12-1989

Regionalization of Flood Data Using Probability Distributions and Their Parameters

Digital Object Identifier: https://doi.org/10.13023/kwrri.rr.173

Nageshwar Rao Bhaskar
University of Kentucky

Carol Alf O'Connor
University of Kentucky

Harold Andrew Myers
University of Kentucky

William Paul Puckett
University of Kentucky

Research Report No. 173

REGIONALIZATION OF FLOOD DATA USING PROBABILITY DISTRIBUTIONS AND THEIR PARAMETERS

By

Nageshwar Rao Bhaskar *
Carol Alf O'Connor **
Principal Co-Investigators

Harold Andrew Myers
William Paul Puckett
Graduate Assistants

Project Number: G 1564-05 (A-115-KY)

Agreement Number: 14-08-0001-G1564

Period of Project: July 1988 - June 1989

Water Resources Research Institute
University of Kentucky
Lexington, Kentucky

The work upon which this report is based was supported in part by funds provided by the United States Department of the Interior, Washington, D.C., as authorized by the Water Resources Research Act of 1984. Public Law 98-242.

December 1989

DISCLAIMER

Contents of this report do not necessarily reflect the views and policies of the United States Department of the Interior, Washington, D.C., nor does mention of trade names or commercial products constitute their endorsement or recommendation for use by the U.S. Government.

ACKNOWLEDGEMENTS

This research study was conducted under a USGS research grant awarded through the Kentucky Water Resources Research Institute, University of Kentucky, Lexington, Kentucky. The authors wish to extend their appreciation to the former Director, Dr. David T. Kao and the members of the Kentucky Water Resources Research Institute for supporting this important study. The authors are deeply indebted to Drs. J. R. Wallis and J. R. M. Hosking of IBM for allowing the use of their flood regionalization computer program and letting us share the method of L-moments.

ABSTRACT

The U. S. Geological Survey recently used the method of residuals to delineate seven flood regions for the State of Kentucky. As an alternative approach, the FASTCLUS clustering procedure of the Statistical Analysis System (SAS) is used in this study to delineate five to six cluster regions in conjunction with statistical properties of the AMF series, like the coefficient of variation as estimated using method of L-moments, LCV, the parameters of the EV1 and GEV flood frequency distributions, and the specific mean annual flood, QSP. For both cluster and USGS flood regions, regionalized flood frequency growth curves are developed and their performance evaluated using Monte Carlo simulation techniques. Flood regions are then evaluated and compared using trends in the hydrological characteristics of important variables, performance of the regionalized flood frequency growth curves, discriminant analysis and regression equations relating flood quantiles to watershed physical characteristics. Results show that the cluster regions are more distinguishable in terms of their flood characteristics than the USGS regions. The regionalized flood frequency growth curves of the EV1 and GEV model are more distinct for the cluster regions than the USGS regions, although their performance in terms of bias and RMSE are comparable. The standard errors associated with the regression equations, developed for predicting the EV1 and GEV flood quantiles, are similar for cluster and USGS regions.

Descriptors: Flood * ; flood frequency * ; simulation * ; regionalization * ;

Identifiers: Cluster Analysis; discriminant analysis; method of residuals; USGS; Kentucky; flood regions.

TABLE OF CONTENTS

List of Tables

Title No. **Page No.**

List of Tables (continued...)

List of Tables (continued...)

List of Tables (continued...)

Title No. **Page No.**

List of Illustrations

Title No. Page No.

List of Illustrations (continued...)

Title No. Page No.

CHAPTER 1

INTRODUCTION

Nature, Scope and Objectives

The problem of estimating flood levels for selected frequencies (or return periods) is fundamental to flood control and mitigation studies. This is often accomplished by the use of flood frequency curves developed from systematic flood records at gauged sites in a watershed. However, due to the short or inadequate flow records at these gauges, the predictive ability of such frequency curves is limited. To overcome this problem, regionalized flood frequency curves are developed using pooled data from all gauges located in a hydrologically homogeneous flood region. The accuracy and reliability of these regionalized curves depends to a large extent on the procedures used to delineate flood regions that have similar flood response. A review of current literature indicates that a limited amount of work has been done in addressing this vital problem of flood regionalization. Furthermore, there exist three distinct methods of regionalization, as described below, that differ fundamentally in the type of approach used.

Method 1: Perform regionalization using specific flood characteristics of original or transformed annual flood data (referred to as response or dependent variables) as recorded at each of the gauged sites. Included in the analysis will be other hydrologic, physical and climatic characteristics of the watershed (referred to as attributes or independent variables affecting flood response) in which the gauged site is located. After identifying the homogeneous flood regions, a regional flood frequency curve is developed.

<u>Method 2</u>: An alternative approach to Method 1 above involves the direct use of the underlying probability distribution and its parameters at each of the gauged sites to accomplish flood regionalization. This is done by first performing a flood-frequency analysis using the annual flood series at each of the gauged sites using commonly accepted probability distributions. Select the most suitable distribution and its parameters describing flood response at each gauge. Perform regionalization using the probability distributions and their parameters. In this context, it must be emphasized that gauges within a homogeneous flood region having similar statistical parameters such as the mean, coefficient of variation, and skewness will not necessarily have similar underlying probability law of flood response.

<u>Method 3</u>: First perform a flood-frequency analysis using annual maximum flood data at each of the gauged sites. A regionalization is then carried out using flood quantiles at selected frequencies (example : the 100-year, 50-year etc. flood levels) as the response variables and other hydrologic, physical and climatic characteristics as the independent variables or attributes.

The three methods of regionalization described above differ in the manner in which they utilize flood data at a gauge. Furthermore, the problem remains as to how different the above methods are in defining homogeneous flood regions.

A secondary objective of the proposed study, therefore, will be to examine this issue in detail.

This study will utilize the systematic flood records available at all the gauges employed by the USGS in deriving the flood frequency curve and its parameters that is most appropriate for a particular gauged site. This information will then be used to classify these sites into homogeneous flood regions. The results from this study will provide a valuable comparison by bringing out the inherent differences in the three methods of flood regionalization described above. In addition, it will examine the most suitable flood probability distribution that can be adopted for the State of Kentucky to accurately describe the flood response of each watershed.

The United States Geological Survey in Louisville is the federal agency primarily responsible for developing regionalized flood information for the State of Kentucky. They have, recently, completed the process of flood regionalization using Method 3 described above. The proposed study will, therefore, examine the problem of regionalization of flood data in the State of Kentucky using Methods 1 and 2. Results from the study should provide a means for comparing these methods of regionalization with an ultimate goal of developing the most accurate and reliable procedure for regionalizing flood data.

With the above discussion in mind, the specific objectives of the proposed study are:

1) Perform flood regionalization using Methods 1 and 2 described above using flood data for the gauges in the State of Kentucky.
2) Identify the probability distribution and its parameters that best fits the annual flood series at each of the gauged sites.
3) Define homogeneous flood regions based upon the

statistical characteristics of the maximum annual flood data and the probability distributions and their parameters.

4) Compare the homogeneous flood regions as obtained using Methods 1, 2 and 3. Results from the USGS study will be used for Method 3.

Related Research

The index-flood method proposed by the U.S. Geological Survey (Darymple, 1960) is a classic example of early attempts to regionalize flood data. The technique involves the derivation of a regionalized frequency curve using the median values of the ratios of flood discharges at various frequencies to the mean annual flood as defined at each gauge. A major requirement for the application of this method is that the gauges used in the analysis must lie in a hydrologically homogeneous region. Thus, the index-flood method is a convenient way to regionalize flood frequency data provided the hydrologically homogeneous regions are defined a priori.

The use of multiple regression analysis is now a widely accepted method adopted by the U.S. Geological Survey for developing regionalized flood frequency prediction equations (McCabe, 1962; Sauer, 1964; Thomas and Benson, 1970; McCain and Jarrett, 1976; and Richter et al, 1984). The technique involves relating flood characteristics (as reflected by flood magnitudes at selected frequency levels) at a particular gauge to the physiographic, climatic and other variables that affect or control flood response of a watershed. Since this relationship is non-linear,a log transformation is utilized to linearize it. Flood characteristics are obtained at each gauge from log-Pearson Type-III flood frequency distribution (in conjunction with a regionalized coefficient of skewness), the latter derived using the procedures recommended by the Water Resources Council (U.S. Department of the Interior, 1982). In order to improve the accuracy of these equations, homogeneous regions are defined using the method of residuals. A residual is the difference between the observed and predicted flood value at a gauged site. This is estimated from the overall regression equation developed for the entire region under investigation. It is assumed that the general trends in these residuals reflect inherent variations in the flood response of various sub-regions. This is the primary basis on which the regionalization of flood data is accomplished. After a detailed analysis of the residuals, a regionalized flood prediction equation is redeveloped using data from all gauges within a homogeneous sub-region. This method of defining homogeneous sub-regions using residuals from an overall regression equation is subjective. This is obvious since the causative factors controlling flood response are not considered explicitly in the process of defining the homogeneous regions. Residuals often reflect statistical variations in the data sample and any trends may be purely incidental. In recognition of this, recent efforts of regionalizing flood data have focussed on the use of more sophisticated statistical methods. For example, DeCoursey and Deal (1974) used discriminant analysis to define homogeneous flood regions using flood and basin characteristics of the watersheds as defined at each streamflow gauge. The basic approach is to classify homogeneous regions using the concepts of cluster analysis. Clusters or groups are formed using flood and basin characteristics at each gauge with the basic premise of maximizing within group similarity while at the same time minimizing between group similarity. A complete linkage algorithm of forming clusters, as proposed by Sokal and Sneath (1963), is used. Discriminant analysis is used to determine any misclassification of points or stations into a cluster or group. Tasker (1982) extended this method of regionalizing flood data to gaging stations in Arizona. It must be pointed out that, in both cases, flood characteristics were obtained from flood frequency curves as defined at each gauge and hence the general approach is similar to Method 3 described in the previous section.

A completely different approach of regionalization of flood data than the one described above was adopted by Wiltshire (1986) in his efforts to define homogeneous flood regions in England (this is similar to Method 1 described in the previous section). Instead of using flood estimates obtained from a flood frequency curve, his approach incorporates specific properties (statistical) of the flood series as the response variable. An iterative search is then employed using the basin characteristics as the independent variables (or attributes) so as to minimize the variance of the response variable within a cluster or group while simultaneously maximizing the variance between groups. The multivariate technique used in the analysis is referred to as Analysis of Variance (ANOVA) for a single response variable and Multivariate Analysis of Variance (MANOVA) for more than one response variable. The main advantage of Wiltshire's approach is that flood data at a gauge are considered explicitly and the use of a fitted flood frequency curve is avoided. However, as pointed out by Wiltshire (1986), there are two weaknesses to his procedure. The first of these is that the annual maximum flood series at each site is characterized by only one response variable, namely, the coefficient of variation. For example, no consideration is given for other flood characteristics like the coefficient of skewness. The second problem is that the resulting solution in terms of basin groupings may not be unique, i.e. different basin characteristics may also produce a statistically significant result. The latter problem could be resolved to a certain extent using physical reasoning and geographic regions.

The use of probability distribution and its parameters for regionalizing flood data has been attempted by several investigators. Such approaches are similar to Method 2 described in the previous section. Houghton (1977) used the Wakeby distribution and its parameters for regionalizing flood experience in the United States and proposed four such distributions for use in flood prediction. Kuczera (1982) examined the relative performance of the Wakeby distribution in estimating extreme flood events in comparison to other more parsimonious probability distributions. The performance was measured using a mean square criterion. In a parallel study Kuczera (1982) shows how empirical Bayes procedures can be used to combine site-specific and regional information to improve upon site-specific estimators. Rossi et al (1984) regionalized annual flood series using the at-site estimates of a two parameter extreme value probability distribution. Synthetic flood data, generated using Monte Carlo techniques, was used to test the relative performance of several regionalization methods by Lettenmaier and Potter (1985). Their results show that for annual flood series having a high coefficient of variation, improvements in regional flood estimation will come from improved estimators of the at-site mean annual flood, rather than the regional (normalized) flood frequency distribution. An overview of recent efforts in flood regionalization is given by Greis (1983).

Although considerable work, as discussed above, has been advanced in developing robust flood frequency probability distributions, little work has been done in addressing the fundamental question of the selection of homogeneous flood regions. This is an extremely vital step in any effort to regionalize flood data based upon such information from specific gauged sites.

CHAPTER 2

RESEARCH PROCEDURES

The accuracy and precision with which flood levels (particularly those associated with large return periods such as the 100-year flood level), can be estimated at gauged and ungauged streamflow sites is primarily influenced by (Cunane, 1987):

1) The form of the underlying flood frequency distribution or model that best describes the underlying law of flood response and the method of estimating its parameters.
2) Amount and type of data used: a) at-site data; b) at-site/regional and c) regional without at site data.
3) Type of flood frequency model: a) Annual maximum (AM) flood series and b) Peaks over threshold (POT) flood series (partial flood series).

The above factors are incorporated into the study procedure as discussed below.

CHOICE OF A FLOOD FREQUENCY MODEL

The choice of a suitable parent probability distribution and the method to estimate its parameters constitutes, by far, the most difficult step in the development of a flood frequency model to best describe the flood response of a watershed. The success of any flood regionalization to estimate flood quantiles accurately is heavily dependent on this choice. The major problem arises from the fact that the true population flood frequency distribution that best fits the AMF data at a site is and will, at least in the near future, be never known. However, numerous efforts by researchers over the past few decades has led to a general consensus that the annual maximum floods come from populations with positively skewed distributions and that these distributions are relatively thick-tailed. Hence, the focus on contending probability distributions has been primarily on a family of skewed distributions. Furthermore, as suggested by Kuczera (1982), a good flood frequency model must possess the following properties: a) it must have the ability to estimate flood quantiles with least bias and, hence, is efficient (measure of accuracy); b) The model must also be resistant by having the capacity to estimate extreme events, irrespective of which contending distribution best represents the real world, without a disastrous loss of performance as indicated by a suitable measure such as low root mean square error (measure of precision); and c) the flood frequency model must perform well even if a misspecification of the underlying parent probability distribution occurs (a property known as robustness). These are the primary criteria that are given due consideration in the present study for testing the performance and suitability of flood frequency models selected for describing flood experience in Kentucky.

a) Flood Frequency or Probability Distributions: Numerous probability distributions have been used to fit AMF data. The following is a list of general forms of probability distributions (refer to Table 2.1) that have been used by various investigators either directly or in a simplified form (example: 2-parameter distributions):

1. Generalized Extreme Value (GEV) and its special case Extreme Value Type-I (EV1 or Gumbel)
2. Generalized Normal and log-Normal

3. Pearson and log-Pearson Type-III
5. Wakeby
6. Generalized Pareto
7. Generalized Lambda
8. Generalized Logistic
7. Kappa

Each of the above distributions require at least three parameters to be estimated which characterize the location, scale and shape of the underlying probability distribution, respectively. They have all been tested by numerous investigators using various procedures for estimating their parameters. Recent studies (Wallis and Wood, 1985, Kuczera, 1982, Lettenmaier et al, 1987, Landwehr et al, 1980) have favored the Generalized Extreme Value, GEV, together with its special case, namely, the Extreme Value Type-I, EV1, (referred to as Gumbel) and the Wakeby, WAK, distributions for modeling AMF data. Furthermore, a relatively new approach called L-moments has been recommended for estimating the parameters of these distributions (Hosking, 1989) over the conventional methods used in the past such as the method of moments and the maximum likelihood method. The method of L-moments is closely linked to the probability weighted moments (PWM) method of estimating parameters as first introduced by Greenwood et al, 1979 and later used by numerous investigators (Landwehr et al, 1979, Landwehr et al, 1980, Hosking et al, 1985, Hosking and Wallis, 1987, Wallis and Wood, 1985, Kuczera, 1982). A brief discussion of the L-moments method, in conjunction with the theory of probability weighted moments (PWM), is given below. This method is chosen as the preferred method for estimating the parameters of the Gumbel(EV1), Generalized Extreme Value (GEV) and the Wakeby (WAK) flood frequency probability distributions in the present study.

TABLE 2.1. Common Probability Distribution Used in Flood Frequency Analysis (Hosking, 1988)

Distribution	Code	Number of parameters	Parameters	$F(x)$, $x(F)$
Generalized extreme-value	GEV	3	ξ α k	$F = \exp[\{1 - k(x - \xi)/\alpha\}^{1/k}]$ $x = \xi + \alpha\{1 - (-\log F)^k\}/k$
Generalized logistic	GLO	3	ξ α k	$F = 1/[1 + \{1 - k(x - \xi)/\alpha\}^{1/k}]$ $x = \xi + \alpha[1 - \{(1 - F)/F\}^k]/k$
Generalized Normal	GNO	3	ξ α k	$F = \Phi[-k^{-1} \log\{1 - k(x - \xi)/\alpha\}]$ $x(F)$ not explicitly defined
Generalized Pareto	GPA	3	ξ α k	$F = 1 - \{1 - k(x - \xi)/\alpha\}^{1/k}$ $x = \xi + \alpha\{1 - (1 - F)^k\}/k$
Gumbel	GUM	2	ξ α	$F = \exp[-\exp\{-(x - \xi)/\alpha\}]$ $x = \xi - \alpha \log(-\log F)$
Kappa	KAP	4	ξ α k h	$F = [1 - h\{1 - k(x - \xi)/\alpha\}^{1/k}]^{1/h}$ $x = \xi + \alpha[1 - \{(1 - F^h)/h\}^k]/k$
Wakeby	WAK	5	ξ α β γ δ	$F(x)$ not explicitly defined $x = \xi + \alpha\{1 - (1 - F)^\beta\}/\beta - \gamma\{1 - (1 - F)^{-\delta}\}/\delta$

b) Method of Estimating Parameters: Probability Weighted Moments and L-Moments: A probability distribution, having a distribution function $F = F(x) = P(X < x)$ of a random variable X, may be characterized by probability weighted moments defined as (Greenwood et al, 1979):

$$M_{p,r,s} = E[X^p\{F(X)\}^r \{1 - F(X)\}^s]$$

$$= \int x^p\{F(x)\}^r \{1 - F(x)\}^s dF(x), \quad (2.1)$$

$$M_{p,r,s} = \int_0^1 \{x(F)\}^p F^r (1 - F)^s dF$$

where p, r and s are real numbers. If r=s=o and p is a non-negative integer then $M_{p,o,o}$ represents the conventional moment of order p as used in the method of moments. If p, r and s are positive integers then the probability weighted moment, $M_{p,r,s}$ can be related to the expected value of the k-order statistic, $X_{k:n}$, of a random sample of size n drawn from the distribution F by the following relationship:

$$M_{p,r,s} = \frac{r!s!}{(r + s + 1)!} EX^p_{r+1:r+s+1} \quad (2.2)$$

In particular, the probability weighted moments, $M_{1,0,r}$, and $M_{1,r,o}$, which are linear functions of the expected value of the k-order statistic, $X_{k:n}$, are sufficient to characterize a distribution and can be defined as follows:

$$\alpha_r = M_{1,0,r} = E[X\{1 - F(X)\}^r],\ r = 0,1,\ldots ,$$
$$= \int_0^1 x(F)\ [1 - F(x)]^k\ dF \tag{2.3}$$
$$= \frac{1}{n} \sum_{i=1}^{n} x_i\ (1 - F_{i,k})^k$$

$$\beta_r = M_{1,r,0} = E[X\{F(X)\}^r],\ r = 0,1,\ldots,$$
$$= \int_0^1 x(F)\ [F(x)]^k \tag{2.4}$$
$$= \frac{1}{n} \sum_{i=1}^{n} x_i (F_{i,k})^k$$

Furthermore, α_r and β_r are related by the following equations:

$$\alpha_r = \sum_{k=0}^{r} (-1)^k \binom{r}{k} \beta_k, \tag{2.5}$$

$$\beta_r = \sum_{k=0}^{r} (-1)^k \binom{r}{k} \alpha_k, \tag{2.6}$$

As stated by Hosking (1986) although the probability weighted moments (PWM's) (Equations 2.3-2.4) can be used to characterize the underlying probability distribution, they are not useful by themselves in defining specific characteristics of a distribution like the scale and shape. Instead, certain linear functions of the PWM's known as L-moments give a better description of the location, scale and shape of a probability distribution. As shown later PWM's and L-moments are closely related.

Consider a real-valued random variable, X, having a distribution function, F(x) and an inverse function, x(F), and let $X_{1:n} < X_{2:n} < \ldots . X_{n:n}$ be the ordered statistic of a random sample drawn from the population distribution of the random variable X. L-moments can then be defined as a linear combination the expected value of the above order statistic as (Hosking, 1986):

$$\lambda_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} EX_{r-k:r}, \quad r = 1,2,\ldots, \tag{2.7}$$

where, the expected value of an ordered statistic, $EX_{j:r}$, is defined as:

$$EX_{j:r} = \frac{r!}{(j-1)!(r-j)!} \int x\{F(x)\}^{j-1}\{1 - F(x)\}^{r-j} \, dF(x) \tag{2.8}$$

Substituting Eq. 2.8 in Eq. 2.7, expanding the binomials in F(x) and summing the coefficients of each power of F(x) gives the following final expression that can be used to calculate the L-moments.

$$\lambda_r = \int_0^1 x(F) \, P^*_{r-1}(F) \, dF, \quad r = 1,2,\ldots, \tag{2.9}$$

where,

$$P^*_r(F) = \sum_{k=0}^{r} p^*_{r,k} F^k \tag{2.10}$$

and

$$p^*_{r,k} = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k}. \qquad (2.11)$$

Note the similarity of Eq. 2.9 with the PWM as defined in Eq. 2.1. The L-moments are simply linear combination of the PWM's, $M_{1,0,r}$ (Eq. 2.3 and 2.5) and $M_{1,r,0}$ (Eq. 2.4 and 2.6) and, hence, are closely related by the following relationships (Hosking 1986):

$$\lambda_{r+1} = (-1)^r \sum_{k=0}^{r} p^*_{r,k}\alpha_k = \sum_{k=0}^{r} p^*_{r,k}\beta_k, \; r = 0, 1, \ldots \qquad (2.12)$$

c) Interpretation and estimation of L-moments: As pointed out by Hosking (1989), the L-moments λ_1, λ_2, λ_3.....λ_r, and L-moment ratios $\tau_3 = \lambda_3/\lambda_2$, $\tau_4 = \lambda_4/\lambda_2$, ...$\tau_r = \lambda_r/\lambda_2$, are useful quantities for summarizing a probability distribution. The L-moments are similar to conventional central moments while the L-moment ratios are similar to the conventional moment ratios. The first L-moment, λ_1, is equal to the mean and is, therefore, regarded as a measure of the central tendency or location, the second L-moment, λ_2, is a measure of scale or dispersion like the variance or standard deviation. The moment ratios, τ_3 and τ_4, which are dimensionless forms of the third and fourth L-moments (λ_3 and λ_4), are measures of skewness and kurtosis, respectively. Thus, these L-moments and ratios together are sufficient to estimate parameters that describe the location, scale, skewness and kurtosis of a flood frequency distribution . Higher order L-moments and ratios have similar interpretation as conventional method of moments for further describing the character of the underlying probability distribution.

The L-moments described above must be estimated from observed maximum annual flood data at a gauged site prior to any flood regionalization effort. A natural estimator of each L-moment (refer to Eq. 2.7 above) based on an observed sample of data is a linear combination of the ordered data values. Such an estimator is known as an L-statistic. In practice, therefore, the L-moments can be estimated from an ordered (lowest to highest value) random sample drawn from an unknown probability distribution. Hosking (1989) presents two such estimation procedures. The one used in this study is referred to as a plotting position estimator. A plotting position, $p_{i:n}$, is a distribution-free estimator of the probability of non-exceedance, $F(x_{i:n})$, of an ordered random variable $X_{i:n}$. Although this estimator is biased, Hosking (1989) has observed in his study that it gives good estimates of the parameters and quantiles when a distribution is fitted to the data. In particular, Hosking (1989) concludes that the plotting position estimator of the form $p_{i:n} = (i - 0.35)/n$, where i is order number (or rank) of observed data value, $x_{i:n}$, of random variable $X_{i:n}$, and n is the sample size, gave good results for generalized extreme value distribution. Thus, the following equations are used in the study to estimate the L-moments and PWM's, α_r and β_r (refer to equations 2.9, 2.3 and 2.4, respectively).

$$a_r[\gamma,\delta] = n^{-1} \sum_{i=1}^{n} (1 - p_{i:n})^r x_i,$$

$$b_r[\gamma,\delta] = n^{-1} \sum_{i=1}^{n} p_{i:n}^r x_i, \qquad (2.13)$$

$$l_r[\gamma,\delta] = n^{-1} \sum_{i+1}^{n} P^*_{r-1}(p_{i:n})\, x_i$$

PROCEDURE FOR FLOOD REGIONALIZATION

Statistical estimates of flood quantiles, based on at-site data only, are highly variable due to modeling and sampling error. Consequently, a process of flood regionalization, whereby flood data from several sites within a homogeneous flood region (defined a priori) are pooled together, is usually recommended. In the present study, an at-site/regional flood data (refer to 1(b) above) approach is adopted using the historical AMF series (systematic record) (refer to 2(a) above) from each of the gauged sites in Kentucky. This data are transformed to a dimensionless form by dividing each observed flood value by the mean annual flood at that site. An index-flood approach for flood regionalization similar to the one used by Hosking and Wallis (1987), and as described below, was used to pool flood data from gauged sites within a homogeneous flood region (the procedure used to delineate such regions is discussed later). An IBM supplied computer program (Hosking, 1988) is used, with some modifications, to accomplish the regionalization. This computer program allows the development of a regionalized flood frequency distribution for commonly used probability distributions. The method of L-moments is used to estimate regionalized parameters. A step-by-step procedure of the index-flood method used in this study is as follows:

1. Define flood regions that have similar underlying flood response. These regions can be delineated either using the statistical moments required to characterize the underlying parent probability distribution or the parameters of this distribution as estimated using the statistical moments. In either case, the basic premise is that regions having similar statistical moments or parameters of the probability distribution must be homogeneous with respect to their flood response.

Alternatively, such regions may be delineated on the basis of the physical characteristics of the watersheds that control flood response. In this study, both approaches are used to identify flood regions.

2. Within each region assume that the regional flood quantile estimate for a given return period, T, is given by q_T. This estimate is derived from the probability distribution of normalized flood data (dimensionless flood variate) and hence is scale independent. The normalized flood variate, X is obtained by dividing each flood observation, Q_i, at a site by an index flood, Q_I. The latter is usually taken as the mean annual flood at the site as is done in this study.
3. Estimate the at-site mean annual flood, $\overline{Q}$, at each site, i, within a region using the average of observed raw flood data as required in step 2 above and the following step.
4. Combine estimates, q_T, and $\overline{Q}$, to obtain the flood quantile estimate, Q_{Ti} at site i within the region.
5. The accuracy and precision of the flood quantile estimate, Q_{Ti}, at site i is then evaluated using Monte Carlo simulation techniques.

DEVELOPMENT OF REGIONALIZED FLOOD FREQUENCY GROWTH CURVES

A frequency growth curve is simply a plot of a cumulative probability density function and can, therefore, be used to compute flood levels at various probability levels of non-exceedance (flood quantiles). In this study these curves are plotted with the normalized flood levels (random variate) on the vertical axis and the probability of non-exceedance on the horizontal axis of a Generalized Extreme Value probability paper. A high value of coefficient of variation and skew prevalent in the AMF data would cause this growth curve to be steeper reflecting more variability in the data. Thus, a given normalized discharge level will be associated with a smaller return period (or probability of exceedance) than a flatter curve. Furthermore, these growth curves can be directly related to the flood response of the watershed (Acreman and Sinclair, 1986). For example, larger watersheds, responding to floods generated from various sub-watershed contributions, may exhibit greater variability in their flood response than smaller watersheds. Hence, flood data from larger watersheds would have a larger coefficient of variation resulting in a steep growth curve. The shape of the growth is, also, influenced by other watershed physical and climatic characteristics like watershed size, slope, landuse, soil and spatial and temporal effects of rainfall inputs. In any event, differences in the shape of these growth curves (regionalized) do reflect variations in flood response, and can, therefore, be used to assess the degree of heterogeneity of flood response between flood regions.

DELINEATION OF FLOOD REGIONS: CLUSTER ANALYSIS

The purpose of cluster analysis in the context of flood regionalization, is to place gauged sites into clusters or groups such that all the gauges within a cluster have similar flood response and those in different groups have dissimilar flood response. Therefore, the success of any clustering technique would greatly depend on the variables used to define similarity of flood response and some sort of measure to cluster gauged sites that are closer than others with respect to these variables. Since the flood response of any watershed is dependent on the underlying probability law of flood response, it is appropriate to use the statistical moments that characterize this distribution and/or the the parameters of the probability distribution (as estimated from the moments) as the variables to measure similarity (referred to as response variables in this study) of gauged sites within a cluster. In order to accomplish this, a criterion to group gauged sites having similar statistical moments or parameters (or response variables) is required. A commonly used method is based on the concept of Euclidean distance. In particular the Mahalanhois distance, as defined in the following equation, has the added advantage when compared to an ordinary Euclidean measure since it explicitly accounts for any correlations that might exist between the variables used in clustering.

$$D^2 = (X_i - X_j)' S^{-1} (X_i - X_j) \tag{2.14}$$

where,

D = Euclidean distance,

X_i and X_j = Vector of the response variables used at a gauged sites i and j, respectively, for measuring similarity of flood response, and

S = pooled within-group covariance matrix.

In this study a clustering technique based on the Euclidean distance measure described above is used to group (or to bring together) gauged sites into homogeneous flood regions or clusters. Several clustering algorithms, such as the average linkage, nearest centroid sorting (referred to as FASTCLUS), complete linkage or Ward's minimum variance can be used to perform the clustering based upon the Eulidean distance given by Equation 2.14 (SAS, 1985). The choice amongst these will depend on the data being analyzed although the FASTCLUS disjoint clustering algorithm has an intuitive appeal over the other methods since its procedure allows for the movement of observations at every step of the clustering process.

a) Choice of the Clustering Algorithm: As mentioned above there are several algorithms that are commonly used in performing cluster analysis. The principal difference between each of these algorithms stems from the manner in which they compute the Euclidean distance measure and the manner in which the clustering is performed. Consequently, the nature of the clusters formed will depend heavily on the variables and their corresponding values. A brief discription of characteristics and biases of the more frequently used clustering algorithms that makes each different or distinct from others is presented in SAS, 1985. These inherent differences are used in this study to make the final choice of the algorithm.

The FASTCLUS clustering technique, as available in the Statistical Analysis System computer software, SAS (SAS Institute 1985), is used to group (or to bring together) gauged sites into distinct flood regions or clusters. This procedure performs disjoint clustering on the basis of Euclidean distances computed from the clustering variables used. The FASTCLUS procedure differs from hierarchical clustering procedures, such as Ward's, by using cluster seeds. Initial cluster seeds are observations which are separated by at least a specified minimum distance. FASTCLUS is an iterative procedure in which cluster seeds are recomputed for each iteration. In each iteration, all observations are assigned to the nearest seed, forming the specified number of clusters, and the seeds are recomputed as the means of the clusters. Observations are then considered as seed replacements using two tests based upon maximizing the distance between seeds. This iteration process continues until a convergence criterion, based upon the maximum distance any seed is changed, is met. Then the final clusters are formed by assigning each observation to the nearest seed. The FASTCLUS procedure is sensitive to outliers.

The FASTCLUS procedure described above is similar to the procedure used by Wiltshire (1986) in his efforts to regionalize flood data in England. In favor of this form of clustering, Wiltshire (1986) points out that "partitioning imposes a certain degree of structure on the data and avoids the undesirable tendency of hierarchical schemes to produce one large dominant cluster located at the centroid of the data with small satellite clusters toward the margins of the data space". A similar situation was observed by the authors when using hierarchical clustering algorithms such as Ward's. This was the primary reason why the FASTCLUS procedure is selected over the other methods in this study. Based on the flood response variables, namely statistical moments required to characterize the underlying probability distribution, the parameters of the probability distribution and the specific mean annual flood, QSP (clustering variables), disjoint clusters or flood regions are successfully delineated.

One of the most difficult problems in cluster analysis is the identification of the optimal number of clusters in a data set that can be clearly distinguished from each other. A review of current literature suggests that several procedures, referred to as stopping rules, available for addressing this vital issue. Such rules are often applied in a subjective manner. To use these rules in the classical "test of hypothesis" setting requires the specification of a null and alternate hypothesis, such as that the data are a random sample from a multivariate normal population. However, it has been shown that there can be large errors associated with these tests if the hypotheses are not stated correctly. Futhermore, there is the additional problem of determining the sampling distribution of the criterion used in the hypothesis testing. Ordinary tests like ANOVA F and t-test are not valid for testing difference between clusters, since clustering methods tend to maximize the separation between clusters and hence violate the basic assumptions of such tests. In view of this, formal tests of hypothesis are not used in this study. Instead, a number of stopping rules are incorporated in a subjective manner while selecting the optimum number of clusters. In doing so, the principal objective of identifying homogeneous cluster or flood regions that can be discriminated easily based upon the attribute variables is given primary emphasis. Milligan and Cooper (1983) used Monte Carlo simulations to evaluate the performance of 30 stopping rules commonly used in cluster analysis. Amongst these, several rules which gave good performance are selected and discussed below. Furthermore, since only 253 gauged sites are being used in the flood regionalization study, it seemed impractical and physically unrealistic to examine more than ten clusters. Cosequently, the following stopping rules, as presented by Milligan and Cooper (1983), are applied to 10 or fewer cluster regions.

1) The goodness of fit criterion, R^2, has the usual interpretation of the proportion of variance accounted for by the clusters. Ward's algorithm attempts to maximize this when deciding on the clusters to merge at each stage of clustering. As clusters are merged R^2 will decrease and, hence, a rule of thumb is to stop clustering whenever there is a significant drop in the value of this criterion.
2) The ratio criterion is defined as the ratio of within cluster sum of squared errors when the data are split into two clusters to the squared errors when only one cluster is used. In general, small ratio of this criterion leads to the regection of the hypothesis of one cluster. This criterion, as first proposed by Duda and Hart (1973), gave the best performance amongst all the other rules examined by Milligan and Cooper (1984). The ratio criterion can be applied at each stage of the clustering to the subpopulations involved. Thus, at

any stage, if the ratio is small, the two clusters being merged should remain separate. In contrast, a larger value of this ratio would support the collapsing of the two clusters into one. The Duda and Hart ratio criterion can be related to the pseudo-t^2 statistic available in the SAS package by a reciprocal relationship (SAS, 1985). The pseudo-t^2 statistic is a measure of the separation between clusters most recently merged. Thus, a rule of thumb while selecting the optimum number of clusters is to look for small values of this statistic.

3) Another stopping rule that performed in the top one-third of the stopping rules studied by Milligan and Cooper (1983), was the pseudo-F statistic. While similar to the F-statistic in ANOVA, the assumptions associated with analysis of variance are not met in the clustering setting and, hence the name "pseudo-F". This statistic provides the measure of separation among all clusters at any step in the clustering process. Ideally, as the number of clusters decreases, the pseudo-F statistic will decrease, then rise at the point where the optimum number of clusters occur, and then fall again (this is referred to as an "elbow" effect). If such is not the case, the pseudo-F will continue to decrease as the clusters are collapsed. In this case, one could look for the largest gap of this statistic in selecting the optimum number of clusters.

4) The cubic clustering criterion (CCC criterion) developed by Searle (SAS, 1985) performed as well as the pseudo-F statistic in the simulation runs by Milligan and Cooper(1983). This criterion is a function of the observed R^2 (refer to stopping rule 1) and the expected R^2 assuming that the clusters, as obtained from a uniform distribution on a hyperbox, are hypercubes of the same size. Guidelines for using CCC criterion include the plotting of CCC statistic versus

the number of clusters with the peaks indicating the possible cutoff point for extracting the optimum number of clusters. Peaks associated with a CCC value greater than or equal to 2 indicate a good number of clusters.

b) Selection of Suitable Flood Characteristics for Clustering: As stated in the previous section, the success of using cluster analysis to delineate homogeneous flood regions depends to a large extent on the variables used to define the flood characteristics at each of the gauged sites (response variables). Since any data set of observations can form clusters, it is imperative to choose variables that reflect the flood experience as accurately as possible in order to ensure flood homogeneity within a cluster. The flood response of a watershed, as measured using the AMF series at a gauge, is stochastic and is, therefore, governed by an underlying probability law (distribution unknown a priori). The latter can be evaluated by fitting the AMF series to an assumed probability distribution using statistical moments of various orders such as the fist order moment. Consequently, it can be postulated that any two gauged sites will have similar flood response if their underlying probability distribution is the same. This would also imply that the statistical moments used to fit the probability distribution (involving the evaluation of its parameters) and/or its parameters must be identical except for the effects of scale. Based upon this premise, the following clustering variables (flood response variables) are initially used to perform cluster analysis using FASTCLUS clustering algorithm. All the clustering variables are standardized prior to clustering in order to suppress any disproportionate effects during clustering.

1) L-moment ratios (dimensionless ratios of L-moments) of normalized maximum annual peak flow data from each gauged sites, namely, coefficient of variation, LCV,

coefficient of skewness, LSK and coefficient of kurtosis, LKUR. All these variables characterize the form of the underlying probability distribution. For two-parameter distributions the first L-moment, LCV is adequate while for probability distributions with more than two parameters higher order L-moment ratios will be required.

2) The specific mean annual flood, QSP, defined as the ratio of the mean annual flood at each site (as estimated using raw flood data) to the watershed size in square miles.
3) The parameters (as estimated using L-moments) of the selected flood probability distributions. The number of parameters used will depend on the distribution selected. Generally, two to three parameters reflecting the location, scale and shape are required for most probability distributions.

The final choice of suitable response variables for obtaining the clusters is based upon the results of cluster analysis, specifically, the ability to extract optimum number of clusters using a cutoff criterion, detailed examination of trends in important hydrological characteristics within and between regions, flood frequency growth curves, discriminant analysis using the attribute variables at each gauged site, and regression analysis relating selected flood levels to watershed physical characteristics. These results are presented in the next chapter.

DELINEATION OF FLOOD REGIONS: USGS METHOD OF RESIDUALS

The U. S. Geological Survey currently employs the method of residuals to perform flood regionalization. The technique involves the use of residuals from a regression equation relating a selected flow quantile (for example the 50-year flood level as obtained from an assumed probability distribution of the AMF series at each gauged site) to the physical and climatic characteristics of the watershed. The probability distribution employed is log-Pearson Type-III. This technique relies on the basic premise that the trends in the residuals reflect regional differences in the flood response of the watersheds. Thus, once a homogeneous region is delineated then the regression equation relating the flood response variable to the watershed characteristics will have residuals that can be attributed to pure chance. Unfortunately, the residuals contain both chance variation (time sampling error) and variation due to basin characteristics (model error) without a measure of the relative amounts of each (Riggs, 1973). This makes the delineation of homogeneous flood regions a difficult, if not an arbitrary, task to accomplish. Nonetheless, this procedure was used to delineate seven homogeneous flood regions for the State of Kentucky (refer to Figure 2-1) using flood data from all gauged sites used in the present study (i.e. both the data sets set aside for gauged and ungauged analysis). A regionalized skewness coefficient was used for estimating the 50-year flood quantile of the log-Pearson Type-III frequency curve fitted to annual peak flow data at each of the gauged sites.

VERIFICATION OF FLOOD REGIONS

a) Hydrological Characteristics of Flood Regions: For each of the clustering scheme and method of residuals the variation in important hydrologic characteristics (response and attribute variables) within and between regions are compared. Tables showing important statistics such as range, minimum and maximum, mean, and median are used for this purpose. These statistical characteristics of the

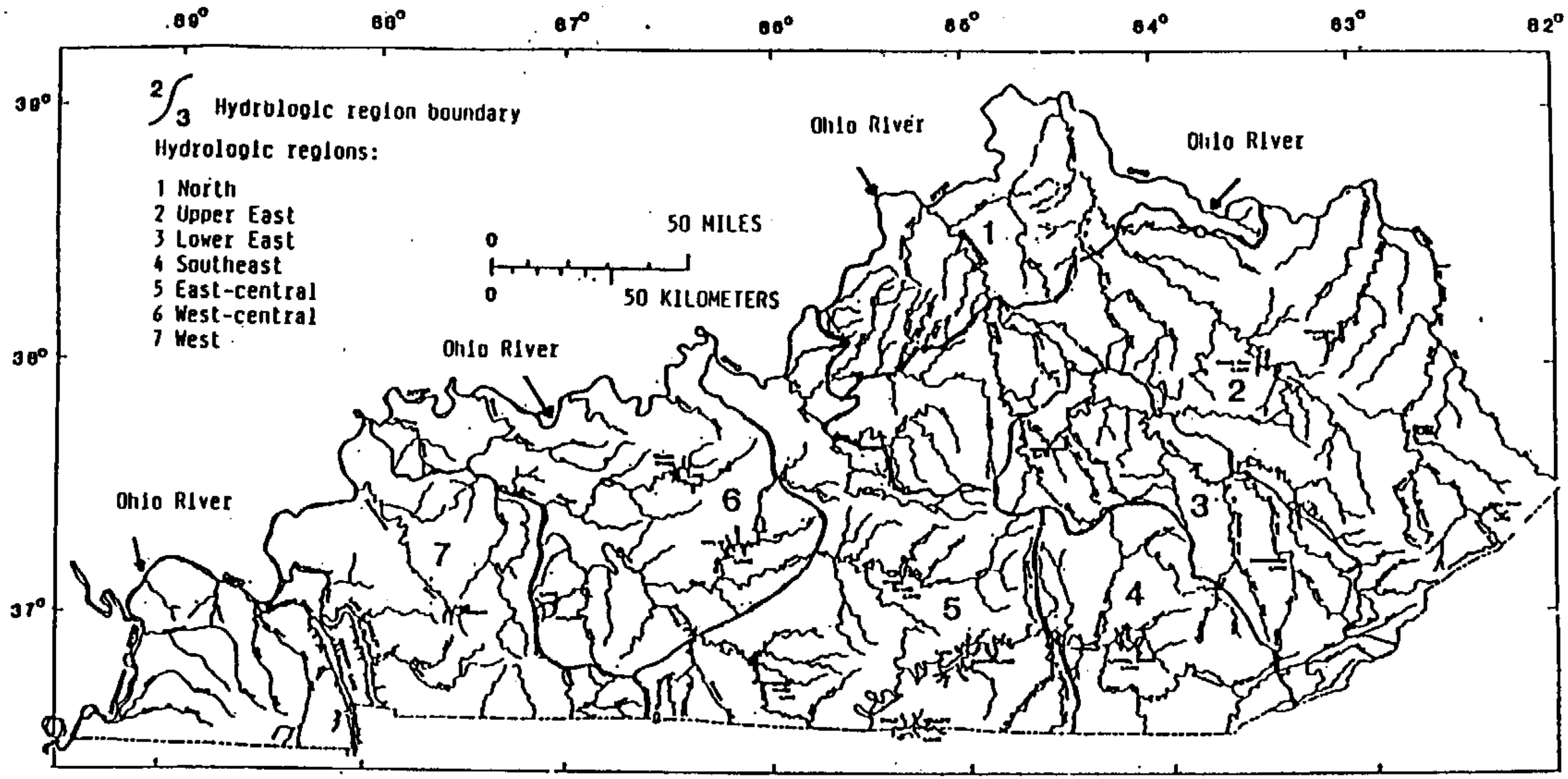

Fig. 2.1 Flood Regions in Kentucky as Defined by Method of Residuals (Choquette 1988)

hydrological attributes at each site within a flood region provide a means to select clusters that may be similar or distinct from others. They will also indicate the type of watersheds that lie within each flood region.

b) Performance of Regionalized Flood Frequency Models: The accuracy of delineation flood regions can be further evaluated by examining the performance of the regionalized flood frequency model. Commonly used measures of performance are scaled values of bias and root mean squared error (RMSE). In this study, this is carried out using Monte Carlo simulation methods as oulined below.

