, river overbank peak flows are computed by indirect means different from the estimation procedures for in-stream flows. Thus, peak flow estimates for major floods are subject to considerable irreducible uncertainty. Improvements in peak flood estimation should also accompany developments in the statistical field.

11. REFERENCES

Aronica, G. T., & Candela, A. (2007). Derivation of flood frequency curves in poorly gauged Mediterranean catchments using a simple stochastic hydrological rainfall-runoff model. *Journal of Hydrology*, *347*(1-2), 132-142.

Bonnin, G. M., Martin, D., Lin, B., Parzybok, T., Yekta, M. & Riley, D. (2006). Precipitation-Frequency Atlas of the United States. *NOAA Atlas 14*, Volume 3, Version 2.0, NOAA, National Weather Service, Silver Spring, Maryland.

Bhaskar, N. R., & O'Connor, C. A. (1989). Comparison of Method of Residual and Cluster Analysis for Flood Regionalization. *Journal of Water Resources Planning and Management*, *115*(6), 793-808.

Cadavid, L., Obeysekera, J., & Shen, H. W. (1991). Flood-Frequency Derivation From Kinematic Wave. *Journal of Hydraulic Engineering*, *117*(4), 489-510.

Chipman, H., & Tibshirani, R. (2006). Hybrid Hierarchical Clustering with Application to Microarray Data. *Biostatistics*, 268-285.

Hosking, J.R.M (1990). L-Moments: Analysis and Estimation of Distribution using Linear Combinations of Order Statistics. *Journal of the Royal Statisticals Society*, 52, 105-124.

Hosking, J. R. M., & Wallis, J. R. (1997). *Regional Frequency Analysis: An approach Based on L-Moments*. *New York*. New York, USA: Cambridge University Press.

J. R. M. Hosking (2008). L-moments, version 1.6. URL: http://CRAN.R-project.org/package=lmom.

J. R. M. Hosking (2012). Regional frequency analysis using L-moments. R package, version 2.4. URL: http://CRAN.R-project.org/package=lmomRFA.

Kottegoda, N. T., & Rosso, R. (1997). Applied Statistics for Civil and Environmental Engineers. New York: McGraw Hill.

López, M. A., & Fields, F. K. (1970). A proposed streamflow data program for Puerto Rico. San Juan, Puerto Rico: U.S Geological Survey unnumbered Open-File Report.

López, M. A., Colón-Diepa, E., & Cobb, E. D. (1979). Floods in Puerto Rico, magnitude and frequency. San Juan, PR: U.S. Geological Survey Water Resources Investigations Report 78-141.

Lim, Y. H., & Voeller, D. L. (2009). Regional Flood Estimations in Red River Using L-Moment-Based Index-Flood and Bulletin 17B Procedures. *Journal of Hydrologic Engineering*.

McCuen, R. H., & Galloway, K. E. (2010). Record Lenght Requirements for Annual Maximmum Flood Series. *Journal of Hydrologic Engineering* , 704-707.

Olsen, J. R., Kiang, J., & Waskom, R. (2010). Workshop on Nonstationarity, Hydrologic Frequency Analysis, and Water Management. Colorado Water Institute Information.

Ramos-Ginés, O. (1999). Estimation of Magnitude and Frequency of Floods for Streams in Puerto Rico: New Empirical Models. San Juan, Puerto Rico: U.S. Geological Survey Water Resources Investigation Report 99-4142.

Rao, R. A., & Srinivas,V. V. (2008). Regionalization of Watersheds: An Approach Based on Cluster Analysis (Vol. 58). Water Science and Technology Library.

Rao, R. A., & Srinivas,V. V. (2006). Regionalization of waterhed by hybrid-cluster analysis. *Journal of Hydrology* 318 , 37-56.

Saf, B. (2009). Regional Flood Frequency Analysis Using L Moments for the Buyuk and Kucuk Menderes River Basins of Turkey. *Journal of Hydrologic Engineering*, 783-794.

Sanchez, J. E. (1995). Flood Frequency estimation for ungaged catchments in Puerto Rico. Mayaguez, PR: University of Puerto Rico.

Segarra-García, R. (1998). Flood Frequency estimation for ungaged catchments in Puerto Rico . *Third International Symposium on Tropical Hydrology and Fifth Caribbean Islands Water Resources Congress* (pp. 65-70). American Water Resources Association.

U.S. Interagency Advisory Committee on Water Data (1982). Guidelines for determining flood flow frequency, Bulletin 17-B of the Hydrology Subcommittee. Reston, Virginia, U.S. Geological Survey, Office of Water Data Coordination.

Yang, T., Hao, Z.-C., & Sun, L.-M. (2009). Using L-moments Approach to Conduct Regional Flood Frequency and Associated Spatial Patterns Analysis in the Pearl River Delta Region. *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery* (pp. 628-632).

APPENDICES

A. *L*-statistics

A.1 *L*-moments values **for sites considered in the study**

Table 13. Summary of *L*-statistics values for selected stations in the study

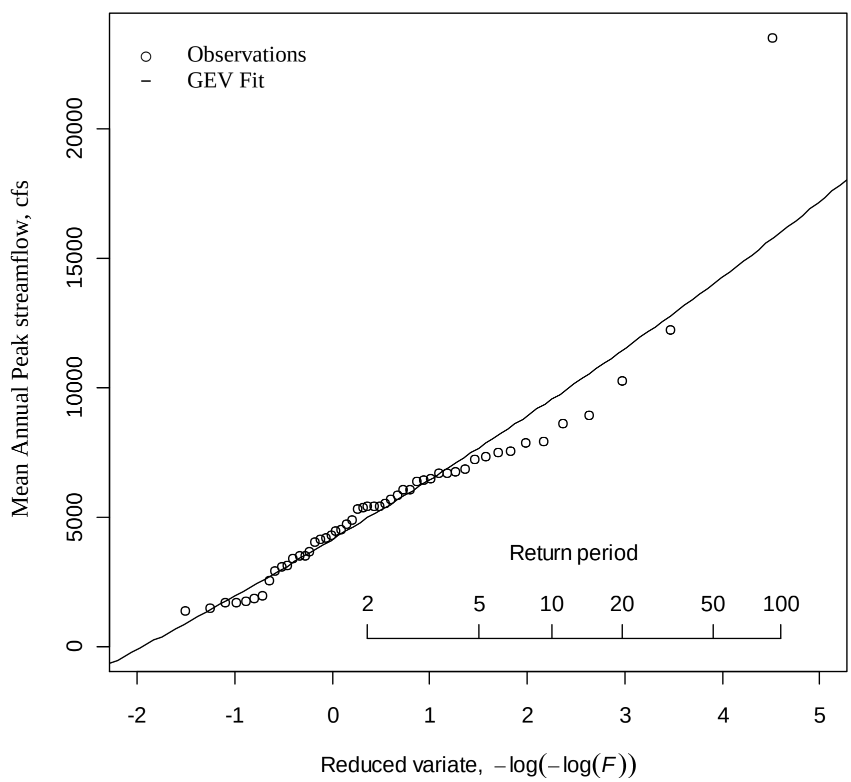
Site number	n	Mean	L-CV	L-skewness	L-kurtosis	t_5	Discordancy
		l_1	T	t_3	t_4		D_i
50028000	51	5569.02	0.29617	0.20777	0.25144	0.21541	1.42
50028400	31	4486.48	0.37890	0.33365	0.18638	0.08602	0.21
50031200	44	13010.82	0.47029	0.37404	0.17192	0.03374	0.24
50034000	35	5380.91	0.54813	0.52798	0.32130	0.17050	0.48
50035000	59	26879.66	0.49182	0.42747	0.23867	0.11786	0.15
50038100	49	37825.10	0.47579	0.28507	0.08060	0.02400	1.09
50038320	35	8318.00	0.32641	0.09325	0.09913	0.10460	1.27
50039500	52	9118.27	0.49765	0.40671	0.13028	0.02354	0.97
50043000	34	18647.35	0.52648	0.44025	0.19170	0.04021	0.6
50047850	29	9964.90	0.49289	0.47211	0.37245	0.33854	0.87
50050900	34	18647.35	0.52648	0.44025	0.19170	0.04021	0.6
50051310	33	6562.06	0.36682	0.13380	0.03269	-0.02624	0.97
50055000	51	25410.59	0.35568	0.19348	0.18716	0.12659	0.82
50056400	40	11524.00	0.37903	0.25601	0.14736	0.06262	0.11
50057000	50	22066.00	0.46875	0.27484	0.04561	0.02049	1.41
50061800	42	6398.10	0.37232	0.21450	0.09969	-0.00227	0.29
50063800	43	8112.09	0.33322	0.18157	0.05008	-0.00541	0.61
50064200	34	7121.18	0.38878	0.36902	0.19075	0.00354	0.43
50065500	34	10115.88	0.27903	0.03393	0.09977	0.05075	1.92
50067000	31	4392.26	0.28203	0.24851	0.12489	0.02820	1.24
50071000	49	8766.94	0.30263	0.27176	0.19660	0.02170	0.74
50075000	29	1228.28	0.28921	0.35666	0.14367	-0.01587	**3.47**
50090500	33	3719.06	0.48895	0.50481	0.35860	0.20820	0.52
50092000	44	6217.36	0.45113	0.44262	0.31344	0.20285	0.28
50112500	47	2119.51	0.48920	0.57703	0.37459	0.27159	0.97
50115000	34	3291.97	0.50299	0.57291	0.44179	0.30883	1.31
50124200	30	4747.27	0.53965	0.51630	0.22447	0.03212	0.84
50138000	36	12319.72	0.69533	0.69272	0.47665	0.31642	2.98
50144000	48	24100.42	0.49549	0.59750	0.42004	0.28244	1.25
50147800	43	24749.53	0.18356	0.01996	0.09839	-0.00002	1.95

A.2 R Code for L-moment estimation and GEV fitting to considered stations

Station 50028000

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(1370, 1480, 1670, 1680, 1730, 1850, 1960, 2570, 2910, 3080, 3120, 3400,
3490, 3490, 3650, 4060, 4120, 4210, 4300, 4470, 4490, 4700, 4910, 5300,
5360, 5410, 5420, 5440, 5540, 5700, 5860, 6040, 6060, 6400, 6420, 6480,
6670, 6670, 6770, 6860, 7240, 7340, 7510, 7530, 7840, 7940, 8590, 8950,
10270, 12200, 23500)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

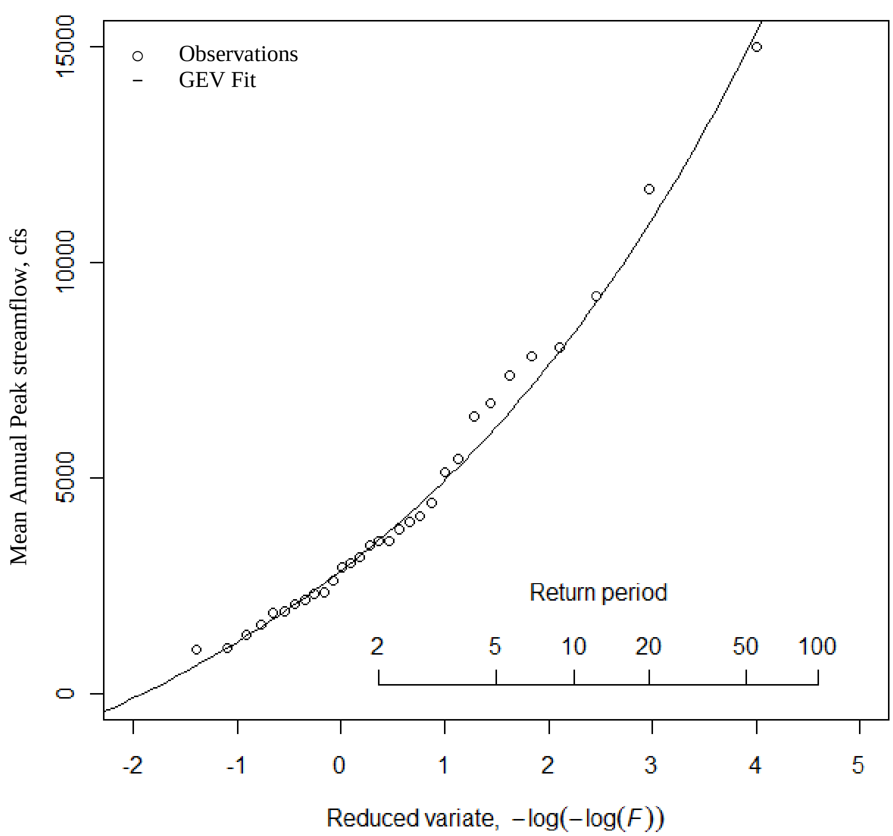
```
           xi         alpha             k
4134.56854991 2248.80294880   -0.05803271
  name  n     l_1          t        t_3        t_4        t_5
1    1 51 5569.02 0.2961749 0.2077667 0.2514385 0.2154058
```



Station 50028400

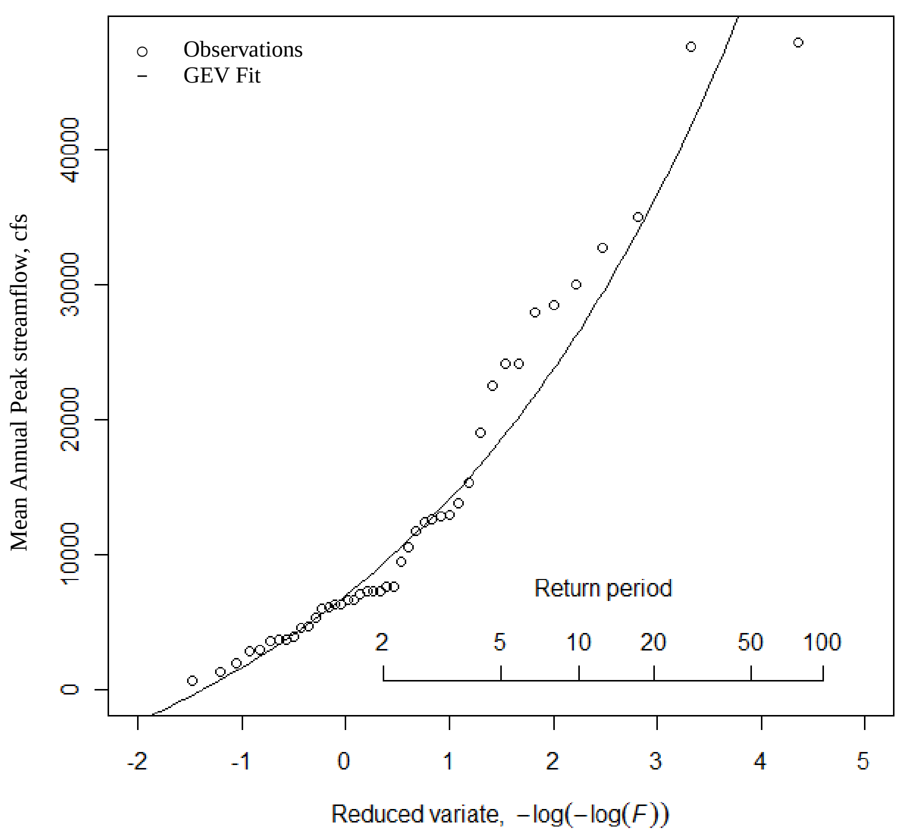
```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(1020, 1060, 1360, 1610, 1880, 1900, 2080, 2160, 2320, 2340, 2600, 2920,
3010, 3150, 3420, 3530, 3540, 3800, 3970, 4120, 4430, 5140, 5450, 6440,
6730, 7380, 7811, 8010, 9200, 11700, 15000)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

```
           xi         alpha            k
4134.56854991 2248.80294880   -0.05803271
  name  n     l_1         t       t_3       t_4       t_5
1    1 51 5569.02 0.2961749 0.2077667 0.2514385 0.2154058
```



Station 50031200

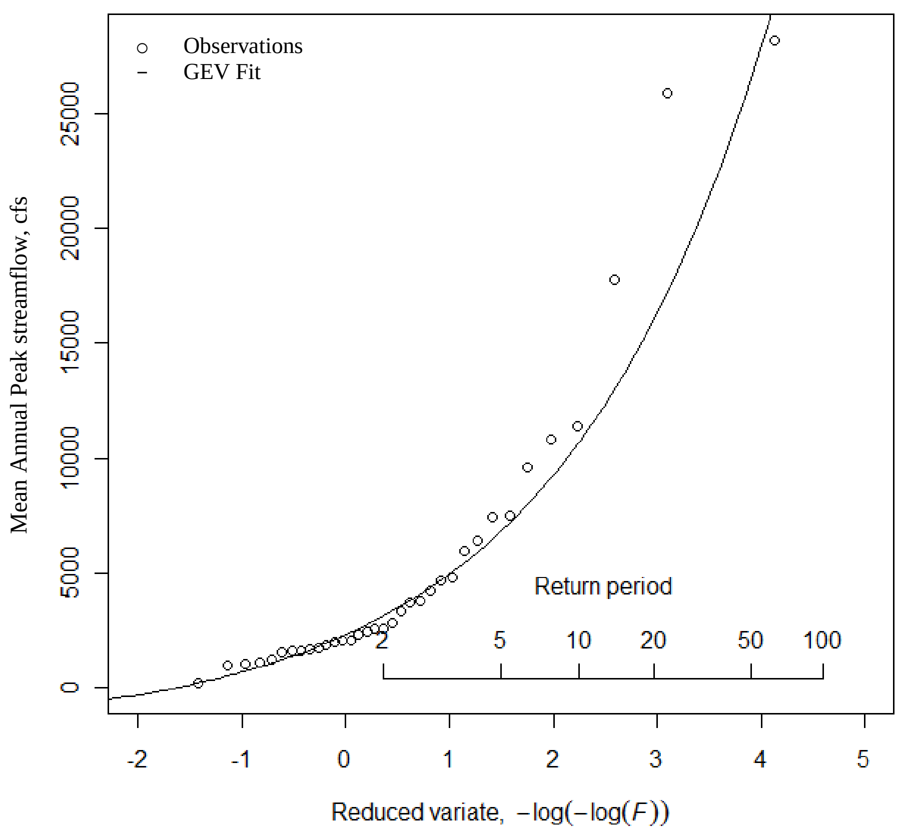
```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(606, 1280, 1920, 2870, 2920, 3630, 3650, 3670, 3910, 4540, 4670,
5320, 5960, 6040, 6270, 6330, 6680, 6680, 7120, 7280, 7290, 7300, 7610,
7650, 9480, 10600, 11700, 12400, 12600, 12800, 12900, 13800, 15300,
19000, 22500, 24100, 24200, 28000, 28500, 30000, 32700, 35000, 47700,
48000)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))