1. For a selected probability distribution, estimate the at-site parameters using method of L-moments.
2. Generate normalized flood flow sequences, having the same record length in years as the historical systematic AMF record at the site, using a suitable a random number generator. A widely used random number generator referred to as RAND is used in this study. This is an IBM function that uses a multiplicative linear congruential method for generating a uniform set of pseudo-random numbers.
3. Using the IBM flood regionalization computer program (Hosking, 1988) a regionalized flood frequency model (for the selected probability distribution) is developed (for each region). A total of 100 simulation runs are made.
4. The regionalized flood frequency model developed in step 3 is used to estimate the flood quantiles at each site using the index-flood method described earlier.
5. The scaled bias in the estimate a flood quantile at each site is computed by taking the difference between the simulated value and the historical estimate (based on regionalized historical flood record) and dividing

this by the historical estimate. The RMSE uses the square of this scaled bias.

6. Steps 2 through 5 are repeated for each of the 100 simulation runs.
7. Based on the 100 simulation runs, regional average values of bias and RMSE are then computed using the corresponding estimates at each of the gauged sites within the region.

c) Discriminant Analysis: The success of any cluster analysis in identifying flood regions that are homogeneous within themselves but are distinct from the others depends to a large extent on the ability to discriminate between them. The variables to be used in discriminating between clusters or regions must be those that control flood response like the physical and climatic characteristics of the watershed(refer to nomenclature). Furthermore, the classification of an ungauged site (does not have observed AMF data) into a particular region can only be carried out using the attribute variables used in the discrimination process.

The power of discriminant analysis is measured by the correct reclassification of the gauged sites into their respective cluster regions that are originally identified in cluster analysis phase. A good discrimination can be obtained when the percentage misclassification of gauged sites is minimal. The success of accomplishing this objective depends on the attribute variables available for discrimination. The overall objectives of discriminant analysis in the context of flood regionalization are:

a) To further explain the differences between cluster regions based upon hydrological variables (referred to as attribute variables) that affect and/or control flood response at each of the gauged site within a cluster region. This would further explain why the flood regions (or clusters) are different with respect to the response variables used in the clustering process.

b) To use results from the discriminant analysis to classify ungauged sites that do not have their flood response variables defined.

d) Regression Analysis: The ultimate objective or purpose of regionalizing flood data is to develop regionalized relationships for predicting the flood response (at selected frequency levels) at both gauged and ungauged sites. For gauged sites, the regionalized relationship can be used together with at site information. The development of a regional equation for predicting flood response or quantiles within a given region can be accomplished using regression analysis by relating the flood level (dependent variable) with important hydrologic variables controlling flood response (independent or attribute variables). In the USGS method of residuals approach, this is accomplished by relating the log-Pearson Type-III flood quantile estimates at each gauged site within a region to hydrologic variables such as the geomorphic characteristics of the watershed. The regression analysis is carried out using log-transformed (base 10) data. The predictive capability of such equations is determined by examining the residual error expressed in percent (Tasker, 1978). Ideally, this error should be as low as possible.

COMPARISON OF FLOOD REGIONALIZATION METHODS

The main focus of this study, as stated earlier, is to compare the two methods of flood regionalization, namely, cluster analysis (Methods 1 and 2) and method of residuals (Method 3). In the following chapter homogeneous flood regions delineated under these two methods are compared with those obtained by the method of residuals (refer to Figure 2.1) using the procedures discussed above. The following specific questions are addressed:

a) How do the homogeneous flood regions delineated in the present study using cluster analysis, differ from those derived by the USGS Method of Residuals in terms of the watersheds and their hydrological characteristics?
b) How well are the regions discriminated by the attribute variables under the two methods of regionalization? What are the most significant attribute variables that provide the maximum discrimination?
c) For the selected probability distributions controlling flood response at each gauged site, how do the results of flood regionalization differ in terms of the performance of the regionalized flood frequency growth curves? What are the differences in the flood quantile estimates at each site? Flood quantile estimates from the log-Pearson Type-III distribution will also be included in this comparison.
d) What are the differences in the regression equations that predict the flood quantiles (at various return periods) for each region using the two methods of regionalization? These regression equations are necessary for predicting flood quantiles at ungauged sites.

SPECIFIC RESEARCH PROCEDURES

Based on the overall procedures presented above, the following specific steps are followed in conducting this study:

a) Hydrologic data, necessary for performing a regional flood frequency analysis, are obtained from the U.S. Geological Survey, Louisville District. These include observed annual flood data as measured at each of the the gauged sites (referred to as response variables) and physical, climatic and hydraulic characteristics of the watersheds that affect flood response (referred to as attributes).

b) Probability distributions recommended for use in flood frequency analysis are selected after a careful review of previous research efforts. The following probability distributions, commonly employed in flood frequency analysis (Kuczera, 1982) are employed:

   a) Generalized Extreme Value (GEV) and its special case, Extreme Value Type-I (EV1)
   b) Wakeby

   The parameters of the probability distribution will be estimated using the method of L-moments.

c) Cluster Analysis is then used to form homogeneous flood regions based upon important statistical properties of the normalized AMF series and the probability distribution selected in step (b). Properties such as the mean, standard deviation, coefficients of variation, skewness and kurtosis (L-moments) and the specific mean annual flood, QSP, and the parameters of the probability distribution, as estimated from L-moments, are used as indices to measure flood response of each watershed. The FASTCLUS procedure available in the Statistical Analysis System (SAS, 1985) is used to obtain clusters or groups. The purpose of this analysis is to place the gauged sites into groups or clusters such that gauges within a cluster have similar flood response and those in different clusters have dissimilar flood response.

d) For the flood regions delineated in step (c) above, determine the most suitable regionalized probability distribution applicable to each of the gauged sites within the region using Monte Carlo simulation. This is based upon a performance criteria, such as the mean squared error and bias, that yield the most reliable estimates of extreme events. The simulation involves a detailed frequency analysis of the AMF series using regional parameters of the underlying probability distribution.

e) For each flood region delineated in step (c) above, summarize and evaluate the trends in the hydrological characteristics and develop a regionalized flood frequency growth curves for a given probability distribution. Evaluate differences in the shapes of these growth curves between regions and relate this to differences in the hydrological characteristics.

f) Perform Discriminant Analysis to distinguish between the clusters formed in step(c) based upon attribute variables such as the physical, climatic and other hydrologic characteristics of the watershed. The discriminant scores, associated with each of the attribute variables, are used to evaluate any misclassification of a gauged site into the homogeneous flood regions defined in step(c). This step will also identify the most important variables that affect or control flood response of a watershed and can later be used for developing flood prediction equations.

g) Within each cluster, perform a stepwise regression analysis with using select flood quantile levels as the dependent variable and other watershed hydrologic attributes as the independent variables. This step will also identify the most significant attributes variables controlling flood response of the watersheds within a cluster, and, additionally, provide a means to compare them with the set of attribute variables that contributed to the discriminant power between clusters as described in step (c) above. Compare the mean square and standard errors associated with the regression equations developed for each cluster region with similar equations obtained for the U.S.G.S. method of residuals flood regions. In this context, it must be emphasized that the actual gauges on each cluster will not be identical to those being used in the method of residuals study since the two methods are quite different in the manner in which the homogeneous flood regions are formed. However, the values of the errors associated with the regression equation within each cluster can be compared overall to those obtained from the method of residuals in order to determine the most suitable method of regionalization.

CHAPTER 3

DATA AND RESULTS

DATA ACQUISITION

Annual maximum floodpeak data (AMF series) was retrieved from WATSTORE by the U.S. Geological Survey, Louisville District office. Additional hydrologic data pertaining to each watershed corresponding to the gauged streamflow sites was provided by the U. S. Geological Survey office in Louisville. This data constitutes a part of the information on the attribute variables (or independent variables) to be used in the regionalization study. Additional geomorphic variables for each watershed may be necessary to further improve the regionalization process. Such data was not readily available at the completion of this report.

The following is a detailed list of hydrologic, physical and meteorologic data that is used in the flood regionalization study.

1) The systematic historic AMF record at each of the gauges in the State of Kentucky. Only gauges located in watersheds with drainage areas less than 1000 square miles and having at least 7 years of flood data is used in the analysis.
2) Physical characteristics affecting or controlling the flood response of the watershed in which the gauge is located. This includes watershed contributing drainage area, A_c, length, B_l, shape index, B_s, average slope, B_s, elevation, soil type, and land use (percent impervious area etc.),

and main channel length, L_c, sinuousity, S_s, and slope, S_c. Geomorphic data such as the number and average length of streams of different orders (for computing geomorphic properties of each watershed such as stream order, stream frequency, drainage density, form factor and bifurcation ratio), and the time of concentration were not readily available at the completion of this study.

3) Climatic data such as seasonal (dry and wet periods) and type of rainfall characteristics experienced in each of the watersheds. The only variable available at the time of this study was the mean annual rainfall.

The list of flood response variables (dependent variables) and the watershed attribute variables (independent variables), to be used in the regionalization study is shown at the end of this report under nomenclature. Pertinent statistical of data corresponding to these variables, as defined at each of the 253 gaging sites, is included in Table A.1, Appendix A. The values of the response variables are derived by computing important statistics of the normalized AMF data for each gauged site. These statistics, either individually or in combination, will be used in defining homogeneous flood regions using cluster analysis as presented in the following sections.

DELINEATION OF CLUSTER FLOOD REGIONS

Using FASTCLUS algorithm, a detailed cluster analysis is carried out using the response variables outlined in the previous section and in Chapter 2 with the following objectives.

1) To obtain optimum number of clusters or regions

that are physically realistic for representing flood experience for the State of Kentucky.

2) The number of clusters selected must satisfy at least one of the several available cutoff criteria. This would ensure that each cluster is homogeneous within itself but heterogeneous with respect to other clusters.
3) The number of gauged sites within a cluster must be sufficiently high in order to permit any statistical analysis.
4) The clusters must lend themselves to maximum possible discrimination based on the attribute variables (hydrological characteristics other than those based on AMF data). This would maintain the hydrologic distinction between the cluster regions.
5) The misclassification of the gauged sites already grouped and the ungauged sites to be assigned to a cluster region must be minimal.

With the above objectives in mind, results from cluster analysis using FASTCLUS algorithm are initially screened for the most suitable response variables to be used for further analysis. These results suggest that independent clusters or flood regions can be successfully formed using the statistical L-moments, LCV, LSK, LKUR of the normalized annual peak flow data, the parameters of the selected probability distribution and the specific mean annual flood, QSP taken individually or in combination. Clustering on physical characteristics of the watershed gave cluster regions that could not be discriminated well based on the flood response variables.

As expected, the composition of each cluster and the optimum number of clusters that can be extracted and discriminated (based upon attribute variables associated with the watersheds in which each of the gauged sites is located) continues to depend heavily on the type and number of response variables used in the analysis. Consequently, the final choice of clustering schemes, incorporating different response or clustering variables, is based upon the overall performance of each flood region. The following sections discuss results of all the clustering schemes and techniques used to delineate and evaluate the flood regions.

a) Clustering Cases: Twelve clustering schemes are adopted initially for further examination. Table 3.1 summarizes the results of the FASTCLUS clustering procedure for the various clustering schemes. Case 13 shown in this table applies to USGS regions, as delineated using method of residuals, and is included for the purpose of comparing the two method of regionalization. These twelve cases, as shown in Table 3.1, involve clustering with the response variables L-moments, namely, coefficients of variation, LCV, skewness, LSK, and kurtosis, LKUR, respectively, the specific mean annual flood, QSP, and the parameters of the EV1 (MEVL and AEVL), GEV (MGVL, AGVL and KGVL) and Wakeby (MWKL, AWKL, BWKL, CWKL, and DWKL) distributions. Each case is included in the study with a specific purpose. For example, for the clustering cases involving L-moments (Cases 1-3), Case 1, with clustering variable, LCV, would be appropriate for 2-parameter flood frequency models that require location and scale parameters to characterize the model completely. It must be emphasized, that the use of normalized AMF data standardizes the first moment (mean), characterizing the location, to 1.0. Since the coefficient of variation, LCV, reflects the dispersion (or scale) effects present in the flood data, this statistic would be totally adequate to describe a flood frequency model involving location and scale parameters. For example, the EV1 distribution used in this study can be characterized completely by LCV. In contrast, a five parameter flood frequency model like the Wakeby would require all L-moments, LCV, LSK and LKUR and

TABLE 3.1. Clustering Characteristics of Cases Examined in the Study

No.	Cluster Variables	No. of Clusters	R^2	CCC	No. of Sites in Each Cluster Region
1	LCV	6	0.953	-6.66	78,33,42,20,70,10
2	LCV, LSKEW	6	0.830	-1.17	66,45,16,57,31,38
3	LCV, LSKEW, LKUR	6	0.766	6.36	38,49,19,43,66,38
4*	LCV, QSP	5	0.759	-4.50	89,16,93,30,25
5*	LCV, LSKEW, QSP	5	0.689	1.97	79,17,75,44,38
6	LCV, LSKEW, LKUR, QSP	5	0.611	4.82	26,26,73,88,40
7	MEVL, AEVL	6	0.953	23.80	79,34,41,20,70,9
8	MGVL, AGVL KGVL	5	0.705	3.65	74,30,29,46,74
9	MWKL, AWKL, BWKL CWKL, DWKL	2	0.215	7.41	44,209
10*	MEVL, AEVL, QSP	5	0.775	12.24	43,10,91,79,30
11*	MGVL, AGVL, KGVL, QSP	6	0.646	4.27	81,21,40,15,68,28
12	MWKL, AWKL, BWKL CWKL, DWKL, QSP	3	0.287	6.57	5,12,236
13#	USGS REGIONS	7	-	-	32,68,26,20,38,31,38

\# Regions delineated by the Method of Residuals. Included for comparative purposes

* Indicates clustering cases selected in the study (referred to as Cases 1-4)

one higher order moment, LBMD. Cases 7-9 correspond to Cases 1-3 with the exception that the actual at-site parameters (as estimated from L-moments) of the appropriate flood frequency model are used as clustering variables. Hence, Case 1 would correspond to Case 7 since the estimation of EV1 parameters require LCV (for normalized AMF flows). Cases 4-6 and 10-12 are similar to the above cases but include an important clustering variable, namely the specific mean annual flood, QSP. Unlike all the other clustering variables, which describe the underlying flood frequency distribution, the specific mean annual flood describes the flood potential of each watershed. An examination of at-site estimates of QSP for the 253 gauged sites in Kentucky indicates that its value decreases as the size of watershed increases.

The relative performance of the above 12 clustering cases is evaluated in detail using the following results.

1. Results of the cutoff criteria for choosing optimum number of clusters,
2. Trends in the hydrological characteristics and regionalized frequency growth curves,
3. Performance of the regional flood frequency model using simulation,
4. Results of discriminant analysis, and
5. Results of regression analysis relating flood quantiles to watershed physical and climatic characteristics.

b) Selection of number of cluster regions and cases: Since one of the main objectives of cluster analysis, in the context of flood regionalization, is to delineate homogeneous flood regions that can be distinguished from each other, the number of clusters obtained must not be too few or large. With this in mind, several cutoff criterion or stopping rules, as discussed in Chapter 2, are used to determine the optimum number of cluster regions. An application of the stopping rules to the 12 clustering schemes gave results shown in Table 3.1. For all schemes the CCC criterion showed a peak or trough value going from larger to smaller number of clusters than the optimum number of clusters (refer to column 6 of Table 3.1) and the R^2 was quite high indicating a clear choice of the optimum number of clusters. The inclusion of QSP as a clustering variable changed the optimum number of cluster regions from 6 to 5 with the exception of the case when the GEV parameters are used in the clustering. Clustering on Wakeby parameters gave only 2-3 regions and gave the worst overall performance compared to all clustering cases examined in this study. Hence, the Wakeby probability distribution is not considered suitable for regionalizing flood data for the State of Kentucky and is dropped from further consideration. Amongst the remaining schemes, the inclusion of QSP as a clustering variable (refer to scheme numbers 4, 5, 10 and 11 in Table 3.1) improved, although marginally, the overall performance. Consequently, all results discussed in the following sections pertain to the following four cases (marked by an asterisk "*" in Table 3.1) that are finally selected from the twelve clustering schemes. These cases incorporate the flood regionalizations Methods 1 and 2.

Case 1 : Clustering with LCV and QSP (Method 1)
Case 2 : Clustering with the Extreme Value Type-1 probability distribution parameters (MEVL and AEVL) and QSP (Method 2)
Case 3 : Clustering with LCV, LSK and QSP (Method 1)
Case 4 : Clustering with the Generalized Extreme Value parameters (MGVL, AGVL and KGVL) and QSP (Method 2)

As mentioned in the previous section, Case 1 and Case 2 are similar since the clustering variable LCV is adequate to estimate the parameters MEVL and AEVL of the EV1 distribution. Also, Case 3 and Case 4 are similar since the clustering variables LCV and LSK are used to estimate the parameters MGVL, AGVL and KGVL of the GEV distribution. All the four clustering cases gave, by and large, disjoint cluster regions as illustrated in the bi-variate plots shown in Figures 3.1-3.13. The numbers shown on these figures are cluster numbers. It is obvious from these figures that the the overlap between cluster regions increases as the number of clustering variables increase (refer to Case 4, Figures 3.8-3.13). The bi-variate plot of EV1 parameters, MEVL versus AEVL (refer to Fig. 3.3), shows an inverse linear relationship suggesting an increase in the location parameter (mode) as the scale parameter decreases. The bi-variate plot involving L-moments, as in Case 3, illustrates that LCV is directly proportional to LSK (refer to Fig. 3.6).

The total number of gauged sites classified into each of the cluster regions for the above four cases is shown in the last column of Table 3.1. The smallest number actual sites within a cluster is 10 (Case 2) which is adequate for performing any statistical analysis within the region.

The number of gauged sites (not the actual gauges) assigned to a particular cluster depends on the clustering variables used in the analysis. This is illustrated in Tables 3.2-3.7. Using Cases 1-4 (clustering with response variables LCV, LSK, EV1 and GEV parameters and QSP), these tables show the number of gauges reassigned when the clustering case is changed to one of the remaining cases. Each row reflects the number of gauged sites reassigned to the cluster numbers shown in the columns when clustering is carried out using any other case in lieu of the one shown on left hand side. For instance, the first row in Table 3.4 shows that of the 89 gauged sites (refer to last column of Table 3.4) assigned to cluster 1 when using clustering variables, LCV and QSP (Case 1), 65 sites are reassigned to Cluster 1 when using LCV, LSK and QSP as clustering variables (i.e. Case 3), 21 gauged sites are reassigned to Cluster 3, and 3 sites to Cluster 4. Thus, there is a clear evidence of movement in the gauged sites between clusters when Case 1 and 3 are compared against each other. A similar comparison of Case 1 (clustering with LCV and QSP) versus Case 2 (clustering with EV1 parameters and QSP) and Case 3 (clustering with LCV, LSK and QSP) versus Case 4 (GEV parameters and QSP) respectively (refer to Tables 3.2 and 3.5), also suggests movement, although to a lesser degree, between cluster regions. Thus, the cluster regions delineated using the L-moments or parameters tend to be dependent on the type and number of clustering variables used. The effect of using different clustering variables (although standardized) on the hydrological composition of cluster regions delineated is illustrated further in the following sections.

c) Comparison of Cluster and USGS Regions: The seven flood regions delineated by the USGS using the method of residuals (refer to Figure 2.1), are quite different in terms of the actual gaged sites when compared to those obtained by cluster analysis. Since cluster regions are not coincident with any geographic or hydrologic boundaries, they can not be illustrated in a convenient manner like the USGS regions of Figure 2-1. Furthermore, the total and the individual gauged sites incorporated within a region vary considerably. This is clearly evident from a comparison of the USGS method of residuals regions with the cluster regions obtained under each of the four cluster schemes (cases 1-4). For example, Tables 3.8-3.11 compares the USGS regions with those obtained under clustering Cases 1-4. In these tables, the rows represent the cluster regions for a particular case with the total gauged sites within each region shown in the last column. In the same manner, the columns represent each of the seven USGS regions (as delineated using method of residuals) with the total gauged sites within a cluster shown in the last row. An examination of these tables indicates, as expected, significant movement of gages between the cluster and USGS regions.

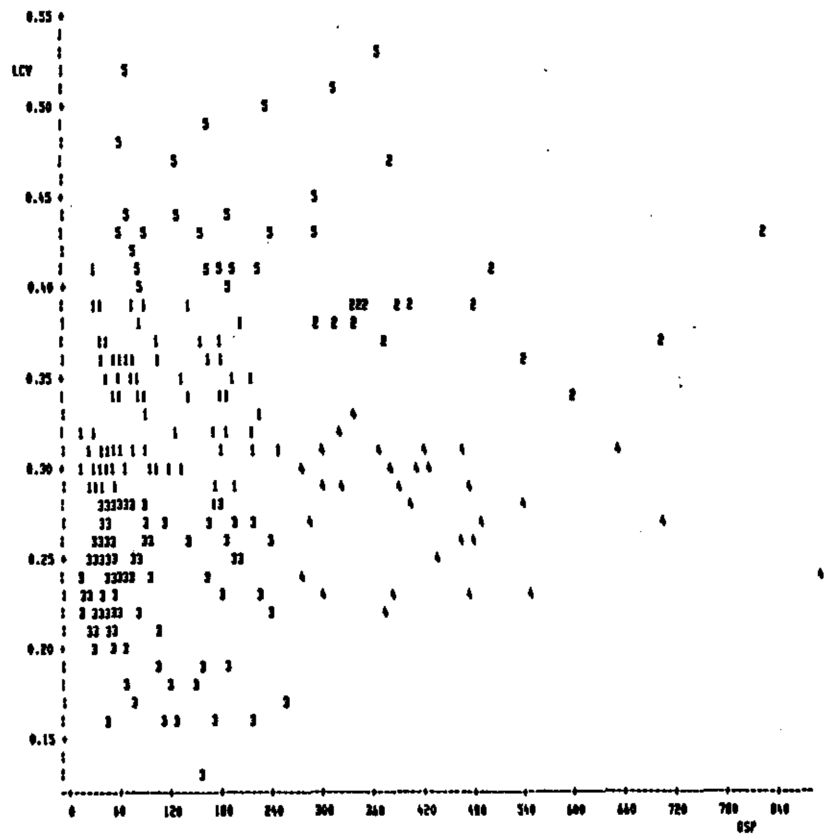

FIG. 3.1 : Bivariate Plot of LCV and OSP
(Case 1: Clustering Variables LCV and OSP)

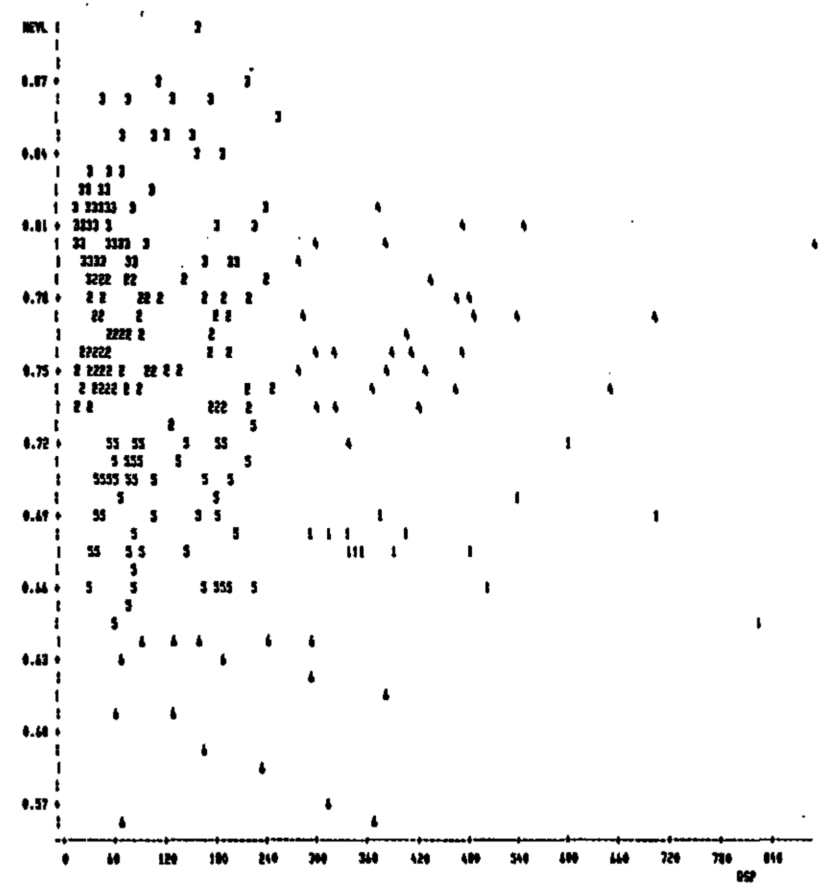

FIG. 3.2 : Bivariate Plot of NEVL and OSP
(Case 2: Clustering on NEVL, AEVL and OSP)

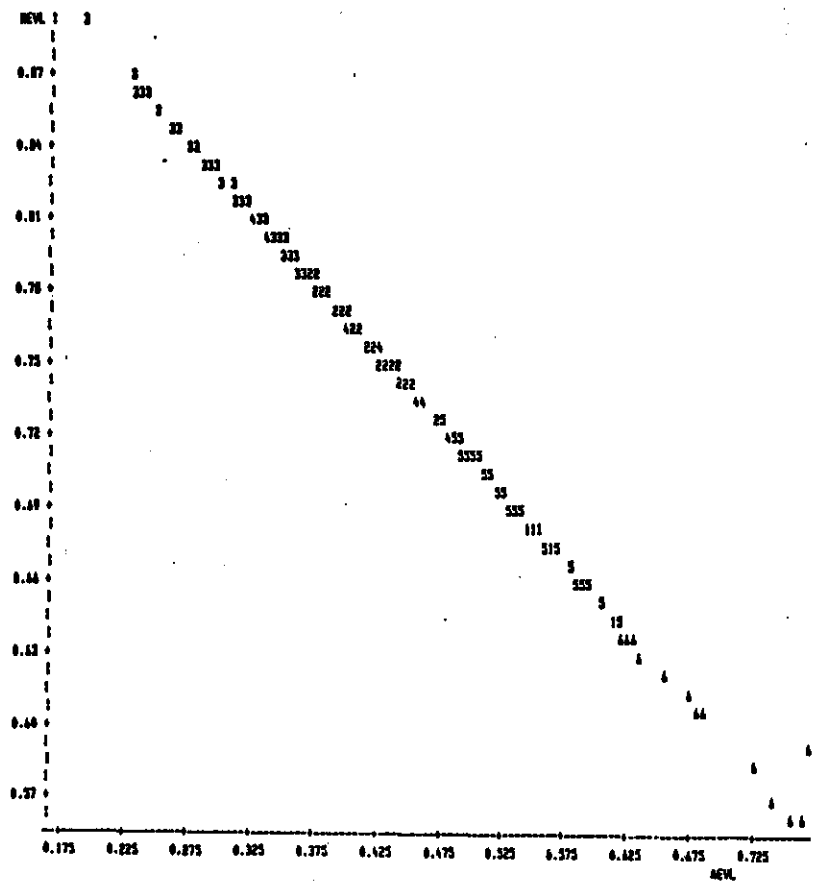

FIG. 3.3 : Bivariate Plot of NEVL and AEVL
(Case 2: Clustering on NEVL, AEVL and BSP)

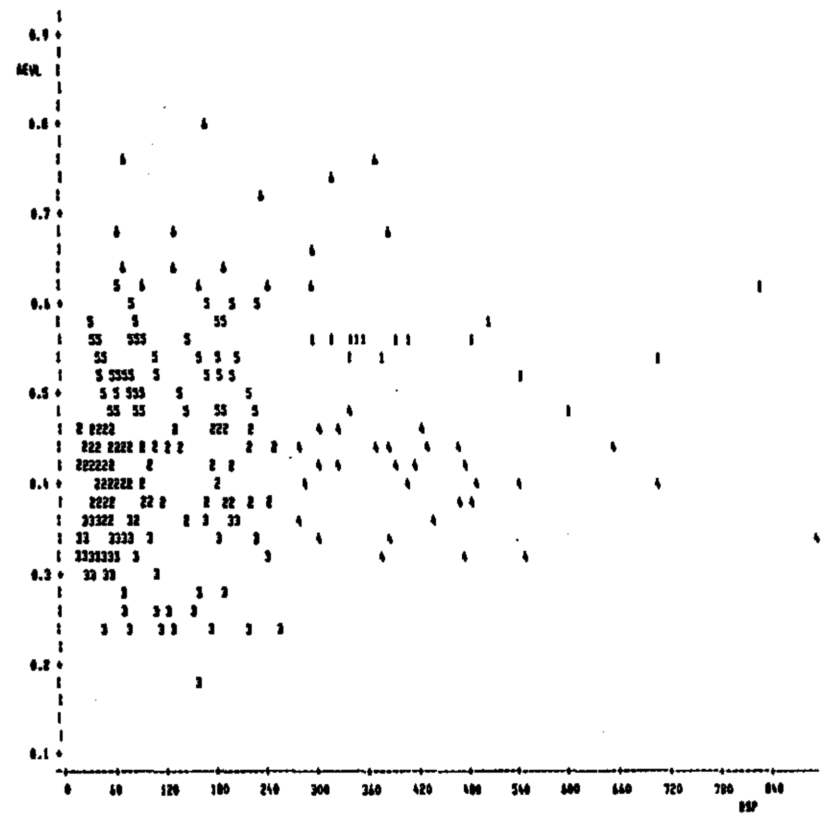

FIG. 3.4 : Bivariate Plot of AEVL and BSP
(Case 2: Clustering on NEVL, AEVL and BSP)

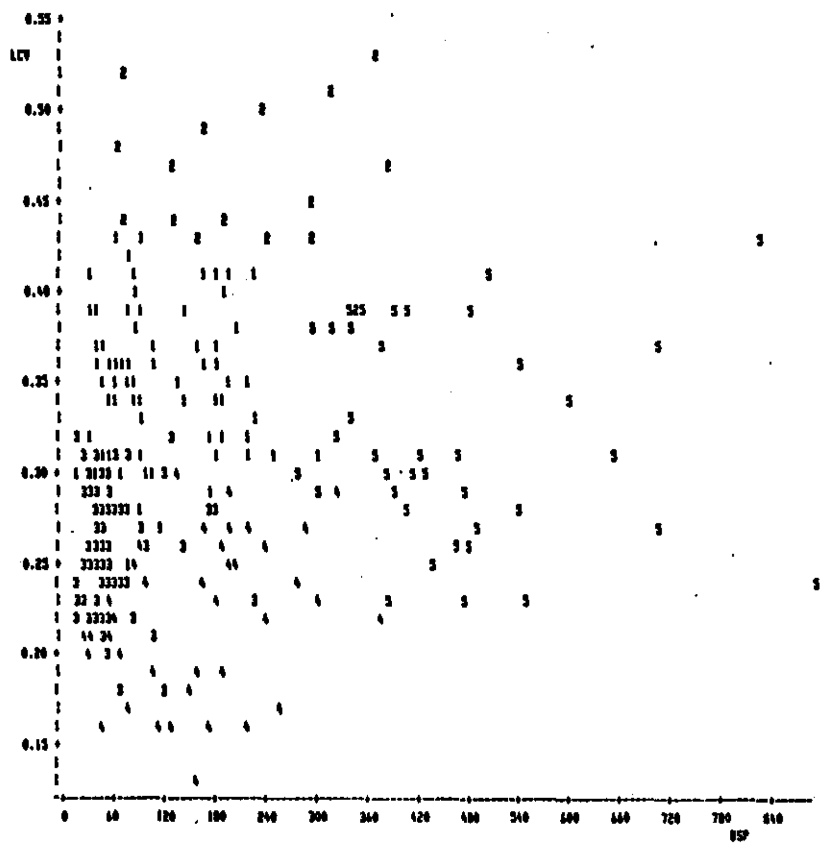

FIG. 3.5 : Bivariate Plot of LCV and BSP
(Case 3: Clustering Variables LCV, LSK and BSP)

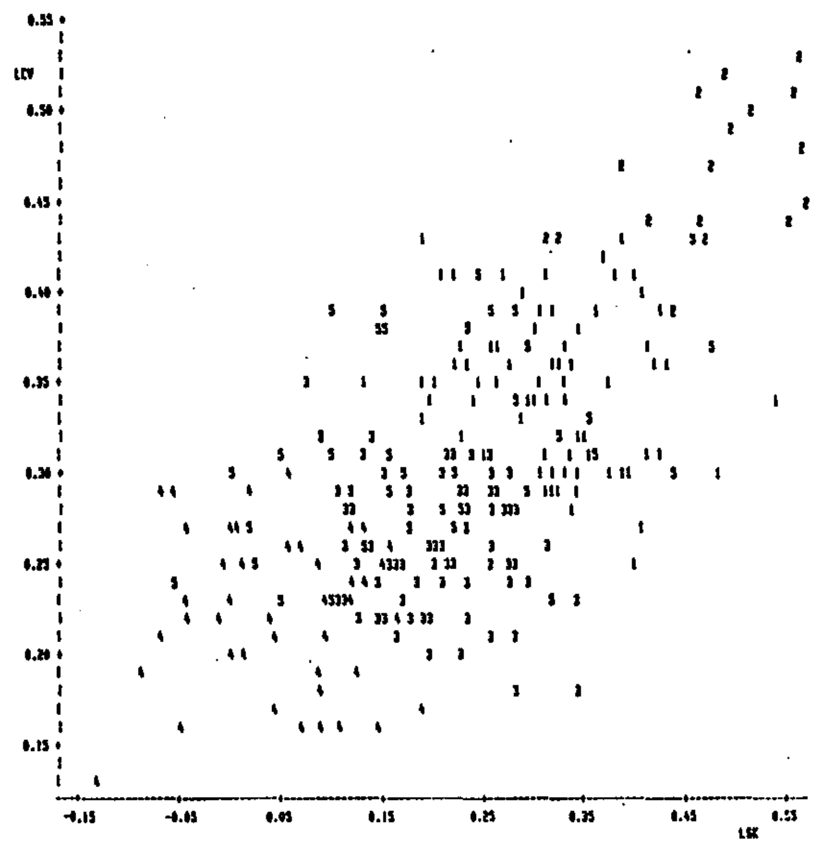

FIG. 3.6 : Bivariate Plot of LCV and LSK
(Case 3: Clustering Variables LCV, LSK and BSP)

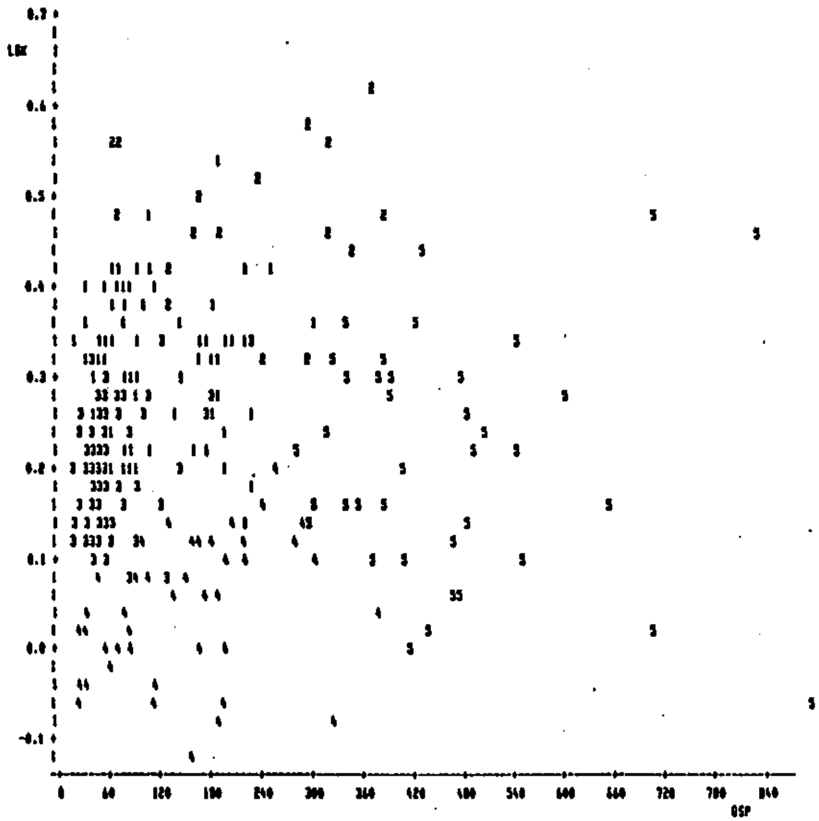

FIG. 3.7 : Bivariate Plot of LSK and OSP
(Case 3: Clustering Variables LCV, LSK and OSP)

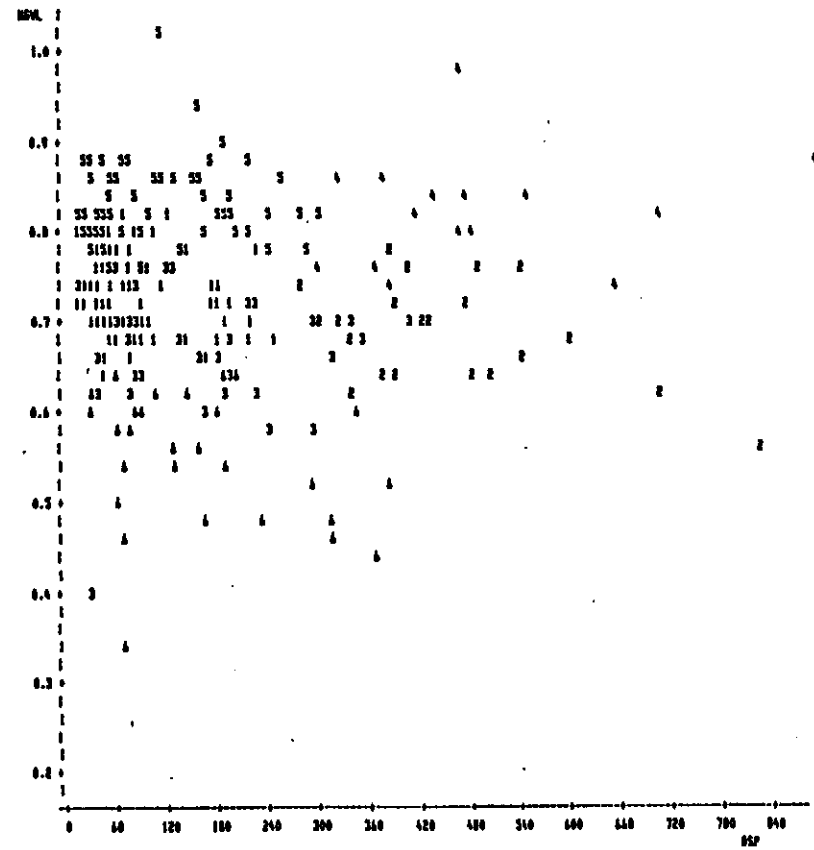

FIG. 3.8 : Bivariate Plot of MGVL and OSP
(Case 4: Clustering on MGVL, AGVL, KGVL and OSP)

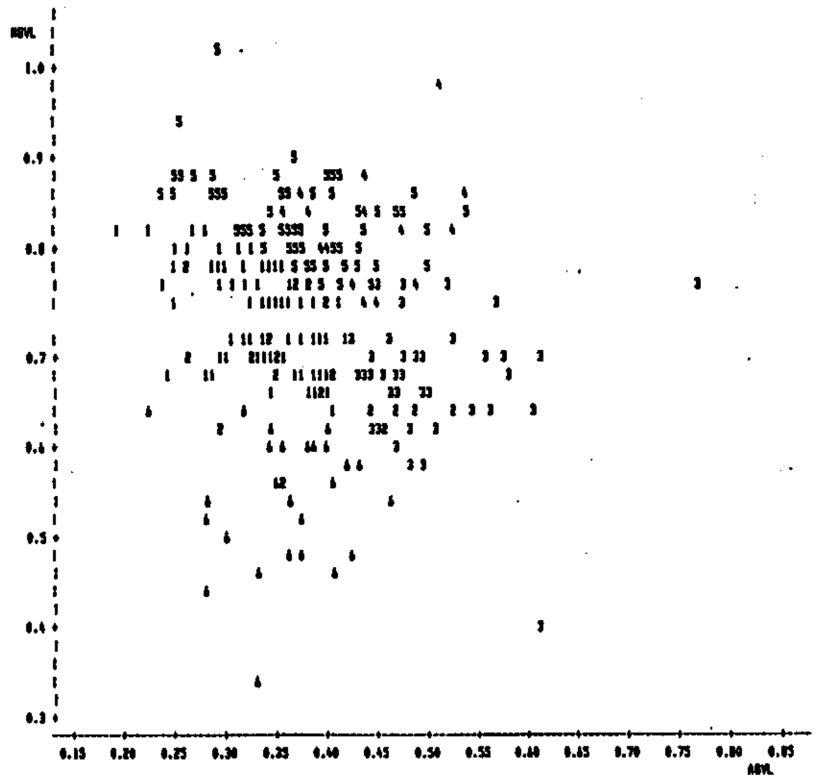

FIG. 3.9 : Bivariate Plot of NGVL and AGVL
(Case 4: Clustering on AGVL, AGVL, KGVL and BSP)

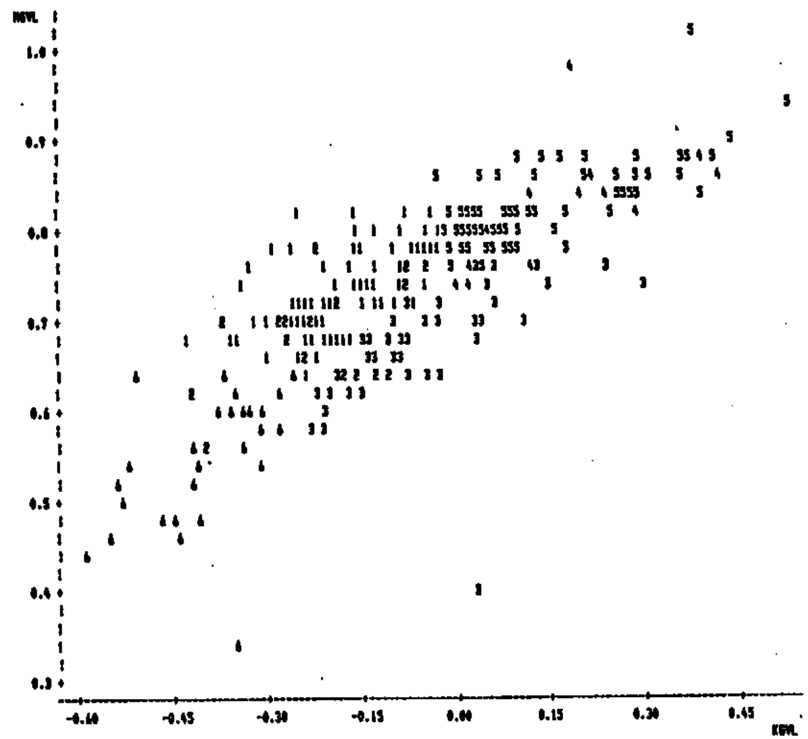

FIG. 3.10 : Bivariate Plot of NGVL and KGVL
(Case 4 : Clustering on NGVL, AGVL, KGVL and BSP)