           xi         alpha             k
6945.0287427 6167.8088742   -0.2948494
  name  n      l_1           t        t_3        t_4        t_5
1    1 44 13010.82 0.4702914 0.3740437 0.1719246 0.0337389
```



Station 50034000

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(182, 924, 996, 1060, 1180, 1560, 1570, 1580, 1650, 1700, 1850, 1960,
2040, 2070, 2300, 2430, 2540, 2560, 2790, 3310, 3710, 3760, 4230, 4650,
4820, 5920, 6410, 7420, 7480, 9580, 10800, 11400, 17800, 25900, 28200)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

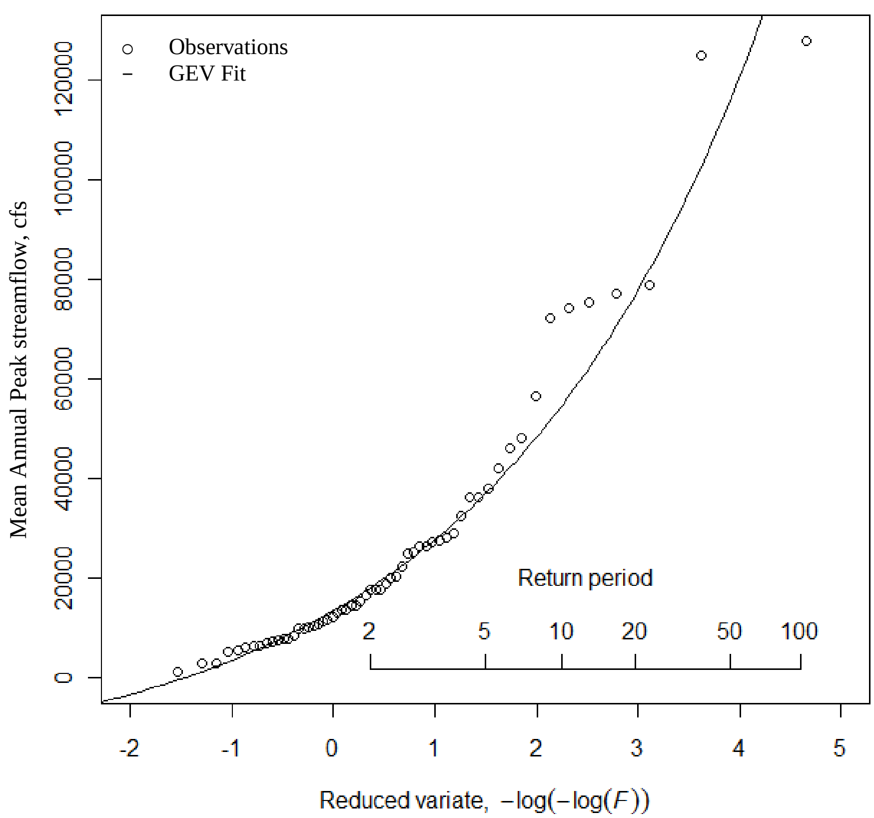
```
         xi        alpha            k
2282.196023 2046.747849   -0.491876
  name  n      l_1          t        t_3        t_4        t_5
1    1 35 5380.914 0.5481346 0.5279825 0.3213005 0.1705034
```



Station 50035000

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(1060, 2760, 2920, 5220, 5620, 6120, 6430, 6470, 6880, 7250,
7490, 7870, 7950, 8360, 9800, 9900, 10200, 10300, 10800, 11300,
11600, 12100, 13000, 13700, 13700, 14400, 14600, 15300, 16600, 17700,
17700, 17800, 18900, 19900, 20200, 22400, 24800, 25200, 26300, 26500,
27300, 27500, 28000, 29100, 32400, 36200, 36400, 38100, 42100, 46000,
48200, 56500, 72100, 74300, 75400, 77300, 78900, 125000, 128000)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

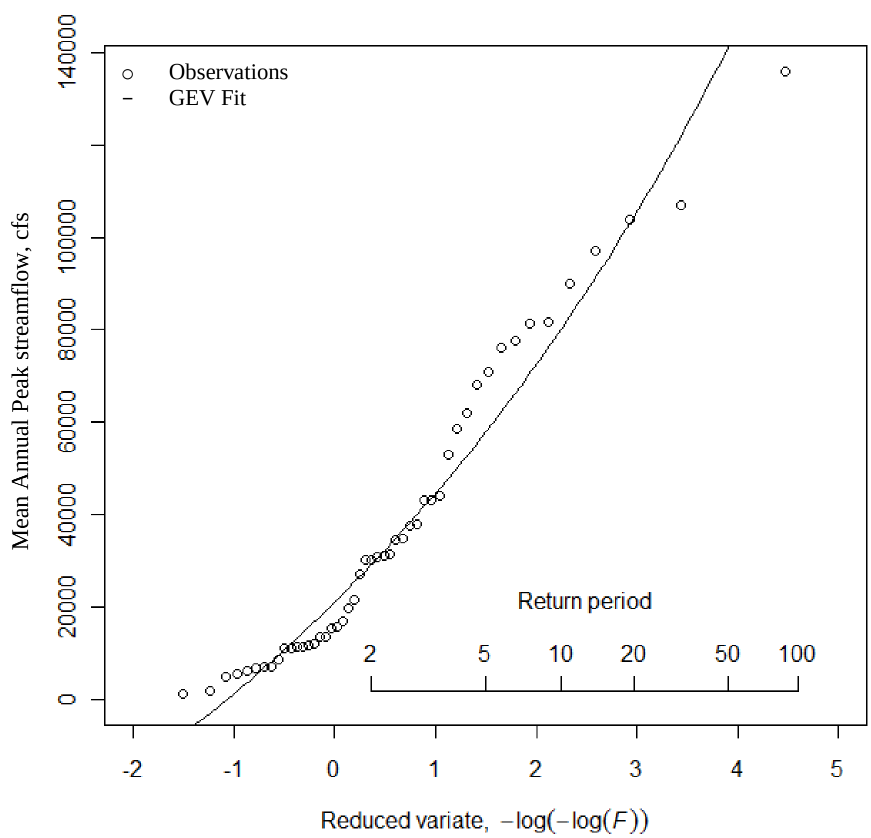
```
           xi         alpha            k
13431.0323536 11844.6651534    -0.3654108
  name  n      l_1         t       t_3        t_4        t_5
1    1 59 26879.66 0.4918223 0.427472 0.2386716 0.1178635
```



Station 50038100

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(1260, 1700, 5000, 5470, 6000, 6610, 6960, 7100, 8550, 11000,
11100, 11300, 11300, 11600, 12100, 13400, 13600, 15500, 15800, 17000,
19800, 21440, 27200, 30190, 30200, 30800, 31030, 31500, 34500, 34800,
37600, 37900, 43000, 43300, 44200, 53000, 58700, 62000, 68110, 70800,
76000, 77600, 81400, 81760, 90000, 97250, 104000, 107000, 136000)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

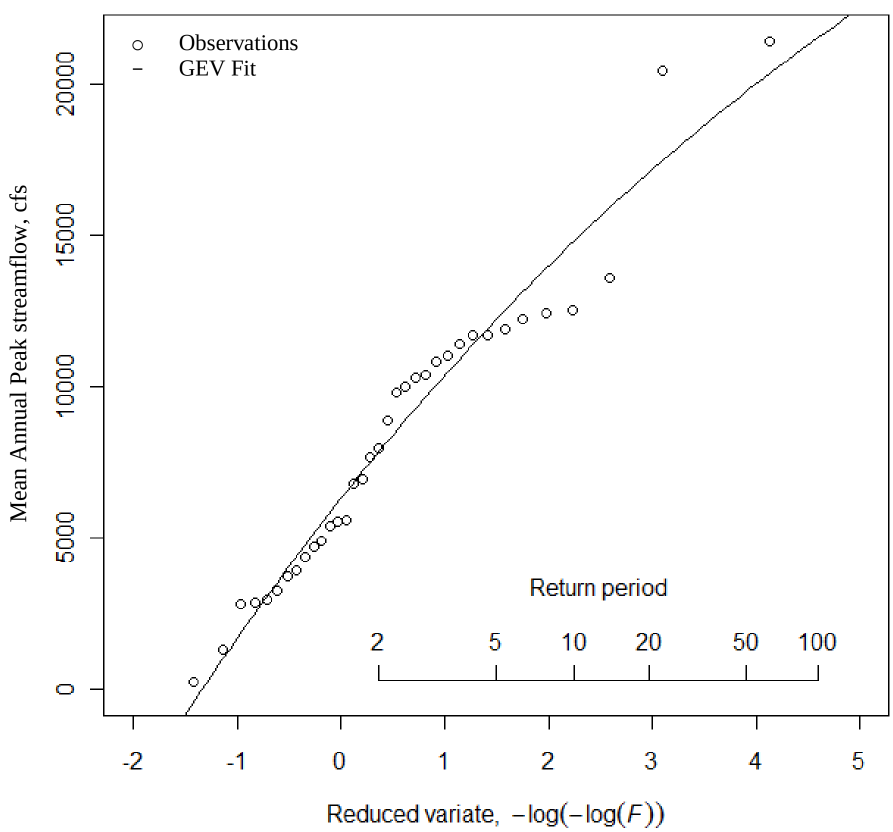
```
           xi          alpha              k
21011.9337738 21565.2004962    -0.1715951
  name  n     l_1          t       t_3      t_4        t_5
1    1 49 37825.1 0.4757868 0.2850661 0.080603 0.02400219
```



Station 50038320

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(220, 1300, 2800, 2850, 2950, 3260, 3720, 3910, 4340, 4700, 4890,
5400, 5550, 5600, 6770, 6930, 7650, 7940, 8860, 9790, 10000, 10300,
10400, 10800, 11000, 11400, 11700, 11700, 11900, 12200, 12400, 12500,
13600, 20400, 21400)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

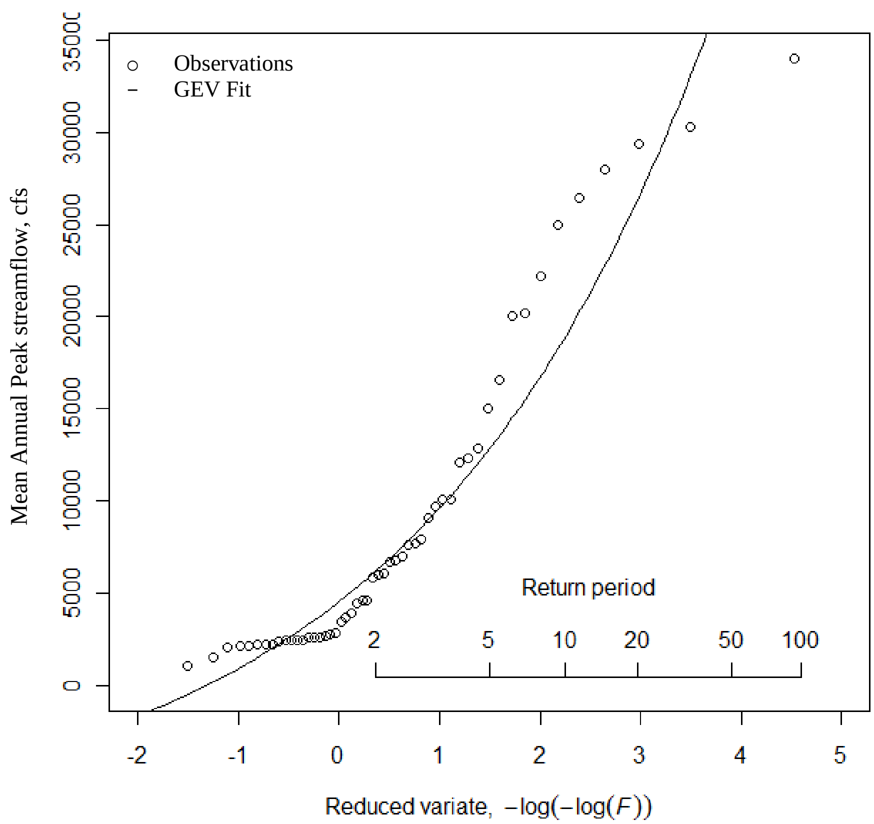
```
           xi         alpha            k
6290.4113289 4335.8473559    0.1230376
  name   n  l_1          t         t_3         t_4       t_5
1    1 35 8318 0.3264056 0.09325012 0.09913084 0.104595
```



Station 50039500

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(1040, 1570, 2050, 2140, 2160, 2210, 2220, 2240, 2400, 2500, 2500,
2500, 2500, 2600, 2640, 2650, 2730, 2790, 2810, 3450, 3670, 3900, 4500,
4600, 4640, 5820, 6000, 6050, 6740, 6800, 7020, 7650, 7710, 7950, 9100,
9700, 10100, 10100, 12100, 12300, 12900, 15000, 16600, 20000, 20200,
22200, 25000, 26400, 28000, 29400, 30300, 34000)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

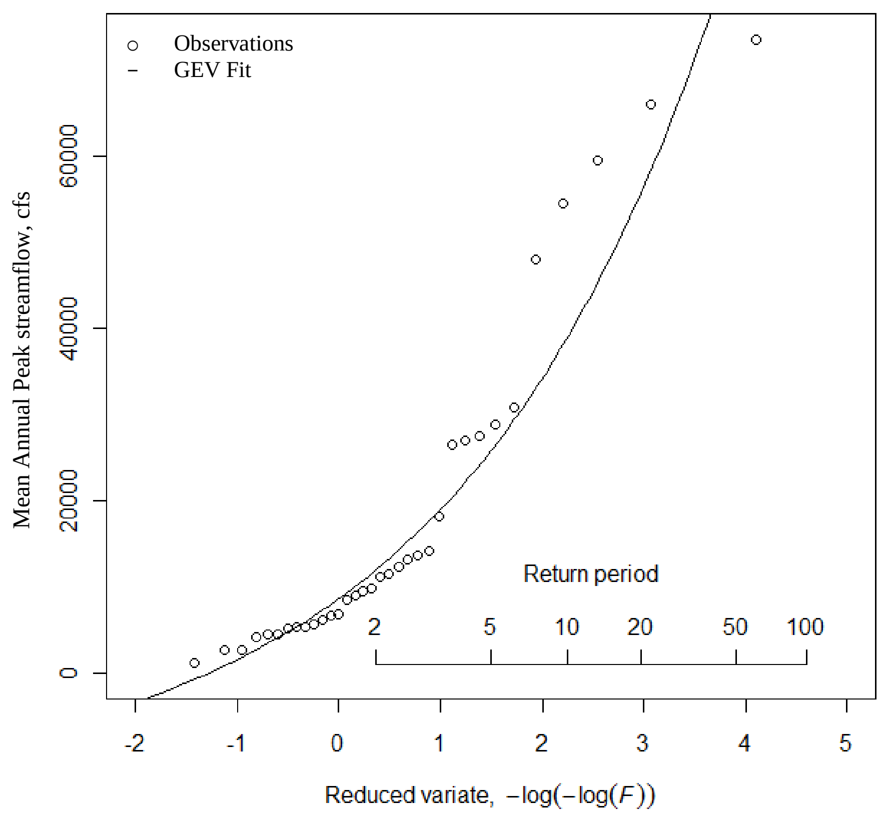
```
          xi        alpha            k
4544.6214125 4261.8981829   -0.3382752
  name  n      l_1         t       t_3        t_4         t_5
1    1 52 9118.269 0.4976546 0.4067054 0.1302836 0.02353612
```



Station 50043000

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(1150, 2680, 2700, 4170, 4420, 4440, 5100, 5370, 5400, 5630, 6200,
6690, 6840, 8490, 8960, 9580, 9760, 11100, 11430, 12400, 13200, 13600,
14100, 18100, 26500, 27000, 27600, 28900, 30800, 48000, 54500, 59600,
66000, 73600)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

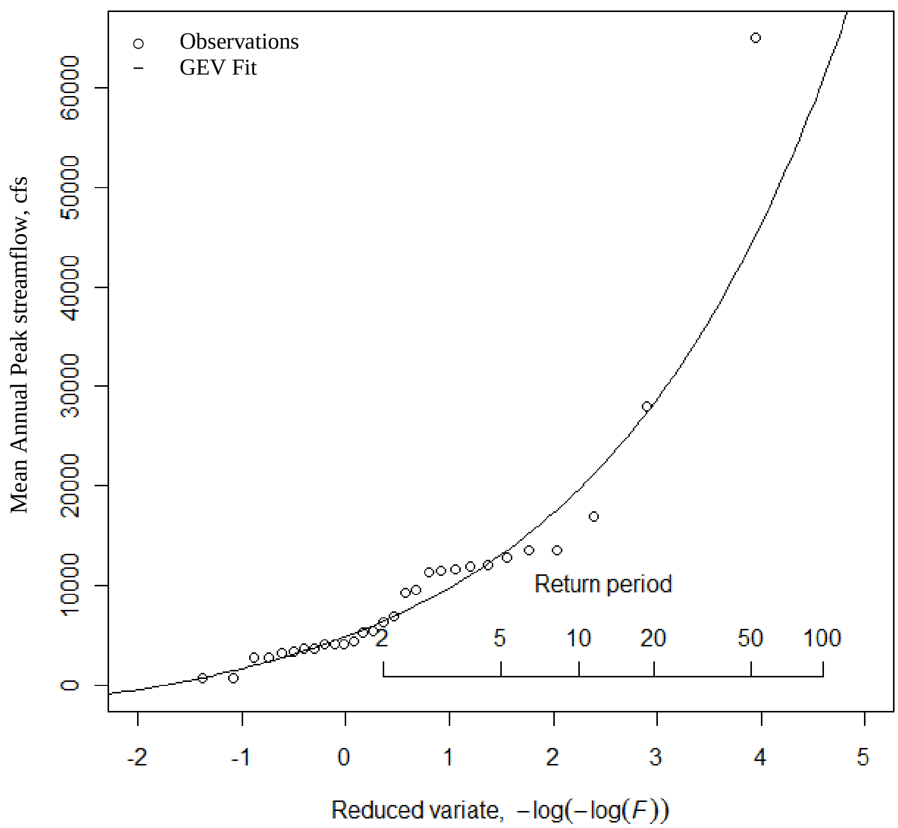
```
          xi         alpha            k
8607.8616707 8536.9758774   -0.3819274
  name  n      l_1          t       t_3        t_4         t_5
1    1 34 18647.35 0.5264753 0.4402481 0.1916981 0.04021034
```



Station 50047850

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(677, 685, 2720, 2810, 3250, 3390, 3640, 3650, 4070, 4120, 4190, 4470,
5240, 5500, 6290, 6980, 9330, 9580, 11400, 11500, 11600, 11900, 12100,
12800, 13490, 13600, 17000, 28000, 65000)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

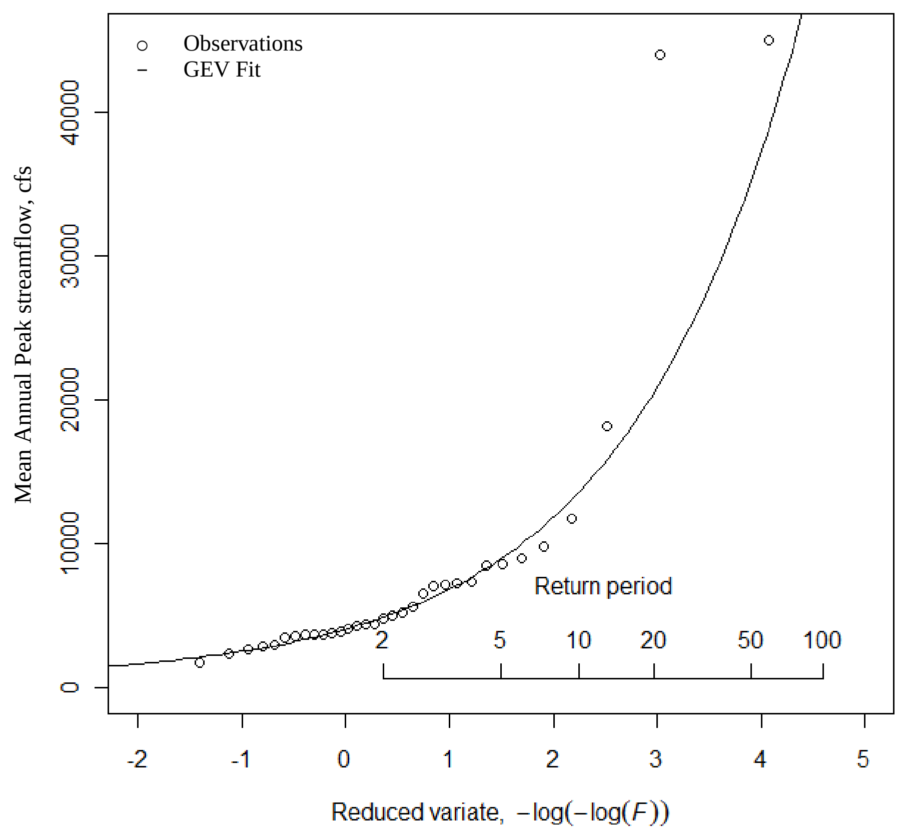
```
          xi         alpha            k
4883.8709707 3951.8037006   -0.4225494
  name  n      l_1         t       t_3        t_4        t_5
1    1 29 9964.897 0.4928871 0.4721144 0.3724457 0.3385429
```



Station 50050900

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(1770, 2320, 2660, 2820, 2960, 3470, 3540, 3620, 3640, 3690, 3780,
3850, 4030, 4290, 4330, 4430, 4760, 5000, 5230, 5650, 6510, 7000, 7170,
7260, 7320, 8470, 8580, 8950, 9830, 11700, 18200, 44000, 45000)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

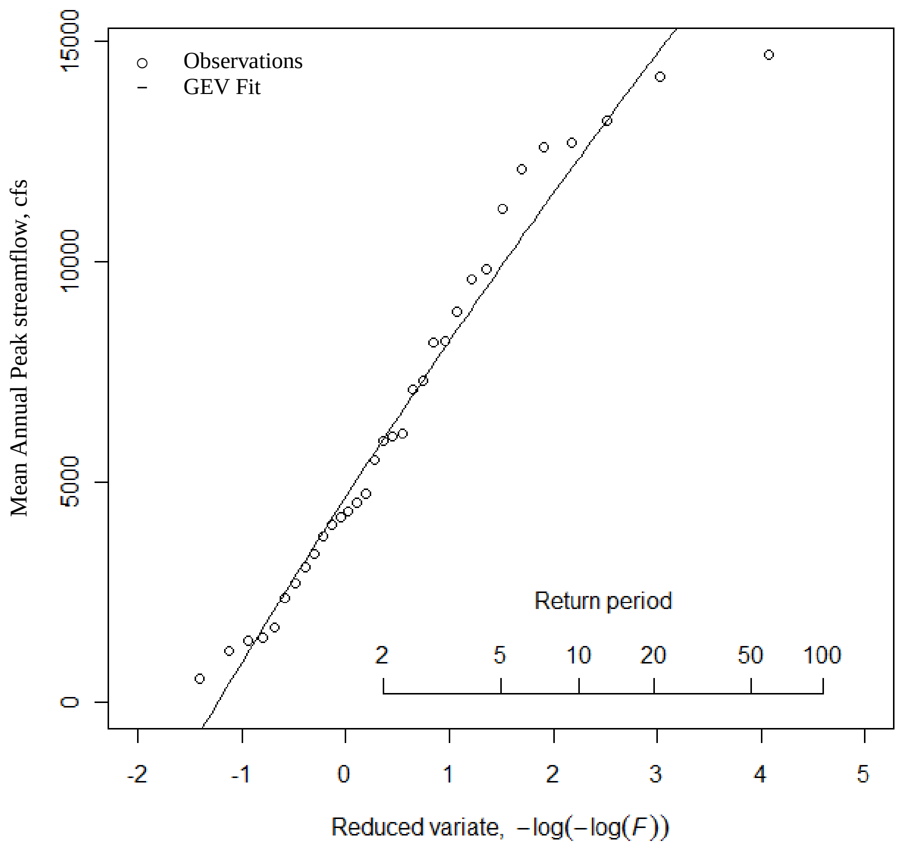
```
           xi        alpha            k
4037.5730816 2036.0649793   -0.5904211
  name  n      l_1         t       t_3       t_4       t_5
1    1 33 8055.455 0.4687065 0.6107171 0.4699207 0.3364215
```



Station 50051310

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(528,  1150,  1400,  1450,  1690,  2380,  2690,  3060,  3380,  3750,  4030,
4190, 4330, 4530, 4740, 5490, 5920, 6020, 6100, 7090, 7300, 8150, 8180,
8860, 9610, 9830, 11200, 12100, 12600, 12700, 13200, 14200, 14700)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

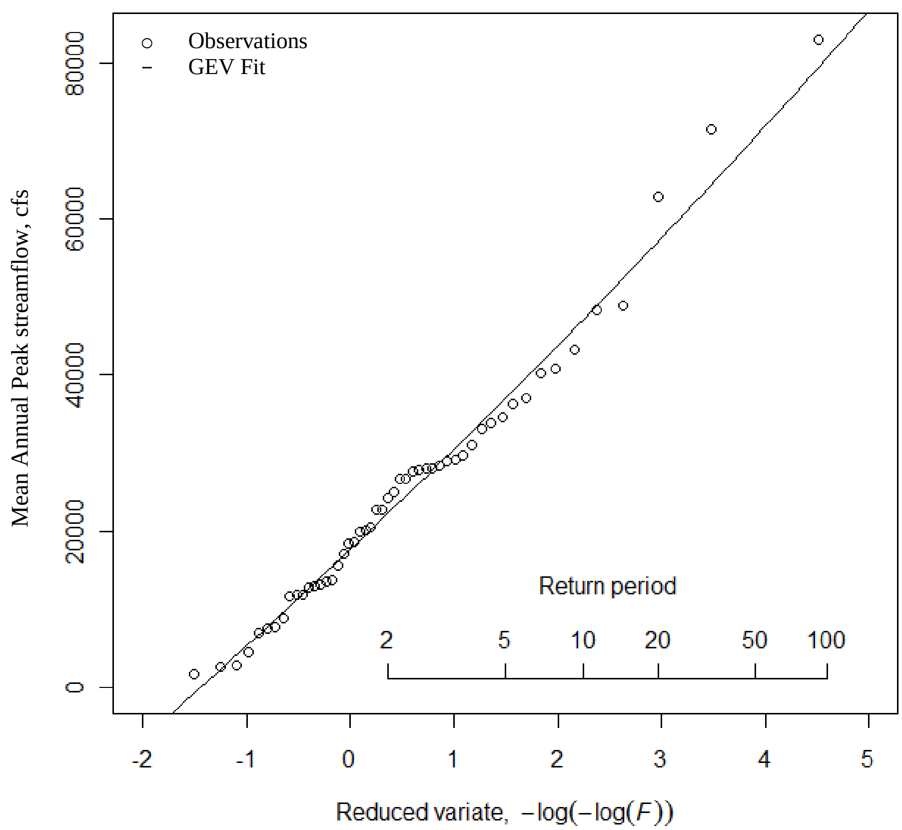
```
          xi        alpha            k
4.650539e+03 3.650788e+03 5.702728e-02
  name  n      l_1         t       t_3        t_4        t_5
1    1 33 6562.061 0.3668188 0.1337961 0.03268929 -0.0262432
```



Station 50055000

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(1720, 2700, 2830, 4470, 6900, 7600, 7780, 8740, 11700, 11800, 11800,
12800, 12900, 13200, 13500, 13700, 15600, 17200, 18400, 18700, 20000,
20100,20500, 22800, 22800, 24300, 25100, 26700, 26800, 27700, 27800,
28000, 28000, 28400, 28900, 29100, 29800, 31100, 33100, 33800, 34700,
36400, 37100, 40200, 40800, 43300, 48300, 49000, 62800, 71500, 83000)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

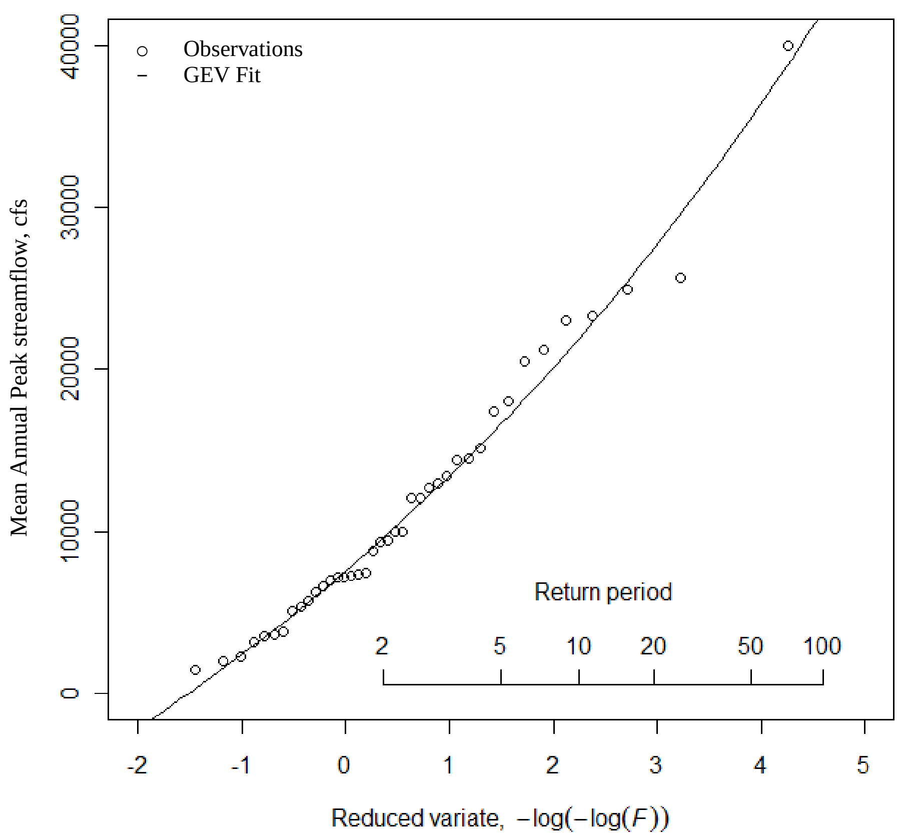
```
           xi        alpha            k
 1.767257e+04  1.259459e+04 -3.632759e-02
  name  n      l_1        t       t_3       t_4       t_5
1    1 51 25410.59 0.355681 0.1934845 0.1871577 0.1265918
```



Station 50056400

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(1480, 1970, 2260, 3180, 3500, 3610, 3820, 5060, 5360, 5710, 6290,
6610, 7000, 7150, 7150, 7220, 7340, 7460, 8780, 9380, 9460, 9930, 9940,
12100, 12100, 12700, 13000, 13400, 14400, 14500, 15100, 17400, 18000,
20500, 21200, 23000, 23300, 24900, 25700, 40000)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

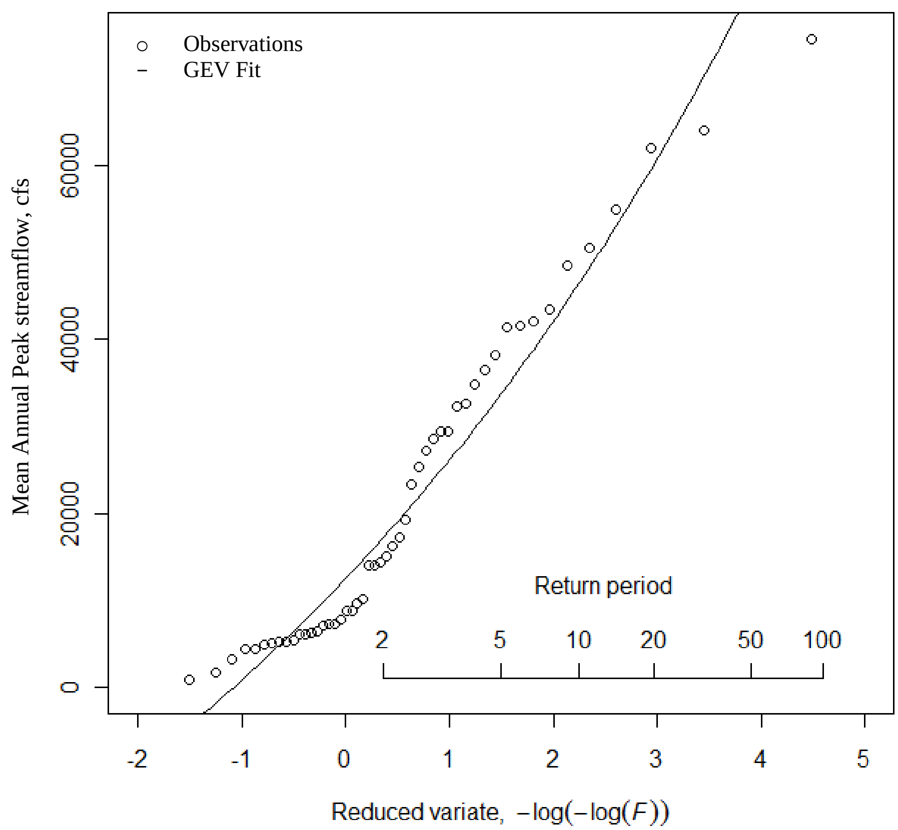
```
          xi        alpha           k
7542.0006493 5506.3315684  -0.1296562
  name  n   l_1         t       t_3       t_4        t_5
1    1 40 11524 0.3790262 0.2560089 0.1473628 0.06261552
```



Station 50057000

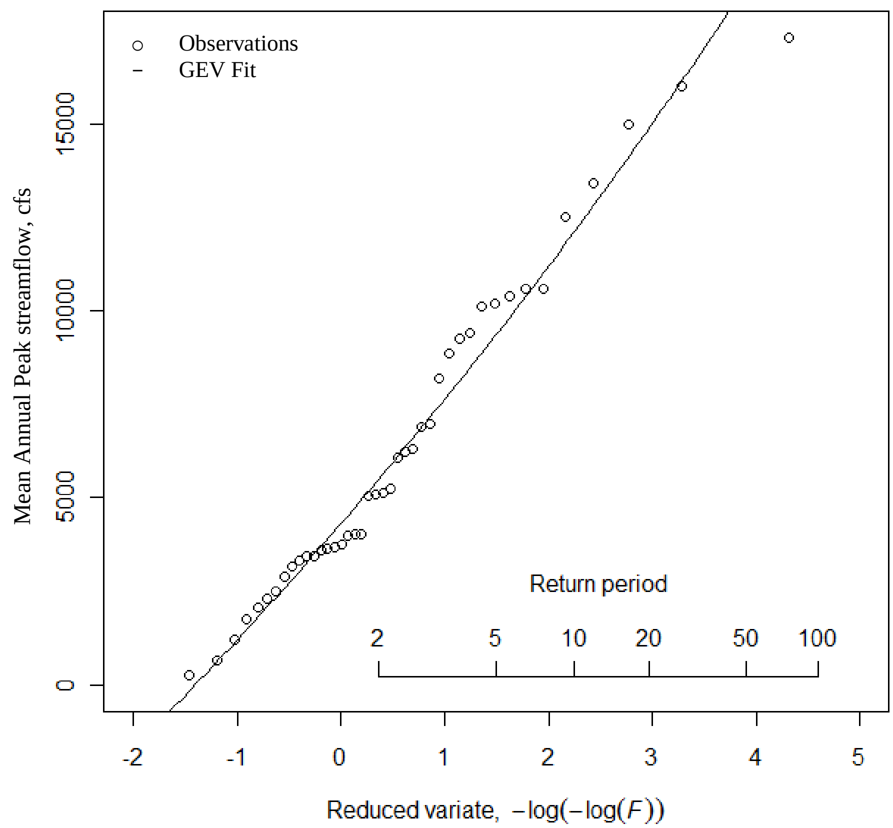
```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(900,  1620,  3220,  4320,  4450,  4910,  5010,  5170,  5190,  5350,  6110,
6140, 6290, 6440, 7130, 7180, 7340, 7840, 8700, 8790, 9700, 10100, 14000,
14100,  14300,  15100,  16200,  17200,  19300,  23300,  25300,  27200,  28600,
29400,  29400,  32300,  32600,  34800,  36500,  38200,  41500,  41600,  42100,
43500, 48600, 50600, 54900, 62100, 64100, 74600)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

```
           xi         alpha            k
12482.6869459 12621.4248390    -0.1569379
  name  n   l_1        t       t_3         t_4        t_5
1    1 50 22066 0.468747 0.274842 0.04560659 0.02048972
```



Station 50061800

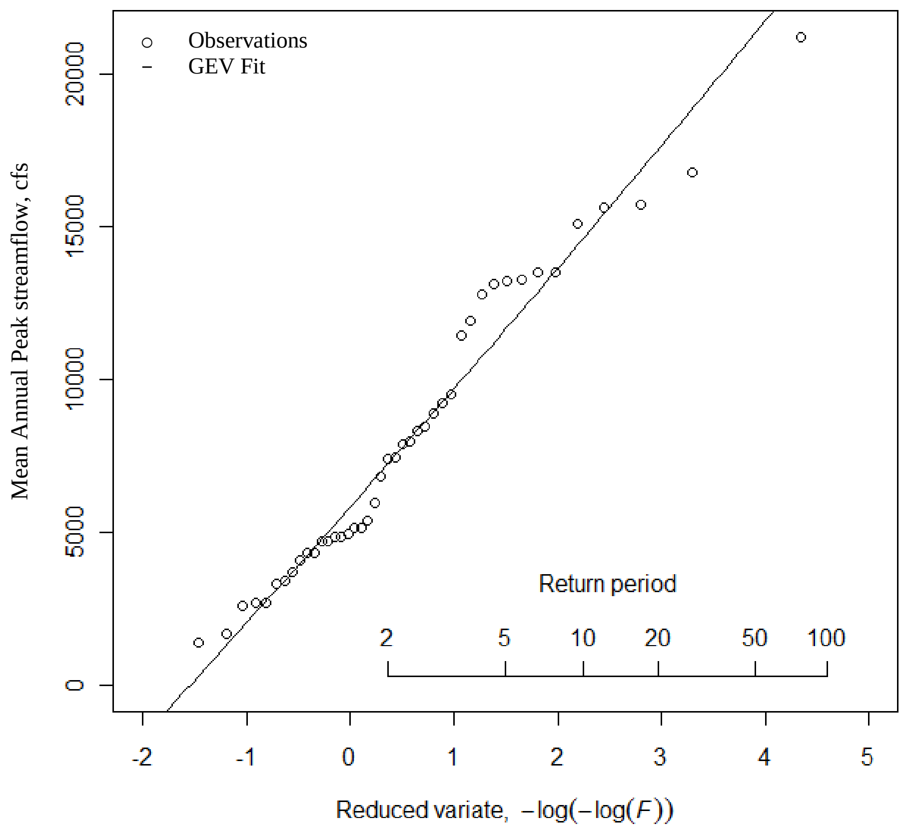
```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(258, 673, 1210, 1760, 2070, 2300, 2510, 2910, 3159, 3320, 3440, 3460,
3590, 3630, 3680, 3770, 4010, 4040, 4050, 5070, 5090, 5140, 5260, 6080,
6220, 6320, 6910, 6960, 8200, 8870, 9260, 9400, 10100, 10200, 10400,
10600, 10600, 12500, 13400, 15000, 16000, 17300)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))

          xi        alpha            k
4311.6277949 3213.9029720   -0.0681788
  name  n      l_1          t       t_3        t_4         t_5
1    1 42 6398.095 0.3723224 0.2144962 0.09968504 -0.00226767
```



Station 50063800

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(1400, 1700, 2580, 2700, 2710, 3290, 3400, 3710, 4100, 4300, 4320,
4700, 4710, 4830, 4830, 4940, 5140, 5150, 5400, 5960, 6800, 7400, 7460,
7880, 7960, 8310, 8440, 8870, 9230, 9520, 11450, 11900, 12790, 13110,
13200, 13260, 13500, 13520, 15080, 15600, 15700, 16770, 21200)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

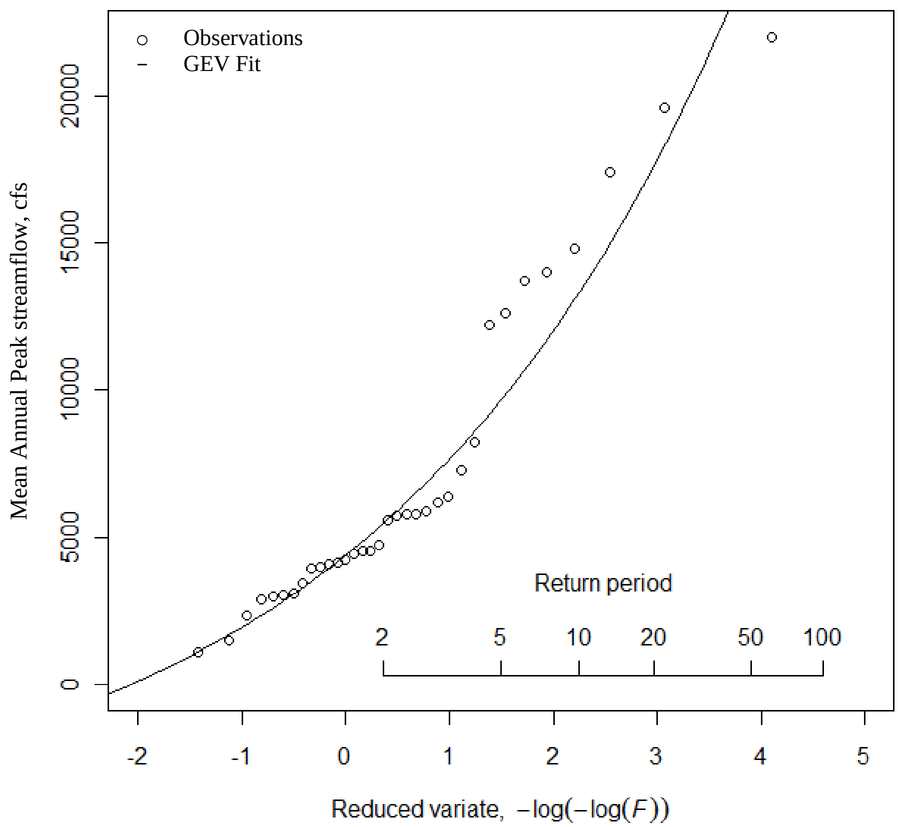
```
           xi          alpha             k
5829.31998967 3834.26011873   -0.01804257
  name  n     l_1         t       t_3        t_4         t_5
1    1 43 8112.093 0.3332214 0.1815725 0.05007878 -0.00541174
```



Station 50064200

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(1070, 1510, 2320, 2870, 2980, 3050, 3100, 3420, 3950, 4000, 4100,
4140, 4240, 4450, 4540, 4540, 4730, 5590, 5710, 5780, 5790, 5880, 6180,
6360, 7290, 8230, 12200, 12600, 13700, 14000, 14800, 17400, 19600, 22000)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