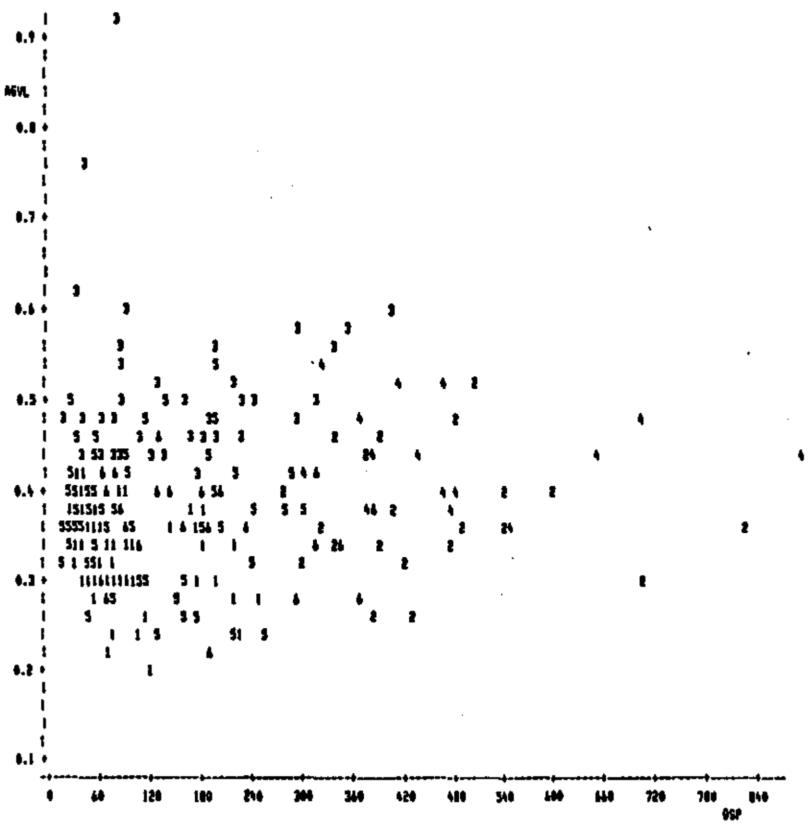

FIG. 3.11 : Bivariate Plot of AGVL and BSP
(Case 4 : Clustering on AGVL, ABVL, KGVL and BSP)

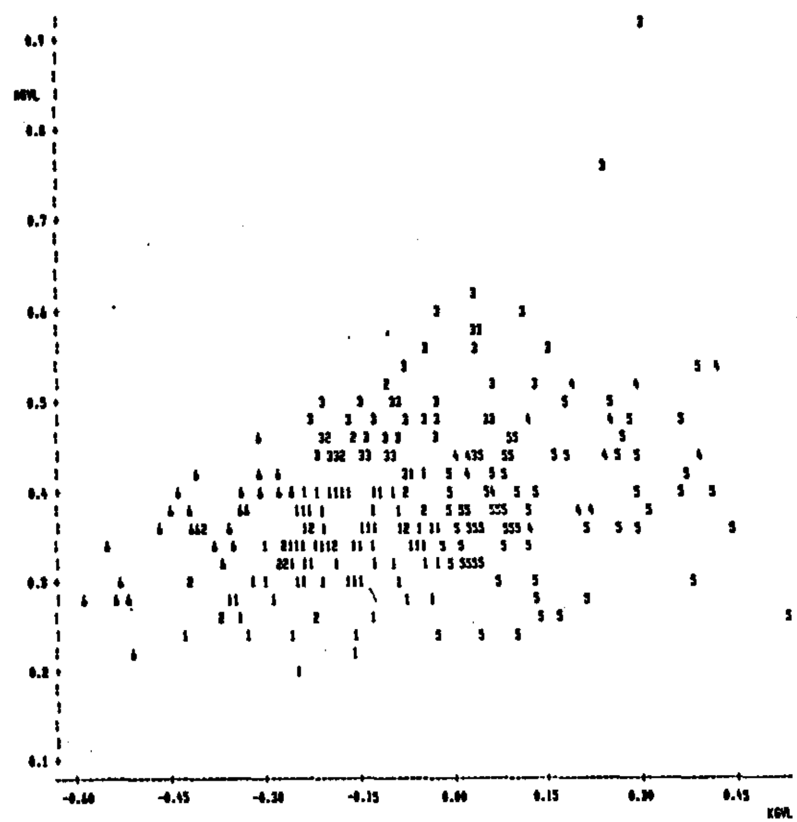

FIG. 3.12 : Bivariate Plot of AGVL and KGVL
(Case 4 : Clustering on NGVL, AGVL, KGVL and BSP)

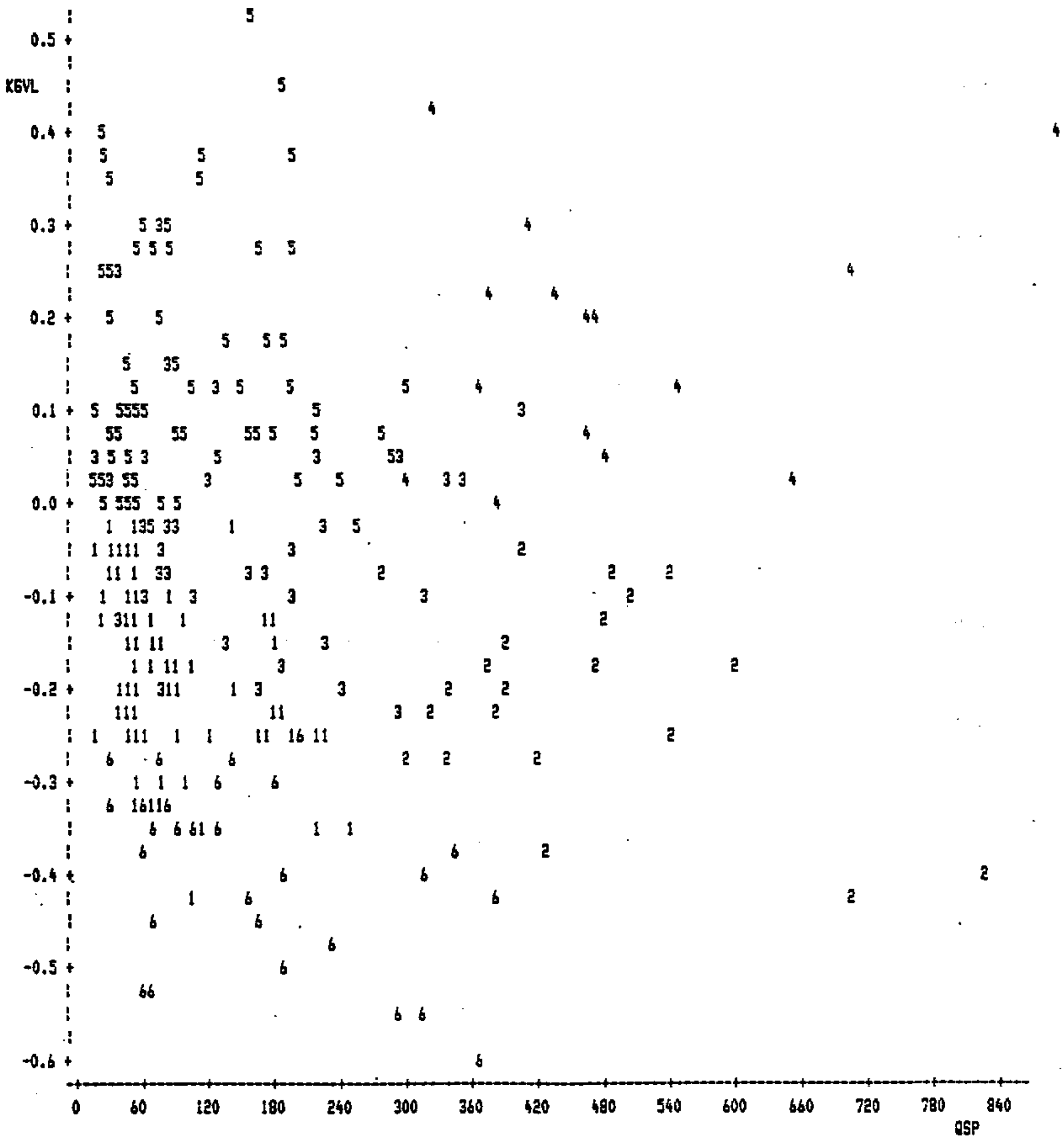

FIG. 3.13 : Bivariate Plot of KGVL and QSP
(Case 4 : Clustering on MGVL, AGVL, KGVL QSP)

TABLE 3.2. Comparison of Actual Number of Gauged Sites Assigned Between Case 1 and 2 Clustering Schemes

LCV & QSP	MEVL, AEVL & QSP 1	2	3	4	5	Total
1	11	0	78	0	0	89
2	7	9	0	0	0	16
3	0	0	13	79	1	93
4	0	1	0	0	29	30
5	25	0	0	0	0	25
Total	43	10	91	79	30	253

TABLE 3.3. Comparison of Actual Number of Gauged Sites Assigned Between Case 1 and 4 Clustering Schemes

LCV & QSP	GEV Parameters and QSP 1	2	3	4	5	6	Total
1	47	0	26	0	7	9	89
2	0	9	5	0	0	2	16
3	34	0	1	0	58	0	93
4	0	12	0	15	3	0	30
5	0	0	8	0	0	17	25
Total	81	21	40	15	68	28	253

TABLE 3.4. Comparison of Actual Number of Gauged Sites Assigned Between Case 1 and 3 Clustering Schemes

LCV & QSP	LCV, LSKEW & QSP 1	2	3	4	5	Total
1	65	0	21	3	0	89
2	0	2	0	0	14	16
3	3	0	54	36	0	93
4	1	0	0	5	24	30
5	10	15	0	0	0	25
Total	79	17	75	44	38	253

TABLE 3.5. Comparison of Actual Number of Gauged Sites Assigned Between Case 3 and 4 Clustering Schemes

LCV, LSKEW & QSP	GEV Parameters and QSP 1	2	3	4	5	6	Total
1	39	1	26	0	0	13	79
2	0	0	2	0	0	15	17
3	42	0	7	0	26	0	75
4	0	0	0	2	42	0	44
5	0	20	5	13	0	0	38
Total	81	21	40	15	68	28	253

TABLE 3.6. Comparison of Actual Number of Gauged Sites Assigned Between Cases 4 and 2 Clustering Schemes

EV1 Parameters and QSP

		1	2	3	4	5	Total
GEV PARAMETERS & QSP	1	0	0	54	27	0	81
	2	1	8	0	0	12	21
	3	17	1	22	0	0	40
	4	0	1	0	0	14	15
	5	0	0	12	52	4	68
	6	25	0	3	0	0	28
	Total	43	10	91	79	30	253

TABLE 3.7. Comparison of Actual Number of Gauged Sites Assigned Between Cases 3 and 2 Clustering Schemes

EV1 Parameters and QSP

		1	2	3	4	5	Total
LCV, LSK & QSP	1	21	0	55	2	1	79
	2	17	0	0	0	0	17
	3	0	0	32	43	0	75
	4	0	0	4	34	6	44
	5	5	10	0	0	23	38
	Total	43	10	91	79	30	253

TABLE 3.8. Comparison of Actual Number of Gauged Sites Assigned Between Cluster Regions (Case 1) and USGS Regions

U. S. G. S. Regions

		1	2	3	4	5	6	7	Total
LCV & QSP	1	5	29	14	6	15	5	15	89
	2	3	0	2	0	3	3	3	16
	3	14	28	6	12	11	15	7	93
	4	6	4	1	0	3	6	10	30
	5	4	7	3	2	6	2	1	25
	Total	32	68	26	20	38	31	38	253

TABLE 3.9. Comparison of Actual Number of Gauged Sites Assigned Between Cluster Regions (Cases 3) and USGS Regions

U. S. G. S. Regions

		1	2	3	4	5	6	7	Total
LCV, LSKEW & QSP	1	7	21	13	6	14	5	13	79
	2	4	4	1	0	6	1	1	17
	3	5	29	8	9	9	9	6	75
	4	7	13	2	5	5	8	4	44
	5	9	1	2	0	4	8	14	38
	Total	32	68	26	20	38	31	38	253

TABLE 3.10. Comparison of Actual Number of Gauged Sites Assigned Between Cluster Regions (Cases 2) and USGS Regions

EV1 PARAMETERS & QSP	U. S. G. S. Regions 1	2	3	4	5	6	7	Total
1	8	9	6	2	12	3	3	43
2	1	0	0	0	0	3	6	10
3	4	32	14	9	14	4	14	91
4	12	23	5	9	9	15	6	79
5	7	4	1	0	3	6	9	30
Total	32	68	26	20	38	31	38	253

TABLE 3.11. Comparison of Actual Number of Gauged Sites Assigned Between Cluster Regions (Cases 4) and USGS Regions

GEV PARAMETERS & QSP	U. S. G. S. Regions 1	2	3	4	5	6	7	Total
1	6	26	10	9	10	7	13	81
2	5	2	0	0	2	4	8	21
3	4	11	7	3	9	2	4	40
4	2	1	1	0	2	3	6	15
5	10	21	4	8	8	13	4	68
6	5	7	4	0	7	2	3	28
Total	32	68	26	20	38	31	38	253

DEVELOPMENT OF FLOOD FREQUENCY GROWTH CURVES

The procedure for developing a regionalized flood frequency growth curve was presented in Chapter 2. For the four clustering cases (Case 1-4), a separate regionalized flood frequency growth curve is developed for the EV1 and GEV probability distributions using historical systematic annual maximum floodpeak series (AMF series) from each of the gauged sites within a cluster region. The index-flood procedure presented in Chapter 2 is applied to accomplish the regionalization. The regionalized weighted (by the record length at each site) average L-moments and the corresponding EV1 and GEV parameters are shown in Tables 3.12 and 3.13 for each of the cluster regions delineated under the four clustering schemes. Similar data for the USGS regions are included for comparative purposes. For the USGS regions, the regionalized EV1 and GEV distributions are fitted using the method of L-moments. The actual gauged sites within each of the seven regions are identical to those contained in the regions delineated by the method of residuals (Choquette, 1988).

The EV1 and GEV regionalized flood frequency growth curves developed from the parameters in Table 3.13, are illustrated in Figures 3.14-3.21. It is important to note that the cluster numbers assigned to each region will change from case to case. Thus, cluster region number 5 in Figure 3.14 for Case 1 is not the same as cluster number 5 in Case 2. These numbers are arbitrarily assigned during the clustering process. Similar curves are developed for the USGS Method of Residuals regions are shown in Figures 3.22-3.23 for the EV1 and GEV distributions, respectively. The vertical scale (showing normalized discharge values) for the EV1 (Gumbel) distribution is drawn to half the scale than the one used for the GEV in order to improve the clarity of these frequency growth curves.

TABLE 3.12. Comparison of Regional Average L-Moments Estimated Using Normalized Historic AMF Data

Region No. *	Regional Average L-Moments: MEAN	M(LCV)	M(LSK)	M(LKUR)	M(LBMD)
Cluster Regions:					
Case 1: Clustering with LCV and QSP					
3	1.0000	0.2383	0.1760	0.1810	0.0781
4	1.0000	0.2823	0.1988	0.1839	0.0786
1	1.0000	0.3242	0.2758	0.1914	0.1016
2	1.0000	0.3862	0.3058	0.1900	0.0809
5	1.0000	0.4432	0.4035	0.2784	0.1657
Case 2: Clustering on EV1 parameters (MEVL, AEVL) and QSP					
4	1.0000	0.2310	0.1641	0.1829	0.0741
5	1.0000	0.2801	0.2008	0.1834	0.0743
3	1.0000	0.3116	0.2637	0.1869	0.1001
2	1.0000	0.3621	0.2837	0.2074	0.1097
1	1.0000	0.4196	0.3703	0.2451	0.1359
Case 3: Clustering with LCV, LSK and QSP					
4	1.0000	0.2229	0.0530	0.1324	0.0519
3	1.0000	0.2613	0.2005	0.1713	0.0784
5	1.0000	0.3224	0.2420	0.1853	0.0809
1	1.0000	0.3383	0.3187	0.2178	0.1176
2	1.0000	0.4615	0.4672	0.3275	0.1847
Case 4: Clustering on GEV parameters (MGVL, AGVL, KGVL) and QSP					
5	1.0000	0.2379	0.1054	0.1379	0.0527
4	1.0000	0.2735	0.0767	0.1239	0.0644
1	1.0000	0.2827	0.2877	0.2218	0.1145
2	1.0000	0.3330	0.3218	0.2319	0.0952
3	1.0000	0.3568	0.2335	0.1385	0.0700
6	1.0000	0.4233	0.4544	0.3236	0.1868
USGS Regions					
6	1.0000	0.2698	0.2230	0.1867	0.0743
1	1.0000	0.2781	0.2728	0.2139	0.0989
4	1.0000	0.2830	0.2037	0.1582	0.0801
2	1.0000	0.2839	0.2265	0.1994	0.1044
7	1.0000	0.3034	0.2691	0.2028	0.1061
3	1.0000	0.3115	0.2695	0.1830	0.0887
5	1.0000	0.3185	0.2852	0.2061	0.1042

* Regions arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK)

TABLE 3.13. Comparison of Regional Average Parameters of EV1 and GEV Probability Distributions Fitted to Normalized Historic AMF Data #

Region No. *	EV1: MEVL	EV1: AEVL	GEV: MGVL	GEV: AGVL	GEV: KGVL
Cluster Regions:					
Case 1: Clustering on LCV and QSP					
3	0.80	0.34	0.80	0.34	-0.01
4	0.76	0.41	0.76	0.39	-0.04
1	0.73	0.47	0.70	0.40	-0.16
2	0.68	0.56	0.63	0.44	-0.20
5	0.63	0.64	0.55	0.42	-0.33
Case 2: Clustering on EV1 parameters (MEVL, AEVL) and QSP					
4	0.81	0.33	0.81	0.34	0.01
5	0.77	0.40	0.76	0.39	-0.05
3	0.74	0.45	0.71	0.39	-0.14
2	0.70	0.52	0.66	0.43	-0.17
1	0.65	0.61	0.58	0.43	-0.29
Case 3: Clustering on LCV, LSK and QSP					
4	0.81	0.32	0.85	0.37	0.19
3	0.78	0.38	0.77	0.36	-0.05
5	0.73	0.47	0.71	0.42	-0.11
1	0.72	0.49	0.68	0.38	-0.22
2	0.62	0.67	0.52	0.38	-0.42
Case 4: Clustering on GEV parameters (MGVL, AGVL, KGVL) and QSP					
5	0.80	0.34	0.82	0.37	0.10
4	0.77	0.39	0.80	0.45	0.15
1	0.76	0.41	0.74	0.34	-0.18
2	0.72	0.48	0.68	0.37	-0.22
3	0.70	0.51	0.68	0.47	-0.10
6	0.65	0.61	0.56	0.36	-0.40
USGS Regions:					
6	0.78	0.39	0.76	0.36	-0.08
1	0.77	0.40	0.74	0.34	-0.15
4	0.76	0.41	0.75	0.39	-0.05
2	0.76	0.42	0.74	0.38	-0.09
7	0.75	0.44	0.72	0.37	-0.15
3	0.74	0.45	0.71	0.38	-0.15
5	0.73	0.46	0.70	0.38	-0.17

\# MEVL, MGVL = location parameters (modes); AEVL, AGVL = scale parameters; and KGVL = shape parameter.
* Regions arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK)

The shapes of the regionalized frequency growth curves for cluster regions (not the actual cluster region numbers) depends on the clustering variables and the underlying probability distribution used. For example, for the EV1 distribution, the regionalized frequency growth curves are different when clustering on LCV and QSP (Case 1) when compared to clustering on LCV, LSK and QSP (Case 3). The differences are more prominent with the GEV probability distribution. It is clear from Figures 3.14-3.21 that EV1 distribution produces straight linear graphs with normalized discharge ratios ranging from 0.0-5.0 since it has only two parameters (as defined by the coefficient of variation, LCV). This distribution would be appropriate for flood data exhibiting a moderate skew close to the EV1 skew of 1.14. In contrast, the GEV distribution produces pronounced non-linear curves with normalized discharge ratios ranging from 0.0-10.0 since it has an additional parameter to capture high skew commonly present in the flood data (as defined by the coefficient of skewness, LSK). Thus, the three-parameter GEV distribution is able to model the upper tail (return periods greater than 20 years) better than the two-parameter EV1 distribution for highly skewed flood data. This is clearly evident for regions that have steeper regionalized flood frequency growth curves, and, hence are characterized by high coefficients of variation, LCV, and skewness, LSK. For example, in Figures 3.14 and 3.15 for Case 1 (cluster regions delineated using variables LCV and QSP) cluster region number 3 has the flattest curve with regionalized EV1 parameters of MEVL = 0.80 (location) and AEVL = 0.34 (scale) and regionalized GEV parameters MGVL = 0.80 (location), AGVL = 0.34 (scale) and KGVL = -0.01 (shape). Since the shape parameter, KGVL, of the GEV distribution is close to zero, the EV1 and GEV flood frequency curves for this cluster region are similar. However, a comparison of the regionalized flood frequency curves for the steepest curves associated with cluster region number 5 (having regionalized parameters for EV1 : MEVL = 0.63; AEVL = 0.64 and for GEV : MGVL = 0.55; AGVL = 0.42 ; KGVL = -0.33) shows considerable difference in the normalized discharge values for return periods greater than 20 years. In all clustering cases, the regionalized flood frequency growth curves are distinct between regions indicating a successful delineation of flood regions (homogeneous within but distinct from other regions) using cluster analysis.

An examination of the regionalized flood frequency curves for the USGS regions, as illustrated in Figures 3.22 and 3.23, shows very little difference between the regions for both EV1 and GEV probability distributions. In both cases, the normalized flood discharge values range from 0.0-5.0 similar to the EV1 distribution for the cluster regions. Thus, at least in terms of their flood frequency growth curves, the USGS regions show more homogeneity across regions than the cluster regions. Furthermore, the frequency growth curves are not very steep for all the seven USGS regions (GEV shape parameter ranges from 0.0 to -0.17 as shown in Table 3.13).

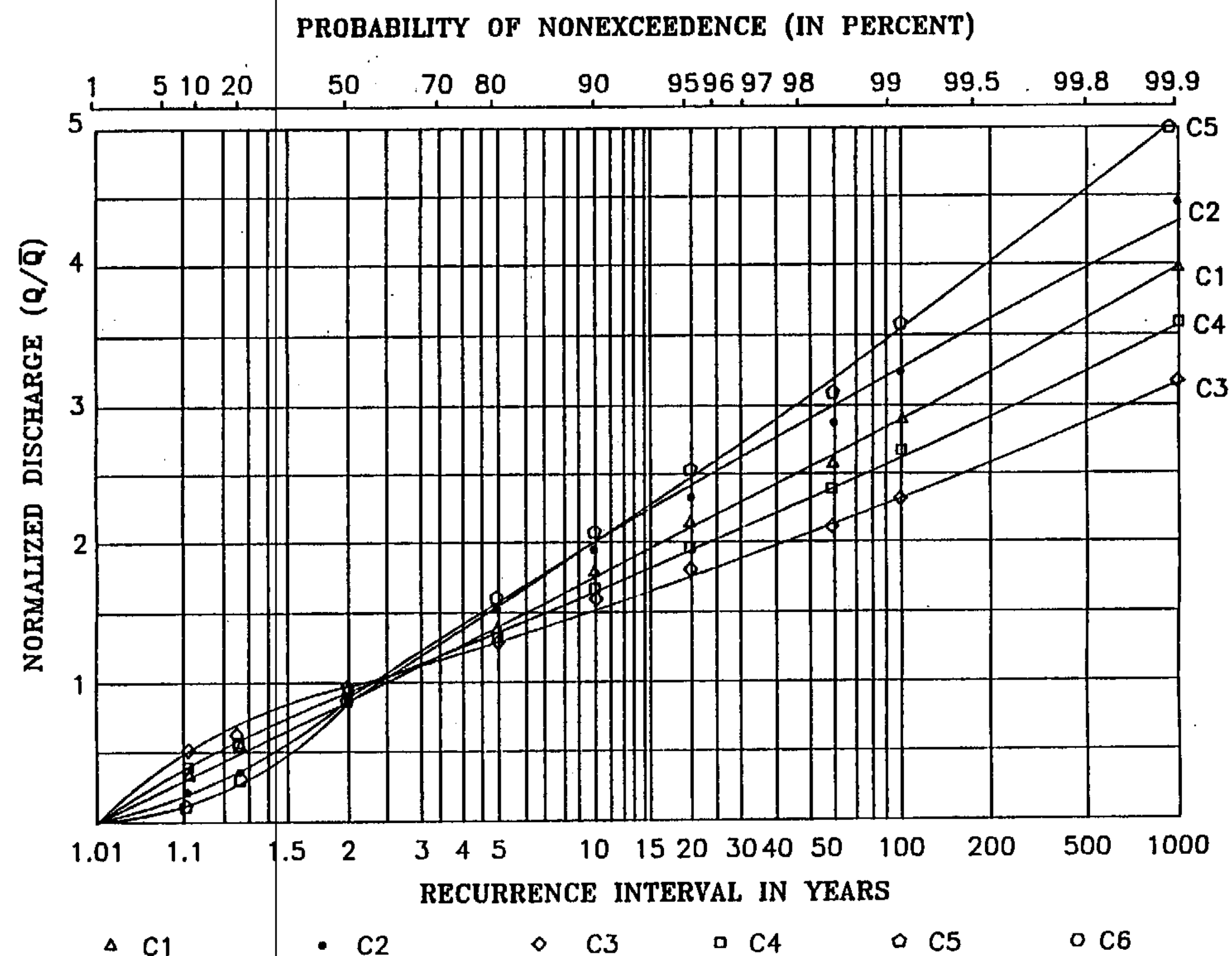

Fig. 3.14. EV1 Regionalized Flood Frequency Growth Curves for Cluster Regions (Case 1: Clustering Variables LCV and QSP)

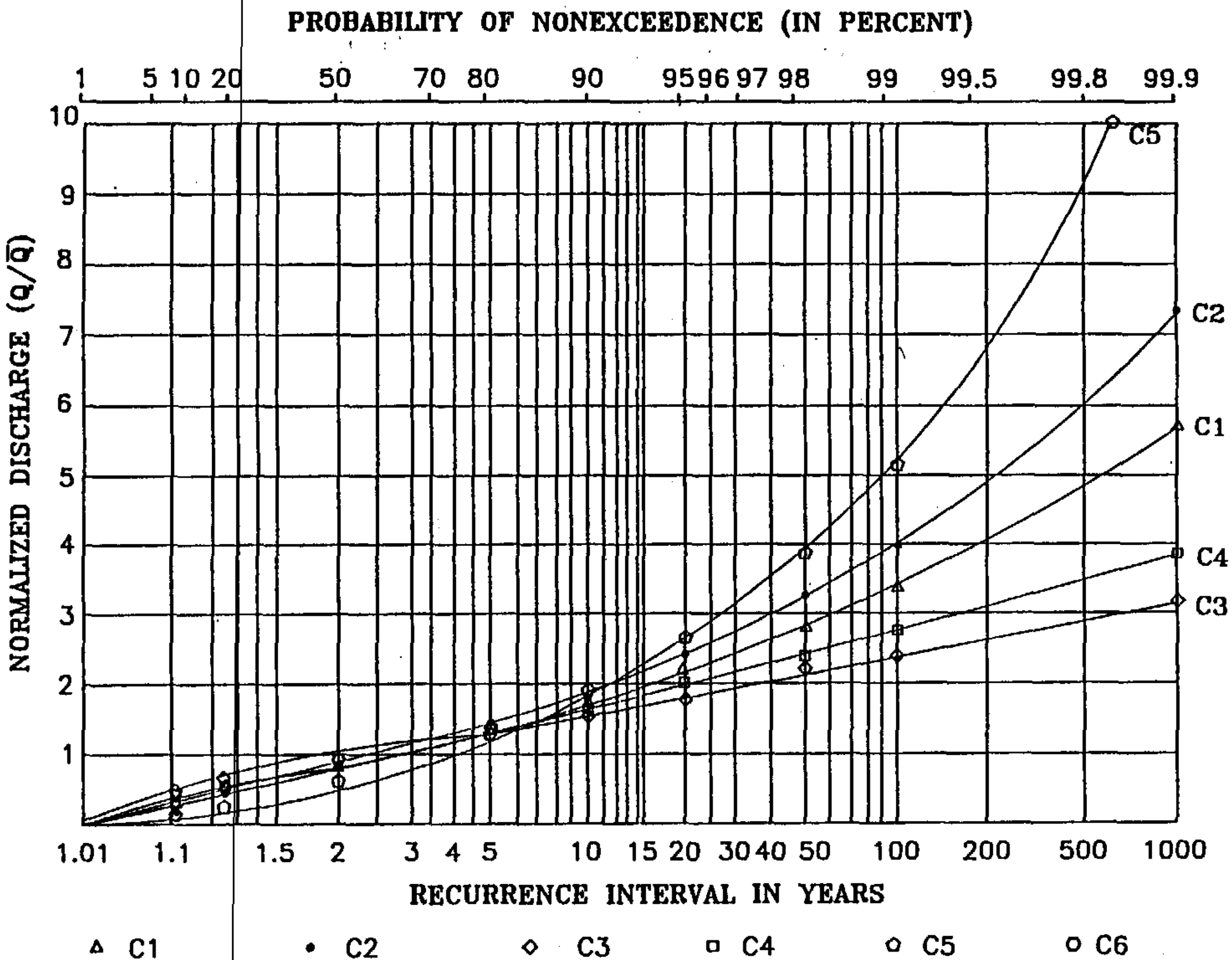

Fig. 3.15. GEV Regionalized Flood Frequency Growth Curves for Cluster Regions (Case 1: Clustering Variables LCV and QSP)

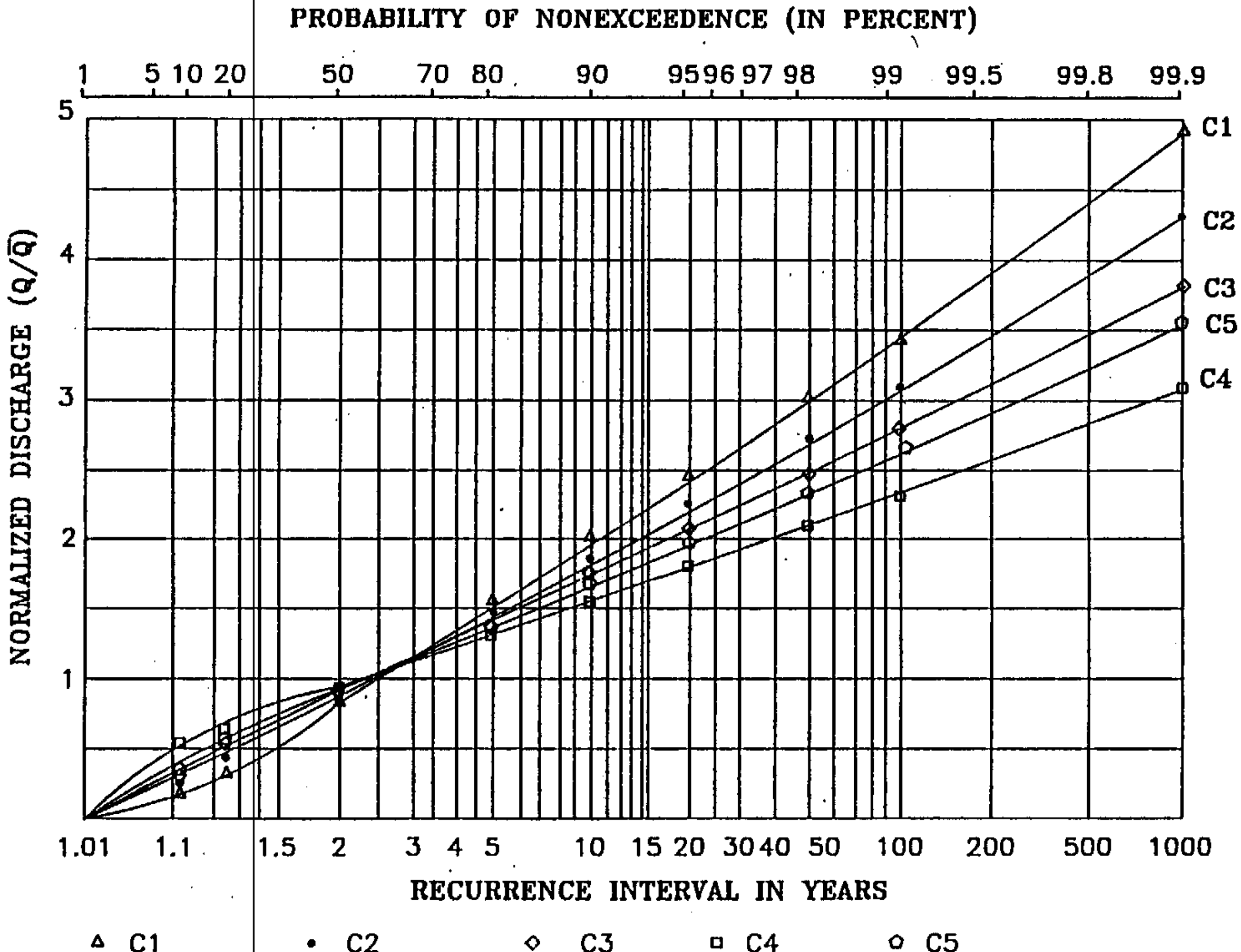

Fig. 3.16. EV1 Regionalized Flood Frequency Growth Curves for Cluster Regions (Case 2: Clustering Variables EV1 Parameters and QSP)

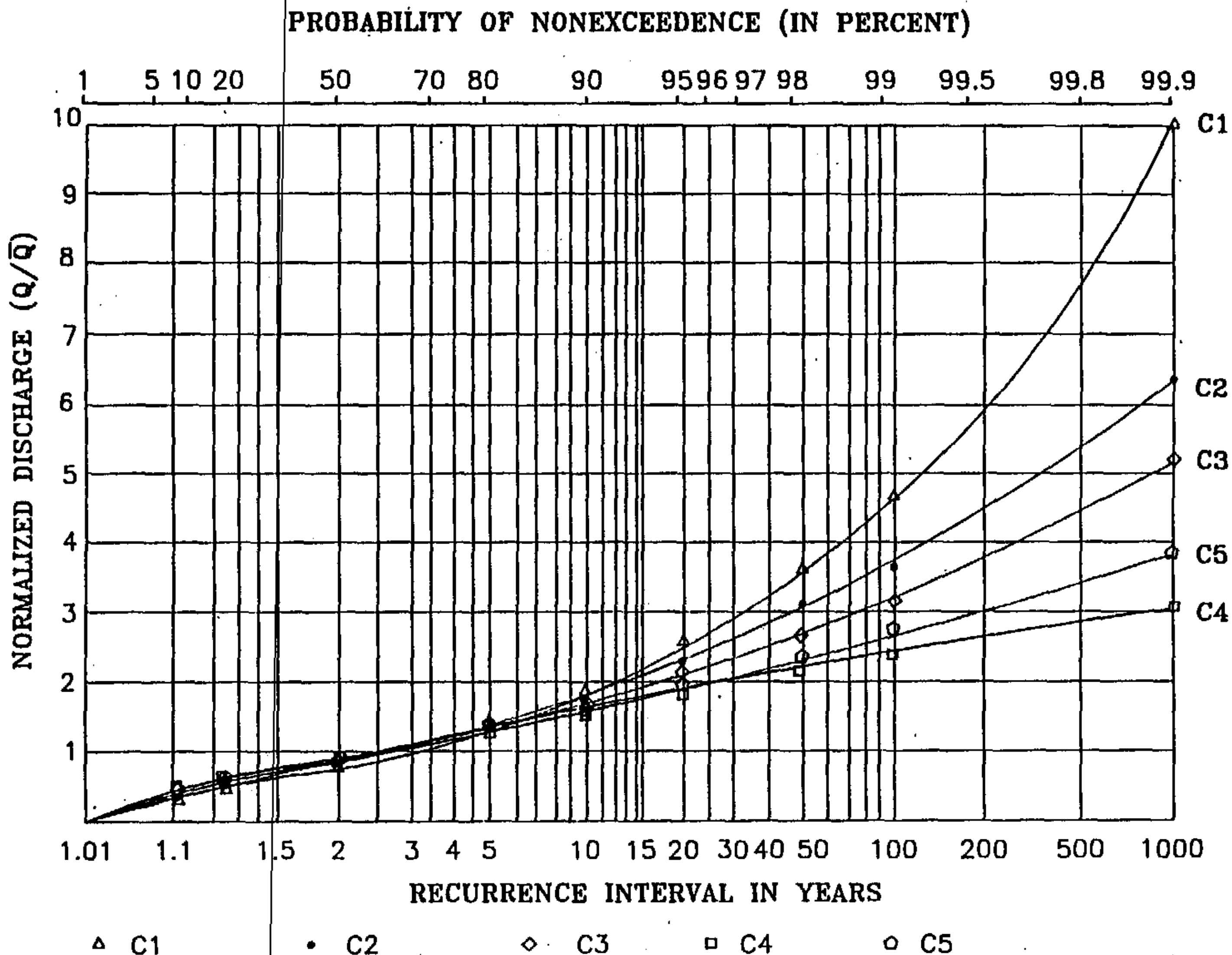

Fig. 3.17. GEV Regionalized Flood Frequency Growth Curves for Cluster Regions (Case 2: Clustering Variables EV1 Parameters and QSP)

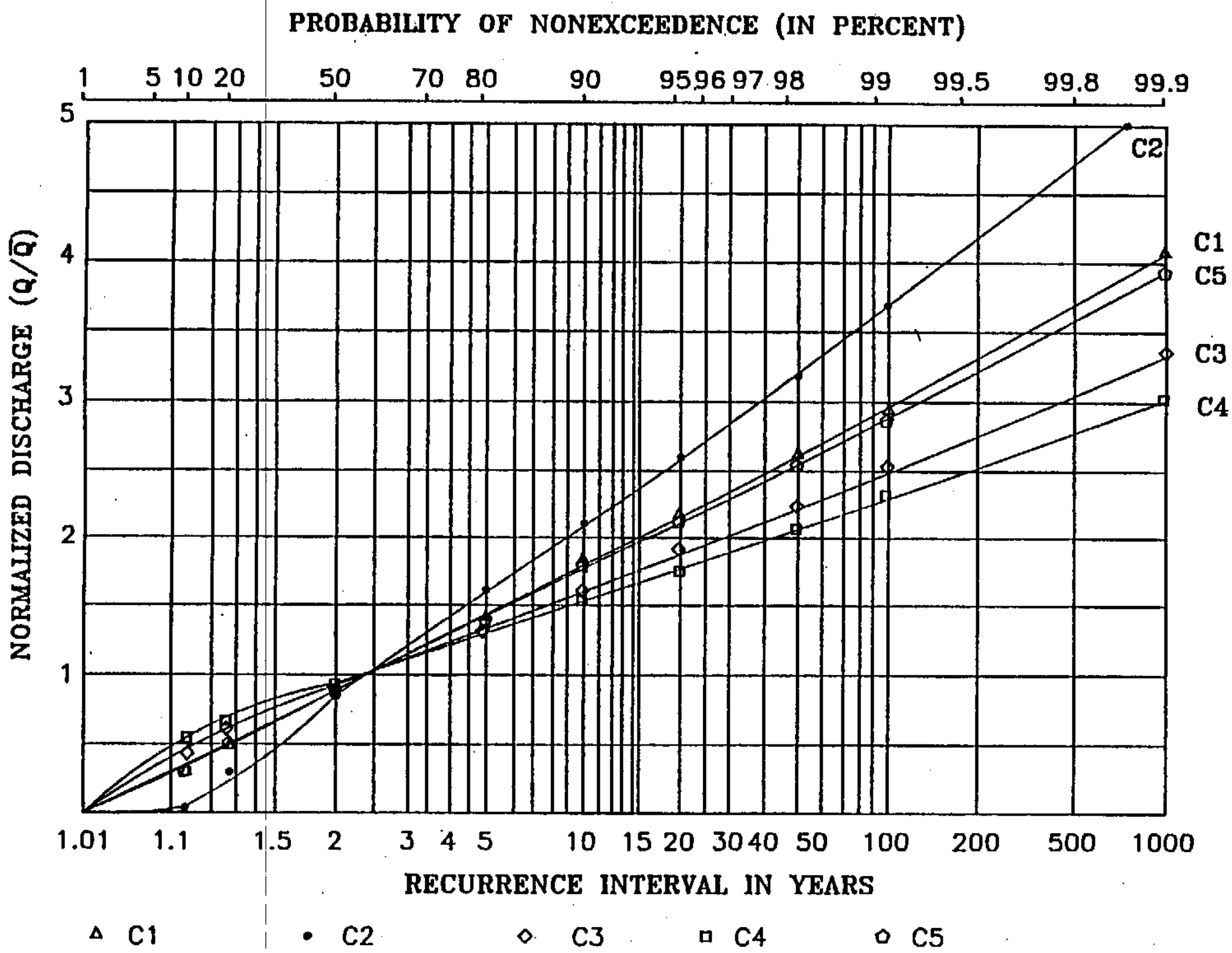

Fig. 3.18. EV1 Regionalized Flood Frequency Growth Curves for Cluster Regions (Case 3: Clustering Variables LCV, LSK and QSP)

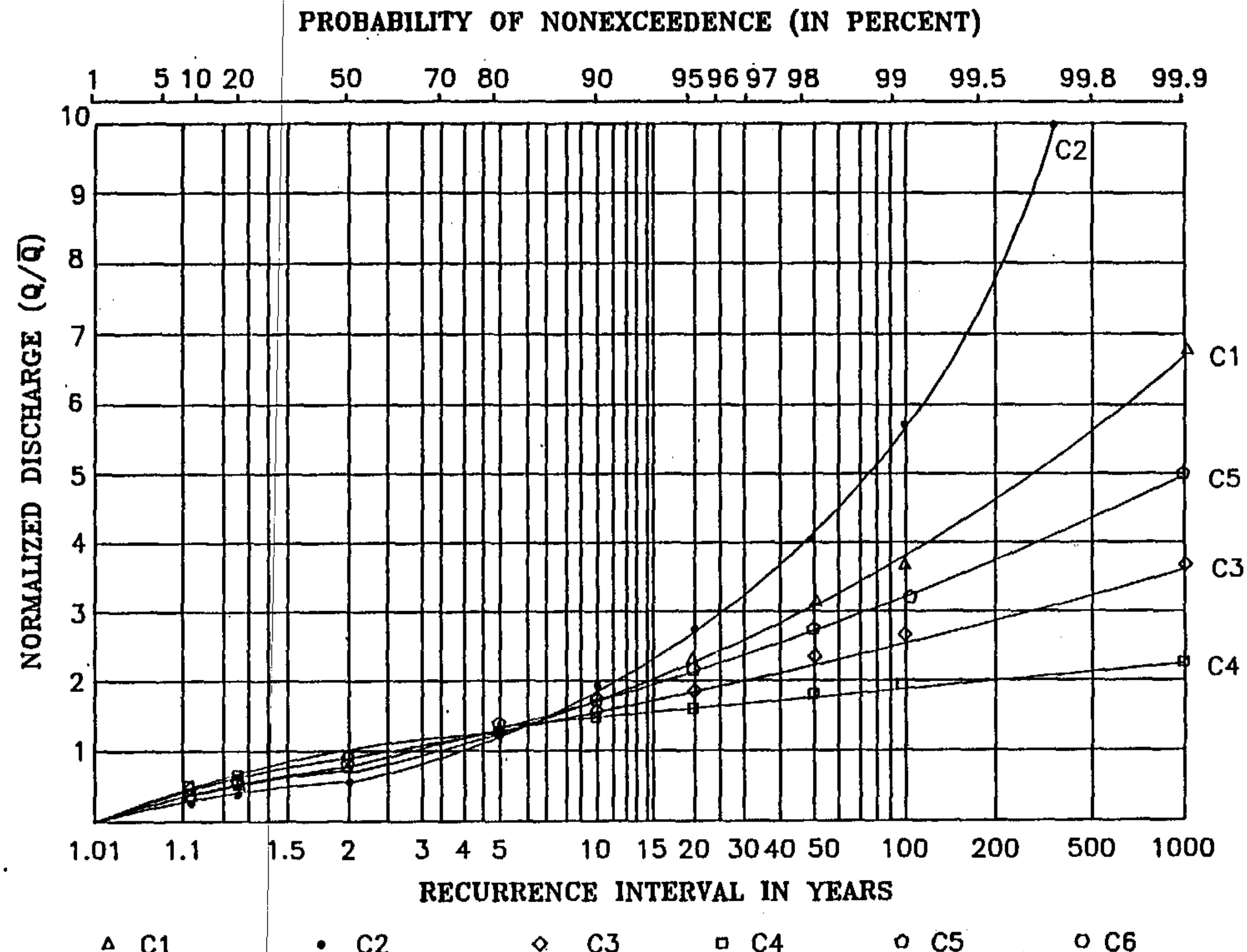

Fig. 3.19. GEV Regionalized Flood Frequency Growth Curves for Cluster Regions (Case 3: Clustering Variables LCV, LSK and QSP)