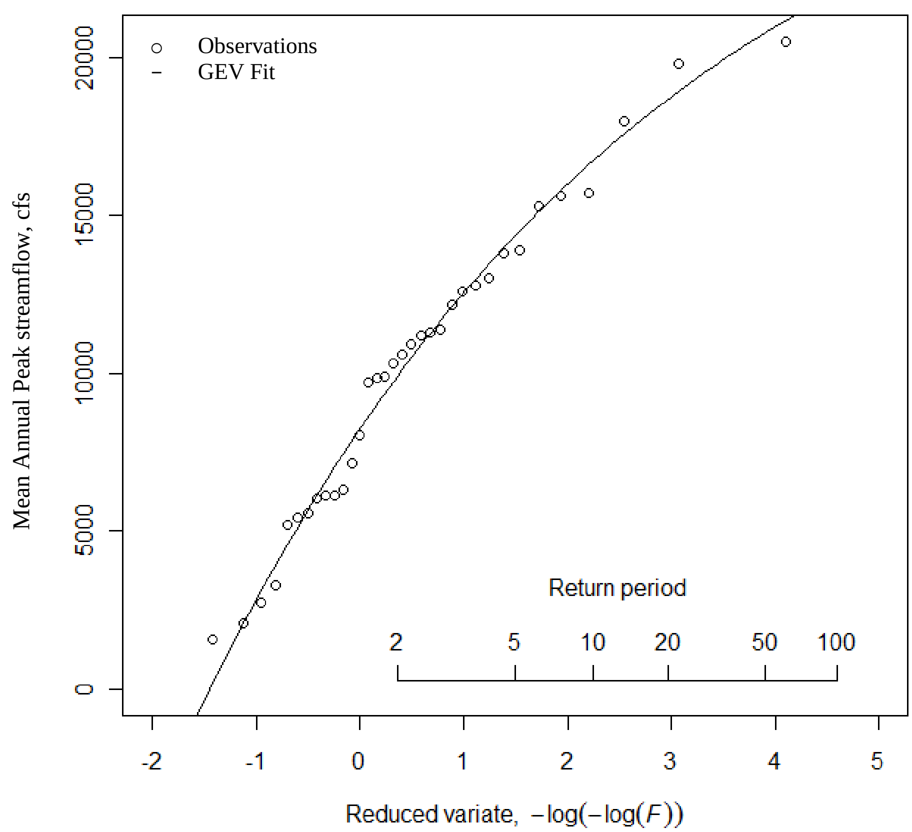
```
          xi        alpha            k
4384.2647902 2820.1880739   -0.2880861
  name  n      l_1         t       t_3       t_4         t_5
1    1 34 7121.176 0.3887779 0.3690195 0.1907469 0.003542907
```



Station 50065500

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(1570, 2110, 2730, 3280, 5190, 5420, 5560, 6020, 6130, 6130, 6320,
7140, 8030, 9700, 9830, 9920, 10300, 10600, 10900, 11200, 11300, 11400,
12160, 12600, 12800, 13000, 13800, 13900, 15300, 15600, 15700, 18000,
19800, 20500)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

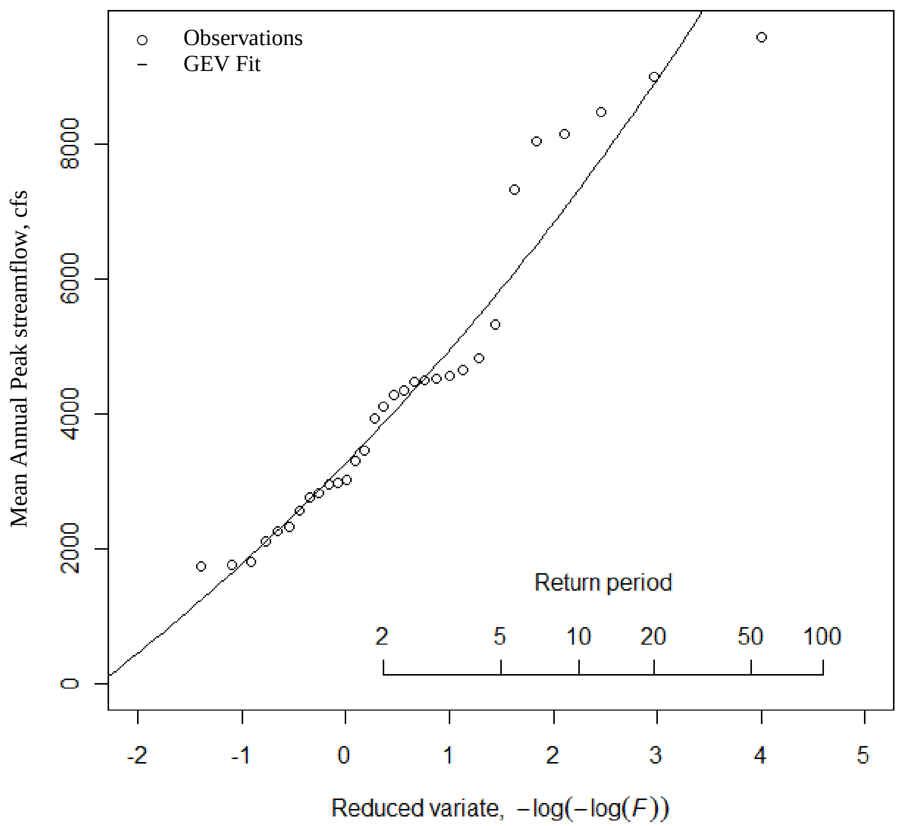
```
         xi       alpha           k
8225.059556 4819.307922    0.223768
  name  n      l_1         t        t_3        t_4        t_5
1    1 34 10115.88 0.2790321 0.03392621 0.09976632 0.05075375
```



Station 50067000

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(1730, 1760, 1810, 2120, 2260, 2320, 2570, 2760, 2830, 2960, 2970,
3030, 3310, 3470, 3940, 4120, 4290, 4350, 4480, 4500, 4520, 4580, 4660,
4840, 5330, 7340, 8050, 8170, 8480, 9010, 9600)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

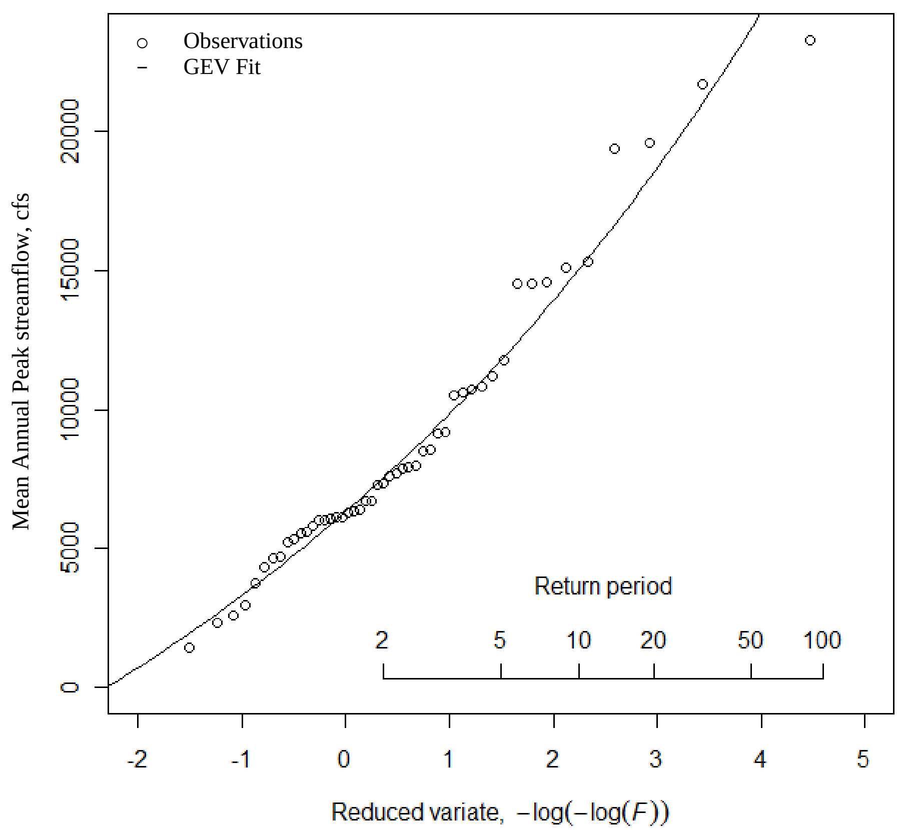
```
          xi        alpha            k
3270.5869614 1581.4666506   -0.1186952
  name  n      l_1        t      t_3      t_4        t_5
1    1 31 4392.258 0.282026 0.248514 0.124889 0.02819502
```



Station 50071000

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(1430, 2300, 2600, 2940, 3770, 4320, 4670, 4710, 5220, 5320, 5530,
5590, 5800, 6000, 6030, 6080, 6110, 6110, 6270, 6360, 6410, 6690, 6710,
7310, 7340, 7600, 7700, 7850, 7900, 7950, 8480, 8570, 9150, 9200, 10500,
10600, 10700, 10800, 11200, 11800, 14500, 14500, 14560, 15100, 15300,
19400, 19600, 21700, 23300)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

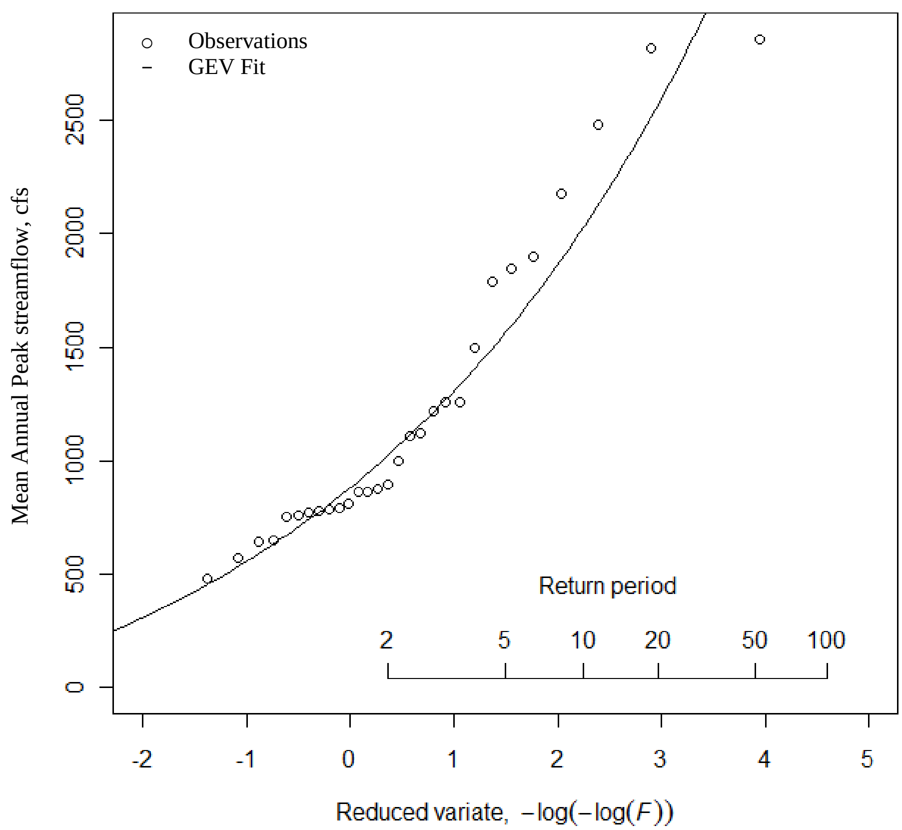
```
           xi         alpha           k
6315.1051988 3254.9817827  -0.1524962
  name  n      l_1        t       t_3       t_4        t_5
1    1 49 8766.939 0.302625 0.2717584 0.1966049 0.02169503
```



Station 50075000

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(482, 567, 641, 650, 752, 760, 771, 779, 783, 791, 811, 859, 859, 872,
893, 1000, 1110, 1120, 1220, 1260, 1260, 1500, 1790, 1850, 1900, 2180,
2480, 2820, 2860)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

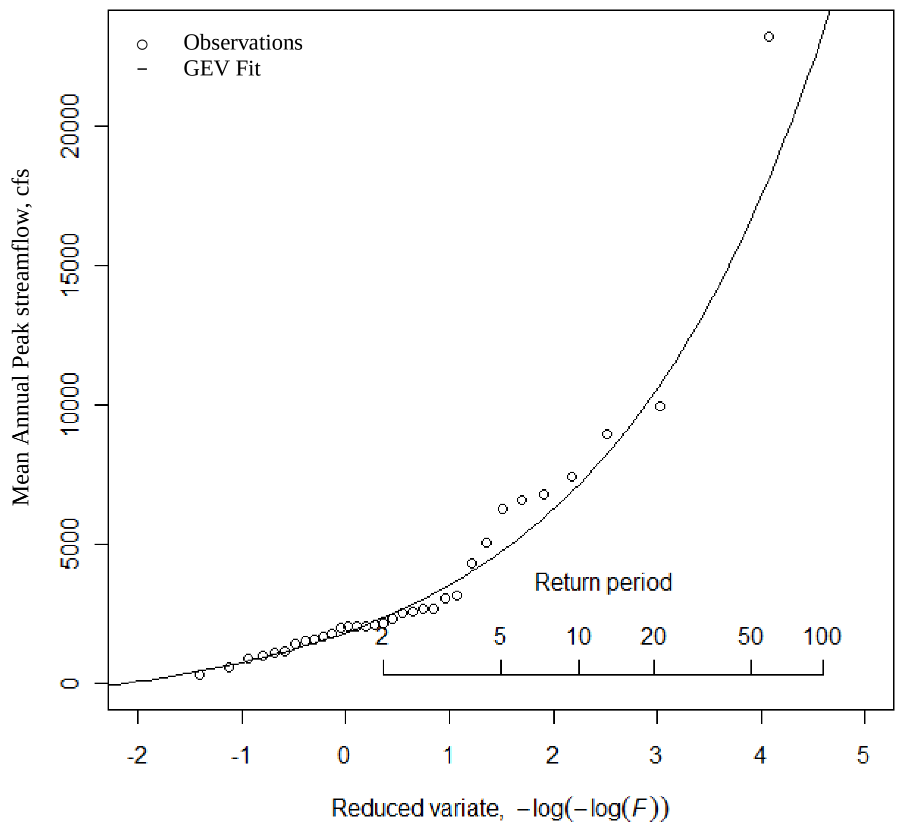
```
         xi       alpha           k
879.5941976 371.1879713  -0.2713537
  name  n      l_1         t       t_3       t_4        t_5
1    1 29 1228.276 0.2892115 0.3566612 0.1436683 -0.01587123
```



Station 50090500

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(297, 549, 883, 1010,  1080,  1160,  1410,  1530,  1550,  1690,  1760,
2010, 2030,  2060,  2070,  2080,  2170,  2300,  2520,  2550,  2680,
2680,  3030, 3140, 4310,  5060,  6280,  6560,  6780,  7400,  8950,  9950,
23200)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

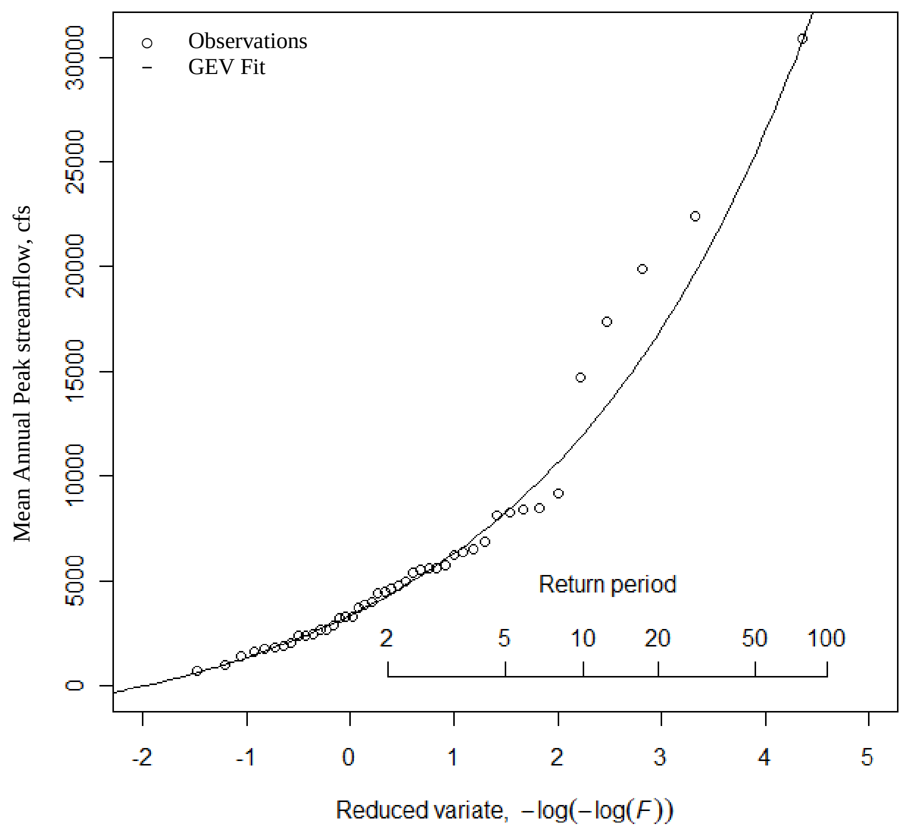
```
         xi        alpha            k
1819.4651256 1344.3461292   -0.4634029
  name  n      l_1         t       t_3       t_4       t_5
1    1 33 3719.061 0.4889487 0.5048077 0.3585974 0.2082034
```



Station 50092000

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(667,   987,  1400,  1600,  1750,  1800,  1890,  2050,  2400,  2410,  2460,
2630,  2640,  2850,  3230,  3270,  3320,  3700,  3820,  3960,  4400,  4500,  4600,
4730,  4950,  5380,  5550,  5600,  5630,  5720,  6230,  6360,  6530,  6830,  8130,
8290,  8370,  8440,  9190,  14700,  17400,  19900,  22400,  30900)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

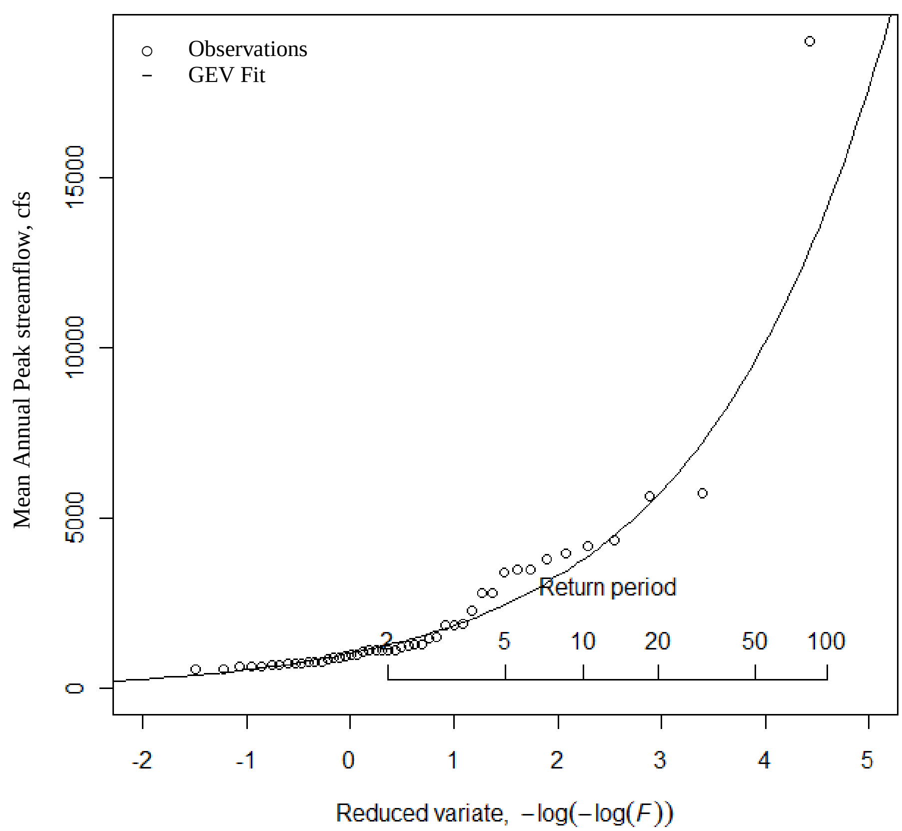
```
           xi         alpha           k
3346.3653130 2425.3694359   -0.3849781
  name  n      l_1         t       t_3       t_4       t_5
1    1 44 6217.364 0.4511339 0.4426193 0.313444 0.202851
```



Station 50112500

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(544, 570, 625, 643, 647, 691, 705, 737, 746, 746, 772, 786,
794, 860, 881, 895, 945, 979, 981, 1080, 1100, 1110, 1110, 1120,
1120, 1200, 1230, 1270, 1310, 1460, 1510, 1830, 1850, 1880, 2300,
2800, 2800, 3410, 3470, 3480, 3780, 3960, 4190, 4330, 5650, 5720, 19000)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