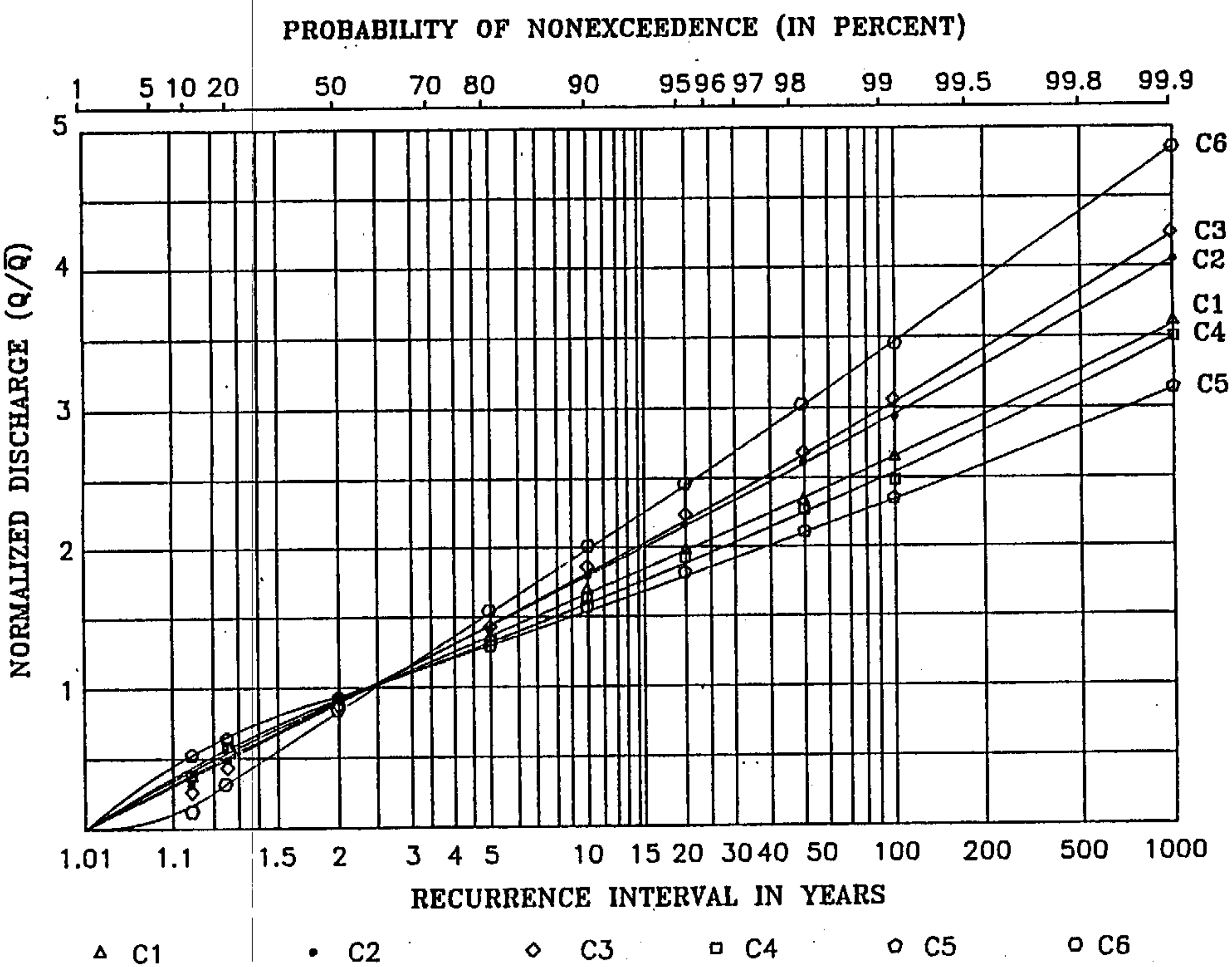

Fig. 3.20. EV1 Regionalized Flood Frequency Growth Curves for Cluster Regions (Case 4: Clustering Variables GEV Parameters and QSP)

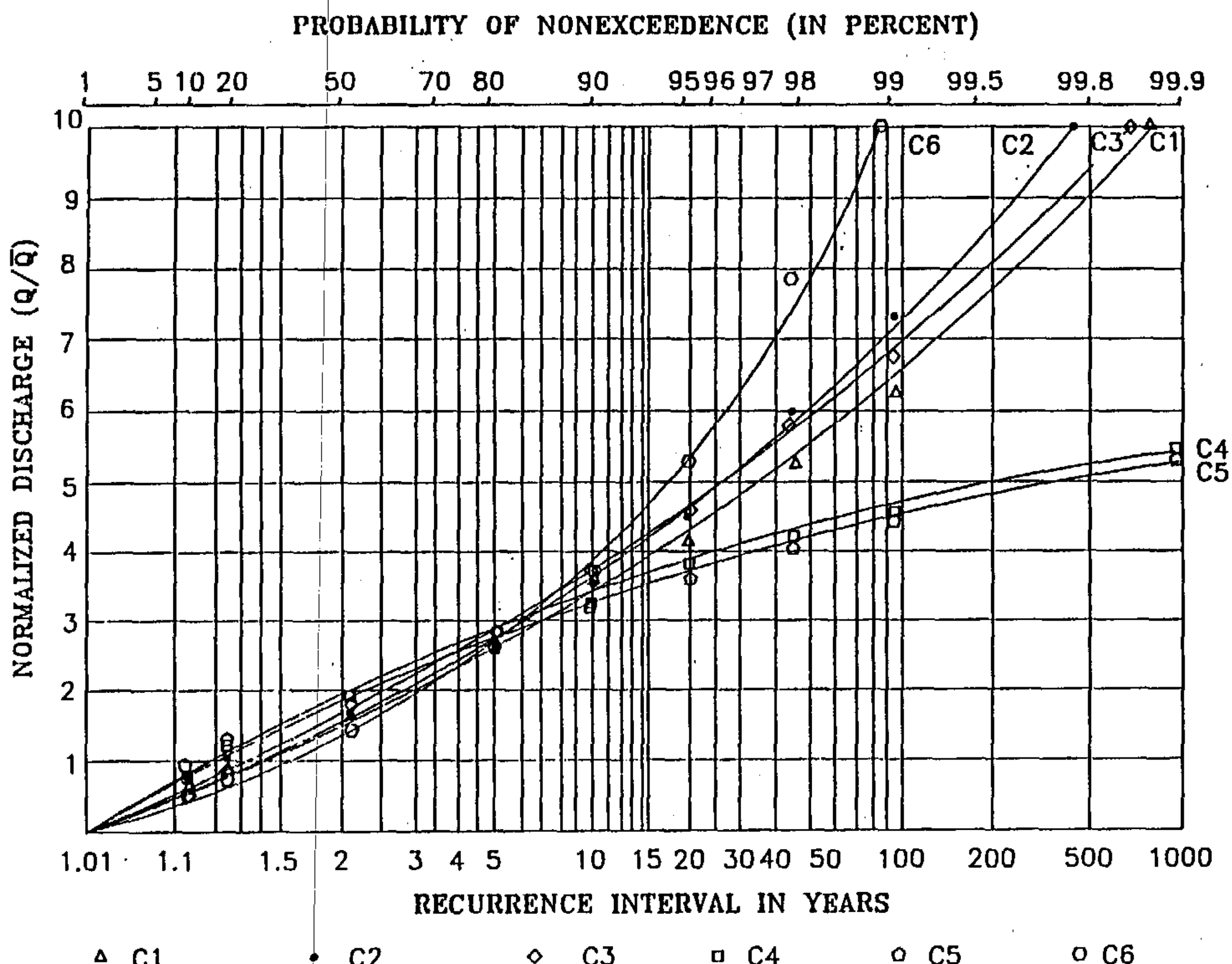

Fig. 3.21. GEV Regionalized Flood Frequency Growth Curves for Cluster Regions (Case 4: Clustering Variables GEV Parameters and QSP)

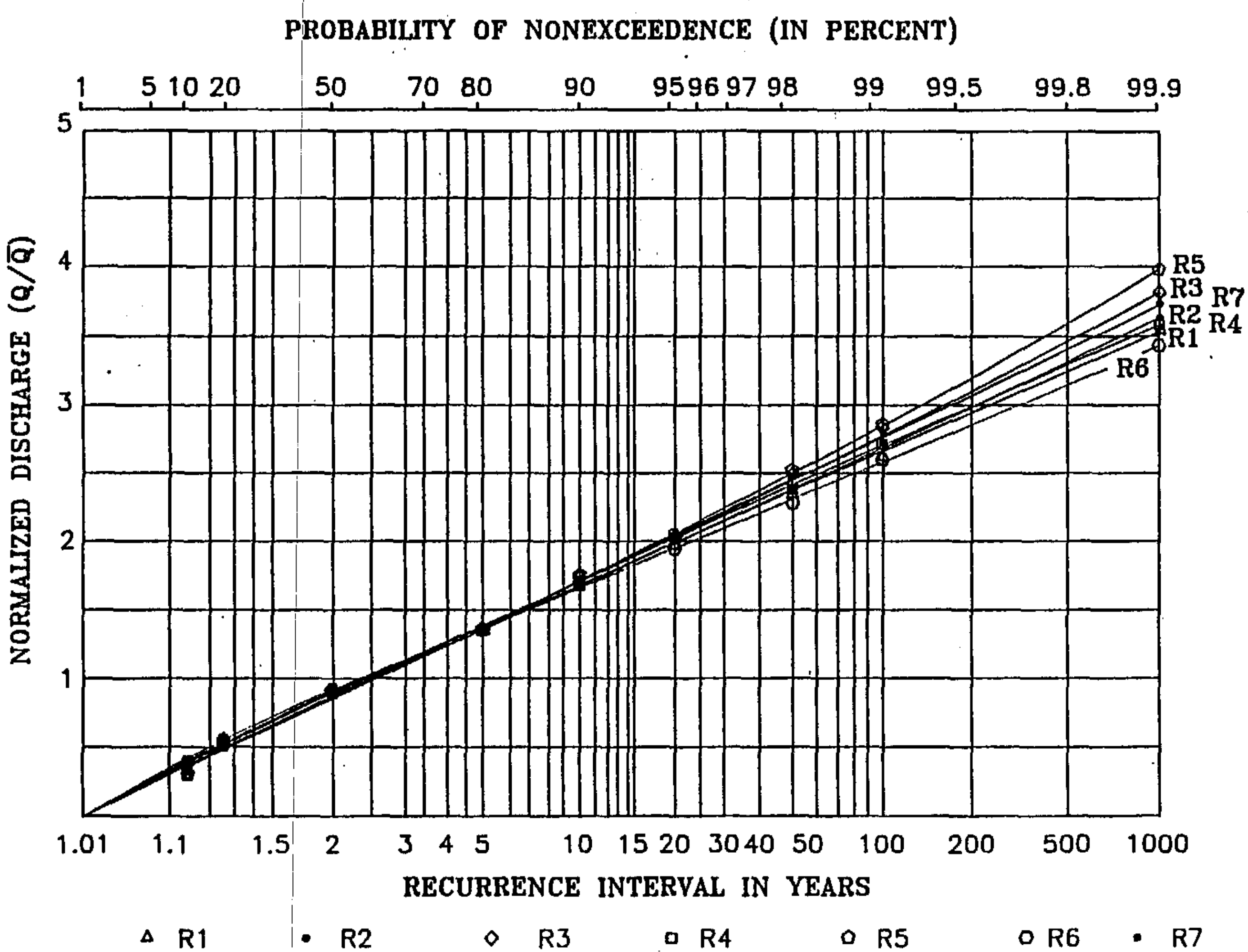

Fig. 3.22. EV1 Regionalized Flood Frequency Growth Curves for USGS Regions

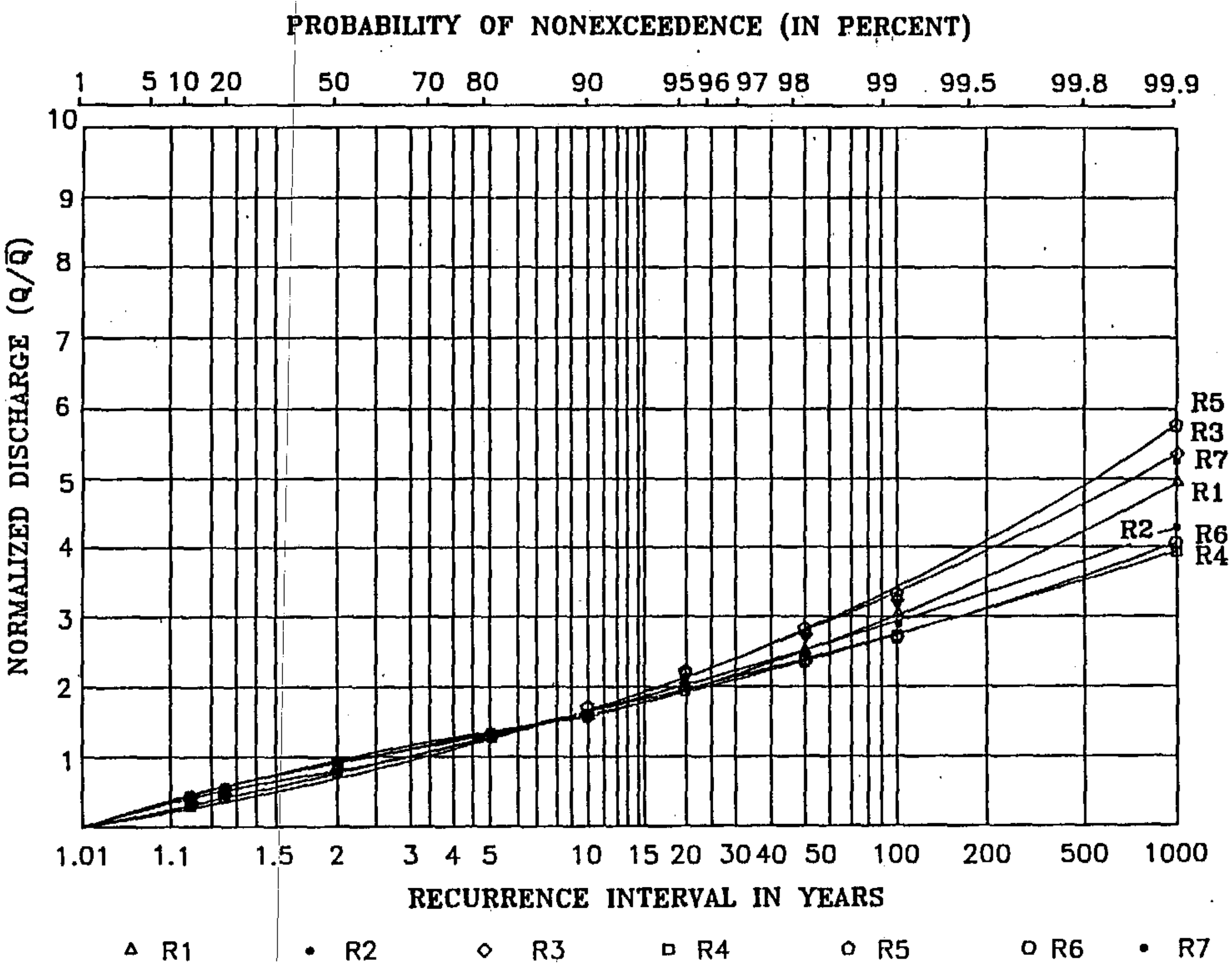

Fig. 3.23. GEV Regionalized Flood Frequency Growth Curves for USGS Regions

VERFICATION AND COMPARISON OF CLUSTER AND USGS FLOOD REGIONS

a) Hydrologic Characteristics of Flood Regions: The presence of a high degree heterogeneity (or the lack of homogeneity) in the flood characteristics between gauged sites within flood regions, as measured by important statistical properties of the AMF series observed at the site, can adversely affect the benefits derived from regionalization. Ideally, one would like to delineate flood regions that are homogeneous within themselves but distinct from others. As pointed out by Lettenmaier et al (1987), an implicit assumption of most index-flood methods of regionalization, similar to the one used in this study, is that the regions are homogeneous. This would imply that statistical moment ratios of the AMF series, like the coefficient of variation, LCV or CV (both measure the scale of a flood frequency distribution and are closely related), are identical at each of the gaged sites within a region. In reality this will never be the case. With this in view, Lettenmaier et al (1987) examined the effects of heterogeneity of the coefficient of variation on various flood regionalization schemes in conjunction with several parent flood probability distribution. They observed that the advantage of using any regionalization method is reduced for large values of regional average mean coefficient of variation, M(CV), and the range, R(CV), of the values of the coefficient of variation of flood data at each of the gaged sites within the region. Thus, these and similar studies clearly indicate the importance of observing the statistical trends of variables controlling flood response within flood regions.

With the above discussion in mind, statistical trends of important hydrologic characteristics, as measured each of the gauged sites within each of the cluster and USGS regions, are developed and examined in detail. For the four clustering cases (Case 1-Case 4), these trends are illustrated in Tables 3.14-3.17. Specifically, trends in the mean, median, maximum, minimum and range statistics of clustering variables (L-moments, QSP and parameters), watershed physical characteristics and other hydrologic variables are included in these tables. The cluster regions in each table are arranged in the order of increasing steepness (i.e. increasing coefficients of variation, LCV and/or skewness, LSK) of the regionalized flood frequency growth curve representing each region (refer to previous section).

The trends in the mean and median values of the clustering variables like the L-moments, parameters of the probability distribution and QSP are quite obvious since cluster analysis will group these variables into regions having small, medium to large values. For instance, Table 3.14(a) shows a clear and distinct mean and median values of regional average L-moment ratios (LCV, LSK and LKUR) and the conventional method of moment ratios (CV, SK and KUR) when clustering with LCV and QSP (Case 1). A similar trend is observed for clustering Cases 2-4 as well.

Table 3.19 shows the variation of the regional median values of the coefficient of variation, M(LCV) / M(CV), including its range within each region, R(LCV) / R(CV). The median value of the coefficient of variation, M(LCV), varies from 0.241-0.434 over the five cluster regions for Case 1 and 0.228-0.467 for Case 2. The trend in the median coefficient of variation, M(CV), (as estimated from the method of moments) varies from 0.438-0.936 for Case 1 and 0.375-1.035 for Case 2. A comparison of M(LCV) and M(CV) for other clustering cases shows similar variation. For all cases, M(LCV) and M(CV) are less than 0.467 and 1.035, respectively. The range in the regional median coefficients of variation, R(LCV) and R(CV), vary from 0.087-0.201 and 0.181-0.725, respectively, over all clustering cases. Thus, each cluster region is fairly homogeneous with respect to the variation of the regional median coefficient of variation. The differences of all regional mean and median L-moments (LCV and LSK in particular), ranging from small to large, make the cluster regions distinct from one another. It is for this reason, as discussed in the previous section, the cluster regions delineated for the four cases in this study are each associated with a distinct regionalized flood frequency growth curve.

Variation in the mean and median values of other physical characteristics, as illustrated in Tables 3.14(b)-3.17(b), suggests that cluster regions for all four cases (Cases 1-4) are grouped into areas having low, medium or high mean annual flood response. Since drainage area, A_c, is highly correlated with the mean annual flood, $\bar{Q}$, it follows a similar trend. Thus, the flood regions delineated have either predominantly small, medium or large watersheds. It is interesting to see that the clustering variable QSP (the specific mean annual flood) shows a reverse trend since it decreases with increasing watershed size. In other words small watersheds tend to generate a greater magnitude of direct runoff per unit area than do larger watersheds. The trends in main channel length, L_c, and slope, S_c, and watershed or basin length, B_c, and slope, B_s, show similar trends as the watershed drainage area, A_c, since they are directly proportional to it. Finally, the watershed shape index, B_s, and main channel sinuousity, S_s, do not show a significant trend between cluster regions for obvious reasons. These two dimensionless variables are ratios of quantities having similar magnitudes, either small or large.

An examination of the maximum and minimum values (range is the difference) of all the hydrologic variables (refer to the third and fourth rows of Tables 3.14(b)-3.17(b) for each cluster region) shows some overlap between cluster regions. For example, cluster region 3 for Case 1 (refer to Table 3.14(b)) contains generally the larger watersheds (a mean and median of 203.6 and 104.0 square miles, respectively) with a maximum watershed size of 960.0 square miles.

However, a minimum watershed size of 0.2 square miles indicates the presence of some small watersheds as well. Since flood response is not entirely a function of watershed size but depends on other physical and climatic factors, these small watersheds are incorporated in cluster region 3 because of the small coefficient of variation, LCV, associated with the floods produced. The presence of this overlap between cluster regions is one of the reasons why the ability to discriminate between them based on physical attributes is not very high. This is demonstrated later in section (c).

Table 3.18 shows the trends in the hydrological characteristics of USGS regions and is used to compare similar variables between cluster and the the USGS regions. A noticeable difference exists in the variation of M(LCV) and M(CV) between USGS regions and cluster regions (refer to Table 3.19). For example, the regional median coefficients of variation, M(LCV) and M(CV), are quite uniform (varying from 0.248-0.321 and 0.443-0.617, respectively) between the USGS regions. However, these regions have a larger range values of the median coefficient of variation, R(CV), when compared to cluster regions indicating a diversity of watersheds (small to large LCV and CV) contained **within** each region.

An examination of the mean and median values of the contributing drainage area, A_c, (associated with each of the gauged sites within a region) suggests that the USGS regions have fairly uniform distribution of small to large watersheds within their regions. A similar trend is observed with the mean annual flood, $\bar{Q}$, since this variable is highly correlated with the contributing drainage area. In contrast, cluster analysis tends to produce regions that have either predominantly small, medium or large watersheds (refer to Tables 3.14(b)-3.17(b)). In this context, it must be emphasized that small watersheds having low mean annual flood, Q, are, generally, associated with high LCV and/or

LSK values while using clustering schemes that included the latter variables as clustering variables.

The distribution of the mean and median values of watershed characteristics such as main channel sinuousity, S_s, and basin shape, both of which involve ratios of similar magnitudes (i.e. either small or large), show similar differences between cluster and USGS regions. Main channel and basin length follow the same trend as the contributing drainage area, A_c, since these variables are highly correlated to it.

An examination of the ranges of the median values of the hydrologic characteristics discussed above indicates that, with the exception of main channel sinuousity, S_s, and basin shape index, B_s (which remain similar for reasons stated in the previous paragraph), the hydrologic characteristics across all cluster regions show more variability than the USGS regions. This is particularly an important asset for discriminating between regions, and, as illustrated later, is the main reason why the USGS regions can not be discriminated easily.

TABLE 3.14(a). Comparison of Important Statistics of Regional Moment Ratios Using L-Moments and Conventional Method of Moments: Clustering with LCV and QSP **

Region No.	No. of Sites	L-Moments* LCV	LSK	LKUR	Method of Moments CV	SK	KUR#
Mean / median / max / min / range							
3	93	0.238	0.176	0.181	0.421	0.811	1.628
		0.241	0.154	0.181	0.438	0.721	0.526
		0.282	0.408	0.468	0.709	4.689	26.467
		0.131	-0.130	0.038	0.202	-0.950	-1.636
		0.151	0.538	0.430	0.507	5.639	28.103
4	30	0.282	0.199	0.184	0.489	0.606	0.520
		0.283	0.148	0.161	0.485	0.428	-0.027
		0.334	0.435	0.317	0.673	2.511	7.464
		0.221	-0.072	-0.021	0.366	-0.728	-2.348
		0.113	0.507	0.338	0.307	3.239	9.812
1	89	0.324	0.276	0.191	0.623	1.390	2.556
		0.318	0.296	0.182	0.618	1.274	1.615
		0.408	0.540	0.384	0.908	3.902	19.933
		0.285	-0.056	-0.008	0.486	-0.455	-1.595
		0.123	0.596	0.392	0.422	4.357	21.528
2	16	0.386	0.306	0.190	0.738	1.182	1.543
		0.386	0.280	0.169	0.701	1.122	0.412
		0.467	0.473	0.500	0.941	3.136	10.954
		0.338	0.102	0.021	0.612	0.131	-1.424
		0.129	0.371	0.479	0.329	3.005	12.378
5	25	0.443	0.404	0.278	0.962	2.077	5.335
		0.434	0.405	0.304	0.936	2.155	5.122
		0.530	0.614	0.475	1.376	3.551	14.323
		0.401	0.190	0.045	0.732	0.471	-0.845
		0.129	0.424	0.430	0.644	3.080	15.168

** Regions arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK)

* Regional averages of L-moment ratios are weighted by the number of years of record at each site within each region. Conventional moment ratios are simple arithmetic averages.

\# The coefficient of kurtosis, KUR, is computed relative to the normal probability distribution which has a KUR = 3.0. Therefore, observed kurtosis is obtained by adding a value of 3.0.

TABLE 3.14(b). Comparison of Important Statistics of Regional Hydrologic Characteristics: Clustering with LCV and QSP **

Reg. No.	Q (cfs)	A_c (sq. mi.)	QSP (csm)	B_s	B_l (mi)	L_c (mi)	S_c (%)	S_s	N (yrs)
Mean / Median / max / min									
3	8539.6	203.6	86.3	2.4	18.7	33.8	0.59	1.6	29.2
	5858.4	104.0	61.6	2.3	16.8	24.9	0.22	1.5	30
	31384.4	960.0	253.1	5.8	66.0	106.9	8.28	3.1	68
	35.2	0.2	14.6	0.2	0.8	1.1	0.05	1.0	7
4	489.1	1.2	430.6	2.3	1.5	1.7	1.78	1.2	13.1
	277.8	0.8	410.3	2.0	1.4	1.7	1.47	1.1	10
	2377.1	5.6	892.7	6.1	3.9	4.4	4.66	1.9	34
	48.6	0.1	275.8	0.9	0.5	0.6	0.53	1.0	7
1	6836.6	152.2	92.4	2.3	15.6	26.2	0.66	1.5	27.1
	4752.0	65.8	73.2	2.2	13.2	19.0	0.35	1.5	27
	27598.9	936.0	251.0	6.3	56.2	102.5	3.83	2.8	63
	40.8	0.2	18.0	0.7	0.6	0.7	0.04	1.0	7
2	414.1	1.1	448.5	1.5	1.4	1.8	2.53	1.3	13.4
	220.4	0.6	384.7	1.4	0.9	1.2	2.14	1.3	10.5
	2699.3	7.8	823.9	2.4	7.2	9.2	6.69	2.0	30
	107.1	0.1	290.9	0.6	0.4	0.6	0.20	1.1	8
5	2301.9	26.1	174.6	2.0	5.0	8.0	2.07	1.3	17.4
	382.4	1.8	163.7	1.9	2.4	2.9	1.40	1.2	15
	16781.4	246.0	368.4	3.6	23.1	58.7	9.66	2.5	32
	67.1	0.6	57.5	0.3	0.5	0.7	0.10	1.0	9

** Regions arranged in increasing steepness of the corre-sponding flood frequency growth curves (i.e. increasing LCV or LSKEW).

TABLE 3.15(a). Comparison of Important Statistics of Regional Moment Ratios Using L-Moments and Conventional Method of Moments: Clustering with EV1 Parameters and QSP **

		L-Moments*			Method of Moments		
Region No:	No: of Sites	LCV	LSK	LKUR	CV	SK	KUR♯
Mean / Median / max / min							
4	79	0.226	0.142	0.182	0.405	0.754	1.537
		0.232	0.143	0.181	0.412	0.695	0.422
		0.268	0.408	0.468	0.709	4.689	21.467
		0.131	-0.130	0.038	0.202	-0.950	-1.636
		0.137	0.538	0.430	0.507	5.639	23.103
5	30	0.276	0.165	0.161	0.485	0.591	0.455
		0.277	0.147	0.161	0.482	0.380	-0.060
		0.334	0.435	0.317	0.673	2.511	7.464
		0.221	-0.072	-0.021	0.366	-0.728	-2.348
		0.113	0.507	0.338	0.307	3.239	9.812
3	91	0.314	0.260	0.182	0.592	1.340	2.542
		0.310	0.262	0.179	0.586	1.265	1.615
		0.374	0.540	0.384	0.852	3.902	19.933
		0.270	-0.056	-0.008	0.446	-0.455	-1.595
		0.104	0.596	0.392	0.406	4.357	21.528
2	10	0.374	0.288	0.206	0.716	1.286	1.963
		0.379	0.280	0.205	0.701	1.133	1.089
		0.426	0.473	0.500	0.928	3.136	10.954
		0.306	0.102	0.021	0.547	0.131	-1.424
		0.120	0.371	0.479	0.381	3.005	12.378
1	43	0.425	0.370	0.237	0.872	1.768	3.873
		0.410	0.366	0.212	0.808	1.662	2.887
		0.530	0.614	0.475	1.376	3.551	14.323
		0.371	0.143	0.033	0.651	0.306	-1.285
		0.159	0.471	0.442	0.725	3.245	15.608

** Regions arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK).

* Regional averages of L-moment ratios are weighted by the number of years of record at each site within each region. Conventional moment ratios are simple arithmetic averages.

♯ The coefficient of kurtosis, KUR, is computed relative to the normal probability distribution which has a KUR = 3.0. Therefore, observed kurtosis is obtained by adding a value of 3.0.

TABLE 3.15(b). Comparison of Important Statistics of Regional Hydrologic Characteristics: Clustering with EV1 Parameters and QSP **

Reg. No.	Q (cfs)	A_c (sq. mi.)	QSP (csm)	B_s	B_1 (mi)	L_c (mi)	S_c (%)	S_s	N (yrs)
Mean / Median / max / min									
4	8518.7	211.0	85.3	2.4	18.8	35.0	0.56	1.7	29.0
	5858.4	104.0	58.0	2.3	16.7	24.4	0.17	1.6	29
	31384.4	960.0	253.1	5.8	66.0	106.9	8.28	3.1	68
	35.2	0.3	14.6	0.2	0.8	1.1	0.05	1.0	7
5	468.5	1.1	416.8	2.4	1.5	1.7	1.75	1.2	12.5
	262.4	0.7	398.5	2.1	1.4	1.7	1.47	1.1	10
	2377.1	1.0	892.6	6.1	3.9	4.4	4.66	1.5	34
	38.6	0.1	241.3	1.1	0.5	0.6	0.53	1.0	7
3	7201.1	153.2	89.3	2.4	16.1	26.6	0.70	1.5	28.6
	4834.5	82.3	63.9	2.2	14.2	21.5	0.32	1.5	29
	27598.9	836.0	251.0	6.3	50.4	102.5	3.84	2.8	63
	40.8	0.2	18.0	0.7	0.6	0.7	0.04	1.0	7
2	254.6	0.5	548.0	1.2	0.7	1.1	2.53	1.4	13.3
	210.7	0.4	522.5	1.1	0.7	0.9	2.44	1.3	10.5
	655.3	1.0	823.9	2.4	1.0	1.9	5.87	2.0	26
	107.1	0.1	377.6	0.6	0.4	0.6	0.20	1.1	8
1	3176.9	59.1	182.6	2.0	7.3	11.8	1.72	1.3	17.9
	537.1	0.5	163.7	1.9	3.0	3.7	1.00	1.3	15
	25299.4	936.0	378.8	3.6	56.2	99.6	9.66	2.5	44
	67.1	0.5	27.0	0.3	0.5	0.7	0.08	1.0	8

** Regions arranged in increasing steepness of the corre-sponding flood frequency growth curves (i.e. increasing LCV or LSKEW).

TABLE 3.16(a). Comparison of Important Statistics of Regional Moment Ratios Using L-Moments and Conventional Method of Moments: Clustering with LCV, LSK and QSP **

Region No:	No: of Sites	L-Moments* LCV	L-Moments* LSK	L-Moments* LKUR	Method of Moments CV	Method of Moments SK	Method of Moments KUR#
Mean / median / max / min / range							
4	44	0.223	0.053	0.132	0.380	0.071	-0.307
		0.228	0.058	0.135	0.375	0.131	-0.388
		0.300	0.190	0.380	0.516	1.346	3.917
		0.131	-0.130	-0.021	0.202	-0.950	-2.348
		0.169	0.320	0.401	0.718	2.296	6.265
3	75	0.261	0.201	0.171	0.476	1.100	1.853
		0.258	0.204	0.183	0.478	1.045	1.313
		0.350	0.345	0.468	0.631	4.423	22.974
		0.180	0.076	-0.001	0.316	0.095	-1.595
		0.170	0.269	0.469	0.315	4.328	24.569
5	38	0.322	0.242	0.185	0.578	0.836	0.936
		0.306	0.222	0.165	0.566	0.807	0.197
		0.426	0.473	0.500	0.928	3.136	10.954
		0.225	-0.056	0.021	0.369	-0.582	-1.571
		0.201	0.529	0.479	0.559	3.718	12.525
1	79	0.338	0.319	0.218	0.669	1.640	3.508
		0.345	0.323	0.197	0.640	1.485	1.903
		0.431	0.540	0.384	0.968	4.689	26.467
		0.249	0.134	0.045	0.502	0.311	-0.958
		0.182	0.406	0.339	0.466	4.378	27.425
2	17	0.462	0.467	0.328	1.054	2.423	6.724
		0.467	0.472	0.343	1.035	2.531	6.773
		0.530	0.614	0.475	1.376	3.551	14.323
		0.393	0.312	0.136	0.824	1.024	-0.483
		0.137	0.302	0.339	0.552	2.527	14.806

** Regions arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK).

* Regional averages of L-moment ratios are weighted by the number of years of record at each site within each region. Conventional moment ratios are simple arithmetic averages.

\# The coefficient of kurtosis, KUR, is computed relative to the normal probability distribution which has a KUR = 3.0. Therefore, observed kurtosis is obtained by adding a value of 3.0.

TABLE 3.16(b). Comparison of Important Statistics of Regional Hydrologic Characteristics: Clustering with LCV, LSKEW and QSP **

Reg. No.	Q (cfs)	A_c (sq. mi.)	QSP (csm)	B_s	B_1 (mi)	L_c (mi)	S_c (%)	S_s	N (yrs)
Mean / Median / max / min									
4	4355.1	97.8	147.5	2.3	10.8	19.9	1.00	1.5	19.9
	2198.8	17.6	152.2	2.0	7.1	9.8	0.37	1.3	15
	24353.3	745.0	373.5	5.8	37.4	92.6	8.28	2.7	68
	35.2	0.1	22.0	0.2	0.5	0.6	0.05	1.0	7
3	10748.2	265.2	48.6	2.5	23.2	40.9	0.38	1.7	34.2
	8958.5	235.0	47.8	2.3	22.5	38.6	0.16	1.6	34
	31384.4	960.0	222.8	5.5	66.0	106.9	2.95	3.1	63
	198.5	1.1	14.6	0.6	1.7	1.9	0.05	1.0	8
5	427.5	1.0	460.6	2.1	1.3	1.5	2.03	1.2	12.9
	254.1	0.6	423.9	1.8	1.0	1.3	1.88	1.1	10
	24353.3	5.6	892.6	6.1	3.9	4.4	5.87	2.0	34
	35.2	0.1	275.8	0.6	0.4	0.6	0.20	1.0	7
1	5456.6	108.7	108.5	2.2	12.1	20.1	1.02	1.5	24.9
	3260.0	40.9	84.9	2.0	8.2	13.2	0.45	1.4	24
	25299.4	936.0	301.8	6.3	56.2	102.5	9.66	2.8	58
	40.8	0.2	18.4	0.7	0.6	0.7	0.04	1.0	7
2	2143.2	21.9	220.5	2.0	5.1	8.7	1.71	1.4	16.4
	382.4	1.6	236.0	2.0	3.5	4.2	1.40	1.4	13
	16446.3	246.0	378.8	3.6	23.1	58.7	6.69	2.8	32
	117.0	0.6	61.8	0.3	0.5	0.7	0.10	1.0	9

** Regions arranged in increasing steepness of the corre-sponding flood frequency growth curves (i.e. increasing LCV or LSKEW).

TABLE 3.17(a). Comparison of Important Statistics of Regional Moment Ratios Using L-Moments and Conventional Method of Moments: Clustering with GEV Parameters and QSP **

Region No.	No. of Sites	L-Moments* LCV	L-Moments* LSK	L-Moments* LKUR	Method of Moments CV	Method of Moments SK	Method of Moments KUR#
Mean / median / max / min / range							
5	68	0.234	0.086	0.142	0.404	0.330	0.071
		0.241	0.111	0.141	0.416	0.356	-0.180
		0.300	0.190	0.400	0.537	1.860	7.104
		0.131	-0.130	-0.008	0.202	-0.950	-2.013
		0.169	0.320	0.408	0.335	2.810	9.117
4	15	0.271	0.067	0.119	0.455	0.027	-0.652
		0.273	0.052	0.107	0.448	-0.031	-0.672
		0.308	0.170	0.277	0.547	0.813	0.853
		0.221	-0.072	-0.021	0.366	-0.728	-2.348
		0.087	0.242	0.298	0.181	1.541	3.201
1	81	0.286	0.292	0.222	0.552	1.459	3.891
		0.293	0.286	0.211	0.563	1.425	2.260
		0.368	0.479	0.468	0.852	4.689	26.467
		0.180	0.190	0.100	0.316	0.695	-0.521
		0.188	0.289	0.368	0.536	3.994	26.988
2	21	0.332	0.311	0.229	0.631	1.421	2.289
		0.318	0.297	0.229	0.612	1.275	1.190
		0.426	0.473	0.500	0.928	3.136	10.954
		0.233	0.205	0.036	0.412	0.488	-1.571
		0.193	0.268	0.464	0.516	2.648	12.525
3	40	0.363	0.219	0.127	0.657	0.913	0.595
		0.356	0.224	0.135	0.653	0.955	0.310
		0.434	0.328	0.223	0.826	1.856	4.772
		0.275	0.076	-0.001	0.506	0.095	-1.595
		0.159	0.252	0.224	0.320	1.761	6.367
6	28	0.433	0.453	0.311	0.950	2.306	6.087
		0.432	0.434	0.316	0.930	2.321	5.759
		0.530	0.614	0.475	1.376	3.551	14.323
		0.337	0.346	0.160	0.698	1.233	0.346
		0.193	0.268	0.315	0.678	2.318	14.669

** Regions arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK).
* Regional averages of L-moment ratios are weighted by the number of years of record at each site within each region. Conventional moment ratios are simple arithmetic averages.
\# The coefficient of kurtosis, KUR, is computed relative to the normal probability distribution which has a KUR = 3.0. Therefore, observed kurtosis is obtained by adding a value of 3.0.

TABLE 3.17(b). Comparison of Important Statistics of Regional Hydrologic Characteristics: Clustering with GEV Parameters and QSP **

Reg. No.	Q (cfs)	A_c (sq. mi.)	QSP (csm)	B_s	B_l (mi)	L_c (mi)	S_c (%)	S_s	N (yrs)
Mean / Median / max / min									
5	8184.3	208.1	102.5	2.3	18.0	34.4	0.66	1.7	26.5
	5609.9	80.9	70.4	2.1	16.1	23.5	0.24	1.5	24
	27598.9	836.0	302.1	5.8	54.8	99.2	8.28	3.1	68
	35.2	0.2	14.6	0.2	0.8	1.0	0.04	1.0	7
4	347.9	0.7	485.5	2.4	1.2	1.4	2.14	1.2	10.8
	251.4	0.7	461.4	2.1	1.0	1.4	1.99	1.1	10
	1167.0	2.5	892.6	6.1	1.9	2.5	4.66	1.9	26
	48.6	0.1	302.9	0.9	0.5	0.6	1.00	1.0	7
1	7585.7	162.2	81.1	2.5	17.4	29.0	0.60	1.6	30.6
	5093.6	104.0	53.3	2.4	16.0	23.4	0.21	1.6	31
	31384.4	960.0	251.0	6.3	66.0	106.9	3.57	2.8	58
	47.3	0.2	15.8	0.7	0.6	0.7	0.04	1.0	7
2	499.3	1.3	453.9	1.9	1.4	1.7	1.91	1.2	14.6
	256.8	0.6	422.2	1.7	1.1	1.3	1.51	1.1	10
	2377.1	5.6	823.9	4.4	3.9	4.4	5.87	2.0	34
	102.0	0.1	275.8	0.6	0.4	0.6	0.20	1.0	8
3	3782.5	74.4	146.2	2.1	9.4	14.8	1.33	1.4	23.5
	1245.0	13.1	114.5	2.0	5.3	6.1	0.62	1.3	23.5
	5026.1	729.0	401.5	4.3	45.8	102.5	9.66	2.2	63
	40.8	0.2	18.0	0.6	0.7	0.8	0.05	1.1	9
6	5125.2	103.1	161.0	2.1	10.3	17.0	1.45	1.4	17.4
	1391.2	7.8	134.3	2.0	4.6	5.6	0.75	1.3	13
	25299.4	936.0	378.8	3.6	56.2	99.6	6.69	2.5	46
	117.0	0.6	27.0	0.3	0.5	0.7	0.07	1.0	8

** Regions arranged in increasing steepness of the corre-sponding flood frequency growth curves (i.e. increasing LCV or LSKEW).

TABLE 3.18(a). Comparison of Important Statistics of Regional Moment Ratios Using L-Moments and Conventional Method of Moments for USGS Regions **

Region No:	No: of Sites	L-Moments* LCV	LSK	LKUR	Method of Moments CV	SK	KUR#
Mean / Median / max / min							
6	31	0.270	0.223	0.187	0.511	0.805	1.264
		0.248	0.166	0.189	0.443	0.887	0.328
		0.475	0.560	0.384	1.215	3.551	14.323
		0.131	-0.130	0.021	0.202	-0.950	-1.424
		0.344	0.690	0.363	1.013	4.501	15.747
1	32	0.278	0.273	0.214	0.589	1.184	2.197
		0.286	0.248	0.209	0.524	1.205	1.164
		0.514	0.571	0.475	1.246	3.310	11.739
		0.161	-0.051	0.043	0.260	-0.412	-1.488
		0.353	0.622	0.432	0.969	3.722	13.227
4	20	0.283	0.204	0.158	0.520	0.765	1.201
		0.278	0.195	0.153	0.493	0.868	0.690
		0.411	0.377	0.279	0.794	1.949	5.492
		0.194	-0.085	0.062	0.331	-0.755	-1.217
		0.217	0.462	0.217	0.463	2.704	6.709
2	40	0.289	0.227	0.199	0.559	1.170	2.368
		0.293	0.220	0.194	0.521	1.076	1.368
		0.525	0.488	0.384	1.228	3.902	19.933
		0.165	-0.072	0.023	0.258	-0.728	-2.013
		0.360	0.560	0.361	0.970	4.630	21.946
7	38	0.303	0.269	0.203	0.581	1.229	2.346
		0.307	0.265	0.186	0.573	1.138	0.946
		0.441	0.473	0.500	0.936	3.185	15.081
		0.180	-0.056	0.027	0.314	-0.582	-1.454
		0.261	0.529	0.527	0.622	3.767	16.535
3	26	0.312	0.270	0.183	0.602	1.147	1.689
		0.321	0.257	0.160	0.614	1.042	0.638
		0.492	0.540	0.343	1.110	2.634	10.443
		0.221	-0.001	-0.021	0.356	-0.456	-2.348
		0.271	0.541	0.364	1.456	3.090	12.790
5	38	0.319	0.285	0.206	0.640	1.418	3.289
		0.301	0.270	0.179	0.617	1.227	1.258
		0.530	0.614	0.468	1.376	4.689	26.467
		0.160	-0.056	-0.008	0.245	-0.455	-1.584
		0.370	0.670	0.476	1.131	5.144	28.051

** Regions delineated using method of residuals.
* Regional averages of L-moment ratios are weighted by the number of years of record at each site within each region. Conventional moment ratios are simple arithmetic averages.
\# The coefficient of kurtosis, KUR, is computed relative to the normal probability distribution which has a KUR = 3.0. Therefore, observed kurtosis is obtained by adding a value of 3.0.

TABLE 3.18(b). Comparison of Important Statistics of Regional Hydrologic Characteristics of U.S.G.S. Regions **

Reg. No.	$\bar{Q}$ (cfs)	A_c (sq. mi.)	QSP (csm)	B_s	B_l (mi)	L_c (mi)	S_c (%)	S_s	N (yrs)
Mean / Median / max / min / range									
6	4056.5	122.8	181.8	2.0	13.8	26.8	1.00	1.6	19.3
	2458.5	30.8	96.0	1.8	5.5	16.3	0.26	1.4	15
	13518.3	637.0	703.3	3.6	34.3	91.4	5.87	2.9	59
	55.2	0.2	15.8	0.8	0.4	0.7	0.05	1.0	9
1	3886.8	60.3	219.7	2.5	8.2	14.8	0.95	1.4	23.7
	1257.6	6.1	188.2	2.3	4.3	5.5	0.61	1.3	17
	23214.5	476.0	541.5	6.1	33.9	80.8	3.90	2.6	58
	38.6	0.2	36.8	0.6	1.0	1.0	0.07	1.0	9
4	8636.6	216.0	79.8	2.2	16.2	27.0	1.36	1.5	28.1
	3985.6	57.2	70.5	1.6	9.5	12.5	0.35	1.5	29.5
	31364.4	960.0	189.9	5.8	66.0	106.9	6.49	2.5	49
	35.2	0.3	30.0	0.9	0.8	1.0	0.07	1.0	9
2	6520.2	175.2	90.6	2.4	15.9	27.2	1.03	1.5	26.2
	4010.1	66.1	58.1	2.4	13.2	18.6	0.39	1.4	27
	27598.9	936.0	326.0	6.3	56.2	99.6	9.66	3.1	68
	40.8	0.2	14.6	0.2	0.5	0.7	0.05	1.0	7
7	3364.5	62.8	275.4	2.2	10.1	16.0	0.71	1.4	24.4
	1861.1	9.1	187.8	2.1	5.5	6.7	0.39	1.3	19
	16214.4	309.0	892.6	5.5	34.3	66.9	2.75	2.2	63
	107.1	0.1	18.0	0.6	0.4	0.6	0.04	1.0	7
3	9088.3	179.3	118.0	2.3	16.3	29.3	1.00	1.6	22.2
	5143.0	65.6	80.6	2.0	12.6	18.7	0.44	1.6	17
	27339.6	721.0	373.5	4.7	50.4	95.0	3.83	2.5	54
	48.6	0.1	32.1	1.1	0.5	0.6	0.08	1.0	7
5	6607.9	109.8	177.8	2.1	12.1	20.7	1.26	1.5	25.6
	2979.8	20.4	167.0	2.0	7.9	10.2	0.50	1.3	24.5
	27219.2	729.0	461.4	3.9	45.8	91.4	8.28	2.8	52
	91.7	0.2	27.9	0.7	0.6	0.7	0.07	1.1	8

** Regions arranged in increasing steepness of the corre-sponding flood frequency growth curves (i.e. increasing LCV or LSKEW).

TABLE 3.19 Comparison of Regional Mean Coefficients of Variation and Their Ranges Within Cluster and USGS Flood Regions

Region No:	No. of Sites	L-Moments# M(LCV)	R(LCV)	R^*(LCV)	Method of Moments@ M(CV)	R(CV)	R^*(CV)
		Case 1: Clustering with LCV and OSP					
3	93	0.241	0.151	0.627	0.438	0.507	1.158
4	30	0.283	0.113	0.399	0.485	0.307	0.633
1	89	0.318	0.123	0.387	0.618	0.422	0.683
2	16	0.386	0.129	0.334	0.701	0.329	0.469
5	25	0.434	0.129	0.297	0.936	0.644	0.688
		Case 2: Clustering with LCV, LSK and OSP					
4	44	0.228	0.169	0.741	0.375	0.314	0.837
3	75	0.258	0.170	0.659	0.478	0.315	0.659
5	38	0.306	0.201	0.657	0.566	0.559	0.988
1	79	0.345	0.182	0.528	0.640	0.466	0.728
2	17	0.467	0.137	0.293	1.035	0.552	0.533
		Case 3: Clustering with Gumbel Parameters and OSP					
4	79	0.232	0.137	0.591	0.412	0.507	1.231
5	30	0.277	0.113	0.408	0.482	0.307	0.637
3	91	0.310	0.104	0.335	0.586	0.406	0.693
2	10	0.379	0.120	0.317	0.701	0.381	0.544
1	43	0.410	0.159	0.388	0.808	0.725	0.897
		Case 4: Clustering with GEV Parameters and OSP					
5	68	0.241	0.169	0.701	0.416	0.335	0.805
4	15	0.273	0.087	0.319	0.448	0.181	0.404
1	81	0.293	0.188	0.642	0.563	0.536	0.952
2	21	0.318	0.193	0.607	0.612	0.516	0.843
3	40	0.356	0.159	0.447	0.653	0.320	0.490
6	28	0.432	0.193	0.447	0.930	0.678	0.729
		USGS Regions					
6	31	0.248	0.344	1.387	0.443	1.013	2.287
1	32	0.286	0.353	1.234	0.524	0.986	1.882
4	20	0.278	0.217	0.781	0.493	0.463	0.939
2	68	0.293	0.360	1.229	0.521	0.970	1.862
7	38	0.307	0.261	0.850	0.573	0.622	1.086
3	26	0.321	0.271	0.844	0.614	0.754	1.228
5	38	0.301	0.370	1.229	0.617	1.131	1.833

\# MLCV is the regional median of LCV, RLCV is the range of LCV's for the region and R^*LCV is the normalized regional LCV (range/median).

@ MCV is the regional median of CV, RCV is the range of CV's for the region and R^*CV is the normalized regional CV (range/median).

b) Performance of Regionalized Flood Frequency Models: The performance of the regionalized flood frequency models is evaluated using the following specific criteria in conjunction with Monte Carlo simulation techniques:

1. The accuracy of the regional flood frequency model to predict the flood levels associated with different return periods as measured by the bias.
2. The precision (as reflected by the overall fit of the model to the flood data) of the flood frequency model as measured by the root mean square error (RMSE).

For each of the cluster regions delineated under the four clustering schemes (Cases 1-4), AMF data is synthetically generated at each of the gauged sites within the region using procedures discussed in Chapter 2. 100 sequences, each having a record length equal to the historic systematic flood record at the gauged site and drawn from both EV1 and GEV populations, are used in the analysis. The regional average L-moments and the corresponding parameters based on synthetically generated flows and the simulation runs compare well with the historical estimates for the flatter regionalized flood frequency growth curves as shown in Tables A.1-A.11, Appendix A. However, the simulated sequences tend to underestimate the higher order L-moments (like LSK and LKUR) with this difference increasing as the regionalized frequency growth curve gets steeper. As pointed below, this is one of the main reasons why the GEV distribution gives larger biases in the flood quantile estimates than the EV1 distribution. The inability of Monte Carlo simulated flood sequences to capture the larger variability associated with historical estimates of higher order moments, like the coefficient of skew, has been widely reported in literature and is referred to as the condition of separation (Matalas, 1975).

The average regional normalized bias and RMSE for select flood quantiles, as estimated using EV1 and GEV flood frequency models, are summarized for the four clustering cases in Tables 3.20-3.23. Similar results for the USGS regions are shown in Table 3.24. The following conclusions are made for the four clustering cases:

a) As expected, the bias and RMSE generally increase with the return period, T, and with the steepness of the regionalized flood frequency growth curve.
The bias changes from positive to negative as the growth curve becomes steeper and, hence, is not uniform across the cluster regions. This is true for both EV1 and GEV distributions over all return periods of interest (10-100 year). Consequently, flood quantiles are overestimated when the growth curves have small slopes and underestimated as the curves become steeper. In a recent study, Landwehr (1980) observed that if the population skew is different (larger or smaller) than the EV1 skew of 1.14, then an EV1 distribution would on the average underestimate the flood quantiles. In this study it appears to hold for a majority of flood regions (particularly those with steep frequency growth curves) indicating regionalized coefficient of skew other than the EV1 skew of 1.14 (refer to Tables 3.14(a)-3.17(a)).

b) The biases and RMSE for the EV1 flood frequency model are lower than the GEV model for all flood frequency growth curves. However, one would expect the GEV model to do better than the EV1 model, at least in terms of the bias, since it has an additional shape parameter to to better characterize the growth curves, in particular the steep ones. Such is not the case in this study.

The larger biases associated with the GEV are partly due to the condition of separation that exists when using Monte Carlo simulated flood data. In other words, the use of a three parameter distribution like the GEV may give larger biases than a more parsimonious distribution like the EV1 due the lower variability of the coefficient of skew and higher order moments observed in simulated flood sequences. Furthermore, as pointed by Wallis (1985), the GEV distribution while having a theoretical appeal for fitting flood data, the asymptotic properties on which it is founded may not be satisfied by the small number of independent flood events commonly encountered in practice.

c) An examination of cluster regions for all four clustering cases (Cases 1-4) indicates that the regional average bias associated with flood levels less than 100 years, ranges from -2.2% to 0.1% for the EV1 distribution and from -14.2% to 0.1% for the GEV distribution while the corresponding RMSE ranges from 9.2% to 21.8% and 9.2% to 43.9%, respectively. These levels are comparable to values reported in previous studies (for example refer to Lettenmaier et al, 1987).

d) Clustering on the parameters of the probability distribution, as opposed to the L-moments used to estimate them, reduces the bias and RMSE, nominally. This occurs inspite of the fact the shape of the growth curves is affected by the clustering variables used (refer to section on development of flood frequency growth curves).

For the USGS regions the biases and RMSE of the regionalized EV1 and GEV distribution are lower than the cluster regions. This is partly due to the relatively flat regionalized flood frequency growth curves associated with all the seven USGS regions. The regional average bias for all seven regions ranges from -0.9% to 0.1% for the EV1 distribution and -6.0% to 0.0% for the GEV distribution. These biases are usually negative at higher return periods (for example the 100 year) indicating an underestimation of flood levels. The RMSE ranges from 11.8% to 15.4% for the EV1 distribution and 12.2% to 29.8% for the GEV distribution. Also note that the bias and RMSE are fairly uniform across the seven USGS regions.

A regionalized log-Pearson Type-III distribution (based on L-moments) is not tested in this study since previous studies have clearly shown that EV1 and GEV outperform the log-Pearson Type-III distribution (Wallis, 1985) in estimating flood quantiles. However, since current practice continues to use this distribution, Tables 3.25-3.29 compare log-Pearson Type-III flood quantile estimates (from the USGS method of residuals study using WRC Bulletin 17-B) to the estimates of the EV1 and GEV distributions at select sites within cluster and USGS regions. These sites are chosen to represent gauged sites that have small to large watershed areas, low to high coefficient of variation and skewness associated with the flood data, and the number of years of systematic historic flood records range from 9 to 58 years. Since the true population flood quantile (for a given return period) is unknown, these tables merely serve the purpose of identifying whether flood quantiles are under or over estimated by the recommended regionalized flood frequency distributions in this study. An examination of these tables suggests that, with the exception of using the EV1 distribution at a few sites, flood quantiles are, generally, overestimated when using log-Pearson Type-III distribution.

TABLE 3.20. Regional Average Normalized Bias and Root Mean Square Error (RMSE) of Quantiles: Clustering with LCV and QSP *

				Quantiles	Probability	Region
Distribution	No:	10 yr.	20 yr.	50 yr.	100 yr.	1000 yr.
				Bias		
EV1	3	0.003	0.004	0.005	0.005	0.006
	4	-0.001	-0.002	-0.002	-0.002	-0.002
	1	-0.001	-0.003	-0.005	-0.007	-0.009
	2	-0.002	-0.010	-0.017	-0.021	-0.028
	5	0.016	0.002	-0.010	-0.017	-0.030
GEV	3	0.004	0.008	0.015	0.020	0.040
	4	0.002	0.008	0.018	0.026	0.061
	1	-0.003	-0.006	-0.009	-0.011	-0.016
	2	-0.009	-0.019	-0.028	-0.033	-0.040
	5	0.007	-0.020	-0.053	-0.078	-0.155
				RMSE		
EV1	3	0.094	0.095	0.095	0.095	0.095
	4	0.156	0.156	0.157	0.157	0.157
	1	0.136	0.136	0.136	0.136	0.136
	2	0.200	0.199	0.199	0.198	0.198
	5	0.200	0.197	0.195	0.195	0.194
GEV	3	0.095	0.096	0.098	0.100	0.113
	4	0.168	0.172	0.179	0.186	0.227
	1	0.162	0.162	0.163	0.164	0.171
	2	0.331	0.333	0.340	0.348	0.390
	5	0.434	0.429	0.428	0.430	0.452

* Regions arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK)

TABLE 3.21. Regional Average Normalized Bias and Root Mean Square Error (RMSE) of Quantiles: Clustering with EV1 Parameters and QSP *

Probability Distribution	Region No:	Quantiles 10 yr.	20 yr.	50 yr.	100 yr.	1000 yr.
				Bias		
EV1	4	0.002	0.003	0.004	0.004	0.005
	5	0.002	0.002	0.002	0.002	0.002
	3	0.000	-0.002	-0.004	-0.005	-0.007
	2	0.017	0.011	0.006	0.003	-0.003
	1	0.005	-0.007	-0.017	-0.022	-0.034
GEV	4	0.004	0.009	0.015	0.020	0.039
	5	-0.002	0.004	0.015	0.024	0.061
	3	-0.006	-0.009	-0.013	-0.015	-0.021
	2	-0.015	-0.027	-0.039	-0.047	-0.064
	1	-0.009	-0.028	-0.053	-0.071	-0.127
				RMSE		
EV1	4	0.092	0.092	0.092	0.092	0.093
	5	0.158	0.158	0.158	0.159	0.159
	3	0.128	0.127	0.127	0.127	0.127
	2	0.201	0.200	0.199	0.199	0.199
	1	0.193	0.191	0.190	0.189	0.189
GEV	4	0.092	0.094	0.096	0.099	0.111
	5	0.163	0.167	0.174	0.183	0.227
	3	0.146	0.146	0.147	0.148	0.155
	2	0.247	0.253	0.265	0.278	0.338
	1	0.385	0.382	0.380	0.381	0.397

* Regions arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK)

TABLE 3.22. Regional Average Normalized Bias and Root Mean Square Error (RMSE) of Quantiles: Clustering with LCV, LSKEW and QSP *

Probability Distribution	Region No:	10 yr.	20 yr.	50 yr.	100 yr.	1000 yr.
		Quantiles				
				Bias		
EV1	4	0.001	0.003	0.004	0.005	0.006
	3	0.002	0.002	0.002	0.002	0.002
	5	0.001	-0.002	-0.005	-0.007	-0.010
	1	0.001	-0.003	-0.006	-0.008	-0.012
	2	0.016	0.000	-0.015	-0.022	-0.038
GEV	4	0.014	0.026	0.043	0.056	0.097
	3	0.001	0.003	0.005	0.007	0.015
	1	-0.001	-0.008	-0.018	-0.025	-0.048
	5	-0.002	-0.003	-0.002	0.000	0.011
	2	-0.015	-0.054	-0.105	-0.142	-0.259
				RMSE		
EV1	4	0.111	0.111	0.111	0.112	0.112
	3	0.096	0.096	0.096	0.096	0.096
	5	0.172	0.171	0.171	0.171	0.171
	1	0.146	0.145	0.145	0.145	0.144
	2	0.218	0.214	0.212	0.211	0.210
GEV	4	0.105	0.109	0.116	0.124	0.154
	3	0.098	0.099	0.100	0.101	0.107
	1	0.188	0.187	0.187	0.188	0.196
	5	0.196	0.198	0.203	0.208	0.233
	2	0.382	0.375	0.375	0.381	0.428

* Regions arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK)

TABLE 3.23. Regional Average Normalized Bias and Root Mean Square Error (RMSE) of Quantiles: Clustering with GEV Parameters and QSP *

Probability Distribution	Region No:	10 yr.	20 yr.	50 yr.	100 yr.	1000 yr.
		Quantiles				
				Bias		
EV1	5	0.003	0.003	0.004	0.004	0.005
	4	0.003	0.003	0.003	0.003	0.003
	1	0.002	0.001	0.000	0.000	-0.001
	2	-0.009	-0.012	-0.016	-0.018	-0.022
	3	0.005	-0.001	-0.005	-0.008	-0.013
	6	0.013	0.000	-0.011	-0.017	-0.029
GEV	5	0.009	0.017	0.029	0.038	0.070
	4	0.009	0.026	0.051	0.071	0.141
	1	-0.003	-0.007	-0.012	-0.016	-0.031
	3	-0.004	-0.004	-0.001	0.003	0.020
	2	-0.008	-0.019	-0.033	-0.044	-0.077
	6	0.016	-0.015	-0.057	-0.089	-0.193
				RMSE		
EV1	5	0.102	0.102	0.103	0.103	0.103
	4	0.165	0.165	0.166	0.167	0.168
	1	0.114	0.114	0.114	0.114	0.114
	2	0.176	0.176	0.176	0.176	0.176
	3	0.159	0.158	0.157	0.157	0.157
	6	0.202	0.200	0.198	0.197	0.196
GEV	5	0.096	0.098	0.103	0.107	0.127
	4	0.154	0.160	0.173	0.187	0.252
	1	0.130	0.130	0.131	0.132	0.139
	3	0.164	0.165	0.168	0.171	0.189
	2	0.210	0.213	0.219	0.227	0.265
	6	0.439	0.434	0.431	0.433	0.459

* Regions arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK)

TABLE 3.24. Regional Average Normalized Bias and Root Mean Square Error(RMSE) of Quantiles: USGS Regions *

Probability Distribution	Region No:	Quantiles 10 yr.	20 yr.	50 yr.	100 yr.	1000 yr.
		Bias				
EV1	6	-0.001	-0.004	-0.006	-0.008	-0.010
	1	0.014	0.011	0.007	0.006	0.002
	2	0.003	0.001	-0.002	-0.003	-0.006
	4	-0.003	-0.004	-0.006	-0.007	-0.009
	7	0.000	-0.002	-0.004	-0.005	-0.007
	3	0.007	0.004	0.002	0.000	-0.003
	5	0.005	0.001	-0.003	-0.005	-0.009
GEV	6	0.000	-0.002	-0.005	-0.007	-0.010
	7	-0.004	-0.007	-0.010	-0.011	-0.015
	4	0.011	0.017	0.026	0.034	0.066
	2	-0.002	-0.004	-0.006	-0.007	-0.010
	1	-0.007	-0.014	-0.023	-0.031	-0.055
	3	0.000	-0.008	-0.017	-0.023	-0.043
	5	-0.010	-0.019	-0.029	-0.037	-0.060
		RMSE				
EV1	6	0.139	0.138	0.138	0.138	0.138
	1	0.156	0.155	0.155	0.154	0.154
	2	0.138	0.137	0.137	0.137	0.137
	4	0.118	0.118	0.118	0.118	0.118
	7	0.142	0.142	0.142	0.142	0.142
	3	0.154	0.154	0.154	0.154	0.153
	5	0.152	0.151	0.151	0.150	0.150
GEV	6	0.162	0.163	0.165	0.168	0.185
	7	0.175	0.176	0.179	0.182	0.198
	4	0.122	0.126	0.134	0.142	0.181
	2	0.154	0.154	0.155	0.157	0.164
	1	0.182	0.182	0.184	0.186	0.202
	3	0.278	0.279	0.281	0.284	0.298
	5	0.178	0.179	0.183	0.187	0.209

* Regions arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK)

TABLE 3.25. Comparison of Quantiles at a Few Selected Stations: Clustering with LCV and QSP (Case 1) #

Cluster Region	Station No.	Area (sq mi)	Yrs of Record	Station * Values LCV	LSK	Quantiles ** 50 yr (cfs)	Diff. (%)	100 yr (cfs)	Diff. (%)
Gumbel/GEV/LP3									
3	322360	14.3	12	0.13	-0.13	4775	57	5310	69
				0.13	-0.13	4801	57	5353	70
				0.20	-0.95	3050	--	3140	--
	435800	309.0	50	0.28	0.27	34746	-11	38638	-14
				0.28	0.27	34934	-10	38950	-13
				0.59	2.47	39000	--	44900	--
4	300065	1.7	10	0.29	-0.07	1312	-23	1471	-24
				0.29	-0.07	1349	-21	1532	-21
				0.50	-0.73	1700	--	1930	--
	321465	0.3	9	0.30	0.43	291	-5	326	-7
				0.30	0.43	299	-2	339	-4
				0.59	1.96	305	--	352	--
1	302085	0.9	10	0.29	-0.06	417	-1	470	2
				0.29	-0.06	462	10	550	20
				0.49	-1.58	420	--	460	--
	309575	3.2	10	0.30	0.48	864	-5	975	-11
				0.30	0.48	958	5	1141	5
				0.64	2.34	909	--	1090	--
2	310385	0.6	11	0.39	0.10	641	-10	729	-11
				0.39	0.10	736	4	900	10
				0.66	-1.42	709	--	815	--
	413202	0.6	10	0.47	0.47	615	-33	700	-41
				0.47	0.47	707	-23	865	-27
				0.94	1.38	915	--	1190	--
5	207962	0.8	10	0.43	0.19	234	-40	267	-47
				0.43	0.19	293	-24	384	-24
				0.76	0.47	387	--	502	--
	302350	0.6	10	0.53	0.61	645	-23	737	-32
				0.53	0.61	809	-3	1059	-3
				1.38	2.96	835	--	1090	--

The first and second station of each region were selected based on a low and a high station skew, respectively.
Regions are arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK).

* Coefficients of variation and skewness computed using L-moments for EV1 and GEV distributions and method of moments using normalized raw AFS for log-Pearson Type-III.

** log-Pearson Type-III estimates computed using Water Resources Council guidelines - Bulletin 17B. "Diff" is the percentage difference of the quantile estimates realitive to the log-Pearson Type-III estimates.

TABLE 3.26. Comparison of Quantiles of a Few Selected Stations: Clustering with EV1 Parameters and QSP (Case 2) #

Cluster Region	Station No.	Area (sq mi)	Yrs of Record	Station * Values LCV	LSK	Quantiles 50 yr (cfs)	Diff. (%)	100 yr (cfs)	Diff. (%)
Gumbel/GEV/LP3									
4	283305	0.6	11	0.17	0.19	310	26	344	28
				0.17	0.19	308	25	341	27
				0.29	1.18	247	--	268	--
	307500	18.3	42	0.27	0.41	4440	-15	4930	-21
				0.27	0.41	4418	-16	4893	-22
				0.71	4.69	5250	--	6240	--
5	277185	0.7	10	0.29	-0.003	661	-16	741	-16
				0.29	-0.003	681	-13	774	-13
				0.50	0.38	785	--	887	--
	290000	5.6	34	0.31	0.37	5570	-15	6241	-19
				0.31	0.37	5735	-12	6519	-16
				0.67	2.51	6540	--	7740	--
3	435500	309.0	50	0.28	0.27	40447	4	45537	1
				0.28	0.27	44266	14	52254	16
				0.59	2.47	39000	--	44900	--
	280728	1.8	10	0.34	0.54	844	-16	950	-23
				0.34	0.54	924	-9	1090	-12
				0.70	1.80	1010	--	1240	--
2	310385	0.6	11	0.39	0.10	615	-13	697	-14
				0.39	0.10	689	-3	830	2
				0.66	-1.42	709	--	815	--
	610820	0.1	9	0.43	0.46	293	-23	332	-29
				0.43	0.46	329	-13	396	-16
				0.86	1.55	379	--	469	--
1	277070	1.5	10	0.39	0.15	1618	95	1845	87
				0.39	0.15	1972	137	2521	155
				0.68	0.50	831	--	987	--
	254400	13.6	15	0.45	0.57	12012	-17	13697	-26
				0.45	0.57	14637	2	18712	2
				1.18	3.31	14400	--	18400	--

The first and second station of each region were selected based on a low and a high station skew, respectively.
Regions are arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK).

* Coefficients of variation and skewness computed using L-moments for EV1 and GEV distributions and method of moments using normalized raw maximum AFS for log-Pearson Type-III.

** log-Pearson Type-III estimates computed using Water Resources Council guidelines - Bulletin 17B. "Diff" is the percentage difference of the quantile estimates realitive to the log-Pearson Type-III estimates.

TABLE 3.27. Comparison of Quantiles at a Few Selected Stations: Clustering with LCV, LSKEW and QSP (Case 3) #

Cluster Region	Station No.	Area (sq mi)	Yrs of Record	Station * Values LCV	LSK	Quantiles ** 50 yr (cfs)	Diff. (%)	100 yr (cfs)	Diff. (%)
Gumbel/GEV/LP3									
4	300065	1.7	10	0.29	-0.07	1153	-32	1279	-34
				0.29	-0.07	1043	-39	1108	-43
				0.50	-0.73	1700	--	1930	--
	283305	0.6	11	0.17	0.19	304	23	337	26
				0.17	0.19	275	11	292	9
				0.29	1.18	247	--	268	--
3	252000	239.0	30	0.23	0.11	19393	13	21658	15
				0.23	0.11	19945	16	22590	20
				0.41	0.78	17200	--	18900	--
	435500	309.0	80	0.28	0.27	36534	-6	40801	-9
				0.28	0.27	37574	-4	42556	-5
				0.59	2.47	39000	--	44900	--
5	277185	0.7	10	0.29	-0.003	718	-9	810	-8
				0.29	-0.003	771	-2	902	2
				0.50	-0.38	785	--	887	--
	610820	0.1	9	0.43	0.46	273	-28	308	-34
				0.43	0.46	293	-23	342	-27
				0.86	1.55	379	--	469	--
1	216505	0.5	9	0.35	0.19	107	-14	121	-17
				0.35	0.19	123	-2	151	3
				0.62	0.69	125	--	146	--
	209575	3.2	10	0.30	0.48	887	-2	1003	-8
				0.30	0.48	1023	13	1252	15
				0.64	2.34	909	--	1090	--
2	305000	22.4	32	0.43	0.31	17538	-18	20075	-24
				0.43	0.31	22941	7	31380	19
				0.82	1.46	21500	--	26300	--
	413202	0.6	10	0.47	0.47	694	-24	794	-33
				0.47	0.47	908	-1	1242	4
				0.94	1.38	915	--	1190	--

\# The first and second station of each region were selected based on a low and a high station skew, respectively. Regions are arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK).

* Coefficients of variation and skewness computed using L-moments for EV1 and GEV distributions and method of moments using normalized raw maximum AFS for log-Pearson-Type III.

** log-Pearson Type-III estimates computed using Water Resources Council guidelines - Bulletin 17B. "Diff" is the percentage difference of the quantile estimates realitive to the log-Pearson Type-III estimates.

TABLE 3.28. Comparison of Quantiles of a Few Selected Stations: Clustering with GEV Parameters and QSP (Case 4) #

Cluster Region	Station No.	Area (sq mi)	Yrs of Record	Station * Values LCV	LSK	Quantiles ** 50 yr (cfs)	Diff. (%)	100 yr (cfs)	Diff. (%)
Gumbel/GEV/LP3									
5	412500	31.3	30	0.20	-0.001	4552	17	5061	31
				0.20	-0.001	4298	11	4656	11
				0.34	-0.07	3880	--	4180	--
	281500	722.0	54	0.29	0.17	58539	-16	65091	-18
				0.29	0.17	55267	-20	59879	-24
				0.51	0.69	69400	--	79200	--
4	300065	1.7	10	0.29	-0.07	1289	-24	1443	-25
				0.29	-0.07	1180	-31	1271	-34
				0.50	-0.73	1700	--	1930	--
	285800	1.0	26	0.31	0.16	1515	-10	1696	-11
				0.31	0.16	1386	-18	1494	-21
				0.55	0.81	1690	--	1900	--
1	403500	960.0	46	0.22	0.19	67939	13	76151	14
				0.22	0.19	75661	26	89972	34
				0.40	0.70	59900	--	66900	--
	209575	3.2	10	0.30	0.48	797	-12	893	-18
				0.30	0.48	887	-2	1055	-3
				0.64	2.34	909	--	1090	--
2	297000	5.2	29	0.33	0.35	4529	-11	5114	-16
				0.33	0.35	5233	3	6408	5
				0.64	1.30	5070	--	6080	--
	610820	0.1	9	0.43	0.46	278	-27	314	-33
				0.43	0.46	321	-15	394	-16
				0.86	1.55	379	--	469	--
3	208500	86.0	58	0.35	0.24	45487	-10	51516	-12
				0.35	0.24	48554	-4	56799	-3
				0.65	1.48	50600	--	58800	--
	313600	1.0	18	0.43	0.33	759	-30	860	-37
				0.43	0.33	810	-26	948	-30
				0.83	1.02	1090	--	1360	--
6	610000	89.7	32	0.39	0.42	23540	-9	26853	-16
				0.39	0.42	30390	17	40995	29
				0.91	2.80	25900	--	31800	--
	303300	39.8	24	0.48	0.56	7450	2	8499	-9
				0.48	0.56	9618	31	12975	39
				1.22	3.55	7320	--	9350	--

\# The first and second station of each region were selected based on a low and a high station skew, respectively. Regions are arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK).

* Coefficients of variation and skewness computed using L-moments for EV1 and GEV distributions and method of moments using normalized raw maximum AFS for log-Pearson Type-III.

** log-Pearson Type-III estimates computed using Water Resources Council guidelines - Bulletin 17B. "Diff" is the percentage difference of the quantile estimates realitive to the log-Pearson Type-III estimates.

TABLE 3.29. Comparison of Quantiles of a Few Selected Stations within USGS Regions #

Cluster Region	Station No.	Area (sq mi)	Yrs of Record	Station * Values LCV	LSK	Quantiles ** 50 yr (cfs)	Diff. (%)	100 yr (cfs)	Diff. (%)
Gumbel/GEV/LP3									
6	322100	323.0	22	0.24	0.21	11685	18	13069	20
				0.24	0.21	12272	24	14073	29
				0.44	1.45	9890	--	10900	--
	315885	0.2	9	0.31	0.41	127	-15	142	-19
				0.31	0.41	133	-11	153	-13
				0.62	2.11	150	--	175	--
1	247100	3.3	31	0.16	0.07	1369	45	1533	53
				0.16	0.07	1504	59	1773	77
				0.28	0.07	946	--	999	--
	298535	0.7	10	0.51	0.55	497	-47	556	-55
				0.51	0.55	546	-40	643	-48
				1.18	2.53	944	--	1240	--
2	283500	362.0	51	0.30	0.20	23649	-8	26536	-9
				0.30	0.20	24955	-3	28774	-1
				0.56	1.18	25600	--	29100	--
	237280	12.2	22	0.44	0.41	3627	-14	4064	-17
				0.44	0.41	3821	-9	4406	-10
				0.99	2.97	4230	--	4920	--
4	402020	3.0	10	0.19	-0.08	1325	25	1485	30
				0.19	-0.08	1368	29	1558	37
				0.33	0.74	1060	--	1140	--
	404900	53.8	29	0.31	0.34	6675	-14	7482	-18
				0.31	0.34	6892	-11	7850	-14
				0.61	1.95	7760	--	9160	--
7	610503	0.8	10	0.24	-0.06	1797	13	2760	60
				0.24	-0.06	1974	24	3827	121
				0.40	-0.58	1590	--	1730	--
	302500	194.0	45	0.35	0.33	21865	-18	24586	-23
				0.35	0.33	24025	-10	28401	-11
				0.71	2.17	26800	--	31900	--
3	284300	28.6	16	0.32	0.09	8857	-16	9971	-18
				0.32	0.09	9745	-7	11540	-5
				0.56	0.18	10500	--	12100	--
	208600	202.0	29	0.40	0.40	41858	-18	47124	-24
				0.40	0.40	46055	-10	54539	-12
				0.86	2.26	51300	--	62200	--
5	415700	4.8	24	0.27	0.01	1949	4	2197	7
				0.27	0.01	2177	16	2604	26
				0.47	0.26	1870	--	2060	--
	307000	173.0	46	0.36	0.42	27952	-22	31500	-29
				0.36	0.42	31219	-13	37336	-16
				0.77	1.88	36000	--	44600	--

\# The first and second station of each region were selected based on a low and a high station skew, respectively. Regions are arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK).

* Coefficients of variation and skewness computed using L-moments for EV1 and GEV distributions and method of moments using normalized raw maximum AFS for log-Pearson Type-III.

** log-Pearson Type-III estimates computed using Water Resources Council guidelines - Bulletin 17B. "Diff" is the percentage difference of the quantile estimates realitive to the log-Pearson Type-III estimates.

c) Discriminant Analysis: In the previous section, homogeneous flood regions are identified using four different clustering schemes (referred to as Cases 1-4) using FASTCLUS clustering algorithm. Although a comparison between the cluster regions is made using important statistics of all hydrological attributes and the performance of the regionalized flood frequency growth curves, it remains to be seen as to what factors, other than the clustering variables employed, cause the fundamental differences between these regions. For instance, if cluster regions are delineated using LCV and QSP (Case 1), the five regions identified can be generally classified as low, medium and high flood regions based on the values of the response variables. Since watershed drainage area, A_c, is highly correlated with one of the clustering variables, namely QSP, the differences between these clusters regions can be further explained on the basis of this or other physical attributes that control flood response of a watershed. With this in mind, a stepwise discriminant analysis is first performed in order to identify the most significant (at 5% level of significance) attribute variables which provide maximum discrimination between the flood regions. Application of this procedure to the cluster regions for all the clustering schemes gave results as summarized in Table 3.30. Results of clustering cases 1-4 and USGS regions are also included this table. The variables listed in column 3 of this table are the significant attribute variables arranged in the order of importance. The following conclusions are drawn:

a) Although the original set of attribute variables defined at each gauge incorporated a broad range of hydrological characteristics of each watershed, the most important variables controlling flood response seem to be the geomorphic properties of the watershed

Table 3.30. Results of Discriminant Results for all Flood Region Delineation Cases Examined in the Study **

Case No.	Cluster Variables	Signif. Discrim. Variables	Overall Discrimination Score	Percent
1	LCV	DAREA, CHANSLOP	62/253	0.245
2	LCV, LSKEW	DAREA, BASLEN	65/242	0.269
3	LCV, LSKEW, LKUR	DAREA, BASLEN	72/242	0.298
4	LCV, QSP	DAREA, BASLEN	111/242	0.459
5*	LCV, LSKEW, QSP	DAREA, BASLEN	105/242	0.434
6*	LCV, LSKEW, LKUR, QSP	DAREA, BASLEN, CHANSLOP	102/242	0.421
7	MEVL, AEVL	DAREA,CHANSLOP	62/253	0.245
8	MGVL, AGVL, KGVL	DAREA, BASLEN SHAPE	79/242	0.326
9	MWKL, AWKL, BWKL, CWKL, DWKL	DAREA	159/253	0.628
10*	MEVL, AEVL, QSP	DAREA, BASLEN, SHAPE, CHANSLOP, STOR, CHANSIN, CHANLEN	114/242	0.471
11*	MGVL, AGVL, KGVL, QSP	DAREA, BASLEN CHANSIN, CHANLEN	88/242	0.364
12	MWKL,AWKL, BWKL, CWKL, DWKL, QSP	DAREA	122/242	0.504
13	USGS REGIONS	DAREA, CHANLEN	41/242	0.169

** Because all physical characteristics are not available for each station, some stations are not included in the discriminant analysis.

* Clustering cases selected in the study (referred to as Cases 1-4)

such as its size and shape and main channel characteristics.

b) Watershed drainage area, A_c, is the most significant attribute for discriminating between clusters for all the clustering schemes and USGS regions.

c) All the significant attributes listed in Table 3.30 describe the physical characteristics that control the magnitude and timing of flood peak response of a watershed. For instance, the magnitude of the flood peak is proportional to the drainage area and its timing is influenced by travel paths such as the basin and main channel lengths.

The next step in discriminant analysis is to perform a classificatory analysis of gauged sites in each cluster region (for a given clustering scheme) in order to determine the percentage gauged sites correctly classified in the original cluster regions. To accomplish this the significant attribute variables are used together with the DISCRIM procedure of SAS (1985) to perform a classificatory discriminant analysis. Tables 3.31-3.34 summarize the results for all the four clustering schemes (Cases 1-4) selected in the study. The horizontal rows in these tables reflect the original cluster groupings while the vertical columns indicate the new cluster groupings into which each site is classified based upon its attributes. If all gauges are correctly classified then the row percentages of the diagonal elements in these tables will be 100%. It is obvious that such is not the case. The low percent classification in some cases indicates that the cluster regions can not be discriminated well based upon the attributes used in the analysis. An overall discriminant score is computed by summing the sites classified correctly (i.e. all sites along the diagonal). This total score divided by the total number of sites being classified gives the overall percent correct classification. This value for each clustering scheme is shown in Column 5 of Table 3.30. Based on these results the following conclusions are drawn:

a) With the exception of the clustering cases involving the Wakeby probability distribution parameters (these cases were dropped due poor performance in estimating flood quantiles), clustering cases 1-4 (labeled as cases 4, 5, 10 and 11 in Table 3.30) provide the best overall percent classification compared to all the other cases considered in the study. The overall percent correct classification ranges from 36.4% to 47.1%.
b) Watershed drainage area, A_c, is the most significant discriminating variable for all the clustering schemes. The remaining variables listed in Table 3.30 are all geomorphic that are closely related to the physical aspects of the watershed.
c) In all clustering schemes there are at least two cluster regions that have a percent classification less than 50%. This occurs with cluster regions that have considerable overlap in their hydrological characteristics.

The results of discriminant analysis for the seven USGS flood regions using all the gauged site data as illustrated in Table 3.35. An examination of the diagonal elements of this table clearly indicates that these regions can not be discriminated between each other easily. In other words, the classification of gauged sites into a region based upon the attribute variables (referred to as discriminating power) can not be achieved with a high degree of certainty. The average discrimination is only 16.9% when compared to a maximum of 47.1% achieved using cluster analysis in conjunction with EV1 parameters and QSP as the clustering variables (refer to Case 10 in Table 3.30). This further supports the observation that each of the USGS flood regions has a mixed composition of watersheds with differing hydrological characteristics. Hence, these regions are not very homogeneous with respect to the characteristics describing flood response. A similar observation was made by Wiltshire (1986) who states that flood regions delineated in a rather arbitrary manner and arranged to coincide with geographical areas are likely to contain drainage basins with a diversity of geomorphology whose flood frequency characteristics may not be comparable. He further states that in this situation a regional average frequency curve will be poorly defined. In contrast, the cluster regions are not only homogeneous with respect to the flood response characteristics (response or clustering variables) but lend themselves to a higher level of discrimination by variables that affect these characteristics (attribute variables). A comparison of the significant variables in the discriminant analysis indicates that for both cluster and USGS regions the geomorphic variables provide good discrimination with contributing drainage area being the most important (refer to Table 3.30). The remaining variables describe the watershed and main channel dimensions.

Table 3.31. Classificatory Discriminant Analysis of Cluster Regions Formed Using Clustering Variables LCV and QSP *

Number and Percentage of Observations Classified into Cluster Region

From Cluster	1	2	3	4	5	Total
1	49 55.1	6 6.7	20 22.5	11 12.4	3 3.4	89 100.0
2	2 14.3	8 57.1	0 0.0	2 14.3	2 14.3	14 100.0
3	38 42.2	3 3.3	31 34.4	13 14.4	5 5.6	90 100.0
4	0 0.0	6 22.2	0 0.0	21 77.8	0 0.0	27 100.0
5	5 22.7	7 31.8	1 4.6	7 31.8	2 9.1	22 100.0
Total Percent	94 38.8	30 12.4	52 21.5	54 22.3	12 5.0	242 100.0
Priors	0.20	0.20	0.20	0.20	0.20	

Overall % correct classification = 111/242 = 46%

* Significant variables at 5% level: drainage area, A_c, basin length, B_l, basin shape, B_s, and main channel slope, S_c.

Table 3.32. Classificatory Discriminant Analysis of Cluster Regions Formed Using Clustering Variables EV1 Parameters and QSP *

Number and Percentage of Observations Classified into Cluster Region

From Cluster	1	2	3	4	5	Total
1	5 12.8	11 28.2	10 25.6	2 5.1	11 28.2	39 100.0
2	2 22.2	6 66.7	0 0.0	0 0.0	1 11.1	9 100.0
3	4 4.4	6 6.6	59 64.8	10 11.0	12 13.2	91 100.0
4	3 3.9	3 3.9	34 44.2	25 32.5	12 15.6	77 100.0
5	2 7.7	5 19.2	0 0.0	0 0.0	19 73.1	26 100.0
Total Percent	16 6.6	31 12.8	103 42.6	37 15.3	55 22.7	242 100.0
Priors	0.20	0.20	0.20	0.20	0.20	

% Correct classification = 114/242 = 47%

* Significant variables at 5% level: drainage area, A_c, basin length, B_l, basin shape, B_s, main channel slope, S_c, main channel sinuousity, S_s, and basin storage, STOR.

Table 3.33. Classificatory Discriminant Analysis of Cluster Regions Formed Using Clustering Variables LCV, LSX and QSP *

From Cluster	Number and Percentage of Observations Classified into Cluster Region					
	1	2	3	4	5	Total
1	21	19	17	3	19	79
	26.6	24.1	21.5	3.8	24.1	100.0
2	2	4	1	0	7	14
	14.3	28.6	7.1	0.0	50.0	100.0
3	15	7	48	1	4	75
	20.0	9.3	64.0	1.3	5.3	100.0
4	12	9	8	0	11	40
	30.0	22.5	20.0	0.0	27.5	100.0
5	0	2	0	0	32	34
	0.0	16.9	30.6	1.7	30.2	100.0
Total	50	41	74	4	73	242
Percent	20.7	16.9	30.6	1.6	30.2	100.0
Priors	0.20	0.20	0.20	0.20	0.20	

% Correct classification = 105/242 = 43%

* Significant variables at 5% level: drainage area, A_c, and basin length, B_1.

Table 3.34. Classificatory Discriminant Analysis of Cluster Regions Formed Using Clustering Variables GEV Parameters and QSP *

From Cluster	Number and Percentage of Observations Classified into Cluster Region						
	1	2	3	4	5	6	Total
1	47	0	13	8	8	5	81
	58.0	0.0	16.1	9.9	9.9	6.2	100.0
2	1	4	2	11	0	0	18
	5.6	22.2	11.1	61.1	0.0	0.0	100.0
3	11	5	7	12	2	2	39
	28.2	12.8	18.0	30.8	5.1	5.1	100.0
4	0	2	2	10	0	0	14
	0.0	14.3	14.3	71.4	0.0	0.0	100.0
5	18	7	11	6	19	3	64
	28.1	10.9	17.2	9.4	29.7	4.7	100.0
6	4	5	5	8	3	1	26
	15.4	19.2	19.2	30.8	11.5	4.6	100.0
Total	81	23	40	55	32	11	242
Percent	33.5	9.5	16.5	22.7	13.2	4.6	100.0
Priors	0.167	0.167	0.167	0.167	0.167	0.167	

% Correct classification = 88/242 = 36%

* Significant variables at 5% level: drainage area, A_c, basin length, B_1, main channel length, L_c, and main channel sinuousity, S_s.