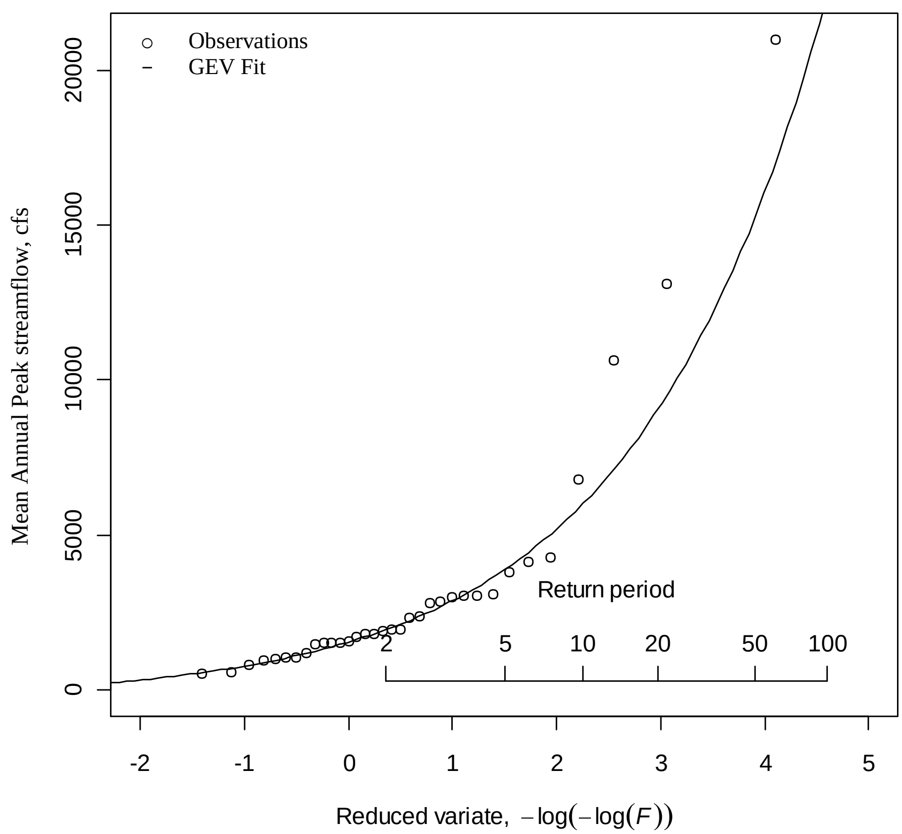
```
          xi         alpha            k
1020.3724859  622.9538200   -0.5508642
  name  n      l_1         t       t_3       t_4       t_5
1    1 47 2119.511 0.4892023 0.5770297 0.3745948 0.2715939
```



Station 50115000

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(500, 574, 800, 923, 980, 1020, 1060, 1200, 1470, 1510, 1530, 1530,
1550, 1700, 1780, 1810, 1880, 1920, 1940, 2330, 2380, 2800, 2850, 3010,
3050, 3050, 3070, 3800, 4120, 4260, 6790, 10640, 13100, 21000)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

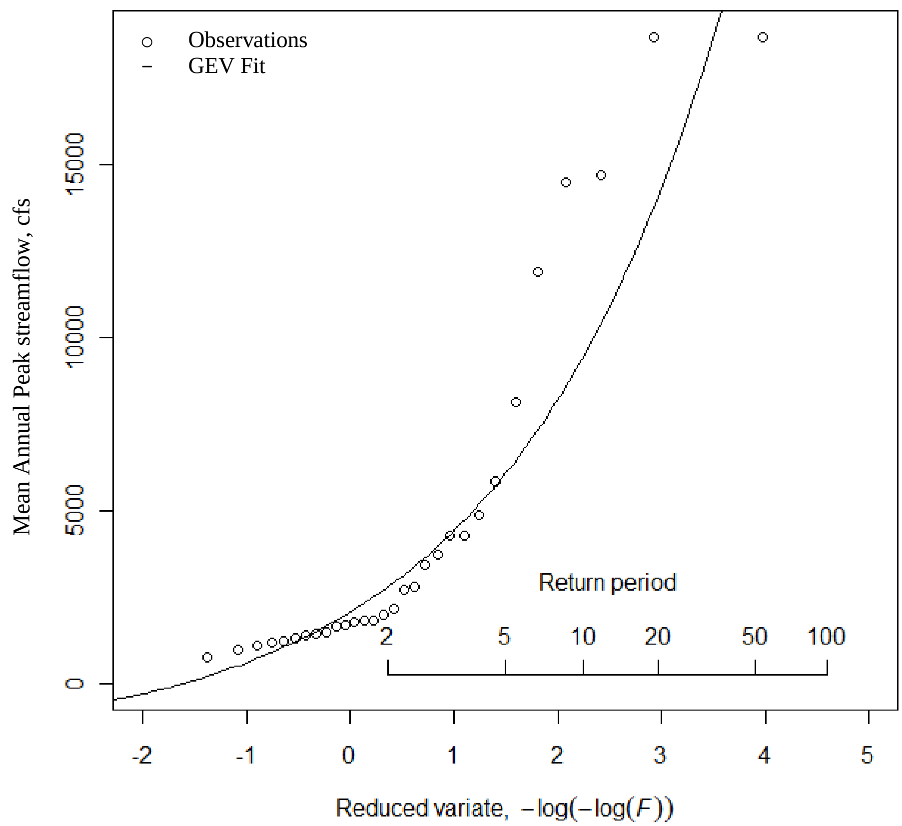
```
         xi        alpha            k
1537.7340516 1007.5147569   -0.5459791
  name  n      l_1         t       t_3       t_4       t_5
1    1 34 3291.971 0.5029942 0.5729141 0.4417894 0.3088251
```



Station 50124200

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(753, 978, 1120, 1190, 1210, 1320, 1400, 1450, 1500, 1640, 1690, 1780,
1800, 1810, 2010, 2170, 2697, 2800, 3440, 3720, 4260, 4300, 4890, 5860,
8130, 11900, 14500, 14700, 18700, 18700)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

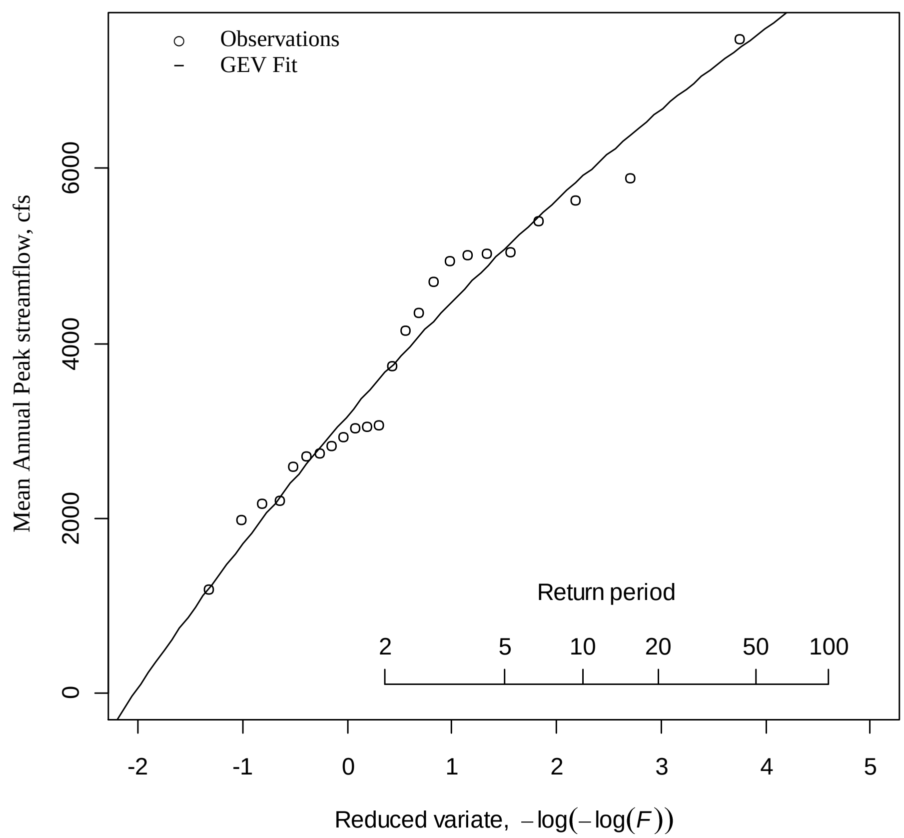
```
         xi       alpha           k
2063.162484 1836.104988   -0.477566
  name  n      l_1          t       t_3       t_4        t_5
1    1 30 4747.267 0.5396494 0.5162951 0.2244669 0.03211737
```



Station 50136400

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(5880, 3020, 7480, 4940, 2920, 4710, 5640, 5000, 2200, 3740, 3040,
4340, 5040, 1970, 2820, 2740, 2710, 5390, 2170, 2590, 4150, 1190, 5020,
3060)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

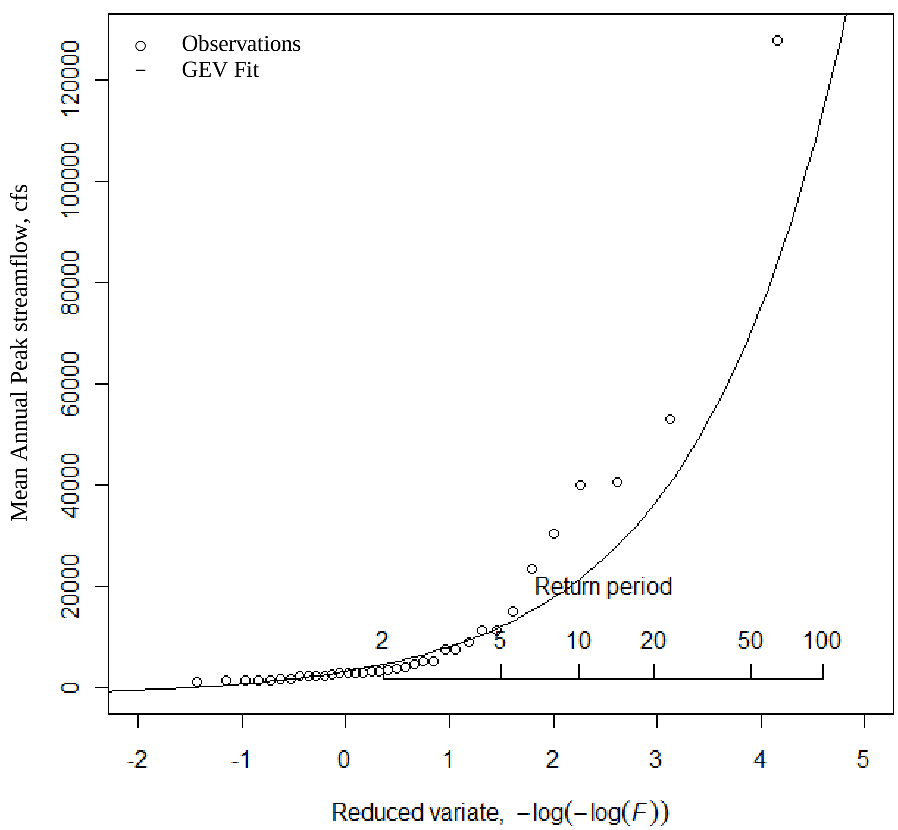
```
          xi        alpha            k
3167.7821129 1382.8525947    0.1150973
  name  n      l_1         t        t_3        t_4         t_5
1    1 24 3823.333 0.2278155 0.09805627 0.06111027 0.006931629
```



Station 50138000

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(1230, 1340, 1340, 1350, 1500, 1700, 1860, 2300, 2320, 2360,
2370, 2650, 2760, 2770, 2780, 2920, 3030, 3310, 3410, 3850,
3960, 4720, 5230, 5310, 7400, 7510, 8930, 11300, 11400, 15000,
23600, 30400, 40000, 40600, 53000, 128000)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

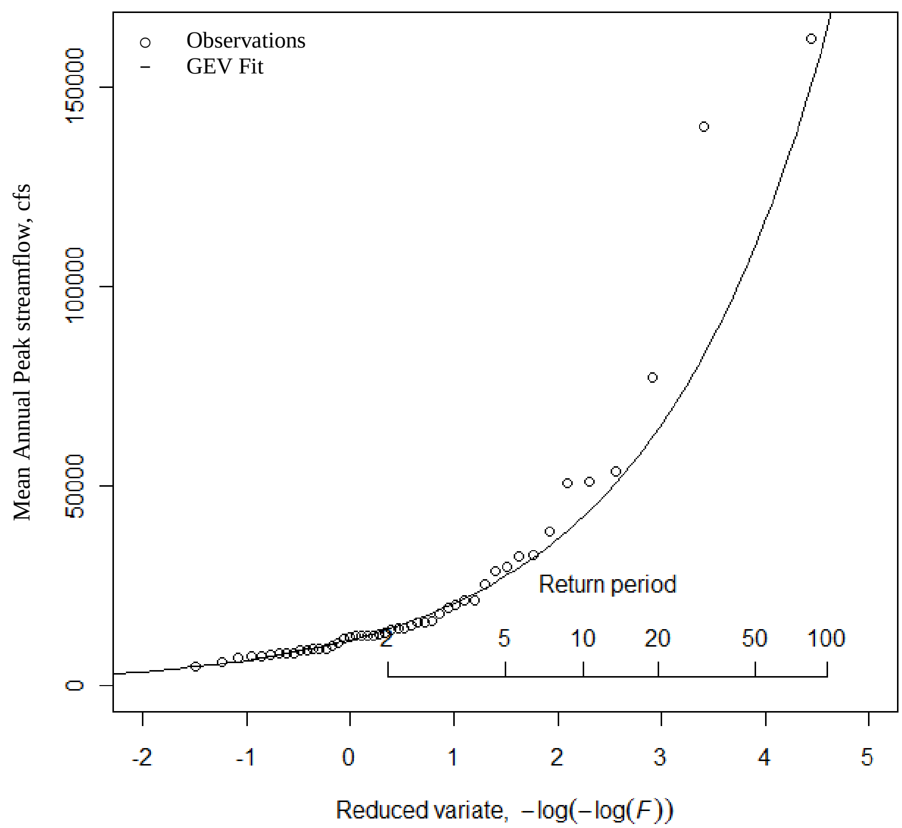
```
          xi        alpha            k
3182.7699802 3414.0809301   -0.6836616
  name  n      l_1         t       t_3       t_4       t_5
1    1 36 12319.72 0.6953266 0.6927246 0.4766538 0.3164175
```



Station 50144000

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(4670, 5950, 6980, 7410, 7450, 7830, 8000, 8020, 8090, 8800, 8960,
9130, 9300, 9320, 9710, 10700, 11600, 12000, 12400, 12400, 12500, 12600,
12700, 13100, 13800, 14200, 14400, 15000, 15600, 15900, 16000, 17800,
19300, 20300, 21300, 21400, 25400, 28600, 29700, 32300, 32700, 38700,
50800, 51200, 53600, 77200, 140000, 162000)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

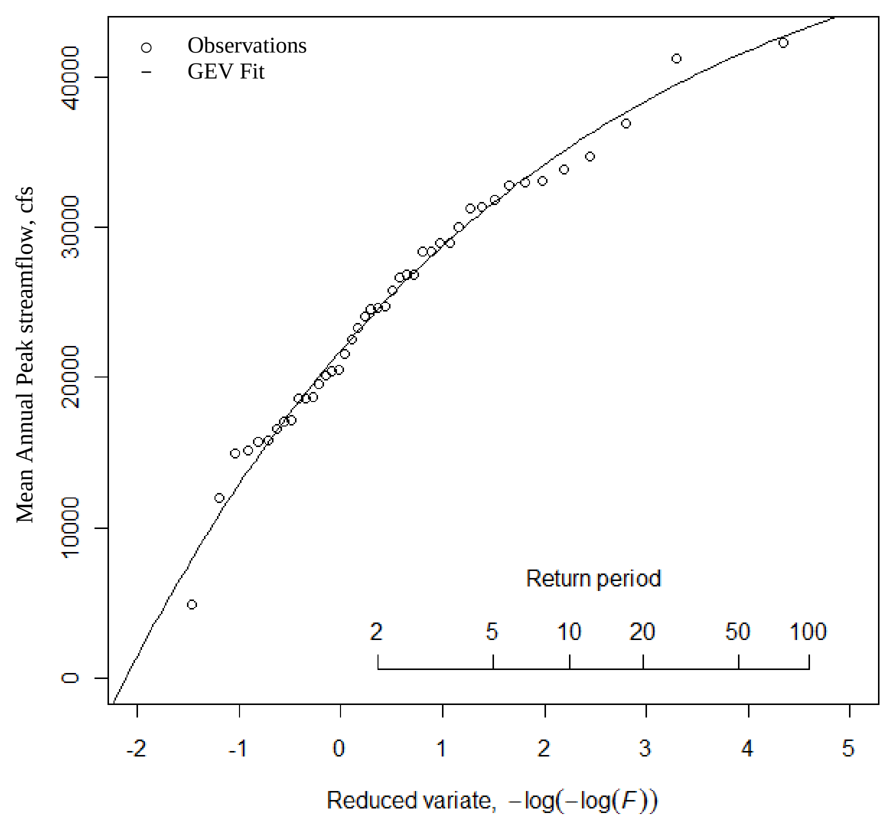
```
              xi           alpha               k
11409.3167332  6724.9077589    -0.5749957
  name  n      l_1          t       t_3       t_4       t_5
1    1 48 24100.42 0.4954893 0.5975035 0.4200445 0.2824378
```



Station 50147800

```
# Call the packages with the functions needed for the analysis.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
# Vector of Peak annual streamflow data for the station.
x=c(4850, 12000, 15000, 15100, 15700, 15800, 16600, 17100, 17130, 18600,
18600, 18700, 19600, 20100, 20400, 20500, 21600, 22500, 23300, 24100,
24500, 24600, 24700, 25800, 26700, 26800, 26800, 28400, 28400, 28950,
29000, 30000, 31300, 31400, 31800, 32800, 33000, 33100, 33800, 34700,
36900, 41200, 42300)
# Computes L-moments
lmom <- samlmu.s(x, nmom=5, sort.data=TRUE)
# Computes GEV Parameters
pelgev(lmom)
regsamlmu(x, nmom = 5, sort.data = TRUE, lcv = TRUE)
# Extreme-value plot
evplot(x)
# Fit a GEV distribution
# Adds the quantile function of a distribution to an extreme-value plot.
evdistq(quagev, pelgev(lmom))
```

```
           xi         alpha             k
2.179453e+04 7.868015e+03 2.482463e-01
  name  n      l_1         t        t_3       t_4            t_5
1    1 43 24749.53 0.1835649 0.01996185 0.0983888 -2.191976e-05
```



B. Cluster Analysis

B.1 Approach *a*: equal importance characteristics R input

```
# Data contains site characteristics
data(DrainageAttributes)
# Take a sequence of vector, matrix or data frames arguments and combine
by columns using cbind
transformed <- cbind(a1 = log10(DrainageAttributes$CDA),
a2 = DrainageAttributes$lat,
a3 = DrainageAttributes$long,
a4 = DrainageAttributes$CS,
a5 = DrainageAttributes$MAR)
#Divide the column by the standard deviation of each value
transformed <- apply(transformed, 2, function(x) x/sd(x))
# Clustering by Hybrid's method
(hyb1<-hybridHclust(transformed, themc=NULL, trace=FALSE))
par(mfrow=c(1,1))
plot(transformed, pch = as.character(1:nrow(transformed)), asp = 1)
plot(hyb1)
```

```
transformed
            a1       a2       a3        a4       a5
 [1,] 2.578375 148.9371 154.2597 0.9883656 2.873245
 [2,] 2.765458 149.8640 154.1018 0.7244713 2.848865
 [3,] 3.578006 148.8987 153.4057 0.6102114 2.432457
 [4,] 2.511504 148.4126 153.5025 0.6340539 2.620995
 [5,] 4.369150 149.1270 153.5141 0.5430705 2.699335
 [6,] 4.554792 149.9996 153.6687 0.3860781 2.673655
 [7,] 2.427549 149.3689 153.2260 1.4008332 2.610593
 [8,] 3.926681 150.1375 153.3171 0.5207137 2.538753
 [9,] 3.698738 147.7955 152.9803 0.4907869 2.196135
[10,] 3.327818 149.2219 152.7679 0.4230092 2.220190
[11,] 1.596863 147.4632 152.4272 2.3658516 3.243493
[12,] 2.071701 147.7729 152.3547 0.3917380 3.250644
[13,] 4.010111 148.4646 152.4734 0.7428661 2.687633
[14,] 2.495333 148.2499 152.2809 0.8914393 2.942483
[15,] 3.653872 148.5935 152.3778 1.2119329 2.626846
[16,] 2.071701 149.0863 152.1955 2.3517017 3.241868
[17,] 1.923633 149.4231 152.0210 2.5788065 6.001015
[18,] 1.778171 149.2988 152.0858 3.1808818 5.615163
[19,] 1.710003 149.1722 151.8767 3.4822732 5.846284
[20,] 1.216352 149.1858 151.8305 2.8731230 3.421303
[21,] 2.403760 148.9235 151.7470 1.9894661 3.875093
[22,] 1.486004 146.7126 152.3130 1.5982233 2.621320
[23,] 2.597981 146.7714 152.5273 1.6060057 2.846264
[24,] 2.025024 147.1919 153.7515 3.3690745 3.149874
[25,] 1.940002 147.1354 153.9151 1.9809762 3.074785
[26,] 2.621899 146.8528 154.2950 1.6880747 2.537128
[27,] 4.270713 147.6576 155.1060 0.5330948 2.302431
[28,] 4.369150 148.8082 154.8801 0.5778083 2.908027
[29,] 3.806313 149.4345 154.9751 0.3778712 3.039353
```

B.2 Approach *b*: area important R input

```
library(lmom)
library(lmomRFA)
library(hybridHclust)
# Data contains site characteristics
data(DrainageAttributes)
# Take a sequence of vector, matrix or data frames arguments and combine
by columns using cbind
transformed <- cbind(a1 = log10(DrainageAttributes$CDA),
a2 = DrainageAttributes$lat,
a3 = DrainageAttributes$long,
a4 = DrainageAttributes$CS,
a5 = DrainageAttributes$MAR)
#Divide the column by the standard deviation of each value
transformed <- apply(transformed, 2, function(x) x/sd(x))
#CDA importance is equal to the other variables together
transformed[,1] <- transformed[,1] *4
# Clustering by Hybrid's method
(hyb1<-hybridHclust(transformed, themc=NULL, trace=FALSE))
par(mfrow=c(1,1))
plot(transformed, pch = as.character(1:nrow(transformed)), asp = 1)
plot(hyb1)
```

```
transformed
             a1       a2       a3        a4       a5
 [1,] 10.313500 148.9371 154.2597 0.9883656 2.873245
 [2,] 11.061830 149.8640 154.1018 0.7244713 2.848865
 [3,] 14.312023 148.8987 153.4057 0.6102114 2.432457
 [4,] 10.046015 148.4126 153.5025 0.6340539 2.620995
 [5,] 17.476600 149.1270 153.5141 0.5430705 2.699335
 [6,] 18.219168 149.9996 153.6687 0.3860781 2.673655
 [7,]  9.710198 149.3689 153.2260 1.4008332 2.610593
 [8,] 15.706722 150.1375 153.3171 0.5207137 2.538753
 [9,] 14.794952 147.7955 152.9803 0.4907869 2.196135
[10,] 13.311272 149.2219 152.7679 0.4230092 2.220190
[11,]  6.387451 147.4632 152.4272 2.3658516 3.243493
[12,]  8.286805 147.7729 152.3547 0.3917380 3.250644
[13,] 16.040445 148.4646 152.4734 0.7428661 2.687633
[14,]  9.981333 148.2499 152.2809 0.8914393 2.942483
[15,] 14.615490 148.5935 152.3778 1.2119329 2.626846
[16,]  8.286805 149.0863 152.1955 2.3517017 3.241868
[17,]  7.694532 149.4231 152.0210 2.5788065 6.001015
[18,]  7.112684 149.2988 152.0858 3.1808818 5.615163
[19,]  6.840013 149.1722 151.8767 3.4822732 5.846284
[20,]  4.865410 149.1858 151.8305 2.8731230 3.421303
[21,]  9.615039 148.9235 151.7470 1.9894661 3.875093
[22,]  5.944016 146.7126 152.3130 1.5982233 2.621320
[23,] 10.391926 146.7714 152.5273 1.6060057 2.846264
[24,]  8.100094 147.1919 153.7515 3.3690745 3.149874
[25,]  7.760006 147.1354 153.9151 1.9809762 3.074785
[26,] 10.487595 146.8528 154.2950 1.6880747 2.537128
[27,] 17.082853 147.6576 155.1060 0.5330948 2.302431
[28,] 17.476600 148.8082 154.8801 0.5778083 2.908027
[29,] 15.225251 149.4345 154.9751 0.3778712 3.039353
```

B.3 Evaluation of clusters heterogeneity and discordancy measure

Example R Input

```
# Call the packages with the functions needed for the analysis of
#regional frequency. First will be obtained the regional L-moments, then
#is identified the homogeneity of the cluster and which distribution is
#the best fit. Finally, the regional quantiles and site quantiles are
#obtained.
# R functions for use with the method of L-moments.
library(lmom)
# R functions for regional frequency analysis using L-moments.
library(lmomRFA)
#Computes a regional weighted average of L-moments.
regavlmom(regdata_cluster) # Weight proportional to record length
regavlmom(regdata_cluster, weight=1) # Equal weights
# Fit a generalized extreme value distribution to the cluster
rfit <- regfit(regdata_cluster, "gev")
rfit # Print details of the fitted distribution
rfit$index # Display the Index flood values for each site
# Plot the regional growth curve
evplot(rfit)
# Discordancy Measure, Heterogeneity measure and Goodness-of-fit measure
regtst(regdata_cluster3_1, nsim=1000)
# Compute cluster quantiles for T=5,10,25,50 and 100
regquant(c(0.8000, 0.9000, 0.9600, 0.9800, 0.99), rfit)
# Compute quantiles for T=5,10,25,50 and 100
sitequant(c(0.8, 0.9, 0.96, 0.98, 0.99), rfit)
```

R Ouputs

Approach *a*: Two Regions

Cluster A_{a2}

```
     l_1       l_2   Column5   Column6   Column7
1.0000000 0.4008452 0.3502243 0.2223710 0.1066801
      l_1       l_2   Column5   Column6   Column7
1.0000000 0.4031775 0.3523911 0.2231883 0.1066351
Regional frequency distribution: gev
Parameters:
        xi      alpha         k
 0.6080621  0.4243489 -0.2625835
 50050900  50051310  50056400  50061800  50063800  50064200  50065500  50067000
 8055.455  6562.061 11524.000  6398.095  8112.093  7121.176 10115.880  4392.258
 50071000  50090500  50092000  50112500  50115000  50124200
 8766.939  3719.061  6217.364  2119.511  3291.971  4747.267
Discordancy measures (critical value 2.97)
1.28 1.11 0.15 0.36 0.63 0.47 2.56 1.86 1.02 0.52 0.23 0.55 1.06 2.20

Heterogeneity measures (based on 1000 simulations)
3.48 5.41 4.27

Goodness-of-fit measures (based on 1000 simulations)
  glo   gev   gno   pe3   gpa
 1.41  0.56 -0.50 -2.33 -2.04
     0.8      0.9     0.96     0.98     0.99
1.388118 1.910002 2.734940 3.494233 4.400175
```

```
             0.8       0.9      0.96      0.98      0.99
50050900 11181.924 15385.937 22031.189 28147.635 35445.410
50051310  9108.917 12533.551 17946.845 22929.369 28874.215
50056400 15996.675 22010.865 31517.452 40267.538 50707.614
50061800  8881.313 12220.375 17498.408 22356.433 28152.736
50063800 11260.545 15494.115 22186.090 28345.541 35694.627
50064200  9885.035 13601.462 19475.991 24883.046 31334.419
50065500 14042.038 19321.353 27666.328 35347.239 44511.640
50067000  6096.974  8389.222 12012.563 15347.572 19326.703
50071000 12169.548 16744.873 23977.055 30633.725 38576.064
50090500  5162.497  7103.415 10171.410 12995.265 16364.518
50092000  8630.437 11875.179 17004.119 21724.917 27357.488
50112500  2942.132  4048.271  5796.736  7406.065  9326.219
50115000  4569.645  6287.672  9003.344 11502.913 14485.248
50124200  6589.768  9067.290 12983.492 16588.056 20888.805
```

Cluster B_{a2}

```
     l_1       l_2   Column5   Column6   Column7
1.0000000 0.4433404 0.3496511 0.2107934 0.1201420
      l_1       l_2   Column5   Column6   Column7
1.0000000 0.4468896 0.3564544 0.2181153 0.1266889
Regional frequency distribution: gev
Parameters:
        xi      alpha          k
 0.5666625  0.4698778 -0.2618006
 50028000  50028400  50031200  50034000  50035000  50038100  50038320  50039500
 5569.020  4486.484 13010.820  5380.914 26879.660 37825.100  8318.000  9118.269
 50043000  50047850  50055000  50057000  50138000  50144000  50147800
18647.350  9964.897 25410.590 22066.000 12319.720 24100.420 24749.530
Discordancy measures (critical value 3.00)
1.04 0.92 0.20 0.32 0.08 0.85 1.35 0.94 0.37 0.64 0.71 1.08 2.77 2.03 1.70

Heterogeneity measures (based on 1000 simulations)
7.82 6.20 4.41

Goodness-of-fit measures (based on 1000 simulations)
  glo   gev   gno   pe3   gpa
 2.15  1.19 -0.01 -2.08 -1.76
     0.8      0.9     0.96     0.98     0.99
1.429872 2.006896 2.918383 3.756800 4.756601
               0.8       0.9      0.96      0.98      0.99
50028000  7962.985 11176.444  16252.53  20921.69  26489.60
50028400  6415.097  9003.907  13093.28  16854.82  21340.41
50031200 18603.805 26111.363  37970.55  48879.05  61887.27
50034000  7694.017 10798.935  15703.57  20215.02  25594.86
50035000 38434.468 53944.683  78445.13 100981.51 127855.81
50038100 54085.044 75911.043 110388.12 142101.34 179918.89
50038320 11893.674 16693.361  24275.11  31249.06  39565.40
50039500 13037.956 18299.418  26610.60  34255.51  43371.96
50043000 26663.320 37423.293  54420.10  70054.37  88698.00
50047850 14248.525 19998.512  29081.38  37436.13  47399.04
50055000 36333.886 50996.412  74157.83  95462.51 120868.03
50057000 31551.551 44284.168  64397.03  82897.55 104959.15
50138000 17615.620 24724.397  35953.66  46282.72  58599.99
50144000 34460.511 48367.037  70334.25  90540.46 114636.07
50147800 35388.655 49669.734  72228.60  92979.04 117723.63
```

Approach *a*: Four Regions

Cluster A_{a4}

```
      l_1       l_2   Column5   Column6   Column7
1.0000000 0.4592085 0.4525098 0.2974913 0.1795529
      l_1       l_2   Column5   Column6   Column7
1.0000000 0.4608294 0.4517979 0.2953517 0.1746265
Regional frequency distribution: gev
Parameters:
        xi      alpha          k
 0.5281938  0.3877682 -0.3976541
 50050900  50051310  50056400  50090500  50092000  50112500  50115000  50124200
 8055.455  6562.061 11524.000  3719.061  6217.364  2119.511  3291.971  4747.267
Discordancy measures (critical value 2.14)
1.20 1.35 0.83 0.55 0.10 0.77 1.16 2.05

Heterogeneity measures (based on 1000 simulations)
-0.30 1.27 1.17


Goodness-of-fit measures (based on 1000 simulations)
  glo   gev   gno   pe3   gpa
 0.28 -0.01 -0.89 -2.38 -1.23
     0.8      0.9     0.96     0.98     0.99
1.323591 1.939232 3.032021 4.154852 5.627516
               0.8       0.9      0.96      0.98     0.99
50050900 10662.124 15621.399 24424.312 33469.225 45332.20
50051310  8685.482 12725.361 19896.309 27264.393 36928.10
50056400 15253.058 22347.714 34941.014 47880.516 64851.49
50090500  4922.514  7212.124 11276.272 15452.149 20929.07
50092000  8229.244 12056.914 18851.180 25832.228 34988.31
50112500  2805.365  4110.224  6426.403  8806.255 11927.58
50115000  4357.222  6383.897  9981.326 13677.653 18525.62
50124200  6283.438  9206.054 14393.815 19724.193 26715.32
```

Cluster B_{a4}

```
       l_1        l_2    Column5    Column6    Column7
1.00000000 0.32722331 0.22118759 0.12756570 0.01482936
       l_1        l_2    Column5    Column6    Column7
1.00000000 0.32633413 0.21988113 0.12696182 0.01608455
Regional frequency distribution: gev
Parameters:
         xi       alpha           k
 0.71141382  0.43682973 -0.07821653
 50061800  50063800  50064200  50065500  50067000  50071000
 6398.095  8112.093  7121.176 10115.880  4392.258  8766.939
Discordancy measures (critical value 1.65)
0.63 0.96 1.07 1.46 1.09 0.79

Heterogeneity measures (based on 1000 simulations)
1.30 1.36 0.92

Goodness-of-fit measures (based on 1000 simulations)
  glo   gev   gno   pe3   gpa
 3.01  1.69  1.26  0.42 -1.47
     0.8      0.9     0.96     0.98     0.99
1.406616 1.786263 2.298939 2.704607 3.129966
               0.8       0.9     0.96     0.98     0.99
50061800  8999.660 11428.679 14708.83 17304.33 20025.82
50063800 11410.596 14490.330 18649.21 21940.02 25390.58
50064200 10016.757 12720.292 16371.15 19259.98 22289.04
50065500 14229.154 18069.621 23255.79 27359.48 31662.36
50067000  6178.218  7845.727 10097.53 11879.33 13747.62
50071000 12331.713 15660.057 20154.66 23711.12 27440.22
```

Cluster C_{a4}

```
     l_1       l_2   Column5   Column6   Column7
1.0000000 0.4425629 0.3302840 0.1823672 0.1015092
      l_1       l_2   Column5   Column6   Column7
1.0000000 0.4440802 0.3363856 0.1897202 0.1084583
Regional frequency distribution: gev
Parameters:
        xi      alpha          k
 0.5727163  0.4873727 -0.2351710
 50028000  50028400  50031200  50034000  50035000  50038100  50038320  50039500
 5569.020  4486.484 13010.820  5380.914 26879.660 37825.100  8318.000  9118.269
 50043000  50047850  50055000  50057000
18647.350  9964.897 25410.590 22066.000
Discordancy measures (critical value 2.76)
1.61 2.09 0.13 1.10 0.19 0.82 1.56 0.81 0.37 1.59 0.76 0.97

Heterogeneity measures (based on 1000 simulations)
4.54 3.80 2.45


Goodness-of-fit measures (based on 1000 simulations)
  glo   gev   gno   pe3   gpa
 3.18  2.09  0.96 -1.02 -1.08
     0.8      0.9     0.96     0.98     0.99
1.449276 2.018455 2.897281 3.688250 4.614033
               0.8       0.9      0.96      0.98      0.99
50028000  8071.045 11240.817  16135.02  20539.94  25695.64
50028400  6502.152  9055.767  12998.61  16547.27  20700.78
50031200 18856.265 26261.757  37696.00  47987.15  60032.35
50034000  7798.428 10861.134  15590.02  19846.16  24827.71
50035000 38956.038 54255.388  77877.93  99138.90 124023.63
50038100 54818.998 76348.268 109589.95 139508.42 174526.25
50038320 12055.075 16789.510  24099.59  30678.86  38379.53
50039500 13214.886 18404.817  26418.19  33630.45  42071.99
50043000 27025.151 37638.840  54026.62  68776.09  86039.49
50047850 14441.883 20113.698  28871.11  36753.03  45978.36
50055000 36826.950 51290.136  73621.63  93720.60 117245.30
50057000 31979.718 44539.231  63931.41  81384.92 101813.25
```

Cluster D_{a4}

```
      l_1       l_2   Column5   Column6   Column7
1.0000000 0.4465239 0.4289497 0.3271842 0.1964339
      l_1       l_2   Column5   Column6   Column7
1.0000000 0.4581269 0.4367300 0.3316957 0.1996111
Regional frequency distribution: gev
Parameters:
        xi      alpha          k
 0.5454715  0.3987028 -0.3673280
50138000 50144000 50147800
12319.72 24100.42 24749.53
Discordancy measures (critical value 3.00)
1.00 1.00 1.00

Heterogeneity measures (based on 1000 simulations)
5.31 4.74 3.12

Goodness-of-fit measures (based on 1000 simulations)
  glo   gev   gno   pe3   gpa
-0.70 -0.90 -1.38 -2.21 -1.66
     0.8      0.9     0.96     0.98     0.99
1.343180 1.940865 2.974470 4.010627 5.341045
              0.8      0.9     0.96     0.98      0.99
50138000 16547.60 23910.91 36644.64 49409.81  65800.18
50144000 32371.20 46775.66 71685.98 96657.80 128721.44
50147800 33243.07 48035.50 73616.73 99261.14 132188.36
```

Approach *b*: Two Regions

Cluster A_{b2}

```
      l_1       l_2   Column5   Column6   Column7
1.0000000 0.3957002 0.3347058 0.2215854 0.1171007
     l_1      l_2  Column5  Column6  Column7
1.000000 0.399638 0.338693 0.221202 0.114994
Regional frequency distribution: gev
Parameters:
        xi      alpha          k
 0.6168512  0.4320202 -0.2412818
 50028000  50028400  50034000  50038320  50050900  50051310  50056400  50061800
 5569.020  4486.484  5380.914  8318.000  8055.455  6562.061 11524.000  6398.095
 50063800  50064200  50065500  50067000  50071000  50090500  50092000  50112500
 8112.093  7121.176 10115.880  4392.258  8766.939  3719.061  6217.364  2119.511
 50115000  50124200
 3291.971  4747.267


Discordancy measures (critical value 3.00)
1.56 0.27 1.26 1.20 1.51 1.15 0.14 0.39 0.80 0.58 1.81 2.00 0.99 0.52 0.21 0.65
1.13 1.85

Heterogeneity measures (based on 1000 simulations)
4.32 5.78 4.31

Goodness-of-fit measures (based on 1000 simulations)
  glo   gev   gno   pe3   gpa
 1.16  0.14 -0.97 -2.90 -2.87
     0.8      0.9     0.96     0.98     0.99
1.397639 1.908024 2.700203 3.416760 4.259046
               0.8       0.9      0.96     0.98      0.99
50028000  7783.478 10625.824 15037.487 19028.00 23718.711
50028400  6270.484  8560.319 12114.420 15329.24 19108.140
50034000  7520.574 10266.913 14529.563 18385.29 22917.559
50038320 11625.559 15870.944 22460.293 28420.61 35426.742
50050900 11258.615 15370.002 21751.368 27523.55 34308.551
50051310  9171.390 12520.570 17718.900 22420.98 27948.118
50056400 16106.388 21988.069 31117.145 39374.74 49081.242
50061800  8942.225 12207.719 17276.158 21860.75 27249.779
50063800 11337.775 15478.068 21904.302 27717.07 34549.775
50064200  9952.831 13587.375 19228.624 24331.35 30329.414
50065500 14138.345 19301.342 27314.934 34563.53 43083.995
50067000  6138.790  8380.534 11859.990 15007.29 18706.827
50071000 12253.013 16727.530 23672.519 29954.52 37338.794
50090500  5197.904  7096.058 10042.221 12707.14 15839.651
50092000  8689.628 11862.880 16788.148 21243.24 26480.037
50112500  2962.311  4044.078  5723.111  7241.86  9027.094
50115000  4600.986  6281.160  8888.992 11247.87 14020.655
50124200  6634.964  9057.900 12818.587 16220.27 20218.827
```

Cluster B_{b2}

```
      l_1       l_2   Column5   Column6   Column7
1.0000000 0.4633968 0.3707916 0.2082162 0.1100022
      l_1       l_2   Column5   Column6   Column7
1.0000000 0.4685206 0.3803787 0.2194071 0.1203463
Regional frequency distribution: gev
Parameters:
        xi      alpha          k
 0.5414467  0.4702969 -0.2904743
 50031200  50035000  50038100  50039500  50043000  50047850  50055000  50057000
13010.820 26879.660 37825.100  9118.269 18647.350  9964.897 25410.590 22066.000
 50138000  50144000  50147800
12319.720 24100.420 24749.530
```

```
Discordancy measures (critical value 2.63)
0.14 0.05 0.67 0.83 0.27 0.58 1.06 0.81 2.28 2.20 2.11

Heterogeneity measures (based on 1000 simulations)
7.19 5.06 4.28

Goodness-of-fit measures (based on 1000 simulations)
  glo   gev   gno   pe3   gpa
 2.58  1.80  0.66 -1.33 -0.68
     0.8      0.9     0.96     0.98     0.99
1.425514 2.035182 3.022193 3.951586 5.082365
              0.8       0.9      0.96      0.98      0.99
50031200 18547.11 26479.39  39321.21  51413.38  66125.74
50035000 38317.33 54705.00  81235.52 106217.30 136612.25
50038100 53920.21 76980.97 114314.75 149469.15 192240.97
50039500 12998.22 18557.34  27557.17  36031.63  46342.37
50043000 26582.06 37950.75  56355.89  73686.61  94772.64
50047850 14205.10 20280.38  30115.84  39377.15  50645.24
50055000 36223.15 51715.18  76795.71 100412.14 129145.89
50057000 31455.39 44908.33  66687.71  87195.70 112147.47
50138000 17561.93 25072.87  37232.57  48682.44  62613.31
50144000 34355.49 49048.74  72836.12  95234.89 122487.13
50147800 35280.80 50369.80  74797.85  97799.90 125786.15
```