Table 3.35. Classificatory Discriminant Analysis of USGS Regions *

Number and Percentage of Observations Classified into U.S.G.S. Region

From Cluster	1	2	3	4	5	6	7	Total
1	22	0	0	1	0	5	2	81
	73.3	0.0	16.1	9.9	9.9	16.7	6.7	100.0
2	29	0	2	13	0	18	6	68
	42.7	0.0	2.9	19.1	0.0	26.5	8.8.	100.0
3	8	0	2	5	0	6	5	26
	30.8	0.0	7.7	19.2	0.0	23.1	19.2	100.0
4	10	2	0	5	0	2	1	20
	50.0	10.0	0.0	25.0	0.0	10.0	5.0	100.0
5	17	0	3	4	0	6	5	35
	48.6	0.0	8.6	11.4	0.0	17.1	14.3	100.0
6	12	0	2	2	1	9	2	28
	42.7	0.0	7.1	7.1	3.6	32.1	7.1	100.0
7	20	0	1	0	1	10	3	35
	57.1	0.0	2.3	0.0	2.3	28.6	8.6	100.0
Total	118	2	10	30	2	56	24	242
Percent	48.8	0.8	4.1	12.4	0.8	23.1	9.9	100.0
Priors	0.142	0.142	0.142	0.142	0.142	0.142	0.142	

% Correct classification = 41/242 = 17%

* Significant variables at 5% level: drainage area, A_c, and main channel length, L_c

d) Regression Analysis: The ultimate objective or purpose of delineating distinct flood regions is to develop regionalized relationships for predicting the flood response (at selected frequency levels) at both gauged and ungauged sites. For gauged sites, such a regionalized relationship can be used together with at-site information for estimating flood levels (Choquette 1988). In contrast, while using cluster analysis, ungauged sites must first be classified into a flood region based on significant physical attributes of the watershed affecting flood response prior to using a regionalized equation. For the method of residuals, this classification is relatively straight forward since an ungauged site is univocally assigned to the geographic region in which it lies.

Overall regression results, pertaining to the equations developed for predicting the 20, 50 and 100 year flood levels within the cluster regions (for all clustering cases examined in this study), are shown in Table 3.36. Cases 1-4 (marked by an asterisk in this table) have the lowest weighted standard error when compared to the remaining cases. For these four cases and USGS flood regions, detailed regionalized regression equations for the EV1 and GEV models are given in Tables 3.37-3.46.

Table 3.47 gives similar equations for the 50 and 100 year flood levels (20 year flood quantile regression equations are not available) and are developed for the USGS method of residuals flood regions using log-Pearson Type-III distribution (Choquette, 1988). This table is provided for the purpose of comparing the performance of the log-Pearson Type-III, EV1 and GEV flood frequency models developed for the seven USGS flood regions. It must be emphasized that the flood levels used in developing these regression equations are estimated from a log-Pearson Type-III flood frequency distribution using a weighted skewness (based on station and a map skew).

TABLE 3.36. Regression Results for all Flood Region Delineation Cases Examined in the Study

Case No.	Cluster Variables	No. of Regions	Weighted Standard Error (%)	Standard Error Range** (%)
1	LCV	6	44.4	32.8 - 53.0
2	LCV, LSKEW	6	44.9	39.8 - 56.5
3	LCV, LSKEW, LKUR	6	45.2	39.2 - 53.1
4*	LCV, QSP	5	36.9	19.3 - 46.6
5*	LCV, LSKEW, QSP	5	36.6	27.0 - 51.0
6	LCV, LSKEW, LKUR, QSP	5	41.1	24.9 - 54.9
7	MEVL, AEVL	6	43.8	32.2 - 53.2
8	MGVL, AGVL, KGVL	5	44.5	39.0 - 56.1
9	MWKL, AWKL, BWKL, CWKL, DWKL	2	45.8	44.0 - 54.1
10*	MEVL, AEVL, QSP	5	39.2	20.1 - 54.9
11*	MGVL, AGVL, KGVL, QSP	6	39.1	23.4 - 52.2
12	MWKL, AWKL, BWKL, CWKL, DWKL, QSP	6	45.7	43.5 - 115.6
13	USGS REGIONS	7	35.0	19.7 - 38.6

* Indicates cases which are selected in the study (referred to as Cases 1-4)

** Based on standard errors of regression equations of each region

An examination of these tables suggests that the standard errors associated with the regression equations are, in general, comparable between the cluster and USGS regions when EV1 and GEV models are used in the regionalization. However, the standard errors are slightly higher when using the log-Pearson Type-III distribution (compare Table 3.46 and 3.47). Hence, even for the USGS flood regions (as delineated using method of residuals), it appears that more accurate regression equations can be developed by using either EV1 or GEV regionalized flood frequency models.

The independent variables and their exponents do not change for a particular flood frequency model **within a flood region** for different return period, T. This is not surprising since the regionalized quantile levels (normalized values) used in estimating the flood levels, Q_T, are scalar multiples of each other. In other words, the 100-year flood quantile can be obtained from the 10-year flood quantile by multiplying the latter with a constant. Hence, the correlation of flood quantiles with the independent variables remains the same from one flood level to the next. Consequently, the exponent term in the regression equations (slope term in the log-relationship) remains unaffected. The effects of scale are absorbed in the intercept term. A similar reasoning applies when comparing the regression equations (for a given flood quantile, Q_T, and flood region) between the EV1 and GEV flood frequency models. Once again, the independent variables and their exponents continue to be identical within a flood region when the return period, T is changed.

For both cluster and USGS regions, the geomorphic variables such as the watershed drainage area, A_c, main channel slope, S_c, and sinuousity, S_s, basin shape, B_s, are the most significant variables. For some cluster regions the exponent of the independent variable (drainage area, A_c) is greater than or equal to 1.0, indicating greater variability in the estimate of the flood quantile as the drainage area increases (true for watersheds greater than 1.0 sq mi). These cluster regions, as compared to other regions have predominantly small watersheds. Furthermore, gauged sites within these cluster regions also have, in general, short flood records (less than 10 years). Thus, a possible explanation for the larger exponent of the drainage area variable, A_c, in the regression equations may be due the fact that small watersheds experience greater variability in their flood response as opposed to larger watersheds due to their inability to dampen temporal effects of rainfall.

In applying the regionalized regressions equations developed for cluster regions (Cases 1-4) for ungauged sites (these sites do not have their flood characteristics defined), one must first assign these sites to a particular cluster region based solely on the physical attributes of the watersheds. Results of classificatory discriminant analysis (refer to Tables 3.31-3.34) of the gauged sites, based on their physical attributes only (i.e. treating them as ungauged sites), show the assignment of watersheds is not with complete certainty. For instance when clustering with LCV and QSP (Case 1), 49 of the 89 sites originally assigned to cluster region 1 are re-assigned to this region while the remaining sites are assigned to the cluster region 2, 3, 4 or 5. The posterior probabilities of these assignments are given in the second row of Table 3.31 for each cluster region. Consequently, in using the regionalized regression equations shown in Table 3.37 for predicting the flood levels at ungauged sites, one must use a weighted predicted flood level as developed from all the regionalized regression equations associated with the cluster regions to which the site is assigned. The weighting can be accomplished using the posterior probabilities of being assigned to each cluster region. Thus, the standard errors of prediction must also be based on the regionalized regressions used. For each cluster region, a weighted standard error may be computed using the following equation:

$$e_j = \sum_{i=1}^{m} e_{ji} * p_{ji} \qquad \text{for } j = 1,2...m \quad ...(3.1)$$

where,

e_j = percent standard error at a site in cluster j,

e_{ji} = percent standard error at a site in cluster j if it was classified into cluster i,

p_{ji} = posterior probability of a site in cluster j being classified into cluster i, and

m = number of cluster regions (equal to 5 or 6 in the present study).

For clustering Cases 1-4, values of e_{ji} can be obtained from the standard errors shown in Column 3 of Table 3.37-3.45 for each cluster region i and the posterior probabilities, p_{ji}, can be obtained from the rows of Tables 3.31-3.34. Based on all the gauged sites that are classified into the cluster regions, a weighted standard error can be computed for each cluster region using Eq. 3.1 above. For USGS regions, the problem of misclassification does not arise since ungauged sites are assigned to a region on the basis of their location in space.

TABLE 3.37. Regression Models for Estimating the Expected EV1 Flood Quantiles for Various Return Periods: Clustering on LCV and QSP (Case 1)

Cluster Region	Regression Equation	% Standard Error	R^2	No. of Sites.
1	$Q_{20} = 546\ A_c^{1.061}\ L_c^{-0.552}$	38.6	0.95	88
2	$Q_{20} = 967\ A_c^{0.716}\ S_c^{-0.168}$	19.3	0.95	15
3	$Q_{20} = 382\ A_c^{0.777}\ S_c^{0.144}$	38.1	0.94	92
4	$Q_{20} = 803\ A_c^{0.960}$	29.0	0.93	29
5	$Q_{20} = 657\ A_c^{0.682}\ S_c^{-0.346}$	46.6	0.93	24
1	$Q_{50} = 659\ A_c^{1.061}\ L_c^{-0.552}$	38.6	0.95	88
2	$Q_{50} = 1183\ A_c^{0.716}\ S_c^{-0.168}$	19.3	0.95	15
3	$Q_{50} = 449\ A_c^{0.777}\ S_c^{0.144}$	38.1	0.94	92
4	$Q_{50} = 958\ A_c^{0.960}$	29.0	0.93	29
5	$Q_{50} = 812\ A_c^{0.682}\ S_c^{-0.346}$	46.6	0.93	24
1	$Q_{100} = 742\ A_c^{1.061}\ L_c^{-0.552}$	38.6	0.95	88
2	$Q_{100} = 1345\ A_c^{0.716}\ S_c^{-0.168}$	19.3	0.95	15
3	$Q_{100} = 499\ A_c^{0.777}\ S_c^{0.144}$	38.1	0.94	92
4	$Q_{100} = 1077\ A_c^{0.960}$	29.0	0.93	29
5	$Q_{100} = 927\ A_c^{0.682}\ S_c^{-0.346}$	46.6	0.93	24

TABLE 3.38. Regression Models for Estimating the Expected EV1 Flood Quantiles for Various Return Periods: Clustering on EV1 Parameters and QSP (Case 2)

Cluster Regions	Regression Equation	% Standard Error	R^2	No. of Sites.
1	$Q_{20} = 550\ A_c^{0.758}$	54.9	0.90	42
2	$Q_{20} = 516\ L_c^{1.596}$	20.1	0.90	8
3	$Q_{20} = 495\ A_c^{0.975}\ L_c^{-0.402}$	37.2	0.95	90
4	$Q_{20} = 397\ A_c^{0.777}\ S_c^{0.169}$	38.6	0.93	78
5	$Q_{20} = 781\ A_c^{0.980}$	29.5	0.93	29
1	$Q_{50} = 676\ A_c^{0.758}$	54.9	0.90	42
2	$Q_{50} = 629\ L_c^{1.596}$	20.1	0.90	8
3	$Q_{50} = 592\ A_c^{0.975}\ L_c^{-0.402}$	37.2	0.95	90
4	$Q_{50} = 466\ A_c^{0.777}\ S_c^{0.169}$	38.6	0.93	78
5	$Q_{50} = 928\ A_c^{0.980}$	29.5	0.93	29
1	$Q_{100} = 773\ A_c^{0.758}$	54.9	0.90	42
2	$Q_{100} = 711\ L_c^{1.596}$	20.1	0.90	8
3	$Q_{100} = 669\ A_c^{0.975}\ L_c^{-0.402}$	37.2	0.95	90
4	$Q_{100} = 517\ A_c^{0.777}\ S_c^{0.169}$	38.6	0.93	78
5	$Q_{100} = 1043\ A_c^{0.980}$	29.5	0.93	29

TABLE 3.39: Regression Models for Estimating the Expected EV1 Flood Quantiles for Various Return Periods: Clustering on LCV, LSK and QSP (Case 3)

Cluster Region	Regression Equation	% Standard Error	R^2	No. of Sites.
1	$Q_{20} = 574\ A_c^{1.069}\ L_c^{-0.605}\ S_c^{-0.119}$	40.4	0.95	78
2	$Q_{20} = 659\ A_c^{0.776}$	51.0	0.90	16
3	$Q_{20} = 395\ A_c^{0.821}\ S_s^{0.227}\ B_s^{-0.194}$	32.7	0.91	74
4	$Q_{20} = 465\ A_c^{0.793}\ S_s^{-0.911}$	40.1	0.95	39
5	$Q_{20} = 887\ A_c^{0.887}$	27.0	0.91	37
1	$Q_{50} = 694\ A_c^{1.069}\ L_c^{-0.605}\ S_c^{-0.119}$	40.4	0.95	78
2	$Q_{50} = 816\ A_c^{0.776}$	51.0	0.90	16
3	$Q_{50} = 467\ A_c^{0.821}\ S_s^{0.227}\ B_s^{-0.194}$	32.7	0.91	74
4	$Q_{50} = 544\ A_c^{0.793}\ S_s^{-0.911}$	40.1	0.95	39
5	$Q_{50} = 1072\ A_c^{0.887}$	27.0	0.91	37
1	$Q_{100} = 784\ A_c^{1.069}\ L_c^{-0.605}\ S_c^{-0.119}$	40.4	0.95	78
2	$Q_{100} = 936\ A_c^{0.776}$	51.0	0.90	16
3	$Q_{100} = 524\ A_c^{0.821}\ S_s^{0.227}\ B_s^{-0.194}$	32.7	0.91	74
4	$Q_{100} = 602\ A_c^{0.793}\ S_s^{-0.911}$	40.1	0.95	39
5	$Q_{100} = 1207\ A_c^{0.887}$	27.0	0.91	37

TABLE 3.40. Regression Models for Estimating the Expected EV1 Flood Quantiles for Various Return Periods: Clustering on GEV Parameters and QSP (Case 4)

Cluster Regions	Regression Equation	% Standard Error	R^2	No. of Sites.
1	$Q_{20} = 480\ A_c^{0.963}\ L_c^{-0.396}$	38.1	0.94	80
2	$Q_{20} = 893\ A_c^{0.850}$	23.4	0.94	20
3	$Q_{20} = 661\ A_c^{1.225}\ L_c^{-0.809}$	52.2	0.93	38
4	$Q_{20} = 904\ A_c^{1.002}$	31.2	0.91	14
5	$Q_{20} = 566\ A_c^{0.962}\ L_c^{-0.456}$	38.1	0.95	63
6	$Q_{20} = 688\ A_c^{0.643}\ S_c^{-0.206}$	42.0	0.95	27
1	$Q_{50} = 584\ A_c^{0.963}\ L_c^{-0.396}$	38.1	0.94	80
2	$Q_{50} = 1080\ A_c^{0.850}$	23.4	0.94	20
3	$Q_{50} = 803\ A_c^{1.225}\ L_c^{-0.809}$	52.2	0.93	38
4	$Q_{50} = 1076\ A_c^{1.002}$	31.2	0.91	14
5	$Q_{50} = 666\ A_c^{0.962}\ L_c^{-0.456}$	38.1	0.95	63
6	$Q_{50} = 847\ A_c^{0.643}\ S_c^{-0.206}$	42.0	0.95	27
1	$Q_{100} = 640\ A_c^{0.963}\ L_c^{-0.396}$	38.1	0.94	80
2	$Q_{100} = 1217\ A_c^{0.850}$	23.4	0.94	20
3	$Q_{100} = 910\ A_c^{1.225}\ L_c^{-0.809}$	52.2	0.93	38
4	$Q_{100} = 1206\ A_c^{1.002}$	31.2	0.91	14
5	$Q_{100} = 741\ A_c^{0.962}\ L_c^{-0.456}$	38.1	0.95	63
6	$Q_{100} = 967\ A_c^{0.643}\ S_c^{-0.206}$	42.0	0.95	27

TABLE 3.41. Regression Models for Estimating the Expected EV1 Flood Quantiles for Various Return Periods for USGS Regions

Cluster Regions	Regression Equation	% Standard Error	R^2	No. of Sites.
1	$Q_{20} = 716\ A_c^{0.963}\ L_c^{-0.396}\ S_c^{0.196}$	38.4	0.95	29
2	$Q_{20} = 341\ A_c^{0.736}$	39.9	0.96	67
3	$Q_{20} = 520\ A_c^{0.744}\ S_c^{0.029}\ B_s^{-0.070}$	19.7	0.99	25
4	$Q_{20} = 289\ A_c^{0.842}\ S_s^{-0.517}$	27.1	0.99	19
5	$Q_{20} = 634\ A_c^{0.720}$	33.2	0.97	37
6	$Q_{20} = 623\ A_c^{0.624}\ S_s^{-0.277}$	38.6	0.96	27
7	$Q_{20} = 818\ A_c^{0.587}$	36.9	0.95	37
1	$Q_{50} = 852\ A_c^{0.963}\ L_c^{-0.396}\ S_c^{0.196}$	38.4	0.95	29
2	$Q_{50} = 408\ A_c^{0.736}$	39.9	0.96	67
3	$Q_{50} = 622\ A_c^{0.744}\ S_c^{0.029}\ B_s^{-0.070}$	19.7	0.99	25
4	$Q_{50} = 344\ A_c^{0.842}\ S_s^{-0.517}$	27.1	0.99	19
5	$Q_{50} = 763\ A_c^{0.720}$	33.2	0.97	37
6	$Q_{50} = 739\ A_c^{0.624}\ S_s^{-0.277}$	38.6	0.96	27
7	$Q_{50} = 982\ A_c^{0.587}$	36.9	0.95	37
1	$Q_{100} = 954\ A_c^{0.963}\ L_c^{-0.396}\ S_c^{0.196}$	38.4	0.95	29
2	$Q_{100} = 457\ A_c^{0.736}$	39.9	0.96	67
3	$Q_{100} = 702\ A_c^{0.744}\ S_c^{0.029}\ B_s^{-0.070}$	19.7	0.99	25
4	$Q_{100} = 385\ A_c^{0.842}\ S_s^{-0.517}$	27.1	0.99	19
5	$Q_{100} = 860\ A_c^{0.720}$	33.2	0.97	37
6	$Q_{100} = 830\ A_c^{0.624}\ S_s^{-0.277}$	38.6	0.96	27
7	$Q_{100} = 1101\ A_c^{0.587}$	36.9	0.95	37

TABLE 3.42: Regression Models for Estimating the Expected GEV Flood Quantiles for Various Return Periods: Clustering on LCV and QSP (Case 1)

Cluster Region No.	Regression Equation	% Standard Error	R^2	No. of Sites.
1	$Q_{20} = 412\ A_c^{1.061}\ L_c^{-0.552}$	38.6	0.95	88
2	$Q_{20} = 1013\ A_c^{0.716}\ S_c^{-0.168}$	19.3	0.95	15
3	$Q_{20} = 415\ A_c^{0.777}\ S_c^{0.144}$	38.1	0.94	92
4	$Q_{20} = 816\ A_c^{0.960}$	29.0	0.93	29
5	$Q_{20} = 696\ A_c^{0.682}\ S_c^{-0.346}$	46.6	0.93	24
1	$Q_{50} = 530\ A_c^{1.061}\ L_c^{-0.552}$	38.6	0.95	88
2	$Q_{50} = 1358\ A_c^{0.716}\ S_c^{-0.168}$	19.3	0.95	15
3	$Q_{50} = 488\ A_c^{0.777}\ S_c^{0.144}$	38.1	0.94	92
4	$Q_{50} = 987\ A_c^{0.960}$	29.0	0.93	29
5	$Q_{50} = 1018\ A_c^{0.682}\ S_c^{-0.346}$	46.6	0.93	24
1	$Q_{100} = 631\ A_c^{1.061}\ L_c^{-0.552}$	38.6	0.95	88
2	$Q_{100} = 1661\ A_c^{0.716}\ S_c^{-0.168}$	19.3	0.95	15
3	$Q_{100} = 544\ A_c^{0.777}\ S_c^{0.144}$	38.1	0.94	92
4	$Q_{100} = 1121\ A_c^{0.960}$	29.0	0.93	29
5	$Q_{100} = 1332\ A_c^{0.682}\ S_c^{-0.346}$	46.6	0.93	24

TABLE 3.43: Regression Models for Estimating the Expected GEV Flood Quantiles for Various Return Periods: Clustering on EV1 Parameters and QSP (Case 2)

Cluster Region No.	Regression Equation	% Standard Error	R^2	No. of Sites.
1	$Q_{20} = 582\ A_c^{0.758}$	54.9	0.90	42
2	$Q_{20} = 537\ L_c^{1.596}$	20.1	0.90	8
3	$Q_{20} = 509\ A_c^{0.975}\ L_c^{-0.402}$	37.2	0.95	90
4	$Q_{20} = 395\ A_c^{0.777}\ S_c^{0.169}$	38.6	0.93	78
5	$Q_{20} = 789\ A_c^{0.980}$	29.5	0.93	29
1	$Q_{50} = 824\ A_c^{0.758}$	54.9	0.90	42
2	$Q_{50} = 704\ L_c^{1.596}$	20.1	0.90	8
3	$Q_{50} = 649\ A_c^{0.975}\ L_c^{-0.402}$	37.2	0.95	90
4	$Q_{50} = 464\ A_c^{0.777}\ S_c^{0.169}$	38.6	0.93	78
5	$Q_{50} = 956\ A_c^{0.980}$	29.5	0.93	29
1	$Q_{100} = 1053\ A_c^{0.758}$	54.9	0.90	42
2	$Q_{100} = 847\ L_c^{1.596}$	20.1	0.90	8
3	$Q_{100} = 766\ A_c^{0.975}\ L_c^{-0.402}$	37.2	0.95	90
4	$Q_{100} = 512\ A_c^{0.777}\ S_c^{0.169}$	38.6	0.93	78
5	$Q_{100} = 1086\ A_c^{0.980}$	29.5	0.93	29

TABLE 3.44. Regression Models for Estimating the Expected GEV Flood Quantiles for Various Return Periods: Clustering on LCV, LSK and QSP (Case 3)

Cluster Region No.	Regression Equation	% Standard Error	R^2	No. of Sites.
1	$Q_{20} = 601\ A_c^{1.069}\ L_c^{-0.605}\ S_c^{-0.119}$	40.4	0.95	78
2	$Q_{20} = 694\ A_c^{0.776}$	51.0	0.90	16
3	$Q_{20} = 399\ A_c^{0.821}\ S_s^{0.227}\ B_s^{-0.194}$	32.7	0.91	74
4	$Q_{20} = 444\ A_c^{0.793}\ S_s^{-0.911}$	40.1	0.95	39
5	$Q_{20} = 912\ A_c^{0.887}$	27.0	0.91	37
1	$Q_{50} = 802\ A_c^{1.069}\ L_c^{-0.605}\ S_c^{-0.119}$	40.4	0.95	78
2	$Q_{50} = 1068\ A_c^{0.776}$	51.0	0.90	16
3	$Q_{50} = 482\ A_c^{0.821}\ S_s^{0.227}\ B_s^{-0.194}$	32.7	0.91	74
4	$Q_{50} = 492\ A_c^{0.793}\ S_s^{-0.911}$	40.1	0.95	39
5	$Q_{50} = 1148\ A_c^{0.887}$	27.0	0.91	37
1	$Q_{100} = 979\ A_c^{1.069}\ L_c^{-0.605}\ S_c^{-0.119}$	40.4	0.95	78
2	$Q_{100} = 1462\ A_c^{0.776}$	51.0	0.90	16
3	$Q_{100} = 544\ A_c^{0.821}\ S_s^{0.227}\ B_s^{-0.194}$	32.7	0.91	74
4	$Q_{100} = 523\ A_c^{0.793}\ S_s^{-0.911}$	40.1	0.95	39
5	$Q_{100} = 1345\ A_c^{0.887}$	27.0	0.91	37

TABLE 3.45. Regression Models for Estimating the Expected GEV Flood Quantiles for Various Return Periods: Clustering on GEV Parameters and QSP (Case 4)

Cluster Region No.	Regression Equation	% Standard Error	R^2	No. of Sites.
1	$Q_{20} = 497\ A_c^{0.963}\ L_c^{-0.396}$	38.1	0.94	80
2	$Q_{20} = 935\ A_c^{0.850}$	23.4	0.94	20
3	$Q_{20} = 679\ A_c^{1.225}\ L_c^{-0.809}$	52.2	0.93	38
4	$Q_{20} = 871\ A_c^{1.002}$	31.2	0.91	14
5	$Q_{20} = 566\ A_c^{0.962}\ L_c^{-0.456}$	38.1	0.95	63
6	$Q_{20} = 724\ A_c^{0.643}\ S_c^{-0.206}$	42.0	0.95	27
1	$Q_{50} = 635\ A_c^{0.963}\ L_c^{-0.396}$	38.1	0.94	80
2	$Q_{50} = 1246\ A_c^{0.850}$	23.4	0.94	20
3	$Q_{50} = 857\ A_c^{1.225}\ L_c^{-0.809}$	52.2	0.93	38
4	$Q_{50} = 988\ A_c^{1.002}$	31.2	0.91	14
5	$Q_{50} = 666\ A_c^{0.962}\ L_c^{-0.456}$	38.1	0.95	63
6	$Q_{50} = 1093\ A_c^{0.643}\ S_c^{-0.206}$	42.0	0.95	27
1	$Q_{100} = 757\ A_c^{0.963}\ L_c^{-0.396}$	38.1	0.94	80
2	$Q_{100} = 1525\ A_c^{0.850}$	23.4	0.94	20
3	$Q_{100} = 1002\ A_c^{1.225}\ L_c^{-0.809}$	52.2	0.93	38
4	$Q_{100} = 1062\ A_c^{1.002}$	31.2	0.91	14
5	$Q_{100} = 741\ A_c^{0.962}\ L_c^{-0.456}$	38.1	0.95	63
6	$Q_{100} = 1476\ A_c^{0.643}\ S_c^{-0.206}$	42.0	0.95	27

TABLE 3.46. Regression Models for Estimating the Expected GEV Flood Quantiles for Various Return Periods for USGS Regions

Region No.	Regression Equation	% Standard Error	R^2	No. of Sites.
1	$Q_{20} = 742\ A_c^{0.963}\ L_c^{-0.396}\ S_c^{0.196}$	38.4	0.95	29
2	$Q_{20} = 348\ A_c^{0.736}$	39.9	0.96	67
3	$Q_{20} = 684\ A_c^{0.744}\ S_c^{0.029}\ B_s^{-0.070}$	19.7	0.99	25
4	$Q_{20} = 292\ A_c^{0.842}\ S_s^{-0.517}$	27.1	0.99	19
5	$Q_{20} = 658\ A_c^{0.720}$	33.2	0.97	37
6	$Q_{20} = 636\ A_c^{0.624}\ S_s^{-0.277}$	38.6	0.96	27
7	$Q_{20} = 846\ A_c^{0.587}$	36.9	0.95	37
1	$Q_{50} = 936\ A_c^{0.963}\ L_c^{-0.396}\ S_c^{0.196}$	38.4	0.95	29
2	$Q_{50} = 430\ A_c^{0.736}$	39.9	0.96	67
3	$Q_{50} = 684\ A_c^{0.744}\ S_c^{0.029}\ B_s^{-0.070}$	19.7	0.99	25
4	$Q_{50} = 354\ A_c^{0.842}\ S_s^{-0.517}$	27.1	0.99	19
5	$Q_{50} = 851\ A_c^{0.720}$	33.2	0.97	37
6	$Q_{50} = 778\ A_c^{0.624}\ S_s^{-0.277}$	38.6	0.96	27
7	$Q_{50} = 1077\ A_c^{0.587}$	36.9	0.95	37
1	$Q_{100} = 1104\ A_c^{0.963}\ L_c^{-0.396}\ S_c^{0.196}$	38.4	0.95	29
2	$Q_{100} = 495\ A_c^{0.736}$	39.9	0.96	67
3	$Q_{100} = 812\ A_c^{0.744}\ S_c^{0.029}\ B_s^{-0.070}$	19.7	0.99	25
4	$Q_{100} = 404\ A_c^{0.842}\ S_s^{-0.517}$	27.1	0.99	19
5	$Q_{100} = 1020\ A_c^{0.720}$	33.2	0.97	37
6	$Q_{100} = 891\ A_c^{0.624}\ S_s^{-0.277}$	38.6	0.96	27
7	$Q_{100} = 1273\ A_c^{0.587}$	36.9	0.95	37

TABLE 3.47. Regression Models for Estimating the Expected log-Pearson Type III Flood Quantiles for Various Return Periods for USGS Regions (Choquette, 1988) **

Region No.	Regression Equation	Standard Error %	No. of Sites.
1	$Q50 = 56\ A_c^{0.959}\ S_c^{0.617}$	44.7	33
2	$Q50 = 670\ A_c^{0.777}\ B_s^{-0.356}\ S_s^{-0.803}$	33.9	77
3	$Q50 = 849\ A_c^{0.714}\ S_s^{-0.392}$	23.4	26
4	$Q50 = 363\ A_c^{0.780}$	26.7	20
5	$Q50 = 940\ A_c^{0.690}$	48.5	40
6	$Q50 = 74\ A_c^{0.873}\ S_c^{0.520}$	36.2	32
7	$Q50 = 1530\ A_c^{0.639}\ B_s^{-0.472}\ S_s^{-0.579}$	37.6	38
1	$Q100 = 51\ A_c^{0.978}\ S_c^{0.669}$	47.8	33
2	$Q100 = 798\ A_c^{0.777}\ B_s^{-0.373}\ S_s^{-0.862}$	35.1	77
3	$Q100 = 1030\ A_c^{0.711}\ S_s^{-0.447}$	24.6	26
4	$Q100 = 420\ A_c^{0.775}$	26.7	20
5	$Q100 = 1100\ A_c^{0.689}$	52.3	40
6	$Q100 = 76\ A_c^{0.882}\ S_c^{0.545}$	38.1	32
7	$Q100 = 1710\ A_c^{0.639}\ B_s^{-0.466}\ S_s^{-0.528}$	39.4	38

** Flood Regions delineated using Method of Residuals using WRC Bulletin 17-B guidelines with a gauged site and regionalized weighted skew.

CHAPTER 4

CONCLUSIONS

Based upon the FASTCLUS algorithm, cluster analysis is used to identify distinct flood regions for the State of Kentucky. Important statistical properties of the annual maximum flood (AMF) series and other watershed hydrologic data from 253 gauged sites in the State of Kentucky are used in the analysis. Clustering variables used in the study are the L-moments, namely the coefficients of variation, LCV, and skewness, LSK of normalized maximum annual flood series, the parameters of the EV1 and GEV probability distributions, and the specific mean annual flood, QSP, based on the raw maximum annual flood series. All clustering variables are further standardized prior to clustering to suppress effects of scale. A comparison of the regions delineated under the two approaches, namely, cluster analysis and method of residuals, is then carried out using the following steps: a) direct comparison of gauged stations assigned to each region; b) comparison of mean, median and range (difference between the maximum and minimum values) of distributional characteristics of all hydrological variables (response and attribute); c) performance of regionalized EV1 and GEV flood frequency models; d) results of discriminant analysis; and e) results of regression analysis relating regionalized estimates flood quantiles of various return periods, Q_{Ti}, to watershed physical and climatic characteristics (referred to as the attribute variables). The following conclusions are made in this study:

1. While the USGS method of residuals regions are or at least made to coincide with geographic or hydrologic boundaries, cluster regions do not. A comparison of actual gauged sites shows considerable difference. Cluster regions differentiate characteristics that control the underlying probability law of flood response, whereas the method of residuals does not address this issue directly.
2. For cluster regions the shape of the regionalized flood frequency growth curve depends on the clustering variables and the underlying probability distribution used. For EV1 distribution these growth curves are linear with normalized discharge ratio ranging from 0.0-5.0. In contrast, the GEV distribution growth curves become increasingly non-linear as the coefficients of variation and skew increase. The normalized discharge levels in this case range from 0.0-10.0. For the USGS regions the growth curves practically plot on one another indicating homogeneity of flood response between flood regions.
3. The regionalized EV1 and GEV flood frequency growth curves show more differences between cluster regions than between the seven USGS regions. This suggests that the cluster regions delineated for Cases 1-4 are homogeneous within themselves but distinct from each other in terms of their flood response when compared to the USGS regions. This property is essential for deriving maximum benefit from any flood regionalization effort.
4. An examination of statistical trends, like the mean and median, of important hydrologic variables (both response and attribute), such as the watershed area, A_c, indicates that the cluster regions have lower variability within each of the with respect to these parameters than the USGS regions. Cluster regions are

generally grouped into low, medium or large watershed drainage areas and flood response areas (as measured by the specific mean annual flood, QSP, and the mean annual flood). In contrast, the USGS flood regions have a mixed population within each of the seven regions, thereby giving similar values across regions. An exception to this are the trends observed for the basin shape, B_s, and channel sinuousity, S_s. These variables, by virtue of their definition, involving ratios of similar magnitude either small or large, show similar variation between regions for both cluster and USGS regions.

5. The performance of these regionalized EV1 and GEV flood frequency models, in terms of regional average bias (computed by taking the difference between simulated and historical estimates of flood quantiles) and RMSE (computed by taking the square of the bias) are comparable for cluster and USGS regions. For both models and all flood regions (cluster and USGS), the bias changes from positive to negative as the return period increases (i.e. the flood frequency growth curve becomes steeper) indicating an underestimation of flood quantiles. This trend in the bias is partly due to the condition of separation commonly found in Monte Carlo simulated flood data. This condition of separation causes simulated flow sequences to have less variability (with the separation increasing with return period) than the historical flood records resulting in an underestimation of flood quantiles when using simulated flows at a gauged site.
6. The **absolute** value of the regional average bias is less than 15% for all flood regions and flood frequency models when predicting flood quantiles having a return period less than 100 years. This indicates a high level of accuracy in the regionalized flood frequency models developed in the study. In some cases, however,

the RMSE is as high as 44% for regions having the steepest flood frequency growth curves indicating a lack of precision or fit. Paradoxically, this occurs with the GEV flood frequency model that should provide a better fit considering the fact that ot has one additional parameter to capture the high skew commonly found in flood data. By and large the RMSE is less than 20% for most cluster and USGS flood regions.

7. In both methods of delineating flood regions, the geomorphic properties of the watersheds such as the drainage area, basin shape, basin length, and main channel length, slope and sinuousity provide the maximum discrimination between flood regions. Discrimination is based on the physical and climatic characteristics of the watersheds (referred to as attribute variables). Watershed contributing drainage area is the most important variable. The USGS flood regions have a low overall discrimination (16.9%) compared to the cluster regions, which have a higher overall discrimination of 47.1%. This further supports the mixed hydrologic composition within the USGS regions.
8. For both methods of regionalization, the significant variables (at a 5% level) in the regression analysis relating EV1 and GEV flood quantiles, Q_{Ti}, of various return periods to watershed physical and climatic characteristics, are geomorphic properties of the watersheds (as was the case with discriminant analysis) with the watershed contributing drainage area, A_c, as the most important variable.
9. The standard errors associated with the regression equations are comparable for both methods. For cluster regions, where the problem of simultaneously classifying gauged and ungauged sites into several cluster regions exists, the weighted standard errors (based on the posterior probabilities and the

corresponding regression equations of the cluster regions to which a site is assigned) are used in making this comparison.

10. The hydrological characteristics of flood regions and their overall performance delineated using Method 1 (clustering on L-moments and QSP) are similar to those of Method 2 (clustering on parameters of the EV1 and GEV probability distributions). However, the actual gauged sites within each region are quite different.
11. A comparison of flood quantile estimates at selected sites indicates that the regionalized EV1 and GEV flood frequency models underestimate flood levels (50 and 100 year return periods) when compared to the log-Pearson Type-III flood frequency model.
12. Overall it appears that regionalized EV1 and GEV flood frequency models, in conjunction with the method of L-moments to estimate their parameters, would better represent flood experience in Kentucky even when using the present flood regions as defined using the method of residuals. This observation is based on the performance of these models in terms of bias and RMSE and not on a direct comparison with a regionalized log-Pearson Type-III flood frequency distribution.

NOMENCLATURE

The following symbols and variables are used in this study:

LCV = L-moment ratio, t_2 (coefficient of variation), of normalized AMF series;

LSK = L-moment ratio, t_3 (coefficient of skewness), of normalized AMF series;

LKUR = L-moment ratio, t_4 (coefficient of kurtosis), of normalized AMF series;

LBMD = L-moment ratio, t_5 (coefficient of bi-modality), of normalized AMF series;

CV = coefficient of variation of raw AMF series;

SK = coefficient of skewness of raw AMF series;

KUR = coefficient of kurtosis of raw AMF series;

M(LCV) = regional weighted mean L-moment ratio, t_2 (coefficient of variation);

M(LSK) = regional weighted mean L-moment ratio, t_3 (coefficient of skewness);

M(LKUR) = regional weighted mean L-moment ratio, t_4 (coefficient of kurtosis);

M(LBMD) = regional weighted mean L-moment ratio, t_5 (coefficient of bi-modality);

AMF = raw or normalized annual maximum floodpeak series;

X = normalized annual maximum flood value = Q/Q;

Q = raw annual maximum flood value in cfs;

$\bar{Q}$ = mean of raw AMF series in cfs (same as STMEAN);

Q_I = index-flood in cfs;

Q_{Ti} = gauged site i estimate of flood level having a return period of T years in cfs.

q_T = regionalized flood quantile estimate of normalized AMF series;

QSP = specific mean annual flood = Q/A_c in cfs/sq.

mile (same as SMDISCH);

F = cumulative probability density function (cdf);

f = probability density function (pdf);

E(X) = expected value of random variable X;

P^*_r (F) = rth shifted Legendre polynomial of function F;

PWM = probability weighted moment;

$M_{p,r,s}$ = probability weighted moment;

α_r = probability weighted moment of order r, $M_{1,0,r}$;

β_r = probability weighted moment of order r, $M_{1,r,0}$;

λ_r = L-moment of order r;

τ_r = L-moment ratio of order r;

a_r = sample estimate of PWM α_r ;

b_r = sample estimate of PWM β_r ;

l_r = sample estimate of the r^{th} L-moment;

t_r = sample estimate of the r^{th} L-moment ratio;

MEVL = sample estimate of the location parameter of the EV1 probability or frequency distribution;

AEVL = sample estimate of the scale parameter of the EV1 probability or frequency distribution;

MGVL = sample estimate of the location parameter of the GEV probability or frequency distribution;

AGVL = sample estimate of the scale parameter of the GEV probability or frequency distribution;

KGVL = sample estimate of the shape parameter of the GEV probability or frequency distribution;

EV1 = Extreme Value Type-1 or Gumbel probability distribution;

GEV = Generalized Extreme Value probability distribution;

WAK = Wakeby probability distribution;

A_c = watershed or basin contributing drainage area in square miles (same as DAREA);

B_w = watershed or basin width in miles;

B_l = watershed or basin length in miles (same as BASLEN);

B_s = watershed shape index = (A_c / B_l) (same as

SHAPE);

BELEV = watershed or basin mean elevation in feet;

PRECIP = watershed or basin mean annual precipitation in inches;

STOR = watershed or basin storage in percent;

SINFL = watershed or basin average soil infiltration in in/hr;

BASIN = watershed or basin designation;

L_c = main channel length in miles (same as CHANLEN);

S_s = main channel sinuousity = (L_c / B_1) (same as CHANSIN);

S_c = main channel slope in percent (same as CHANSLOP);

ELEV = gauged site mean elevation in feet;

ISTN = gauged site USGS Station Number;

T = return period in years;

N = number of years of AMF data at a gauged site;

USREG = region assigned to gauged site using method of residuals;

REFERENCES

1. Ackerman, M. C., and Sinclair, C. D. (1986). "Classification of Drainage Basins According to Their Physical Characteristics; An Application for Flood Frequency Analysis in Scotland." Journal of Hydrology, Elsevier Science Publishers, Vol. 84, 365-380.

2. Choquette, A. F. (1988). "Regionalization of Peak Discharges Based for Streams in Kentucky." Water Resources Investigations Report 87-4209, U.S. Geological Survey, Louisville District, Louisville, Kentucky, pp. 105.

3. Cunane, C. (1987). "Review of Statistical Models for Flood Frequency Estimation." Hydrologic Frequency Modeling, Proceedings of the International Symposium on Flood Frequency and Risk Analysis, Edited by V. P. singh, Reidel Publishing Company, Boston, 49-95.

4. Darymple, T. (1960). "Flood Frequency Analysis." In the Manual of Hydrology, Part 3, U.S. Geological Survey Water-Supply Paper 1543-A, Washington, D.C.: Government Printing Office.

5. DeCoursey, D. G., and Deal, R. B. (1974). "General Aspects of Multivariate Analysis With Applications to Some Problems in Hydrology." Proceedings, Symposium on Statistical Hydrology, Misc. Publication No. 1275, USDA-ARS, 47-68.

6. Duda, R. O., and Hart, P. E. (1973). "Pattern Classification and Scene Analysis." John Wiley and Sons, New York.

7. Fiorentino, M., Versace, P., and Rossi, F. (1985). "Regional Flood Frequency Estimation Using the Two-Component Extreme Value Distribution." Hydrological Sciences Journal, Vol. 30, 51-64.

8. Greis, N. P. (1983). "Flood Frequency Analysis: A Review of 1979-1982." Rev. Geophysical Space Physics, Vol. 21 (3), 699-706.

9. Greenwood, A. J., Landwehr, J. M., Matalas, N. C., and Wallis, J. R. (1979). "Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressable in Inverse Form." Water Resources Research, American Geophysical Union, Vol. 15, No. 5, 1049-1054.

10. Hosking, J. R. M. (1986). "The Theory of Probability Weighted Moments." IBM Research Report, No. RC 12210 (# 54860), 1-160.

11. Hosking, J. R. M., and Wallis, J. R. (1987). "An Index-Flood Procedure for Regional Rainfall Frequency Analysis." Abstract, American Geophysical Union, EOS, Vol. 68, No. 16.

12. Hosking, J. R. M. (1988). "Fortran Routines for Use With the Method of L-Moments." IBM Research Report, No. RC 13844 (# 62140), 1-80.

13. Hosking, J. R. M. (1989). "L-Moments: Analysis and Estimation of Distributions Using Linear Combination of Ordered Statistics." accepted for publication in J. R. Statistical Society.

14. Houghton, J. C. (1977). "Robust Estimation of the Frequency of Extreme Events in a Flood Frequency

Context." Ph.D. Dissertation, Div. of Applied Science, Harvard University, Cambridge, Massachusetts.

15. Kuczera, G. (1982). "Robust Flood Frequency Models." Water Resources Research, American Geophysical Union, Vol. 18, No. 2, 315-324.

16. Landwehr, J. M., Matalas, N. C. and Wallis, J. R. (1979 a). "Probability Weighted Moments Compared with Some Traditional Techniques in Estimating Gumbel Parameters and its Quantiles." Water Resources Research, American Geophysical Union, Vol. 15, No. 5, 1055-1064.

17. Landwehr, J. M., Matalas, N. C. and Wallis, J. R. (1979 b). "Estimation of Parameters and Quantiles of Wakeby Distributions 1. Known Lower Bounds." Water Resources Research, American Geophysical Union, Vol. 15, No. 6, 1361-1372.

18. Landwehr, J. M., Matalas, N. C. and Wallis, J. R. (1980). "Quantile Estimation with More or Less Floodlike Distributions." Water Resources Research, American Geophysical Union, Vol. 16, No. 3, 547-555.

19. Letttenmaier, D. P. and Potter, K. W. (1985). "Testing Flood Frequency Estimation Methods Using a Regional Flood Generation Model." Water Resources Research, American Geophysical Union, Vol. 21, No. 12, 1903-1914.

20. Lettenmaier, D. P., Wallis, J. R. and Wood, E. F. (1987). "Effect of Regional Heterogeneity on Flood Frequency Estimation." Water Resources Research, American Geophysical Union, Vol. 23, No. 2, 313-323.

21. Matalas, N. C., Slack, J. R. and Wallis, J. R. (1975). "Regional Skew in Search of a Parent." Water Resources

Research, American Geophysical Union, Vol. 11, No. 6, 815-826.

22. McCabe, J. A. (1962). "Floods in Kentucky - Magnitude and Frequency." Kentucky Geological Survey, Information Circular No: 9, Series X.

23. McCain, J. F., and Jarrett, R. D. (1976). "Manual for Estimating Flood Characteristics of Natural-Flow Streams in Colorado." Colorado Water Conservation Board, Technical Manual No. 1, pp. 68.

24. Milligan, G. W., and Cooper, M. C. (1983). "An Examination of Procedures for Determining the Number of Clusters in a Data Set." College of Administrative Science Working Paper Series 83-51, The Ohio State University, Columbus, Ohio, pp. 38.

25. Mosley, M. P. (1981). "Delimitation of New Zealand Hydrological Regions." Journal of hydrology, Vol. 49, 173-192.

26. Richter, B. D., Kircher, J. E., Remmers, M. A., and Forst, B. A. (1984). "Summary of Basin and Streamflow Characteristics for Selected Basins in Western Colorado and Adjacent Streams." U. S. Geological Survey, Open-File Report 84-137, pp. 226.

27. Riggs, H. C. (1973). "Regional Analysis of Streamflow Characteristics." Techniques of Water Resources Investigations, Chapter B3, Book 4, U.S. Geological Survey, pp. 15.

28. Sauer, V. B. (1964). "Floods in Louisiana, Magnitude and Frequency." Second Edition. Louisiana Department of Highways, pp. 402.

29. Sokol, R. R., and Sneath (1963). "Principles of Numerical Taxonomy." W. H. Freeman and Company, San Franscisco, California.

30. Statistical Analysis System. (1985). SAS User's Guide: Statistics: Version 5 Edition, SAS Institute Inc., North Carolina, pp. 956.

31. Stedinger, J. R. (1983). "Estimating a Regional Flood Frequency Distribution." Water Resources Research, American Geophysical Union, Vol. 19, No. 2, 503-510.

32. Tasker, G. D. (1982). "Comparing Methods of Hydrologic Regionalization." Water Resources Bulletin, Journal, American Water Resources Association, Vol. 18(6), 965-970.

33. Thomas, D. M., and Benson, M. A. (1970). "Generalization of Streamflow Characteristics From Drainage-Basin Characteristics." U. S. Geological Survey Water-Supply Paper 1975, pp. 55.

34. U. S. Department of the Interior. (1982). Guidelines for Determining Flood Flow Frequency, Bulletin 17-B. Hydrology Subcommittee, Interagency Advisory Committee on Water Data, U. S. Geological Survey, Reston, Virginia, pp. 28.

35. Wallis, J. R., and Wood, E. F. (1985). "Relative Accuracy of Log Pearson Type-III Procedures." Journal of Hydraulic Engineering, ASCE, Vol. 111, No. 7, 1043-1056.

36. Wiltshire, S. E. (1986). "Regional Flood Frequency Analysis II: Multivariate Classification of Drainage

Basins in Britain." Hydrological Sciences Journal, 31(3), 335-346.

APPENDIX A

TABLE A1. Important Statistics of Hydrological Characteristics of Gauged Sites Used in the Study.

OBS	ISTN	BASIN	USRES	N	STMEAN	DAREA	SHDITCH	SHAPE	BASLEN	CHANLEN	CHANSIN	CHANSLOP	LCV	LSKEW	LKUR
1	22500	JPKY	7	31	818.7	1.72	476.013	2.033	1.87	2.1000	1.123	0.53228	0.2932	0.2940	0.3167
2	23000	JPKY	7	40	7506.0	212.00	25.409	5.549	34.30	41.0000	1.219	0.10038	0.2584	0.3144	0.2358
3	23040	JPKY	7	11	215.1	0.53	405.832	1.447	0.94	1.6200	1.005	1.10966	0.2786	0.2052	0.2078
4	23500	JPKY	7	38	3943.4	36.80	107.159	3.330	11.07	13.9000	1.250	0.20781	0.1051	0.0073	0.0897
5	23935	JPKY	7	11	125.8	0.23	547.036	1.726	0.63	0.9500	1.500	1.12235	0.2254	0.0980	0.1902
6	24000	JPKY	7	43	3610.2	60.70	52.551	3.319	15.10	18.2000	1.205	0.15152	0.2581	0.2070	0.1894
7	24500	JPTN	7	22	4654.1	110.00	60.472	2.557	16.77	24.3000	1.449	0.07045	0.3381	0.3343	0.1536
8	202400	BSHV	2	16	11580.0	306.00	37.843	2.939	29.99	37.5000	1.250	0.66667	0.3743	0.2572	0.2226
9	203000	USHV	2	50	21894.6	750.00	29.805	2.054	39.46	86.9000	2.202	0.24432	0.2644	0.1372	0.0799
10	203600	BSHV	2	19	27590.9	836.00	33.013	2.887	49.13	99.2000	2.019	0.19897	0.2920	0.1163	0.0801
11	204500	BSHV	2	43	6849.3	256.00	26.755	3.574	30.25	48.5000	1.603	0.07765	0.2513	0.2106	0.1715
12	206600	BSHV	2	20	1890.1	38.50	49.095	4.692	13.44	18.2800	1.360	0.41098	0.2408	0.2910	0.2757
13	207000	BSHV	2	32	7406.6	891.00	26.139	4.416	36.40	52.3000	1.437	0.10795	0.3057	0.2560	0.1824
14	207020	BSHV	2	36	7590.0	300.00	25.303	4.200	35.53	52.3000	1.472	0.10795	0.2940	0.2312	0.1877
15	207400	BSVA	2	27	1359.7	19.88	68.674	1.130	4.73	6.3000	1.332	2.25379	0.5247	0.4884	0.3717
16	207500	BSVA	2	34	12283.5	235.00	52.270	1.932	21.31	23.5000	1.103	0.61318	0.2969	0.1472	0.1678
17	207962	BSKY	2	10	74.8	0.82	91.228	1.641	1.16	1.4444	1.262	9.65678	0.4313	0.1902	0.0445
18	207965	BSKY	2	9	877.0	6.20	141.452	4.065	5.02	5.3900	1.074	2.43580	0.2564	0.1918	0.2040
19	208000	BSKY	2	31	13446.5	392.00	34.200	3.980	39.50	52.7000	1.234	0.32008	0.2745	0.2332	0.1952
20	208560	BSVA	2	38	16781.4	284.00	58.676	1.012	17.01	23.5000	1.282	0.36553	0.3458	0.2434	0.1850
21	208950	BSVA	2	21	3825.3	66.50	57.523	2.048	13.81	17.0000	1.231	0.80492	0.4292	0.3880	0.3040
22	209000	BSVA	2	39	11173.0	221.00	50.560	3.021	25.84	41.3000	1.598	0.19318	0.2895	0.2615	0.2315
23	209300	BSKY	2	8	24462.5	554.00	47.766	1.239	26.20	40.2700	1.537	0.40152	0.2957	0.2729	0.1962
24	209575	BSKY	2	10	330.3	3.17	166.719	0.682	1.47	2.5900	1.762	3.42530	0.3045	0.4707	0.2525
25	210000	BSKY	2	46	3044.0	56.30	54.067	3.095	13.20	21.5000	1.429	0.46023	0.2800	0.1073	0.0918
26	211500	BSKY	2	14	4983.4	206.00	24.191	5.096	32.40	55.1000	1.701	0.12121	0.2084	0.0192	0.0611
27	212000	BSKY	2	31	6267.4	183.00	60.049	2.047	14.59	21.000	1.439	0.15625	0.3413	0.1922	0.1721
28	212515	BSKY	2	7	181.0	2.10	86.190	0.756	1.26	2.190	1.738	1.72727	0.3450	0.3070	0.1712
29	213700	BSHV	2	17	25299.4	936.00	27.029	3.378	56.23	99.600	1.771	0.19318	0.3083	0.3640	0.3402
30	214500	LSKY	2	30	10911.0	400.00	27.277	1.756	26.50	47.200	1.701	0.07000	0.2499	0.2175	0.2015
31	214505	LSKY	2	9	40.0	0.51	79.956	1.807	0.96	1.020	1.062	3.20989	0.3524	0.1902	0.1229
32	214540	LSKY	2	13	861.9	12.20	70.649	3.110	6.16	8.730	1.417	0.34716	0.1047	0.2801	0.2375
33	214563	LSKY	2	16	206.0	0.94	219.149	2.394	1.50	1.570	1.047	1.60947	0.2467	0.1217	0.2305
34	214564	LSKY	2	9	205.2	1.61	177.157	0.172	2.26	2.430	1.075	1.40114	0.2219	0.1160	0.1107
35	216000	TYKY	2	29	5217.6	59.60	87.543	1.386	9.09	12.000	1.320	0.33674	0.2647	0.0703	0.0381
36	216901	TYKY	2	10	198.5	1.11	178.029	2.403	1.66	1.070	1.187	1.96534	0.2750	0.2774	0.3144
37	217000	TYKY	2	45	8293.3	242.00	34.270	3.300	20.60	61.200	2.140	0.00712	0.2961	0.3061	0.2998
38	237200	LKON	2	22	1516.9	12.20	124.333	1.034	4.73	6.913	1.462	1.26094	0.4374	0.4106	0.3508
39	237500	LKON	2	53	23302.8	387.00	60.214	2.262	29.59	45.900	1.551	0.15720	0.2413	0.2336	0.2176
40	237900	LKKY	1	14	4230.0	22.40	193.304	1.014	4.77	5.770	1.810	0.61269	0.2488	0.0901	0.1046
41	238030	LKKY	1	11	348.0	1.90	183.158	1.247	1.68	2.280	1.425	1.00568	0.3370	0.3284	0.1898
42	238500	LKON	1	57	10873.7	218.00	49.879	2.925	25.25	44.400	1.750	0.15000	0.2230	0.1768	0.1183
43	246500	LKON	1	20	11163.0	237.00	47.101	3.406	28.41	52.700	1.855	0.09981	0.2136	0.1615	0.1089
44	247100	LKON	1	31	586.6	3.34	175.623	.	.	.	.	0.60417	0.1637	0.0677	0.2141
45	247500	LKON	1	57	21211.9	476.00	44.563	2.419	33.93	80.300	2.367	0.13125	0.2165	0.1420	0.1306
46	248500	LKKY	2	47	4834.5	140.00	34.532	2.915	20.20	36.400	1.802	0.08902	0.3125	0.2191	0.1919
47	249500	LKKY	2	36	12082.2	827.00	14.610	3.631	34.80	81.020	1.470	0.06231	0.2172	0.1484	0.2328
48	250000	LKKY	2	41	7982.2	47.50	160.046	1.402	8.16	9.700	1.189	0.55682	0.3557	0.3400	0.2792
49	250080	LKKY	2	10	43.0	0.19	226.316	2.728	0.72	0.800	1.111	3.11326	0.3281	0.1097	0.0778
50	250100	LKKY	2	18	6136.1	84.70	72.445	1.444	11.80	18.500	1.568	0.29924	0.1680	0.0413	0.2113
51	250150	LKKY	2	11	733.3	2.43	301.833	1.867	2.13	2.760	1.296	0.86174	0.3140	0.3591	0.1773
52	250320	LKKY	2	9	1106.0	4.01	275.810	1.100	2.10	3.050	1.452	1.01250	0.2985	0.2213	0.0361

OBS	CV	SK	KUR	MEVL	AEVL	MGVL	AGVL	KGVL	MWKL	AWKL	BWKL	CWKL	DWKL
1	0.60646	2.32171	7.4049	0.7550	0.4230	0.7240	0.3452	-0.1854	0.1744	2.575	4.284	0.1882	0.4405
2	0.49755	1.52084	2.4518	0.7848	0.3726	0.7532	0.2933	-0.2131	0.0000	20.683	37.583	0.4154	0.1049
3	0.40473	0.80158	0.9047	0.7600	0.4019	0.7584	0.3013	-0.0541	-0.2637	20.228	25.868	0.5267	-0.0309
4	0.31750	0.24681	-0.9546	0.0458	0.2471	0.0631	0.2978	0.1328	0.3353	4.142	16.135	0.6192	-0.4617
5	0.34947	0.31147	-0.6120	0.0123	0.3252	0.0304	0.3579	0.1153	0.1949	3.149	5.357	0.2839	0.0678
6	0.47232	1.17804	2.0501	0.7851	0.3723	0.7757	0.3523	-0.0569	0.1795	2.927	9.823	0.4634	-0.0125
7	0.63090	1.27413	0.8056	0.7184	0.4878	0.6727	0.3495	-0.2408	0.3245	0.000	6.000	0.6740	0.0023
8	0.72459	1.78950	4.7719	0.4883	0.5400	0.6504	0.4709	-0.1315	0.0834	1.560	1.191	0.0679	0.6615
9	0.46292	0.53818	-0.3953	0.7798	0.3015	0.7090	0.3992	0.0515	0.1854	4.182	20.288	0.8757	-0.4172
10	0.49911	0.31923	-0.8100	0.7568	0.4213	0.7739	0.4531	0.0053	0.1895	1.078	2.687	0.4506	-0.2560
11	0.45617	1.04108	0.9135	0.7908	0.3123	0.7808	0.3410	-0.4623	0.2683	2.452	7.057	0.4721	-0.0381
12	0.48017	2.27279	7.2908	0.7994	0.3475	0.7739	0.2855	-0.2910	0.3264	2.262	5.780	0.2439	0.2027
13	0.57118	1.42515	2.4456	0.7454	0.4410	0.7211	0.3049	-0.1308	0.2643	1.024	1.991	0.3100	0.2116
14	0.54095	1.41192	2.8933	0.7552	0.4242	0.7301	0.3463	-0.0931	0.2538	1.277	1.962	0.2208	0.2991
15	1.22030	2.88579	9.5689	0.5431	0.7570	0.4544	0.4050	-0.4430	-0.3742	16.485	28.179	0.4053	0.4004
16	0.53704	0.87311	1.7476	0.7528	0.4283	0.7599	0.4422	0.0357	0.1368	1.968	2.401	0.2050	0.2797
17	0.75951	0.47105	-0.8446	0.6408	0.6223	0.6321	0.6040	-0.0313	-0.8103	1.386	0.361	0.4000	0.0000
18	0.44314	0.69518	1.4824	0.7845	0.3699	0.7869	0.3582	-0.0338	-0.7296	44.198	32.422	0.3734	0.1003
19	0.50614	1.24831	1.8114	0.7714	0.3960	0.7551	0.7450	0.2437	0.2222	2.415	4.707	0.4400	0.0525
20	0.65452	1.48468	3.1523	0.7120	0.4909	0.6884	0.4452	-0.1113	0.0894	2.004	5.075	0.5444	0.0590
21	0.94837	2.99727	11.1595	0.6426	0.6192	0.5714	0.4199	-0.3135	0.1178	1.103	0.412	0.0377	0.8108
22	0.54651	1.34650	2.0026	0.7589	0.4177	0.7348	0.3616	-0.1377	-0.0344	10.055	10.015	0.4460	0.0783
23	0.51289	1.01475	0.1727	0.7538	0.4266	0.7263	0.3621	-0.1541	0.0000	12.099	26.421	0.5556	0.0057
24	0.63790	2.33813	5.8712	0.7465	0.4392	0.6844	0.2409	-0.4308	0.0000	30.904	58.811	0.2058	0.2671
25	0.50651	0.54100	0.0233	0.7402	0.4155	0.7801	0.4528	0.0999	0.1951	1.493	1.356	0.1239	0.2747
26	0.48609	-0.12032	-1.1186	0.7590	0.4161	0.8127	0.4999	0.2498	0.0953	1.973	1.750	0.1750	0.0707
27	0.42086	1.02411	1.2133	0.7150	0.4923	0.7083	0.4763	-0.0344	0.0032	1.750	2.145	0.2627	0.2660
28	0.60644	1.10873	0.5458	0.7127	0.4470	0.6720	0.3966	-0.2038	0.0000	19.327	60.440	0.4944	-0.0160
29	0.42753	2.41118	7.2866	0.6767	0.5601	0.6171	0.3990	-0.2024	0.0549	2.199	3.398	0.2249	0.4947
30	0.45109	0.94061	0.8217	0.7919	0.3605	0.7805	0.3356	-0.0726	-0.5742	63.320	56.352	0.4903	-0.0511
31	0.61702	0.60579	-0.4476	0.7065	0.5085	0.6994	0.4936	-0.0313	-0.7655	339.512	160.012	0.9599	-0.2513
32	0.31563	1.28467	1.4430	0.8462	0.2664	0.8282	0.2233	-0.1644	0.1312	21.806	39.443	0.3032	0.0001
33	0.45518	0.35491	0.3312	0.7779	0.3840	0.7919	0.4110	0.0765	-0.0493	5.995	7.965	0.3450	0.0933
34	0.26955	0.21182	-1.1260	0.8069	0.2346	0.8205	0.3600	0.0056	0.2765	2.074	6.180	0.3953	-0.2840
35	0.45295	0.13855	-1.2171	0.7796	0.3819	0.8099	0.4363	0.1613	0.0000	34.011	110.742	0.0302	0.0381
36	0.51486	1.59357	3.7199	0.7703	0.3779	0.7439	0.3350	-0.1606	0.0406	6.269	9.103	0.2732	0.2969
37	0.60950	2.26524	7.4337	0.7534	0.4272	0.7189	0.3415	-0.2054	0.0803	5.040	9.453	0.3161	0.2776
38	0.99450	2.96870	18.7294	0.6358	0.6310	0.5584	0.4072	-0.3434	-0.0061	2.373	4.573	0.3236	0.4424
39	0.45788	1.42052	2.4270	0.7990	0.3482	0.7845	0.3158	-0.0966	0.3362	1.624	3.844	0.2500	0.2145
40	0.42539	0.42191	0.4225	0.7928	0.3589	0.8152	0.3988	0.1282	0.1799	2.419	2.939	0.1349	0.3447
41	0.62273	1.30902	0.8174	0.7194	0.4961	0.6751	0.3724	-0.2326	0.0000	22.319	63.087	0.6292	0.0346
42	0.39513	0.76238	0.0509	0.8143	0.3217	0.8128	0.3185	-0.0107	0.3748	1.520	8.128	0.5600	-0.2209
43	0.36161	0.56501	-0.5269	0.8221	0.3082	0.8240	0.3119	0.0131	0.4026	1.213	6.710	0.5500	-0.2500
44	0.28300	0.06525	0.3319	0.8636	0.2362	0.8830	0.2695	0.1657	0.0745	12.601	18.626	0.3224	-0.1374
45	0.30010	0.60266	-0.1267	0.8197	0.3123	0.8259	0.3244	0.0427	0.2559	4.399	13.682	0.5563	-0.2516
46	0.50320	1.29910	2.1980	0.7398	0.4509	0.7258	0.4186	-0.0750	0.1465	1.922	3.842	0.3956	0.1334
47	0.39046	0.71375	0.5263	0.8192	0.3133	0.8241	0.3229	0.0337	0.1257	6.107	10.359	0.3161	0.0612
48	0.85170	3.90249	19.9326	0.7038	0.5132	0.6544	0.3843	-0.2486	0.2204	1.167	0.470	0.0116	0.9034
49	0.56136	0.51796	-0.9575	0.7267	0.4734	0.7203	0.4599	-0.0306	0.0000	0.320	0.190	0.0770	0.0199
50	0.27777	-0.05217	0.0928	0.8601	0.2424	0.8858	0.2847	0.2110	0.2240	5.487	9.463	0.2630	-0.0516
51	0.57470	1.23541	0.4088	0.7385	0.4530	0.6914	0.3265	-0.2746	0.3099	0.000	0.000	0.5756	0.0560
52	0.50999	0.48799	-1.5710	0.7514	0.4307	0.7367	0.3904	-0.0784	0.3208	0.866	0.275	0.0000	0.0000

TABLE A1 (cont..). Important Statistics of Hydrological Characteristics of Gauged Sites Used in the Study.

OBS	ISTN	BASIN	USREG	N	STMEAN	DAREA	SHDISCH	SHAPE	BASLEN	CHANLEN	CHANSIN	CHANSLOP	LCV	LSKEW	LKUR
53	250620	LKKY	2	9	94.3	0.33	285.859	2.509	0.91	1.030	1.132	2.46402	0.2734	0.1310	0.0229
54	251000	LKKY	2	40	6075.0	117.00	51.850	2.711	17.96	32.200	1.793	0.04924	0.2021	0.1766	0.2679
55	251008	LKKY	2	9	117.0	0.93	125.806	0.293	0.53	0.650	1.226	2.33464	0.4708	0.3850	0.3037
56	251015	LKKY	2	11	99.0	0.45	220.000	6.347	1.09	1.760	1.041	1.53708	0.3008	0.4232	0.3030
57	252000	LKKY	2	30	8607.0	239.00	36.013	1.401	18.30	54.200	3.071	0.04545	0.2282	0.1091	0.2259
58	252500	LKKY	2	68	18099.3	615.00	29.430	1.592	31.44	85.900	2.732	0.04564	0.2031	0.0154	0.1735
59	254400	LFKY	1	15	3987.3	13.60	293.186	1.541	4.57	5.710	1.244	0.52504	0.4547	0.5705	0.4748
60	260010	KYKY	1	12	256.0	0.68	377.574	1.304	0.97	1.320	1.362	1.53570	0.3730	0.2714	0.2313
61	260012	KYKY	1	13	382.4	1.62	236.040	1.934	1.77	2.220	1.254	2.61627	0.5013	0.5130	0.4244
62	277070	KYKY	1	10	537.1	1.54	348.766	0.636	0.99	1.400	1.414	0.99011	0.3708	0.1509	0.1402
63	277105	KYKY	1	10	282.1	0.68	414.853	2.261	1.24	1.720	1.387	3.90152	0.2950	-0.0028	0.0460
64	277250	KYIN	1	10	38.6	0.16	241.250	.	.	.	.	1.74621	0.2590	0.1563	0.0430
65	277259	KYKY	2	9	221.0	3.03	72.937	2.606	2.81	3.690	1.313	5.22022	0.4108	0.3662	0.2528
66	277300	KYKY	2	27	2491.2	66.40	37.518	2.429	12.70	17.290	1.354	0.38932	0.3571	0.3105	0.2198
67	277400	KYKY	3	19	3064.6	40.90	74.929	1.161	6.87	13.590	1.957	0.96970	0.3541	0.2030	0.1185
68	277450	KYKY	2	12	3342.5	60.60	55.157	2.662	12.70	19.810	1.560	0.32336	0.2038	0.2250	0.1484
69	277500	KYKY	3	36	20600.1	466.00	44.295	1.588	27.29	61.600	2.265	0.14015	0.2617	0.1102	0.1813
70	277630	KYKY	3	10	286.2	1.32	216.818	1.939	1.60	2.330	1.456	3.56894	0.3170	0.3412	0.0998
71	278000	KYKY	2	28	272.0	2.21	123.061	2.542	2.37	2.600	1.097	2.36742	0.3029	0.1530	0.0992
72	278500	KYKY	2	32	10636.9	177.00	60.095	2.170	19.60	35.700	1.821	0.16667	0.3112	0.1230	0.0010
73	280600	KYKY	3	29	16781.4	202.00	83.076	1.448	17.10	27.200	1.571	0.46402	0.4011	0.4045	0.2072
74	280700	KYKY	3	28	5303.2	61.30	86.512	2.800	13.10	18.400	1.405	0.84470	0.3431	0.2964	0.1592
75	280720	KYKY	3	10	330.3	1.84	183.059	1.850	1.39	1.880	1.353	3.03045	0.3373	0.5402	0.2806
76	280935	KYKY	3	11	257.0	1.57	163.694	3.430	2.39	2.730	1.142	2.45133	0.4916	0.4944	0.3432
77	281000	KYKY	3	26	17234.9	537.00	32.099	4.730	50.40	75.000	1.085	0.08902	0.2164	0.3516	0.2436
78	281040	KYKY	2	13	12663.8	155.00	81.702	2.021	17.70	29.300	1.655	0.31061	0.2231	0.2303	0.2298

OBS	CV	SK	KUR	MEVL	AEVL	MGVL	AGVL	KGVL	MAKL	AAKL	BAKL	CAKL	DAKL
53	0.45917	0.23170	-2.0132	0.7724	0.3944	0.7830	0.4161	0.0415	0.3066	1.065	0.537	0.0000	0.0000
54	0.37633	1.16070	2.0366	0.8317	0.2915	0.8264	0.2603	-0.0411	0.2248	5.265	9.980	0.2439	0.1752
55	0.97382	2.14015	5.5013	0.6079	0.6792	0.5305	0.4637	-0.3095	-0.1060	2.706	5.762	0.4570	0.3387
56	0.66708	2.56760	7.5621	0.7429	0.4455	0.6065	0.2794	-0.2578	-0.2029	21.933	28.375	0.2762	0.3946
57	0.41277	0.77705	2.0765	0.8099	0.3293	0.8252	0.3574	0.0971	0.1403	3.083	5.074	0.2421	0.1609
58	0.35202	0.10911	0.1538	0.8309	0.2927	0.8693	0.3533	0.2564	0.1917	3.190	5.272	0.2529	0.0028
59	1.18089	3.30943	11.7392	0.6214	0.6557	0.5185	0.2787	-0.5431	-1.0225	117.048	84.263	0.2944	0.5403
60	0.70594	1.31622	1.1905	0.6894	0.5382	0.6478	0.4419	-0.1806	-0.0553	4.767	10.398	0.5586	0.1228
61	1.24621	3.08500	10.3377	0.5825	0.7232	0.4752	0.3626	-0.4735	-0.0767	3.644	8.565	0.3436	0.3059
62	0.68232	0.59033	0.2367	0.6766	0.5638	0.6824	0.5792	0.0297	-0.1329	2.510	3.382	0.5432	0.0304
63	0.50116	-0.37524	-1.3165	0.7544	0.4254	0.8170	0.5222	0.2807	0.0000	2.134	6.313	1.2207	-0.7244
64	0.43285	0.35679	-1.4084	0.7843	0.3737	0.7805	0.3320	0.0245	0.3778	-0.308	6.568	0.9957	-0.5060
65	0.80795	1.70542	3.2103	0.6513	0.6042	0.5067	0.4271	-0.2943	-0.3981	29.027	43.154	0.4164	0.1677
66	0.69571	1.38471	2.6151	0.7026	0.5142	0.6580	0.4023	-0.2168	0.1508	1.730	6.214	0.5315	0.1335
67	0.62042	0.02420	0.1562	0.7051	0.5109	0.6936	0.4844	-0.0508	0.1373	1.107	1.060	0.2473	0.1779
68	0.33947	0.79651	-0.4159	0.8303	0.2940	0.8196	0.2704	-0.0037	0.0098	43.766	74.694	0.4048	-0.1494
69	0.46954	0.72129	1.0329	0.7821	0.3775	0.7992	0.4091	0.0958	0.1408	2.474	3.849	0.2402	0.1547
70	0.57626	1.04024	-0.4655	0.7360	0.4573	0.6918	0.3417	-0.2503	0.3716	0.000	0.000	0.6173	0.0177
71	0.52754	0.54277	-0.5053	0.7578	0.4369	0.7532	0.4475	0.0265	0.1626	1.425	5.914	0.0086	-0.2987
72	0.54094	0.45134	-0.7135	0.7409	0.4418	0.7532	0.4725	0.0583	0.1192	1.621	7.958	0.9706	-0.3076
73	0.95717	2.26416	5.2661	0.6660	0.5787	0.5761	0.3706	-0.3354	0.1547	1.278	4.953	0.4435	0.2967
74	0.63461	1.10953	0.7970	0.7143	0.4750	0.6766	0.4028	-0.1877	0.0000	21.707	78.007	0.7282	-0.0408
75	0.70170	1.00456	2.3779	0.7191	0.4866	0.6447	0.2261	-0.5060	0.4614	0.000	0.000	0.3216	0.4030
76	1.10967	2.59006	7.4059	0.5906	0.7903	0.4079	0.3736	-0.4505	0.0000	4.881	22.140	0.5007	0.3668
77	0.62399	1.05039	3.0942	0.7365	0.4565	0.6903	0.3341	-0.2645	0.2867	1.604	7.109	0.4370	0.1845
78	0.38056	1.16303	1.5435	0.8142	0.3219	0.8014	0.2935	-0.0930	0.3528	1.968	4.938	0.2522	0.2002

OBS	ISTN	BASIN	USREG	N	STMEAN	DAREA	SHDISCH	SHAPE	BASLEN	CHANLEN	CHANSIN	CHANSLOP	LCV	LSKEW	LKUR
79	281100	KYKY	4	24	11925.4	163.00	73.142	1.371	16.00	24.400	1.525	0.25947	0.2460	0.3771	0.2666
80	281200	KYKY	3	26	23126.3	408.00	47.585	1.569	27.61	45.500	1.448	0.14432	0.2713	0.1739	0.1459
81	281500	KYKY	3	54	27339.6	722.00	37.867	1.815	36.20	77.000	2.127	0.09659	0.2875	0.1749	0.0710
82	282190	KYKY	3	11	171.6	0.59	298.909	1.906	1.06	1.190	1.123	2.02197	0.3848	0.1436	0.0513
83	282500	KYKY	2	21	2784.9	65.00	42.324	3.891	16.00	10.700	1.169	0.10750	0.3134	0.3115	0.2703
84	283000	KYKY	2	29	2543.0	24.00	105.958	1.633	6.26	7.400	1.182	0.14697	0.3554	0.2202	0.1208
85	283305	KYKY	2	11	146.8	0.58	253.135	1.022	0.77	1.090	1.416	3.03030	0.1708	0.1104	0.3776
86	283500	KYKY	2	51	9905.5	362.00	27.363	4.186	38.90	60.900	1.771	0.11344	0.3025	0.2040	0.1590
87	284300	KYKY	3	16	3550.8	28.60	124.159	2.020	8.98	13.460	1.521	0.42045	0.3246	0.0072	-0.0012
88	284310	KYKY	3	18	4241.7	53.30	79.581	1.672	9.45	16.000	1.693	0.20033	0.3350	0.2944	0.1037
89	284340	KYKY	3	10	347.4	1.83	189.036	1.993	1.91	2.270	1.241	1.17235	0.4039	0.2073	0.1965
90	284550	KYKY	3	9	1536.1	10.90	142.762	1.924	4.60	5.579	1.213	0.36210	0.2803	0.3640	0.1600
91	285000	KYKY	3	43	14720.7	319.00	46.292	1.903	24.60	50.500	2.053	0.07765	0.2536	0.2586	0.1579
92	285100	KYKY	3	7	40.6	0.13	373.504	2.243	0.54	0.560	1.037	2.71534	0.2210	0.0367	-0.0210
93	285500	KYKY	3	7	19100.0	379.00	50.376	2.576	31.90	70.880	2.473	0.10114	0.2290	-0.0012	0.1300
94	287120	KYKY	2	10	199.6	1.26	157.937	2.658	1.83	2.050	1.120	1.09060	0.4339	0.4692	0.3200
95	287534	KYKY	2	10	564.3	4.47	126.284	3.597	4.01	5.530	1.379	0.22917	0.1169	0.1430	0.0920
96	288000	KYKY	2	33	4344.1	111.00	39.134	2.603	17.60	31.400	1.386	0.07197	0.2182	0.1220	0.1959
97	288500	KYKY	2	27	143.3	1.93	75.379	2.183	2.35	2.700	1.149	0.72917	0.3911	0.3085	0.1344
98	289000	KYKY	2	25	1003.2	21.00	47.770	1.368	5.73	9.100	1.475	0.31250	0.2745	0.2288	0.1087
99	289190	KYKY	2	10	1027.0	5.30	193.774	2.725	3.80	4.550	1.197	0.54735	0.2700	0.0087	0.0775
100	289500	KYKY	2	51	12914.3	403.00	32.045	2.261	32.70	70.200	2.391	0.06818	0.2454	0.1274	0.2162
101	290000	KYKY	1	34	2377.1	5.63	422.213	1.409	3.01	3.400	1.130	0.75379	0.2142	0.3450	0.3058
102	290530	KYKY	1	10	496.0	4.97	101.840	1.926	2.29	1.490	1.122	0.70453	0.3735	0.4126	0.2514
103	291000	KYKY	1	43	4419.0	42.90	107.607	1.594	0.27	16.000	1.135	0.17065	0.2942	0.2003	0.1025
104	291050	KYKY	1	10	226.9	0.50	391.207	2.442	1.19	1.210	1.017	1.39962	0.2916	0.2960	0.3170

OBS	CV	SK	KUR	MEVL	AEVL	MGVL	AGVL	KGVL	MAKL	AAKL	BAKL	CAKL	DAKL
79	0.69320	-1.92377	3.9966	0.7119	0.4992	0.6566	0.3463	-0.2990	-0.1964	27.33	44.9	0.4837	0.1957
80	0.48105	0.03226	0.6660	0.7741	0.3914	0.7730	0.3891	-0.0061	0.2737	1.30	1.9	0.2108	0.2250
81	0.50047	0.60644	-0.2571	0.7606	0.4147	0.7591	0.4118	-0.0078	0.2975	0.18	6.0	0.9114	-0.3476
82	0.66249	0.30551	-1.2045	0.6795	0.5552	0.6905	0.5765	0.0423	0.0000	1.69	36.9	1.4234	-0.4079
83	0.64613	2.10605	5.7755	0.7390	0.4521	0.7013	0.3570	-0.2090	0.2237	1.60	2.9	0.2136	0.4117
84	0.65132	1.11074	1.2970	0.7039	0.5130	0.6842	0.4694	-0.0887	0.0000	3.04	22.5	0.8035	-0.1755
85	0.29398	1.10149	3.0477	0.8578	0.2404	0.8543	0.2391	-0.0316	-0.1193	20.04	21.4	0.1544	0.3099
86	0.55796	1.17074	1.0258	0.7401	0.4365	0.7300	0.4149	-0.0523	0.7215	1.40	3.0	0.4115	0.0760
87	0.56122	0.10997	-1.5969	0.7294	0.4604	0.7591	0.5210	0.1297	0.1324	1.45	0.7	0.0000	0.0000
88	0.60601	1.01320	0.0312	0.7203	0.4845	0.6839	0.3957	-0.1049	0.2980	0.77	0.1	0.0000	0.0000
89	0.74055	1.03637	0.6094	0.6636	0.5827	0.6220	0.4821	-0.1740	0.0000	2.16	6.7	0.6616	0.0672
90	0.72105	1.25063	0.4550	0.6767	0.5601	0.6171	0.3990	-0.2823	0.2503	0.00	0.0	0.6979	0.0671
91	0.59339	1.90630	5.7945	0.7008	0.3658	0.7603	0.3103	-0.1334	0.2794	2.17	4.5	0.2210	0.3149
92	0.36626	-0.15297	-2.3400	0.8160	0.3103	0.8511	0.3762	0.2190	0.4294	-1.04	5.8	1.4980	-1.0442
93	0.35753	-0.45570	-0.6218	0.8093	0.3394	0.8580	0.4048	0.2068	-0.0390	7.48	11.8	0.6536	-0.4365
94	0.92493	2.33044	6.0164	0.6307	0.6260	0.5515	0.3517	-0.4109	0.0000	9.96	33.8	0.4826	0.3239
95	0.25761	0.43817	-1.6363	0.8627	0.2379	0.8474	0.2470	0.0423	0.0000	73.62	118.2	0.5329	-0.3935
96	0.30049	0.67463	1.0507	0.8183	0.3147	0.8297	0.3360	0.0759	0.1986	3.94	7.5	0.3364	0.0046
97	0.72007	1.19633	0.7050	0.6743	0.5643	0.6280	0.4490	-0.2048	0.1955	0.85	0.1	0.0000	0.0000
98	0.49352	0.82902	-0.3622	0.7697	0.3910	0.7542	0.3647	-0.0876	0.0000	58.28	148.3	0.7468	-0.2252
99	0.44581	-0.29413	-1.0534	0.7751	0.3094	0.8318	0.4744	0.2025	0.0550	2.54	4.7	0.7353	-0.4409
100	0.45168	0.95555	1.0582	0.7956	0.3541	0.8060	0.3753	0.0670	0.1726	2.89	4.2	0.2090	0.2431
101	0.67336	2.51146	7.4653	0.7384	0.4532	0.6901	0.3227	-0.2826	0.2294	2.10	5.6	0.2902	0.3400
102	0.72122	1.54418	1.9001	0.6870	0.5389	0.6225	0.3462	-0.3460	0.0000	31.94	95.4	0.5393	0.1134
103	0.37997	1.21212	1.1303	0.8283	0.2975	0.8081	0.2492	-0.1648	0.0000	51.38	85.5	0.4135	-0.0245
104	0.54669	1.50371	3.6800	0.7572	0.4206	0.7251	0.3421	-0.1882	-4.7632	52.74	39.4	0.3568	0.2265

TABLE A1 (cont..). Important Statistics of Hydrological Characteristics of Gauged Sites Used in the Study.

OBS	ISTN	BASIN	USREG	N	STMEAN	DAREA	SMDISCH	SHAPE	BASLEN	CHANLEN	CHANSIN	CHANSLOP	LCV	LSKEW	LKUR
105	291500	KYKY	1	58	23214.5	437.00	53.122	2.171	30.00	80.800	2.623	0.06610	0.2230	0.1895	0.2031
106	292200	SAKY	1	9	471.1	0.87	541.507	4.416	1.96	2.060	1.051	2.63258	0.2762	0.2220	0.2691
107	292460	SAKY	1	18	3580.6	24.10	148.571	1.902	6.77	9.660	1.427	0.22216	0.1787	0.0889	0.1030
108	292472	SAKY	1	11	303.7	0.97	313.121	2.445	1.54	1.830	1.188	2.06439	0.3829	0.2342	0.0484
109	292500	SAKY	1	44	1434.7	17.20	83.413	2.849	7.00	8.900	1.271	0.34742	0.3816	0.3001	0.2193
110	292785	SAKY	1	23	895.4	6.51	137.548	2.368	3.95	5.340	1.352	0.45398	0.3455	0.2616	0.1535
111	293000	SAKY	1	40	1606.2	18.40	87.292	1.737	5.73	9.600	1.675	0.37689	0.2811	0.3360	0.2189
112	294000	SAIN	1	30	7341.3	189.00	38.843	1.383	16.17	25.100	1.552	0.10417	0.3024	0.3215	0.2195
113	295500	SAKY	2	46	9401.7	192.00	48.967	1.972	19.46	25.300	1.287	0.07008	0.2138	0.0938	0.1694
114	295845	SAKY	1	10	428.5	1.36	315.074	2.280	1.76	1.800	1.030	1.42235	0.5133	0.4613	0.2004
115	296500	SAKY	1	27	3134.5	19.10	164.111	4.073	0.82	9.940	1.127	0.27955	0.2432	0.1289	0.2005
116	297000	SAKY	1	29	1743.6	5.15	338.540	2.878	3.85	4.400	1.143	0.98674	0.3335	0.3546	0.1619
117	297500	SAKY	1	24	6347.1	31.80	199.594	4.364	11.78	14.860	1.261	0.28371	0.2492	0.1499	0.2188
118	298000	SAKY	1	41	10790.7	138.00	78.194	2.400	18.20	34.700	1.907	0.10417	0.2490	0.3991	0.2540
119	298535	SAKY	1	10	212.8	0.68	312.941	3.250	1.51	1.600	1.060	2.47159	0.5136	0.5532	0.3756
120	299900	SAKY	5	47	16552.3	237.00	69.257	2.764	25.70	38.600	1.502	0.17140	0.2821	0.2688	0.2198
121	300065	SAKY	2	16	557.5	1.71	326.023	1.791	1.75	1.940	1.109	1.43182	0.2880	-0.0718	0.1578
122	300400	SAKY	2	14	18695.7	436.00	42.880	0.640	16.70	29.850	1.787	0.07121	0.2444	0.1430	0.0682
123	301000	SAKY	2	46	21055.0	669.00	31.472	1.800	34.70	92.600	2.669	0.07008	0.2146	0.0423	0.1205
124	302000	SAKY	5	41	2912.9	64.00	45.513	2.520	12.70	15.200	1.197	0.22159	0.2572	0.1976	0.1896
125	302085	GRKY	5	10	163.2	0.85	192.000	2.739	1.57	2.000	1.274	1.14583	0.2891	-0.0558	-0.0081
126	302220	GRIN	5	15	3920.0	37.10	159.569	1.938	11.24	14.400	1.281	0.35227	0.3733	0.2262	0.0704
127	302300	GRIN	5	16	3046.8	16.11	189.122	3.864	7.09	10.200	1.293	0.35985	0.2612	0.0573	0.1155
128	302350	GRIN	5	10	206.3	0.56	368.393	.	.	.	.	2.65152	0.5295	0.6143	0.4606
129	303000	GRIN	6	59	13510.3	284.00	47.600	2.307	33.14	77.100	2.326	0.07197	0.2474	0.1559	0.1519
130	303300	GRIN	6	24	2458.3	39.80	61.770	2.749	10.46	14.000	1.338	0.29167	0.4752	0.5601	0.3838
131	303400	GRIN	6	15	1422.6	7.86	180.992	2.181	4.14	4.350	1.051	0.44886	0.4055	0.3835	0.1822
132	303440	GRIN	6	10	96.5	0.26	371.154	.	.	.	.	1.97917	0.3066	0.0981	0.1991
133	303900	GRIN	6	10	69.5	0.25	278.000	.	.	.	.	1.55303	0.2440	0.1185	0.2161
134	304500	GRKY	5	28	482.3	2.14	225.384	1.447	1.76	2.100	1.193	2.84091	0.4097	0.2678	0.1643
135	305000	GRKY	5	32	5457.0	22.40	243.616	1.263	5.32	8.300	1.568	0.60852	0.4322	0.3115	0.1707
136	305500	GRKY	5	30	6587.0	36.30	181.460	1.185	6.56	9.100	1.387	0.53788	0.3718	0.2654	0.1473
137	305559	GRKY	5	10	406.0	0.88	461.364	3.890	1.85	1.950	1.050	1.98580	0.2624	0.1294	0.1601
138	305835	GRKY	5	10	273.4	0.71	385.070	3.127	1.49	1.710	1.148	2.92045	0.2331	0.3177	0.2655
139	306000	GRKY	5	8	22235.0	682.00	32.603	2.405	40.50	79.138	1.954	0.07339	0.4077	0.4004	0.2077
140	306500	GRKY	5	29	20328.6	729.00	27.886	2.050	45.80	102.500	2.238	0.06629	0.2941	0.3202	0.2166
141	307000	GRKY	5	46	11050.5	173.00	63.922	1.860	18.78	28.440	1.521	0.17765	0.3644	0.4192	0.2886
142	307100	GRKY	5	19	16446.3	246.00	66.855	2.014	23.10	58.700	2.541	0.09659	0.4403	0.5531	0.4058
143	307500	GRKY	5	42	2106.4	18.30	115.101	2.080	4.17	9.000	1.459	0.31061	0.2673	0.4077	0.3702
144	309500	GRKY	5	29	1055.4	5.34	197.637	3.288	4.19	4.900	1.169	0.45368	0.3544	0.2428	0.0751
145	310000	GRKY	5	40	4751.9	35.60	133.482	2.020	8.57	10.700	1.250	0.37879	0.2997	0.0583	0.0239
146	310300	GRKY	6	26	8958.5	237.00	37.799	1.587	23.80	46.300	1.945	0.07955	0.2941	0.2532	0.1072
147	310385	GRKY	6	11	224.8	0.56	401.461	0.754	0.65	0.850	1.308	2.22027	0.3858	0.1021	0.0208
148	310400	GRKY	6	26	2168.1	54.40	39.855	2.709	15.21	27.200	1.788	0.15133	0.3947	0.3172	0.2119
149	310500	GRKY	6	26	9796.2	378.00	25.916	1.633	31.30	91.400	2.920	0.04924	0.2347	0.1661	0.2159
150	310800	GRKY	6	10	121.1	0.31	390.645	1.766	0.74	0.790	1.068	5.86894	0.3066	0.2786	0.1685
151	311000	GRKY	6	15	10513.3	480.00	21.903	2.153	38.90	75.300	1.941	0.06061	0.2481	0.1519	0.2480
152	311600	GRKY	6	13	1996.5	10.90	183.162	2.114	4.80	6.710	1.398	0.05398	0.3597	0.3229	0.2114
153	312000	GRKY	6	34	4910.0	30.80	159.416	0.947	5.40	7.700	1.426	0.50758	0.1906	0.1243	0.1566
154	312500	GRKY	5	26	23812.9	514.00	46.329	1.076	23.90	65.800	2.753	0.08144	0.3681	0.3324	0.1306
155	312795	GRKY	5	9	171.7	0.89	192.884	0.666	0.77	0.860	1.117	3.52424	0.2871	0.3414	0.1457
156	313020	GRKY	5	9	91.7	0.24	381.944	1.602	0.62	0.660	1.065	2.30265	0.3030	0.1698	0.1065