Approach *b*: Four Regions

Cluster A_{b4}

```
     l_1       l_2   Column5   Column6   Column7
1.0000000 0.4061591 0.3423380 0.2276285 0.1253744
      l_1       l_2   Column5   Column6   Column7
1.0000000 0.4127310 0.3456598 0.2239698 0.1206823
Regional frequency distribution: gev
Parameters:
        xi      alpha          k
 0.6047995  0.4368115 -0.2517863
 50028000  50028400  50034000  50038320  50051310  50056400  50061800  50071000
 5569.020  4486.484  5380.914  8318.000  6562.061 11524.000  6398.095  8766.939
 50092000  50112500  50115000  50124200
 6217.364  2119.511  3291.971  4747.267
Discordancy measures (critical value 2.76)
1.50 0.69 1.12 1.42 1.11 0.15 0.41 1.67 0.23 0.85 1.41 1.46

Heterogeneity measures (based on 1000 simulations)
3.38 4.29 3.00

Goodness-of-fit measures (based on 1000 simulations)
  glo   gev   gno   pe3   gpa
 0.93  0.13 -0.81 -2.43 -2.29
     0.8      0.9     0.96     0.98     0.99
1.400876 1.927251 2.751630 3.503742 4.394355
               0.8       0.9      0.96      0.98      0.99
50028000  7801.505 10732.900 15323.882 19512.412 24472.250
50028400  6285.007  8646.581 12345.144 15719.485 19715.203
50034000  7537.992 10370.372 14806.284 18853.337 23645.645
50038320 11652.485 16030.875 22888.058 29144.130 36552.243
50051310  9192.632 12646.739 18056.363 22991.772 28836.024
50056400 16143.693 22209.642 31709.783 40377.128 50640.545
50061800  8962.936 12330.736 17605.190 22417.277 28115.500
50071000 12281.393 16896.093 24123.372 30717.097 38525.041
50092000  8709.755 11982.422 17107.885 21784.042 27321.303
50112500  2969.172  4084.830  5832.110  7426.221  9313.883
50115000  4611.643  6344.455  9058.286 11534.219 14466.089
50124200  6650.331  9149.176 13062.722 16633.201 20861.176
```

Cluster B_{b4}

```
      l_1        l_2    Column5    Column6    Column7
1.00000000 0.37201677 0.31742319 0.20790141 0.09836562
      l_1       l_2   Column5   Column6   Column7
1.0000000 0.3734521 0.3247595 0.2156665 0.1036175
Regional frequency distribution: gev
Parameters:
        xi      alpha          k
 0.6439550  0.4199373 -0.2172917
 50050900  50063800  50064200  50065500  50067000  50090500
 8055.455  8112.093  7121.176 10115.880  4392.258  3719.061
Discordancy measures (critical value 1.65)
1.17 0.80 0.35 1.60 1.19 0.89

Heterogeneity measures (based on 1000 simulations)
2.88 4.10 3.50

Goodness-of-fit measures (based on 1000 simulations)
  glo   gev   gno   pe3   gpa
 0.90  0.24 -0.38 -1.46 -1.65
     0.8      0.9     0.96     0.98     0.99
1.388605 1.862771 2.583778 3.223281 3.962465


               0.8       0.9      0.96     0.98     0.99
50050900 11185.848 15005.465 20813.506 25965.00 31919.46
50063800 11264.496 15110.968 20959.846 26147.56 32143.88
50064200  9888.503 13265.117 18399.536 22953.55 28217.41
50065500 14046.965 18843.564 26137.186 32606.33 40083.82
50067000  6099.113  8181.769 11348.619 14157.48 17404.17
50090500  5164.308  6927.757  9609.227 11987.58 14736.65
```

Cluster C_{b4}

```
     l_1        l_2        t_3        t_4        t_5
1.00000000 0.42321821 0.29945388 0.15747254 0.07053799
       l_1        l_2        t_3        t_4        t_5
1.00000000 0.42790019 0.31162856 0.17107216 0.08329827
Regional frequency distribution: gev
Parameters:
        xi      alpha          k
 0.6001912  0.4940580 -0.1920423
 50031200  50039500  50043000  50047850  50055000  50057000  50147800
13010.820  9118.269 18647.350  9964.897 25410.590 22066.000 24749.530
Discordancy measures (critical value 1.92)
0.15 0.84 0.28 1.48 1.14 1.27 1.84

Heterogeneity measures (based on 1000 simulations)
7.08 5.07 3.47

Goodness-of-fit measures (based on 1000 simulations)
  glo   gev   gno   pe3   gpa
 2.97  1.99  1.22 -0.14 -0.70
     0.8      0.9     0.96     0.98     0.99
1.459011 1.990942 2.782513 3.470235 4.251214
              0.8      0.9     0.96     0.98      0.99
50031200 18982.93 25903.79 36202.78 45150.61  55311.78
50039500 13303.65 18153.95 25371.70 31642.54  38763.71
50043000 27206.69 37125.80 51886.49 64710.70  79273.88
50047850 14538.89 19839.53 27727.46 34580.54  42362.91
50055000 37074.33 50591.02 70705.30 88180.73 108025.86
50057000 32194.53 43932.13 61398.93 76574.22  93807.29
50147800 36109.83 49274.89 68865.89 85886.70 105215.55
```

Cluster D_{b4}

```
     l_1       l_2       t_3       t_4       t_5
1.0000000 0.5268037 0.4833716 0.2882961 0.1722818
      l_1       l_2       t_3       t_4       t_5
1.0000000 0.5396063 0.5006915 0.3039932 0.1851802
Regional frequency distribution: gev
Parameters:
        xi      alpha          k
 0.4530612  0.4119288 -0.4367087
50035000 50038100 50138000 50144000
26879.66 37825.10 12319.72 24100.42
Discordancy measures (critical value 3.00)
1.00 1.00 1.00 1.00

Heterogeneity measures (based on 1000 simulations)
1.82 2.34 2.12

Goodness-of-fit measures (based on 1000 simulations)
  glo   gev   gno   pe3   gpa
 0.98  0.78  0.02 -1.27 -0.12
     0.8      0.9     0.96     0.98     0.99
1.325776 2.030006 3.322786 4.693895 6.542071
              0.8      0.9      0.96      0.98      0.99
50035000 35636.40 54565.86  89315.37 126170.30 175848.65
50038100 50147.60 76785.17 125684.73 177547.05 247454.50
50138000 16333.19 25009.10  40935.80  57827.47  80596.49
50144000 31951.75 48923.99  80080.55 113124.84 157666.67
```

C. Selected regions for further analysis

Table 14. *L*-moments and site characteristics of Region 1

Region 1													
id	Site number	n	Mean	L-CV	L-skewness	L-kurtosis	t5	Lat	Long	CDA	MAR	CS	D_i
			L1	t	t3	t4				(mi^2)	(in)	(ft/mi)	
1	50028000	51	5569.02	0.296	0.208	0.251	0.215	18.301	66.783	18	72.32	139.7	2.0
2	50028400	31	4486.484	0.379	0.334	0.186	0.086	18.414	66.714	22.2	63.11	102.4	1.4
4	50034000	35	5380.914	0.548	0.528	0.321	0.171	18.236	66.455	16.7	74.61	89.62	1.1
11	50050900	33	8055.455	0.469	0.611	0.470	0.336	18.119	65.989	5.99	72.76	334.4	1.4
22	50090500	33	3719.061	0.489	0.505	0.359	0.208	18.027	65.940	5.29	66.10	225.9	0.2
23	50092000	44	6217.364	0.451	0.443	0.313	0.203	18.034	66.033	18.4	60.94	227	0.1
24	50112500	47	2119.511	0.489	0.577	0.375	0.272	18.086	66.563	9.68	60.51	476.2	0.7
25	50115000	34	3291.971	0.503	0.573	0.442	0.309	18.079	66.634	8.8	58.15	280	0.7
26	50124200	30	4747.267	0.540	0.516	0.224	0.032	18.044	66.798	18.9	50.83	238.6	1.4

Table 15. *L*-moments and site characteristics of Region 2

Region 2													
id	Site number	n	Mean	L-CV	L-skewness	L-kurtosis	t5	Lat	Long	CDA	MAR	CS	D_i
			L1	t	t3	t4				(mi^2)	(in)	(ft/mi)	
12	50051310	33	6562.061	0.367	0.134	0.033	-0.026	18.158	65.958	10.2	77.56	55.37	0.9
14	50056400	40	11524.000	0.379	0.256	0.147	0.063	18.216	65.926	16.4	69.17	126	0.6
16	50061800	42	6398.095	0.372	0.214	0.100	-0.002	18.319	65.889	10.2	66.97	332.4	0.3
17	50063800	43	8112.093	0.333	0.182	0.050	-0.005	18.360	65.814	8.64	64.80	364.5	0.9
18	50064200	34	7121.176	0.389	0.369	0.191	0.004	18.345	65.842	7.34	65.59	449.6	1.1
19	50065500	34	10115.880	0.279	0.034	0.100	0.051	18.329	65.751	6.8	62.60	492.2	1.9
20	50067000	31	4392.258	0.282	0.249	0.125	0.028	18.331	65.731	3.91	60.28	406.1	1.4
21	50071000	49	8766.939	0.303	0.272	0.197	0.022	18.299	65.695	14.8	52.95	281.2	0.9

Table 16. *L*-moments and site characteristics of Region 3

Region 3													
id	Site number	n	Mean	L-CV	L-skewness	L-kurtosis	t5	Lat	Long	CDA	MAR	CS	D_i
			L1	t	t3	t4				(mi^2)	(in)	(ft/mi)	
3	50031200	44	13010.82	0.470	0.374	0.172	0.034	18.296	66.413	55.2	68.50	86.25	0.4
5	50035000	59	26879.66	0.492	0.427	0.239	0.118	18.324	66.460	134	68.19	76.76	0.2
6	50038100	49	37825.1	0.476	0.285	0.081	0.024	18.431	66.527	165	59.76	54.57	1.2
8	50039500	52	9118.269	0.498	0.407	0.130	0.024	18.448	66.375	81.6	59.33	73.6	1.2
9	50043000	34	18647.35	0.526	0.440	0.192	0.040	18.160	66.229	63.2	59.37	69.37	0.5
10	50047850	29	9964.897	0.493	0.472	0.372	0.339	18.336	66.137	41.7	59.33	59.79	1.7
13	50055000	51	25410.59	0.356	0.193	0.187	0.127	18.243	66.009	89.6	62.60	105	2.1
15	50057000	50	22066	0.469	0.275	0.046	0.020	18.258	65.968	60.1	66.34	171.3	0.8

Table 17. *L*-moments and site characteristics of Region 4

Region 4													
id	Site number	n	Mean	L-CV	L-skewness	L-kurtosis	t5	Lat	Long	CDA	MAR	CS	D_i
			L1	t	t3	t4				(mi^2)	(in)	(ft/mi)	
27	50138000	36	12,320	0.695	0.693	0.477	0.316	18.143	67.149	120	62.36	75.35	1
28	50144000	48	24,100	0.495	0.598	0.420	0.282	18.285	67.051	134	74.84	81.67	1
29	50147800	43	24,750	0.184	0.020	0.098	-0.00002	18.362	67.093	71.3	65.91	53.41	1

C.1 *L*-moment Ratio Diagrams for regions defined

Example R Input

```
library(lmom)
library(lmomRFA)
#Draw the L-moment Diagram for the clusters identified.
lmrd (regdata, xaxs="i", yaxs="i",las=1, twopar=FALSE)
```

R Outputs

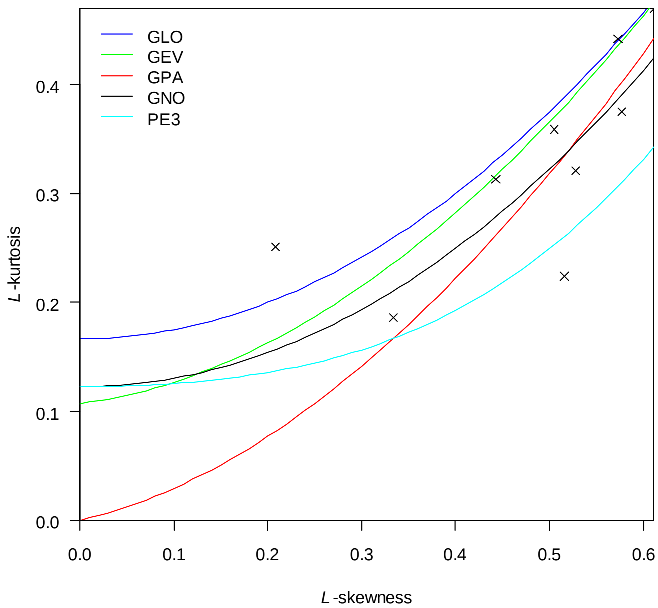


Figure 10. *L*-moment ratio diagram Region 1

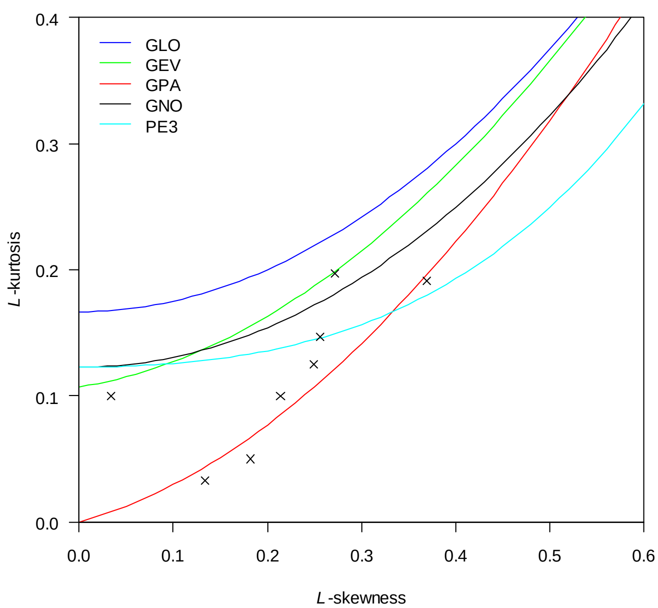


Figure 11. *L*-moment ratio diagram Region 1

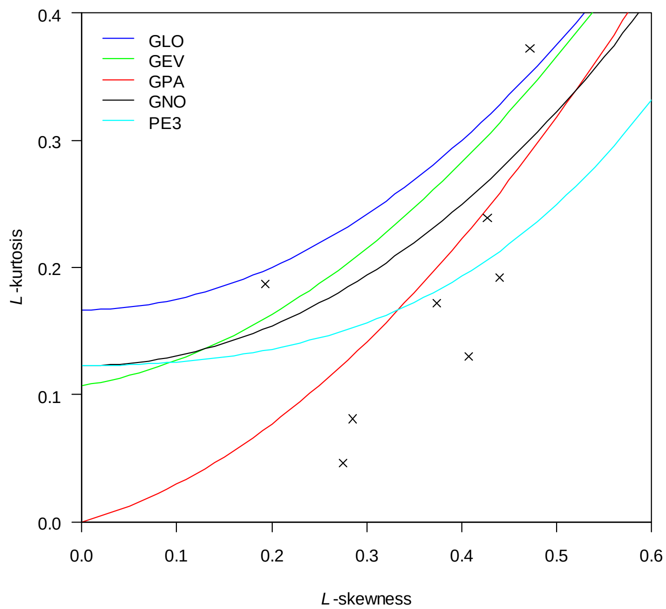


Figure 12. *L*-moment ratio diagram Region 3

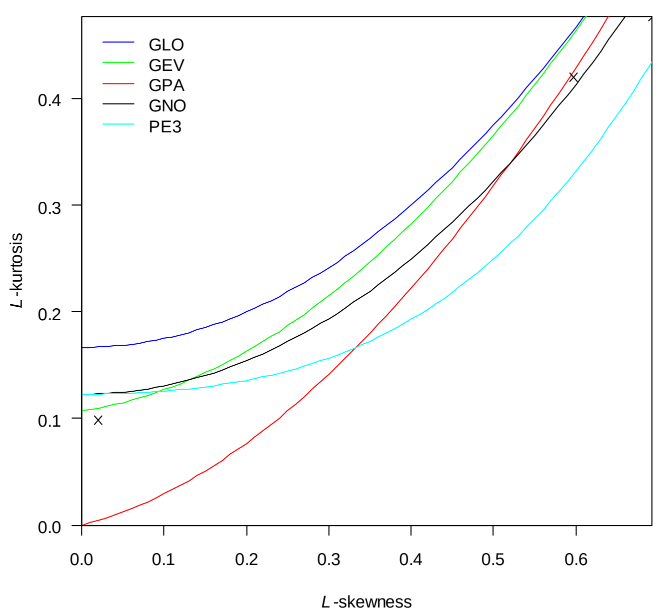


Figure 13. *L*-moment ratio diagram Region 4

C.2 Relative Regional RMSE

Example R Input

```
library(lmom)
library(lmomRFA)
rfit <- regfit(regdata_cluster, "gev")
evplot(rfit) # Plot the regional growth curve
# Compute error bounds for quantile estimates. We will
# (optimistically) generate bounds for a homogeneous region
# with the same frequency distribution as the one fitted to
# the regdata_cluster4 data.
fval <- seq(.2, .99, by=.01) # A lot of quantiles
simq <- regsimq(rfit$qfunc, nrec=regdata_clusterb$n, nrep=100, f=fval,
fit=rfit$dist)
# Regional growth curve, and bounds
rbounds <- regquantbounds(simq, rfit)
evplot(rfit, rbounds)
```

Outputs

Table 18. Regional relative RMSE and error bounds at 90% confidence level

Region 1

F	q(F)	Relative RMSE	rel.bound. 0.05	rel.bound. 0.95	F	q(F)	Relative RMSE	rel.bound. 0.05	rel.bound. 0.95
0.2	0.370	0.024	0.323	0.415	0.6	0.818	0.049	0.727	0.855
0.21	0.379	0.025	0.332	0.423	0.61	0.834	0.049	0.743	0.871
0.22	0.389	0.025	0.340	0.432	0.62	0.851	0.050	0.759	0.887
0.23	0.398	0.026	0.348	0.440	0.63	0.868	0.050	0.776	0.904
0.24	0.408	0.027	0.356	0.450	0.64	0.886	0.050	0.793	0.921
0.25	0.417	0.027	0.365	0.458	0.65	0.904	0.051	0.812	0.939
0.26	0.426	0.028	0.373	0.466	0.66	0.923	0.051	0.830	0.958
0.27	0.436	0.029	0.381	0.476	0.67	0.943	0.051	0.850	0.977
0.28	0.445	0.030	0.389	0.485	0.68	0.964	0.051	0.869	0.997
0.29	0.454	0.030	0.397	0.495	0.69	0.985	0.052	0.890	1.018
0.3	0.464	0.031	0.406	0.505	0.7	1.008	0.052	0.913	1.039
0.31	0.473	0.032	0.415	0.514	0.71	1.031	0.052	0.937	1.062
0.32	0.483	0.033	0.423	0.524	0.72	1.055	0.052	0.963	1.085
0.33	0.493	0.033	0.432	0.533	0.73	1.081	0.052	0.989	1.110
0.34	0.502	0.034	0.441	0.543	0.74	1.107	0.052	1.017	1.135
0.35	0.512	0.035	0.449	0.553	0.75	1.136	0.051	1.045	1.162
0.36	0.522	0.035	0.458	0.563	0.76	1.165	0.051	1.076	1.191
0.37	0.532	0.036	0.467	0.573	0.77	1.197	0.051	1.108	1.223
0.38	0.542	0.037	0.476	0.583	0.78	1.230	0.050	1.141	1.256
0.39	0.553	0.037	0.486	0.593	0.79	1.265	0.049	1.177	1.291
0.40	0.563	0.038	0.496	0.604	0.8	1.302	0.048	1.216	1.328
0.41	0.574	0.039	0.506	0.614	0.81	1.343	0.047	1.259	1.368
0.42	0.584	0.039	0.516	0.625	0.82	1.386	0.046	1.302	1.410
0.43	0.595	0.040	0.526	0.636	0.83	1.432	0.044	1.351	1.456
0.44	0.606	0.040	0.537	0.647	0.84	1.482	0.043	1.404	1.506
0.45	0.618	0.041	0.547	0.658	0.85	1.537	0.040	1.462	1.560
0.46	0.629	0.042	0.558	0.669	0.86	1.597	0.038	1.526	1.619
0.47	0.641	0.042	0.568	0.681	0.87	1.663	0.035	1.597	1.684
0.48	0.653	0.043	0.579	0.693	0.88	1.736	0.032	1.676	1.761
0.49	0.665	0.043	0.590	0.705	0.89	1.819	0.028	1.759	1.841
0.50	0.677	0.044	0.601	0.717	0.9	1.912	0.026	1.848	1.945
0.51	0.690	0.044	0.612	0.729	0.91	2.019	0.025	1.955	2.062
0.52	0.702	0.045	0.623	0.742	0.92	2.144	0.029	2.088	2.211
0.53	0.716	0.046	0.635	0.755	0.93	2.293	0.038	2.231	2.389
0.54	0.729	0.046	0.647	0.768	0.94	2.475	0.055	2.401	2.613
0.55	0.743	0.047	0.659	0.782	0.95	2.706	0.081	2.602	2.907
0.56	0.757	0.047	0.672	0.796	0.96	3.011	0.120	2.888	3.298
0.57	0.772	0.048	0.685	0.810	0.97	3.448	0.183	3.273	3.887
0.58	0.786	0.048	0.698	0.824	0.98	4.158	0.301	3.894	4.855
0.59	0.802	0.048	0.712	0.839	0.99	5.681	0.601	5.198	7.138

Table 19. Regional relative RMSE and error bounds at 90% confidence level Region 2

F	q(F)	Relative RMSE	rel.bound. 0.05	rel.bound. 0.95	F	q(F)	Relative RMSE	rel.bound. 0.05	rel.bound . 0.95
0.2	0.470	0.016	0.443	0.506	0.6	1.022	0.014	0.990	1.048
0.21	0.483	0.016	0.454	0.519	0.61	1.039	0.013	1.009	1.064
0.22	0.496	0.016	0.466	0.532	0.62	1.056	0.013	1.028	1.081
0.23	0.509	0.016	0.479	0.547	0.63	1.074	0.013	1.048	1.097
0.24	0.522	0.016	0.491	0.561	0.64	1.092	0.012	1.067	1.114
0.25	0.535	0.017	0.504	0.575	0.65	1.110	0.012	1.085	1.131
0.26	0.548	0.017	0.516	0.589	0.66	1.129	0.011	1.105	1.149
0.27	0.561	0.017	0.529	0.601	0.67	1.148	0.011	1.124	1.169
0.28	0.574	0.017	0.541	0.614	0.68	1.167	0.010	1.144	1.188
0.29	0.586	0.017	0.553	0.628	0.69	1.188	0.010	1.165	1.209
0.3	0.599	0.017	0.565	0.642	0.7	1.208	0.010	1.186	1.229
0.31	0.612	0.017	0.577	0.654	0.71	1.229	0.009	1.208	1.248
0.32	0.625	0.017	0.589	0.667	0.72	1.251	0.009	1.231	1.269
0.33	0.637	0.017	0.601	0.679	0.73	1.274	0.009	1.254	1.291
0.34	0.650	0.017	0.614	0.691	0.74	1.297	0.009	1.278	1.313
0.35	0.663	0.017	0.626	0.704	0.75	1.321	0.009	1.303	1.337
0.36	0.676	0.017	0.639	0.716	0.76	1.345	0.009	1.328	1.362
0.37	0.689	0.017	0.651	0.729	0.77	1.371	0.009	1.353	1.389
0.38	0.702	0.017	0.664	0.742	0.78	1.398	0.009	1.378	1.417
0.39	0.715	0.017	0.677	0.755	0.79	1.425	0.010	1.403	1.444
0.4	0.728	0.017	0.690	0.768	0.8	1.454	0.010	1.429	1.473
0.41	0.742	0.017	0.703	0.781	0.81	1.484	0.011	1.456	1.503
0.42	0.755	0.017	0.716	0.794	0.82	1.516	0.012	1.487	1.537
0.43	0.769	0.017	0.729	0.806	0.83	1.549	0.014	1.516	1.573
0.44	0.782	0.017	0.743	0.819	0.84	1.583	0.015	1.547	1.611
0.45	0.796	0.017	0.756	0.832	0.85	1.620	0.017	1.581	1.651
0.46	0.810	0.017	0.770	0.845	0.86	1.659	0.019	1.619	1.693
0.47	0.824	0.016	0.785	0.859	0.87	1.701	0.021	1.657	1.741
0.48	0.838	0.016	0.799	0.872	0.88	1.746	0.023	1.697	1.792
0.49	0.852	0.016	0.814	0.885	0.89	1.794	0.026	1.738	1.847
0.5	0.866	0.016	0.829	0.899	0.9	1.846	0.029	1.781	1.906
0.51	0.881	0.016	0.844	0.914	0.91	1.904	0.033	1.832	1.971
0.52	0.896	0.016	0.859	0.928	0.92	1.968	0.037	1.888	2.044
0.53	0.911	0.015	0.874	0.942	0.93	2.039	0.041	1.948	2.125
0.54	0.926	0.015	0.890	0.957	0.94	2.121	0.046	2.016	2.221
0.55	0.941	0.015	0.906	0.972	0.95	2.217	0.053	2.099	2.333
0.56	0.957	0.015	0.922	0.986	0.96	2.334	0.061	2.200	2.470
0.57	0.973	0.015	0.938	1.002	0.97	2.482	0.071	2.328	2.646
0.58	0.989	0.014	0.955	1.017	0.98	2.687	0.087	2.505	2.893
0.59	1.005	0.014	0.972	1.032	0.99	3.032	0.114	2.795	3.312

Table 20. Regional relative RMSE and error bounds at 90% confidence level Region 3

F	q(F)	Relative RMSE	rel.bound . 0.05	rel.bound . 0.95	F	q(F)	Relative RMSE	rel.bound . 0.05	rel.bound . 0.95
0.2	0.274	0.014	0.247	0.314	0.6	0.904	0.025	0.857	0.956
0.21	0.285	0.015	0.257	0.328	0.61	0.928	0.024	0.882	0.978
0.22	0.296	0.016	0.266	0.342	0.62	0.952	0.024	0.907	1.002
0.23	0.307	0.016	0.276	0.354	0.63	0.977	0.023	0.933	1.026
0.24	0.318	0.017	0.286	0.367	0.64	1.003	0.023	0.960	1.050
0.25	0.330	0.018	0.296	0.380	0.65	1.030	0.022	0.987	1.075
0.26	0.342	0.018	0.307	0.393	0.66	1.057	0.022	1.016	1.101
0.27	0.354	0.019	0.317	0.406	0.67	1.085	0.021	1.045	1.128
0.28	0.366	0.020	0.328	0.420	0.68	1.114	0.020	1.076	1.155
0.29	0.378	0.020	0.339	0.433	0.69	1.144	0.019	1.107	1.184
0.3	0.391	0.021	0.351	0.446	0.7	1.175	0.018	1.140	1.213
0.31	0.404	0.021	0.363	0.460	0.71	1.208	0.018	1.174	1.243
0.32	0.417	0.022	0.375	0.474	0.72	1.241	0.017	1.210	1.275
0.33	0.430	0.022	0.387	0.488	0.73	1.275	0.015	1.246	1.307
0.34	0.443	0.023	0.400	0.502	0.74	1.311	0.014	1.284	1.342
0.35	0.457	0.023	0.413	0.516	0.75	1.349	0.013	1.324	1.377
0.36	0.471	0.024	0.426	0.532	0.76	1.387	0.012	1.365	1.414
0.37	0.485	0.024	0.440	0.547	0.77	1.428	0.011	1.407	1.451
0.38	0.500	0.024	0.454	0.563	0.78	1.470	0.010	1.450	1.491
0.39	0.515	0.025	0.468	0.577	0.79	1.515	0.009	1.496	1.534
0.4	0.530	0.025	0.483	0.592	0.8	1.562	0.008	1.542	1.579
0.41	0.545	0.025	0.497	0.607	0.81	1.611	0.008	1.590	1.628
0.42	0.560	0.026	0.512	0.622	0.82	1.662	0.009	1.641	1.680
0.43	0.576	0.026	0.528	0.637	0.83	1.717	0.011	1.691	1.739
0.44	0.592	0.026	0.544	0.653	0.84	1.776	0.013	1.743	1.803
0.45	0.609	0.026	0.560	0.669	0.85	1.838	0.016	1.798	1.870
0.46	0.626	0.026	0.576	0.686	0.86	1.904	0.019	1.856	1.943
0.47	0.643	0.026	0.593	0.703	0.87	1.975	0.023	1.919	2.023
0.48	0.661	0.026	0.611	0.720	0.88	2.053	0.027	1.987	2.110
0.49	0.678	0.026	0.628	0.737	0.89	2.137	0.031	2.061	2.204
0.5	0.697	0.027	0.646	0.755	0.9	2.229	0.037	2.142	2.308
0.51	0.715	0.026	0.665	0.773	0.91	2.331	0.043	2.231	2.423
0.52	0.734	0.026	0.684	0.791	0.92	2.445	0.050	2.330	2.553
0.53	0.754	0.026	0.704	0.810	0.93	2.574	0.058	2.443	2.701
0.54	0.774	0.026	0.724	0.829	0.94	2.724	0.068	2.573	2.872
0.55	0.794	0.026	0.745	0.849	0.95	2.902	0.080	2.722	3.075
0.56	0.815	0.026	0.766	0.869	0.96	3.119	0.095	2.906	3.326
0.57	0.837	0.026	0.788	0.890	0.97	3.400	0.116	3.148	3.651
0.58	0.859	0.025	0.811	0.911	0.98	3.796	0.145	3.491	4.112
0.59	0.881	0.025	0.834	0.933	0.99	4.476	0.198	4.075	4.903

Table 21. Regional relative RMSE and error bounds at 90% confidence level
Region 4

F	q(F)	Relative RMSE	Rel.Bound. 0.05	Rel.Bound. 0.95	F	q(F)	Relative RMSE	Rel.Bound. 0.05	Rel.Bound. 0.95
0.2	0.415	0.049	0.361	0.621	0.6	0.828	0.059	0.709	0.900
0.21	0.424	0.049	0.368	0.628	0.61	0.844	0.060	0.725	0.918
0.22	0.433	0.049	0.375	0.634	0.62	0.861	0.061	0.741	0.936
0.23	0.442	0.048	0.381	0.639	0.63	0.879	0.063	0.754	0.953
0.24	0.450	0.048	0.387	0.644	0.64	0.897	0.064	0.768	0.970
0.25	0.458	0.048	0.393	0.647	0.65	0.915	0.065	0.785	0.989
0.26	0.467	0.047	0.399	0.649	0.66	0.935	0.067	0.803	1.012
0.27	0.475	0.047	0.406	0.651	0.67	0.955	0.068	0.821	1.036
0.28	0.483	0.046	0.412	0.653	0.68	0.976	0.070	0.841	1.059
0.29	0.491	0.046	0.417	0.658	0.69	0.998	0.071	0.863	1.085
0.3	0.499	0.046	0.425	0.662	0.7	1.021	0.072	0.885	1.110
0.31	0.507	0.045	0.432	0.665	0.71	1.045	0.074	0.908	1.131
0.32	0.515	0.045	0.439	0.667	0.72	1.070	0.075	0.932	1.157
0.33	0.523	0.045	0.446	0.670	0.73	1.096	0.076	0.956	1.185
0.34	0.532	0.045	0.455	0.673	0.74	1.124	0.078	0.979	1.215
0.35	0.540	0.045	0.464	0.676	0.75	1.153	0.079	1.002	1.247
0.36	0.549	0.045	0.470	0.679	0.76	1.184	0.080	1.028	1.280
0.37	0.558	0.045	0.477	0.682	0.77	1.216	0.081	1.059	1.318
0.38	0.567	0.045	0.484	0.686	0.78	1.251	0.082	1.095	1.362
0.39	0.576	0.045	0.490	0.689	0.79	1.287	0.083	1.129	1.406
0.4	0.585	0.045	0.497	0.694	0.8	1.326	0.084	1.165	1.448
0.41	0.595	0.045	0.503	0.699	0.81	1.368	0.085	1.203	1.492
0.42	0.604	0.045	0.510	0.705	0.82	1.412	0.085	1.245	1.543
0.43	0.614	0.046	0.518	0.711	0.83	1.460	0.086	1.287	1.600
0.44	0.624	0.046	0.528	0.717	0.84	1.512	0.087	1.336	1.660
0.45	0.635	0.047	0.538	0.723	0.85	1.569	0.087	1.390	1.729
0.46	0.645	0.048	0.548	0.733	0.86	1.630	0.087	1.448	1.803
0.47	0.656	0.048	0.558	0.738	0.87	1.698	0.087	1.505	1.877
0.48	0.667	0.049	0.567	0.744	0.88	1.773	0.086	1.583	1.956
0.49	0.679	0.050	0.577	0.756	0.89	1.857	0.086	1.661	2.047
0.5	0.691	0.051	0.588	0.762	0.9	1.952	0.086	1.753	2.148
0.51	0.703	0.051	0.598	0.768	0.91	2.060	0.087	1.866	2.258
0.52	0.715	0.052	0.608	0.777	0.92	2.186	0.090	1.987	2.402
0.53	0.728	0.053	0.619	0.791	0.93	2.334	0.096	2.096	2.562
0.54	0.741	0.054	0.630	0.805	0.94	2.514	0.110	2.273	2.808
0.55	0.754	0.054	0.641	0.820	0.95	2.738	0.136	2.497	3.091
0.56	0.768	0.055	0.653	0.835	0.96	3.032	0.179	2.768	3.489
0.57	0.782	0.056	0.665	0.850	0.97	3.445	0.254	3.142	4.154
0.58	0.797	0.057	0.679	0.866	0.98	4.096	0.398	3.756	5.270
0.59	0.812	0.058	0.693	0.883	0.99	5.433	0.790	4.649	7.912

C.3 At-Site quantiles

Table 22. Site quantiles estimated from regional quantiles

Region	Site	Quantile (cfs)				
		5	10	25	50	100
		0.8	0.9	0.96	0.98	0.99
Region 1	50028000	7,253.27	10,647.97	16,769.53	23,153.72	31,635.21
	50028400	5,843.34	8,578.16	13,509.78	18,652.97	25,485.79
	50034000	7,008.27	10,288.32	16,203.10	22,371.65	30,566.66
	50050900	10,491.68	15,402.04	24,256.73	33,491.30	45,759.58
	50090500	4,843.82	7,110.85	11,198.90	15,462.34	21,126.39
	50092000	8,097.69	11,887.61	18,721.84	25,849.27	35,318.17
	50112500	2,760.52	4,052.51	6,382.31	8,812.06	12,040.03
	50115000	4,287.57	6,294.25	9,912.84	13,686.68	18,700.27
	50124200	6,182.99	9,076.78	14,295.06	19,737.20	26,967.18
Region 2	50051310	9,340.18	11,900.09	15,335.58	18,037.68	20,856.49
	50056400	16,402.81	20,898.42	26,931.66	31,676.98	36,627.24
	50061800	9,106.80	11,602.75	14,952.39	17,586.98	20,335.35
	50063800	11,546.44	14,711.03	18,958.01	22,298.38	25,783.03
	50064200	10,136.00	12,914.03	16,642.23	19,574.57	22,633.55
	50065500	14,398.55	18,344.84	23,640.87	27,806.36	32,151.75
	50067000	6,251.77	7,965.22	10,264.73	12,073.36	13,960.11
	50071000	12,478.52	15,898.57	20,488.39	24,098.41	27,864.35
Region 3	50031200	18,917.95	26,869.91	39,444.98	51,024.16	64,844.64
	50035000	39,083.47	55,511.81	81,491.23	105,413.19	133,965.56
	50038100	54,998.32	78,116.31	114,674.59	148,337.61	188,516.55
	50039500	13,258.11	18,831.03	27,643.91	35,758.86	45,444.55
	50043000	27,113.55	38,510.46	56,533.29	73,128.78	92,936.55
	50047850	14,489.12	20,579.48	30,210.64	39,079.05	49,664.06
	50055000	36,947.42	52,477.89	77,037.44	99,651.99	126,643.87
	50057000	32,084.33	45,570.65	66,897.63	86,535.60	109,974.76
Region 4	50138000	16,336.89	24,047.07	37,357.93	50,465.45	66,938.45
	50144000	31,958.99	47,042.02	73,081.36	98,722.91	130,948.17
	50147800	32,819.76	48,309.03	75,049.70	101,381.87	134,475.07

Table 23. At-site quantiles estimated from the GEV distribution

Region	Site	Quantile (cfs)				
		5	10	25	50	100
		0.8	0.9	0.96	0.98	0.99
Region 1	50028000	7,166.83	8,152.25	10,039.50	12,425.41	16,236.97
	50028400	6,555.87	8,914.71	12,036.68	14,401.21	16,768.20
	50034000	7,164.08	11,490.99	19,287.60	27,291.89	37,724.10
	50050900	8,949.71	13,610.59	23,381.32	35,115.51	52,730.70
	50090500	4,731.70	7,149.31	11,690.70	16,612.62	23,372.31
	50092000	8,269.70	12,029.03	18,629.61	25,342.15	34,068.24
	50112500	2,473.27	3,795.96	6,475.48	9,592.38	14,143.57
	50115000	3,877.78	5,997.22	10,272.70	15,226.51	22,435.15
	50124200	6,088.15	9,480.04	15,929.88	23,000.82	32,808.78
Region 2	50051310	9,898.84	12,360.84	15,324.64	17,422.10	19,422.51
	50056400	16,658.97	21,930.42	29,367.99	35,507.42	42,181.16
	50061800	9,387.40	12,128.41	15,798.38	18,678.43	21,677.12
	50063800	11,659.00	14,635.37	18,454.11	21,329.58	24,220.14
	50064200	9,675.50	13,314.99	19,195.03	24,721.04	31,433.22
	50065500	14,365.67	16,745.65	19,234.04	20,767.32	22,068.36
	50067000	5,866.96	7,350.07	9,423.15	11,119.05	12,948.44
	50071000	11,800.88	15,053.87	19,733.88	23,670.17	28,018.06
Region 3	50031200	18,577.42	26,634.25	39,726.06	52,097.12	67,194.39
	50035000	37,112.55	54,809.46	85,347.08	115,897.18	155,053.79
	50038100	57,914.03	80,263.77	112,943.97	140,862.71	172,114.20
	50039500	12,871.06	18,922.96	29,134.21	39,134.60	51,719.73
	50043000	25,894.01	39,038.02	62,049.36	85,386.96	115,653.55
	50047850	13,160.68	19,739.36	31,668.21	44,173.83	60,862.63
	50055000	37,105.86	47,226.15	60,405.78	70,474.56	80,721.38
	50057000	33,830.84	46,559.21	64,944.42	80,467.32	97,669.58
Region 4	50138000	12,531.39	22,817.50	45,664.25	74,353.94	118,979.26
	50144000	28,408.28	45,017.60	77,888.30	114,841.33	166,987.58
	50147800	31,864.22	35,690.53	39,184.13	40,997.82	42,312.83

Table 24. Percent errors between at-site quantile and regional estimated site quantile

Region	Site	Quantile				
		5	**10**	**25**	**50**	**100**
		0.8	**0.9**	**0.96**	**0.98**	**0.99**
Region 1	50028000	1%	31%	67%	86%	95%
	50028400	-11%	-4%	12%	30%	52%
	50034000	-2%	-10%	-16%	-18%	-19%
	50050900	17%	13%	4%	-5%	-13%
	50090500	2%	-1%	-4%	-7%	-10%
	50092000	-2%	-1%	0%	2%	4%
	50112500	12%	7%	-1%	-8%	-15%
	50115000	11%	5%	-4%	-10%	-17%
	50124200	2%	-4%	-10%	-14%	-18%
Region 2	50051310	-4%	-2%	0%	1%	2%
	50056400	1%	-3%	-8%	-13%	-17%
	50061800	-1%	-3%	-5%	-8%	-10%
	50063800	1%	2%	3%	2%	2%
	50064200	7%	-1%	-13%	-23%	-31%
	50065500	2%	12%	23%	31%	39%
	50067000	9%	10%	9%	6%	3%
	50071000	8%	8%	4%	0%	-5%
Region 3	50031200	9%	9%	2%	-5%	-13%
	50035000	13%	9%	-2%	-12%	-22%
	50038100	2%	5%	4%	2%	-2%
	50039500	11%	7%	-2%	-12%	-21%
	50043000	12%	6%	-6%	-17%	-28%
	50047850	18%	13%	-2%	-14%	-27%
	50055000	7%	20%	31%	37%	41%
	50057000	2%	6%	6%	4%	1%
Region 4	50138000	30%	5%	-18%	-32%	-44%
	50144000	12%	4%	-6%	-14%	-22%
	50147800	3%	35%	92%	147%	218%