OBS	CV	SK	KUR	MEVL	AEVL	MGVL	AGVL	KGVL	MWKL	AWKL	BWKL	CWKL	DWKL
105	0.41191	1.19710	2.3495	0.8143	0.3218	0.8099	0.3127	-0.0302	0.2964	2.65	6.5	0.3250	0.0718
106	0.48597	1.09405	2.1399	0.7700	0.3985	0.7562	0.3683	-0.0794	-0.1741	12.53	16.0	0.3783	0.0703
107	0.29183	0.22558	-1.0376	0.8512	0.2578	0.8675	0.2868	0.1302	0.4368	1.41	5.2	0.4340	-0.2873
108	0.66999	0.63583	-1.0966	0.6011	0.5524	0.6579	0.5006	-0.0976	0.1419	1.06	0.2	0.0000	0.0000
109	0.74305	1.66199	3.1074	0.6022	0.5506	0.6393	0.4450	-0.1930	0.0320	2.59	7.2	0.5736	0.1213
110	0.62912	1.03538	0.3511	0.7123	0.4984	0.6835	0.4316	-0.1377	-0.0175	10.12	31.1	0.7668	-0.0914
111	0.56674	2.13283	5.8666	0.7659	0.4056	0.7274	0.3057	-0.2441	0.4033	0.56	1.3	0.2501	0.2972
112	0.57868	1.48442	1.7802	0.7482	0.4363	0.7098	0.3387	-0.2229	-0.1948	35.54	54.9	0.5152	0.0791
113	0.37778	0.57141	0.6813	0.8220	0.3084	0.8402	0.3411	0.1221	0.2828	2.29	6.1	0.2436	0.0884
114	1.07219	2.15532	5.1219	0.5726	0.7405	0.4710	0.4242	-0.4088	0.0004	43.98	400.8	0.6553	0.2641
115	0.45354	1.34649	3.9169	0.7975	0.3509	0.8082	0.3713	0.0649	0.2528	1.95	2.3	0.0755	0.5198
116	0.63678	1.29760	0.5174	0.7223	0.4811	0.6731	0.3499	-0.2606	0.3488	0.00	0.0	0.6205	0.0471
117	0.44805	0.90289	1.2451	0.7925	0.3595	0.7977	0.2678	0.0314	0.1747	2.95	4.7	0.2492	0.2001
118	0.50202	1.86145	3.1723	0.7927	0.3592	0.7500	0.2378	-0.3283	0.0000	73.43	130.9	0.3061	0.1881
119	1.17717	2.53112	6.7730	0.5723	0.7410	0.4576	0.3316	-0.5525	0.0000	434.58	2271.8	0.4649	0.4251
120	0.55195	1.94337	5.8055	0.7650	0.4071	0.7399	0.3480	-0.1482	0.2664	1.66	4.4	0.3523	0.1764
121	0.50136	-0.72779	-0.0418	0.7601	0.4155	0.8531	0.5398	0.4172	-0.1271	3.75	2.5	0.0080	0.8619
122	0.40954	0.39548	-0.9942	0.7964	0.3526	0.8034	0.3642	0.0423	0.0000	46.42	112.2	0.8553	-0.4497
123	0.37222	0.23161	-0.2546	0.8213	0.3096	0.8538	0.3632	0.2093	0.2965	1.84	2.5	0.1435	0.1857
124	0.47991	1.40318	3.4233	0.7858	0.3711	0.7788	0.3562	-0.0426	0.2950	1.49	2.5	0.2052	0.2741
125	0.48974	-0.45458	-1.5843	0.7593	0.4170	0.8451	0.5354	0.3867	0.0000	1547.49	24329.7	2.0925	-1.2365
126	0.65986	0.72157	-0.7506	0.6891	0.5386	0.6671	0.4944	-0.0857	0.1554	1.07	0.3	0.0000	0.0000
127	0.43902	0.07299	-0.7345	0.7825	0.3768	0.8168	0.4348	0.1834	0.1756	2.00	2.2	0.1482	0.2360
128	1.37591	2.96145	9.0592	0.5590	0.7639	0.4363	0.2821	-0.5946	0.0000	14.98	63.8	0.3553	0.5380
129	0.43955	0.63884	-0.0517	0.7940	0.3569	0.7976	0.3640	0.0219	-0.1382	20.12	30.8	0.6148	-0.2163
130	1.21511	3.55100	14.3228	0.6043	0.6856	0.4976	0.3006	-0.5307	0.0000	345.74	1341.1	0.4179	0.4371
131	0.77882	1.47277	1.5153	0.6623	0.5850	0.5960	0.4007	-0.3075	0.2331	0.00	0.000	0.6835	0.1088
132	0.52496	0.12869	-0.1065	0.7467	0.4423	0.7492	0.4967	0.1150	-0.1704	6.02	7.719	0.4982	-0.0379
133	0.40335	0.26448	0.1333	0.7968	0.3528	0.8105	0.3775	0.0817	-0.1607	11.70	15.130	0.4405	-0.0604
134	0.76071	1.23856	1.3004	0.6389	0.5910	0.6227	0.5061	-0.1467	0.0586	0.94	1.841	0.5423	0.1122
135	0.82353	1.45522	2.2235	0.6401	0.6235	0.5881	0.4934	-0.2090	0.0000	2.99	19.545	0.8621	-0.0087
136	0.68110	1.04196	0.2087	0.6911	0.5352	0.6590	0.4602	-0.1432	0.0000	6.76	27.349	0.8369	-0.0989
137	0.43229	0.39426	-0.2278	0.7815	0.3785	0.7929	0.4002	0.0640	0.1491	2.83	5.645	0.4476	-0.0546
138	0.41155	1.37805	1.3590	0.8058	0.3364	0.7768	0.2631	-0.2176	0.0000	21.80	25.313	0.3365	0.1582
139	0.79071	1.73995	3.3357	0.6605	0.5881	0.5904	0.3883	-0.3300	0.2433	0.00	0.000	0.6480	0.1437
140	0.57708	1.85617	4.3200	0.7551	0.4243	0.4031	0.6148	0.0299	0.3214	1.07	4.912	0.4289	0.1423
141	0.76666	1.87735	3.0191	0.6965	0.5257	0.3374	0.3327	-0.3547	0.0000	13.96	35.561	0.4609	0.2546
142	1.03468	2.31187	4.6993	0.6333	0.6352	0.5351	0.2843	-0.5223	0.0000	16.80	49.607	0.3605	0.4603
143	0.70871	4.68932	26.4673	0.7774	0.3856	0.7305	0.2505	-0.3395	0.3819	1.38	3.500	0.1486	0.5243
144	0.66007	0.92559	0.2680	0.7048	0.5113	0.6807	0.4568	-0.1104	0.2137	0.96	0.218	0.0000	0.0000
145	0.51617	0.12424	-1.1508	0.7504	0.4323	0.7894	0.4984	0.1817	0.0000	14.32	77.876	1.4342	-0.7522
146	0.52817	0.88737	-0.1757	0.7551	0.4244	0.7325	0.3726	-0.1255	0.3553	0.77	0.192	0.0000	0.0000
147	0.66129	0.13109	-1.4244	0.6788	0.5565	0.7070	0.6094	0.1084	0.0000	-1.42	1.315	2.9435	-0.8261
148	0.76533	1.63117	2.7101	0.6713	0.5694	0.6223	0.4457	-0.2169	0.1147	1.04	2.952	0.5117	0.1779
149	0.42472	1.04461	1.9720	0.8046	0.3386	0.8055	0.3404	0.0059	0.2287	2.92	5.574	0.2813	0.1409
150	0.69953	0.96799	0.1176	0.6788	0.5578	0.6607	0.4687	-0.1623	0.0000	3.99	15.552	0.7816	-0.0355
151	0.44315	1.02910	2.4467	0.7934	0.3579	0.7981	0.3671	0.0282	0.1438	3.41	4.998	0.2070	0.2786
152	0.67847	1.38926	1.2166	0.7004	0.5190	0.6544	0.4018	-0.2250	0.0000	8.07	23.234	0.6171	0.0750
153	0.32884	0.49993	-0.2777	0.8413	0.2749	0.8507	0.2926	0.0722	0.3445	2.84	7.978	0.3847	-0.1346
154	0.68875	1.28359	1.2153	0.6935	0.5310	0.6442	0.4037	-0.2380	0.2631	0.74	0.002	0.0000	0.0000
155	0.51848	1.13102	0.0246	0.7610	0.4141	0.7208	0.3092	-0.2508	0.4312	0.00	0.000	0.5583	0.0195
156	0.51267	0.46152	-0.6721	0.7477	0.4371	0.7477	0.4371	0.0001	0.2157	0.97	2.158	0.5199	-0.0860

TABLE A1 (cont..). Important Statistics of Hydrological Characteristics of Gauged Sites Used in the Study.

OBS	ISTN	BASIN	USREG	N	STMEAN	DAREA	SHDISCH	SHAPE	BASLEN	CHANLEN	CHANSIN	CHANSLOP	LCV	LSKEW	LKUR
157	312500	GRKY	5	33	1666.0	7.47	222.762	2.441	4.27	4.000	1.124	0.09432	0.2310	0.3451	0.4602
158	313600	GRIN	5	10	290.1	0.95	294.795	.	.	.	.	1.40000	0.4337	0.3277	0.1357
159	313700	GRKY	5	17	8092.4	91.00	97.710	2.627	17.00	22.010	1.342	0.17157	0.2970	0.3773	0.3774
160	313800	GRKY	5	25	2699.3	7.80	346.067	2.420	7.23	9.200	1.272	0.36932	0.3929	0.4389	0.3233
161	314000	GRKY	5	44	21092.3	250.00	50.917	1.676	20.30	41.700	1.473	0.12500	0.3581	0.4294	0.3124
162	314750	GRKY	6	9	270.6	0.50	540.009	1.345	0.82	1.090	1.329	4.29394	0.3605	0.2348	0.2452
163	315885	GRKY	6	9	55.2	0.22	251.010	1.638	0.44	0.730	1.141	1.72254	0.3101	0.4120	0.2031
164	316000	GRKY	6	44	5611.2	80.00	69.445	1.993	13.43	24.400	1.017	0.13447	0.2302	0.1039	0.1423
165	317000	GRKY	6	22	9156.8	158.00	57.953	1.263	16.06	39.000	2.361	0.09290	0.2167	-0.0136	0.1270
166	317500	GRKY	6	26	2169.2	22.60	95.982	1.325	7.46	11.300	1.515	0.29053	0.2354	0.1214	0.1251
167	318000	GRKY	6	17	8736.5	344.00	25.377	1.715	27.90	49.200	1.763	0.06437	0.2260	-0.0460	0.1956
168	318200	GRKY	6	22	2395.5	20.10	168.920	1.577	5.63	6.300	1.119	0.39209	0.2877	0.3433	0.1750
169	318500	GRKY	6	11	8669.1	394.00	22.003	1.435	26.09	73.300	2.726	0.04602	0.2143	-0.0601	0.1015
170	318505	GRKY	6	10	154.7	0.22	703.182	2.425	0.76	0.820	1.079	4.65014	0.2725	0.0166	0.0663
171	318800	GRKY	6	31	5858.4	124.00	47.245	3.004	19.30	29.400	1.534	0.05602	0.2453	0.2781	0.1781
172	319000	GRKY	6	20	10630.8	627.00	16.700	2.135	40.20	74.500	1.053	0.05492	0.2304	0.1101	0.2090
173	319520	GRKY	6	10	112.8	0.26	433.846	3.047	0.89	0.930	1.045	2.30303	0.2529	0.0266	0.0860
174	320500	GRKY	7	45	8906.0	194.00	45.907	1.039	10.09	40.300	2.133	0.08749	0.3519	0.3317	0.2432
175	321275	GRKY	7	8	464.0	0.95	488.421	2.213	1.45	1.490	1.028	0.98163	0.2746	0.2217	0.1860
176	321350	GRKY	6	15	2706.0	58.20	46.495	3.613	14.50	19.500	1.345	0.14205	0.1606	0.0044	0.1070
177	321465	GRKY	6	9	123.4	0.29	425.670	1.540	0.67	0.750	1.119	1.17609	0.3016	0.4349	0.2811
178	322100	TWIN	6	22	5093.6	323.00	15.770	1.707	23.40	42.000	1.789	0.04565	0.2371	0.2055	0.2077
179	322360	TWKY	6	12	2228.3	14.30	155.828	2.296	5.73	5.910	1.031	0.25644	0.1306	-0.1296	0.2046
180	366200	TWIN	1	16	1080.5	9.31	116.058	3.514	5.72	8.900	1.556	0.34659	0.1605	-0.0511	0.1613
181	366400	TWIN	1	10	76.0	0.16	475.000	6.126	0.99	0.980	0.990	1.34470	0.2271	0.0502	0.2772
182	378350	TWIN	6	19	4415.8	104.00	42.460	1.087	14.01	18.600	1.328	0.07197	0.2096	0.2574	0.2110

OBS	CV	SK	KUR	MEVL	AEVL	MGVL	AGVL	KGVL	MWKL	AWKL	BWKL	CWKL	DWKL
157	0.630689	4.42330	22.9741	0.8076	0.3333	0.7740	0.2471	-0.2556	0.3032	2.69	4.201	0.0477	0.7462
158	0.825571	1.02415	-0.4834	0.6380	0.6258	0.5819	0.4001	-0.2316	0.1270	0.88	0.813	0.0000	0.0000
159	0.631734	2.44792	7.2949	0.7520	0.4296	0.7043	0.2981	-0.2993	0.0700	6.22	10.918	0.2370	0.4166
160	0.856355	2.57541	7.9820	0.6720	0.5660	0.5984	0.3428	-0.3002	0.0000	8.12	21.100	0.4311	0.3195
161	0.792951	2.76445	9.4696	0.7018	0.5166	0.6354	0.3194	-0.3480	0.1567	4.14	14.742	0.4026	0.3059
162	0.689300	1.67847	3.4337	0.6998	0.5200	0.6510	0.3935	-0.2616	0.1427	1.86	5.746	0.4666	0.2022
163	0.615346	2.11419	5.1233	0.7418	0.4474	0.6067	0.2879	-0.3452	-1.0276	464.57	246.247	0.4071	0.2363
164	0.425493	0.73317	0.3392	0.8017	0.3436	0.7983	0.3367	-0.0216	-2.9795	1078.53	309.360	0.4164	-0.2220
165	0.367096	-0.31756	-0.2964	0.8196	0.3126	0.8696	0.3874	0.3083	0.0601	5.43	9.286	0.5857	-0.4230
166	0.404983	0.48835	-0.2232	0.8040	0.3396	0.8144	0.3429	0.0769	0.3070	1.46	3.040	0.3434	-0.0200
167	0.390055	-0.54334	0.6696	0.8118	0.3261	0.8753	0.4153	0.3682	-0.0930	6.73	7.443	0.3229	-0.0907
168	0.543117	1.47752	1.9903	0.7604	0.4150	0.7199	0.3088	-0.2530	0.0000	0.29	0.363	0.1750	0.1010
169	0.349079	-0.58842	-0.6316	0.8215	0.3092	0.8894	0.4006	0.4101	0.1451	3.02	4.305	0.3704	-0.2997
170	0.448356	-0.23242	-1.2185	0.7731	0.3731	0.8241	0.4735	0.2542	0.1343	1.13	1.920	0.1970	0.0336
171	0.450487	1.25437	1.2964	0.7957	0.3540	0.7721	0.2976	-0.1617	0.2461	0.15	28.653	0.4914	-0.0261
172	0.400093	0.34513	0.1673	0.8081	0.3324	0.8233	0.3603	0.0953	0.0763	5.47	8.989	0.3573	0.0009
173	0.414150	-0.14408	-1.1312	0.7894	0.3440	0.8332	0.4350	0.2366	0.1763	1.94	2.710	0.3370	-0.1205
174	0.713395	2.17470	6.4722	0.7069	0.5077	0.6599	0.3065	-0.2372	0.1733	1.35	4.707	0.4390	0.2133
175	0.462625	0.82238	-0.0773	0.7713	0.3962	0.7577	0.3663	-0.0789	-1.6192	262.92	125.350	0.5836	-0.0041
176	0.256305	0.19894	-1.1531	0.8662	0.2317	0.8818	0.2592	0.1377	0.5087	1.07	2.908	0.2333	-0.0734
177	0.588970	1.96046	4.0377	0.7409	0.4351	0.6923	0.2656	-0.3750	0.0000	55.68	115.187	0.3910	0.2493
178	0.438379	1.44735	3.4670	0.8025	0.3421	0.7943	0.3266	-0.0546	0.3497	1.42	2.361	0.1440	0.3709
179	0.281512	-0.95043	0.3203	0.8912	0.1885	0.9464	0.2539	0.5317	0.3012	4.39	6.945	0.1472	-0.0607
180	0.260342	-0.41184	-0.6276	0.8464	0.2215	1.0149	0.2961	0.3778	0.3415	2.60	3.908	0.1062	0.1752
181	0.304317	-0.03120	0.0520	0.8109	0.3276	0.8429	0.3811	0.1956	-0.0566	4.82	7.491	0.1915	0.2373
182	0.375813	1.10585	0.4410	0.8254	0.3024	0.8087	0.2637	-0.1316	0.0000	26.83	42.989	0.3856	0.0111

OBS	ISTN	BASIN	USREG	N	STMEAN	DAREA	SHDISCH	SHAPE	BASLEN	CHANLEN	CHANSIN	CHANSLOP	LCV	LSKEW	LKUR
183	378590	TWIN	6	10	102.0	0.32	318.750	.	.	.	.	1.50560	0.3178	0.3269	0.2349
184	301600	TWIL	7	17	112.8	0.16	705.167	.	.	.	.	1.70000	0.3692	0.4728	0.5000
185	302520	TWIL	7	13	210.2	1.10	191.119	.	.	.	.	0.53504	0.4414	0.4617	0.3200
186	302975	TWKY	7	10	437.1	0.91	482.527	0.910	0.91	1.360	1.495	0.20341	0.3091	0.2576	0.1079
187	303000	TWKY	7	44	4530.5	246.00	18.416	2.546	25.57	53.300	2.084	0.03780	0.2963	0.3449	0.2773
188	303605	TWKY	7	10	1167.0	2.52	463.095	1.463	1.92	2.510	1.307	2.11191	0.3077	0.0520	0.0890
189	304000	TWKY	7	20	634.3	2.10	302.063	1.526	1.79	2.200	1.229	0.54545	0.2329	0.0933	0.1560
190	305000	CLIL	7	36	2684.6	19.10	140.542	2.357	6.71	12.010	1.790	0.40606	0.3393	0.2970	0.1050
191	305500	CLKY	7	26	655.3	1.00	655.269	0.933	0.99	1.860	1.879	2.75000	0.3061	0.1572	0.1455
192	400500	CLKY	2	46	4194.8	82.30	50.970	3.308	16.50	25.000	1.564	0.53220	0.3134	0.2581	0.1970
193	400700	CLKY	3	24	6992.5	82.40	84.860	4.060	18.29	24.400	1.334	0.73466	0.3396	0.2367	0.1009
194	400800	CLKY	3	11	4323.6	56.20	76.933	1.075	10.23	19.000	1.857	0.62462	0.3567	0.2192	0.0860
195	401000	CLKY	3	46	19916.1	374.00	53.252	3.945	30.41	51.300	1.336	0.24621	0.2903	0.2549	0.2211
196	401500	CLKY	4	42	3540.2	35.30	100.290	2.551	9.49	9.900	1.043	2.40644	0.2642	0.2558	0.1650
197	402000	CLKY	4	43	4430.9	60.60	73.118	1.499	9.53	12.700	1.333	1.40909	0.2778	0.2793	0.1499
198	402020	CLKY	4	10	562.1	2.96	189.899	5.762	4.13	5.870	1.421	4.40826	0.1935	-0.0847	0.2794
199	403000	CLKY	4	45	31304.4	809.00	38.794	4.538	60.59	94.900	1.566	0.15530	0.2510	0.3993	0.1396
200	403500	CLKY	4	46	28839.1	960.00	30.041	4.535	65.98	106.900	1.620	0.14015	0.2246	0.1912	0.1047
201	403538	CLKY	4	9	938.4	11.50	81.604	0.958	3.32	5.990	1.804	0.43769	0.3495	0.0762	0.1007
202	403910	CLKY	4	17	18852.4	231.00	38.829	1.114	19.20	27.700	1.443	0.29167	0.2775	0.1171	0.0739
203	404820	CLKY	4	12	7079.2	140.00	50.565	1.438	14.19	30.819	2.172	0.07008	0.3380	0.3104	0.2349
204	404867	CLKY	4	10	35.2	0.31	113.548	3.226	1.00	1.220	1.220	1.86136	0.2675	-0.0408	0.0444
205	404900	CLKY	4	29	2831.3	53.80	52.626	0.932	7.08	10.000	1.412	0.19500	0.3098	0.3379	0.2561
206	405000	CLKY	4	33	7543.6	201.00	37.520	1.421	16.90	28.400	1.680	0.10985	0.2931	0.2267	0.1922
207	405856	CLKY	3	10	642.3	1.91	336.283	1.070	1.89	2.240	1.185	1.94943	0.3704	0.1529	0.0327
208	406000	CLKY	4	32	327.3	3.89	84.150	1.018	1.99	2.600	1.307	0.93239	0.2536	-0.0053	0.1359

OBS	CV	SK	KUR	MEVL	AEVL	MGVL	AGVL	KGVL	MWKL	AWKL	BWKL	CWKL	DWKL
183	0.584008	1.87516	0.7555	0.7353	0.4586	0.6930	0.3523	-0.2305	0.0000	11.262	26.995	0.4976	0.1229
184	0.927860	3.12509	10.9534	0.6925	0.5327	0.6179	0.2966	-0.4235	-0.2009	10.764	13.375	0.1855	0.5076
185	0.935634	1.91715	2.0060	0.6324	0.6368	0.5450	0.3644	-0.4094	0.0000	7.098	24.662	0.4900	0.3216
186	0.702739	1.17060	1.6217	0.6760	0.5613	0.6440	0.4091	-0.1320	0.0291	1.642	3.658	0.5520	0.1074
187	0.634840	2.66515	9.1318	0.7533	0.4274	0.7113	0.3170	-0.2553	0.2767	1.434	4.067	0.2075	0.3161
188	0.516720	-0.11446	-0.8275	0.7637	0.4439	0.9710	0.5154	0.1925	-0.2635	11.160	19.364	1.0970	-0.5353
189	0.407307	0.50991	0.2505	0.8061	0.3359	0.8261	0.3718	0.1229	0.2501	1.992	2.892	0.1029	0.2055
190	0.661976	1.07251	4.9806	0.7173	0.4878	0.6797	0.3975	-0.1097	0.2612	0.908	0.400	0.0469	0.6347
191	0.547425	0.81257	0.4060	0.7451	0.4416	0.7492	0.4496	0.0199	0.1707	1.503	1.359	0.0771	0.5120
192	0.506225	1.26545	1.5724	0.7390	0.4521	0.7130	0.3936	-0.1327	-0.0037	7.902	17.753	0.5790	0.0050
193	0.408739	0.92572	0.3543	0.7172	0.4099	0.6959	0.4420	-0.1014	0.0000	133.682	547.415	0.9302	-0.2301
194	0.610957	0.67069	-0.7371	0.7030	0.5145	0.6061	0.4776	-0.0752	0.1063	1.042	0.201	0.0000	0.0000
195	0.573076	1.67699	4.2032	0.7516	0.4303	0.7203	0.3767	-0.1200	0.1050	2.212	5.179	0.3037	0.1604
196	0.507423	1.02642	5.4919	0.7800	0.3812	0.7591	0.3332	-0.1294	0.3603	0.853	3.040	0.4544	0.0190
197	0.523140	1.39195	1.7003	0.7607	0.4007	0.7300	0.3255	-0.1890	0.4130	0.174	1.285	0.5141	-0.0065
198	0.331361	-0.75542	0.7418	0.8309	0.2792	0.9055	0.3659	0.4421	0.0352	5.209	5.045	0.0463	0.5512
199	0.455259	1.12603	1.7890	0.7914	0.3621	0.7837	0.3467	-0.0452	0.3541	0.939	2.562	0.3741	0.0214
200	0.397700	0.69726	-0.5211	0.8129	0.3241	0.8002	0.3161	-0.0320	0.0000	49.686	98.062	0.6463	-0.2065
201	0.596178	0.09456	-0.5313	0.7090	0.5042	0.7464	0.5695	0.1515	0.0505	1.630	0.717	0.0000	0.0000
202	0.477162	0.29040	-1.2167	0.7609	0.4003	0.7069	0.4302	0.0039	0.0000	12.323	35.828	0.9826	-0.4760
203	0.637019	1.55014	2.7210	0.7179	0.4600	0.6774	0.3075	-0.2074	0.1930	1.395	3.555	0.3031	0.2340
204	0.462712	-0.45690	-1.0743	0.7772	0.3039	0.8502	0.4095	0.3584	0.0031	2.306	1.054	0.0605	0.4004
205	0.612843	1.94855	4.0272	0.7420	0.4470	0.6994	0.3342	-0.2456	0.1010	6.541	16.796	0.4395	0.1719
206	0.540742	1.22230	1.6476	0.7551	0.4229	0.7401	0.3870	-0.0065	0.2367	1.424	2.469	0.2650	0.2467
207	0.651257	0.32041	-1.1227	0.6849	0.5459	0.6917	0.5592	0.0267	0.0656	1.373	0.470	0.0000	0.0000
208	0.444384	0.05597	0.0257	0.7080	0.3658	0.8443	0.4499	0.2933	0.1036	2.650	2.332	0.0460	0.5542

TABLE A1 (cont..). Important Statistics of Hydrological Characteristics of Gauged Sites Used in the Study.

OBS	ISTN	BASIN	USREG	N	STMEAN	DAREA	SHDISCH	SHAPE	BASLEN	CHANLEN	CHANSIN	CHANSLOP	LCV	LSKEW	LKUR
209	406500	CLKY	4	49	23580.0	604.00	39.040	1.018	24.90	57.30	2.310	0.06818	0.2403	0.1671	0.1674
210	407000	CLKY	4	9	24353.3	745.00	32.689	1.079	37.41	74.70	1.997	0.06818	0.2234	-0.0418	0.1558
211	407100	CLKY	4	29	109.2	0.67	163.047	1.710	1.07	1.31	1.220	3.90152	0.4105	0.3102	0.1829
212	407200	CLKY	4	29	47.3	0.26	181.963	2.650	0.83	1.00	1.205	3.53542	0.3140	0.2508	0.1789
213	407300	CLKY	4	30	67.1	0.85	78.980	1.176	1.00	1.55	1.550	4.23500	0.4066	0.2200	0.1636
214	407500	CLKY	4	33	10159.7	165.00	61.574	2.305	19.50	30.00	1.538	0.19129	0.2779	0.1121	0.0623
215	408500	CLTN	5	52	27219.2	382.00	71.255	1.492	23.87	46.20	1.935	0.13371	0.2382	0.2721	0.2279
216	409000	CLTN	5	22	2325.9	13.50	172.290	1.527	4.54	5.70	1.256	1.03258	0.2845	0.2543	0.2238
217	412500	CLKY	4	30	2125.9	31.30	67.921	2.914	9.55	12.20	1.277	0.40360	0.1984	-0.0007	0.0882
218	413200	CLKY	5	15	4018.7	43.40	92.596	2.589	10.60	17.50	1.651	0.38258	0.3332	0.2859	0.1786
219	413202	CLKY	5	10	215.9	0.57	378.772	1.320	0.87	1.15	1.322	6.68561	0.4645	0.4722	0.1696
220	413425	CLKY	5	10	164.3	0.76	216.104	1.563	1.09	1.25	1.147	0.28371	0.1595	0.1093	0.2225
221	414102	CLKY	5	10	610.5	3.52	198.409	1.691	2.44	3.13	1.283	0.92523	0.4118	0.2051	0.1387
222	414500	CLTN	5	43	17946.5	196.00	91.564	3.564	26.83	37.60	1.401	0.70076	0.2754	0.1734	0.1024
223	415000	CLTN	5	32	7474.9	81.00	92.307	2.504	16.97	20.50	1.208	0.63636	0.2644	0.1215	0.1065
224	415700	CLTN	5	24	771.3	4.77	161.688	1.406	3.35	4.40	1.312	1.29697	0.2661	0.0076	0.1012
225	416000	CLTN	5	42	7555.7	106.00	71.280	2.293	15.59	23.40	1.501	0.23295	0.2548	0.1631	0.1936
226	417700	CLTN	5	30	165.2	0.49	337.211	.	.	.	.	3.05739	0.3940	0.3035	0.1506
227	418000	CLTN	5	43	3847.2	51.60	74.558	1.390	10.46	20.10	1.922	0.27652	0.3053	0.2130	0.1791
228	435140	CLKY	7	13	3030.5	20.00	104.541	0.935	4.41	7.37	1.671	0.26042	0.3196	0.3243	0.1520
229	435500	CLTN	7	50	16214.4	309.00	52.474	1.631	33.93	66.90	1.972	0.08409	0.2824	0.2719	0.2809
230	435600	CLTN	7	9	715.2	3.50	204.349	2.503	2.96	3.52	1.189	0.97992	0.3798	0.3455	0.1607
231	436000	CLTN	7	47	8257.9	165.00	50.048	2.768	22.69	36.30	1.600	0.12424	0.2930	0.3108	0.2372
232	436700	CLTN	7	27	6542.2	124.00	52.760	2.716	18.35	25.70	1.400	0.23295	0.3099	0.2376	0.1607
233	437300	CLKY	7	7	251.4	0.83	302.926	2.970	1.57	1.77	1.127	1.43845	0.2920	0.1592	0.0660
234	437390	CLKY	7	8	196.6	0.39	504.167	1.050	0.64	0.85	1.320	2.66439	0.4079	0.2423	0.2288

OBS	CV	SK	KUR	MEVL	AEVL	MGVL	AGVL	KEVL	WAKL	AWKL	BWKL	CWKL	DWKL
209	0.443734	0.75219	0.1305	0.7932	0.3502	0.7940	0.3597	0.0044	0.1609	4.7697	11.264	0.5136	-0.1108
210	0.363745	-0.61890	-0.2924	0.8140	0.3223	0.8752	0.4092	0.3602	0.1060	3.3762	3.726	0.1474	0.1792
211	0.794229	1.71809	3.9263	0.6501	0.5922	0.6091	0.4698	-0.2071	0.0786	1.1353	6.559	0.7428	0.0370
212	0.579746	1.26126	1.6905	0.7306	0.4529	0.7151	0.3993	-0.1221	0.2329	1.1043	2.411	0.3737	0.1570
213	0.736935	0.98335	0.6307	0.6616	0.5063	0.6621	0.5435	-0.0764	-0.0254	1.4269	3.731	0.7573	-0.0463
214	0.470532	0.35261	-0.9196	0.7485	0.4010	0.7862	0.4336	0.0722	0.2440	0.7238	2.525	0.7530	-0.3676
215	0.460572	1.70618	3.0822	0.8016	0.3437	0.7798	0.2921	-0.1530	0.3369	2.2940	7.352	0.3363	0.1360
216	0.526271	1.21558	1.4359	0.7631	0.4105	0.7410	0.3597	-0.1272	0.0404	7.3673	14.933	0.4639	0.0670
217	0.335500	-0.06546	-0.7529	0.8348	0.2862	0.8769	0.3505	0.2851	0.3243	1.7595	2.737	0.2194	-0.0707
218	0.603301	1.02791	0.2265	0.7225	0.4807	0.6805	0.3987	-0.1728	0.0000	16.9729	30.453	0.6637	-0.0193
219	0.940800	1.37688	0.3460	0.6115	0.6731	0.5174	0.3753	-0.4226	0.1989	0.0000	0.000	0.5745	0.2829
220	0.244453	0.26470	-0.3038	0.8472	0.2300	0.8779	0.2497	0.0967	-0.3008	43.7167	42.268	0.3269	-0.1255
221	0.732047	0.80240	0.3982	0.6570	0.5942	0.6420	0.5638	-0.0540	-0.0509	1.5584	2.038	0.6796	-0.0129
222	0.409451	0.82749	0.5579	0.7707	0.3973	0.7697	0.3953	-0.0054	0.3001	0.8362	1.412	0.3924	-0.0333
223	0.460958	0.45744	-0.4763	0.7781	0.3844	0.7921	0.4107	0.0769	0.2205	1.4381	3.563	0.5363	-0.1753
224	0.473198	0.26113	1.0676	0.7764	0.3839	0.8316	0.4665	0.2703	0.0066	3.4309	2.156	0.0980	0.4162
225	0.491666	1.85954	7.1041	0.7870	0.3670	0.7896	0.3712	0.0107	0.2329	2.0663	3.470	0.2621	0.2078
226	0.733685	1.14850	0.4772	0.6719	0.5485	0.6266	0.4566	-0.1977	0.0000	24.4077	125.033	0.8536	-0.0573
227	0.565129	1.16095	1.3135	0.7458	0.4405	0.7330	0.9101	0.2775	0.1823	1.6254	3.247	0.3793	0.1281
228	0.579763	1.10536	0.0722	0.7338	0.4618	0.6926	0.3561	-0.2269	0.3542	0.6590	0.020	0.0000	0.0000
229	0.507450	2.46941	8.9989	0.7649	0.4074	0.7390	0.3463	-0.1527	0.2311	1.9945	3.346	0.1788	0.4230
230	0.697753	1.23332	0.3900	0.6844	0.5468	0.6305	0.4051	-0.2561	0.2522	-0.0000	0.000	0.7275	0.0270
231	0.634295	3.18513	15.0811	0.7560	0.4227	0.7209	0.3349	-0.2079	0.3261	0.9747	1.797	0.2009	0.3027
232	0.542656	1.09121	0.9001	0.7419	0.4472	0.7222	0.4029	-0.1026	0.2463	0.9876	1.969	0.3617	0.1375
233	0.470475	0.20040	-1.4542	0.7568	0.4213	0.7600	0.4278	0.0167	0.0000	40.0377	136.627	1.0028	-0.4259
234	0.757195	1.09535	2.1036	0.6603	0.5885	0.6328	0.5261	-0.1096	-0.1014	2.5486	3.276	0.3595	0.2875

OBS	ISTN	BASIN	USREG	N	STMEAN	DAREA	SHDISCH	SHAPE	BASLEN	CHANLEN	CHANSIN	CHANSLOP	LCV	LSKEW	LKUR
235	437490	CLKY	7	9	311.0	1.41	220.547	3.035	2.82	3.39	1.202	0.51345	0.3480	0.1340	0.1165
236	437500	CLKY	7	33	3260.0	35.30	92.351	2.650	11.10	17.40	1.540	0.13447	0.3116	0.3341	0.2609
237	438000	CLKY	7	46	7744.3	150.00	31.427	2.707	25.70	48.00	1.848	0.04818	0.2910	0.3273	0.1852
238	438070	CLKY	7	15	3742.0	30.50	123.344	0.999	5.52	6.70	1.214	0.37879	0.1804	0.3423	0.3400
239	529500	TNVA	2	30	6074.7	112.00	54.239	1.560	13.22	23.10	1.770	0.70030	0.3926	0.3152	0.3606
240	530000	TNVA	2	27	2643.4	40.00	66.006	2.735	10.46	15.60	1.396	2.95155	0.2904	0.2512	0.2624
241	530504	TNVA	3	37	4912.7	70.00	71.181	1.990	11.80	16.50	1.400	1.12500	0.3012	0.3305	0.3155
242	531000	TNVA	3	11	14522.7	290.00	50.070	2.330	26.04	45.90	1.763	0.40720	0.2533	0.2933	0.2563
243	531500	TNVA	3	53	14073.4	319.00	44.180	3.443	33.14	50.90	1.777	0.31818	0.2763	0.3129	0.2482
244	610000	TNKY	7	32	7747.8	87.70	86.598	1.062	9.74	13.29	1.352	0.16269	0.3737	0.4244	0.3375
245	610470	TNKY	7	10	461.3	0.96	480.521	2.013	1.39	1.66	1.196	1.09337	0.2482	0.1398	0.1552
246	610500	TNKY	7	45	11031.6	227.00	52.122	4.280	31.20	37.50	1.242	0.11742	0.3561	0.2731	0.1774
247	610503	TNKY	7	10	731.9	0.82	892.541	1.449	1.09	1.21	1.116	1.00246	0.2574	-0.0564	0.0749
248	610545	TNKY	7	15	5408.7	18.70	81.440	2.422	12.90	13.00	1.070	0.21970	0.2471	0.0114	0.1427
249	610820	TNKY	7	9	107.1	0.13	823.932	2.412	6.56	0.62	1.107	1.42330	0.4257	0.4519	0.2233
250	611264	3PKY	7	14	2555.1	14.60	175.010	2.049	5.47	6.13	1.121	0.33767	0.3170	0.2221	0.1704
251	612030	3PIL	7	43	4384.0	244.00	17.967	2.073	22.49	49.31	2.192	0.05095	0.3151	0.1402	0.0261
252	612230	3PIL	7	21	141.5	0.27	598.236	0.593	0.40	0.80	2.000	2.67000	0.3304	0.1013	0.1114
253	614602	3PIL	7	14	469.3	1.95	240.659	.	.	.	.	0.45200	0.2178	0.1637	0.1457

OBS	CV	SK	KUR	MEVL	AEVL	MGVL	AGVL	KEVL	WAKL	AWKL	BWKL	CWKL	DWKL
235	0.592405	0.31061	-0.7322	0.7102	0.5021	0.7225	0.5277	0.0566	-0.1444	5.4057	12.0731	0.9475	-0.2951
236	0.616050	1.44153	2.2576	0.7405	0.4495	0.6580	0.3392	-0.2432	-0.0512	13.5977	25.2994	0.4455	0.1109
237	0.553985	1.56988	1.9031	0.7577	0.4198	0.7197	0.3224	-0.2300	0.0000	41.0900	93.1579	0.5455	0.0321
238	0.339651	1.90631	3.8554	0.8498	0.2602	0.8245	0.1940	-0.2518	0.1703	16.0516	27.9539	0.1921	0.2992
239	0.675270	2.03043	10.9623	0.7490	0.4364	0.6956	0.2915	-0.3231	0.0447	7.5459	14.3639	0.2748	0.3751
240	0.547734	1.91263	5.7679	0.7445	0.4045	0.7440	0.3546	-0.1270	0.1626	3.0907	5.3777	6.2950	0.2499
241	0.630012	2.20376	5.7362	0.7492	0.4315	0.6902	0.2931	-0.3196	0.0000	13.7123	25.6205	0.3445	0.2849
242	0.442046	1.22734	1.2943	0.7891	0.3654	0.7637	0.3045	-0.1690	0.1171	11.9444	27.9440	0.4532	0.0363
243	0.618823	2.65449	10.4423	0.7532	0.4275	0.7134	0.3247	-0.2388	0.2344	2.9053	9.9529	0.4129	0.1740
244	0.968110	2.79660	8.2635	0.6722	0.5680	0.6000	0.3553	-0.3616	0.1636	1.7810	3.5226	0.3022	0.4309
245	0.437634	0.36651	-0.5324	0.7916	0.3733	0.7962	0.3950	0.0487	0.1796	2.4252	5.7154	0.4703	-0.0725
246	0.467283	1.30229	1.4151	0.7035	0.5137	0.6706	0.4359	-0.1544	-0.3375	31.7733	48.2057	0.7264	-0.0377
247	0.375118	-0.53238	-1.2230	0.8023	0.3424	0.8732	0.4398	0.3078	0.0090	4.9393	10.2071	0.9947	-0.7787
248	0.414901	-0.12327	-1.4687	0.7743	0.3564	0.8423	0.4315	0.2631	0.0000	10.4444	39.1445	1.2031	-0.8069
249	0.959130	1.34807	0.9872	0.6453	0.6161	0.5619	0.3556	-0.4033	0.2570	0.0000	0.0000	0.5553	0.2545
250	0.565367	1.03213	1.1082	0.7354	0.4585	0.7195	0.4234	-0.0796	0.2904	1.2626	1.6561	0.2293	0.2808
251	0.550578	0.45160	-0.9180	0.7374	0.4530	0.7473	0.4742	0.0467	0.2388	-0.6355	3.2212	1.5315	-0.4511
252	0.412300	1.00169	0.1579	0.7182	0.4881	0.6848	0.4082	-0.1462	0.2821	0.8053	0.1218	0.0000	0.0000
253	0.356704	0.45056	-0.7618	0.8186	0.3142	0.8208	0.3187	0.0154	0.0000	19.2198	34.6402	0.5545	-0.2283

TABLE A2. Comparison of Regional Average Historical and Simulated L-Moments: Clustering on LCV and QSP. *

Region No:		Historic L-Moments			
3	1.0000	0.2383	0.1760	0.1810	0.0781
4	1.0000	0.2823	0.1988	0.1839	0.0786
1	1.0000	0.3242	0.2758	0.1914	0.1016
2	1.0000	0.3862	0.3058	0.1900	0.0809
5	1.0000	0.4432	0.4035	0.2784	0.1657
Region No:		**Average Simulated L-Moments using EV1**			
3	1.0000	0.2403	0.1781	0.1676	0.0653
4	1.0000	0.2816	0.1817	0.1644	0.0560
1	1.0000	0.3153	0.1910	0.1547	0.0717
2	1.0000	0.3536	0.2122	0.1488	0.0686
5	1.0000	0.3786	0.2239	0.1472	0.0789
Region No:		**Average Simulated L-Moments using GEV**			
3	1.0000	0.2397	0.1870	0.1747	0.0722
4	1.0000	0.2796	0.2148	0.1883	0.0733
1	1.0000	0.3177	0.2739	0.2023	0.1047
2	1.0000	0.3633	0.2982	0.2033	0.1012
5	1.0000	0.4122	0.3655	0.2382	0.1348

* Regions arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK)

TABLE A3. Comparison of Regional Average Historical and Simulated L-Moments: Clustering on EV1 Parameters and QSP. *

Region No:		Historic L-Moments			
4	1.0000	0.2310	0.1641	0.1829	0.0741
5	1.0000	0.2801	0.2008	0.1834	0.0743
3	1.0000	0.3116	0.2637	0.1869	0.1001
2	1.0000	0.3621	0.2837	0.2074	0.1097
1	1.0000	0.4196	0.3703	0.2451	0.1359
Region No:		**Average Simulated L-Moments using EV1**			
4	1.0000	0.2329	0.1767	0.1687	0.0638
5	1.0000	0.2798	0.1863	0.1694	0.0564
3	1.0000	0.3053	0.1863	0.1554	0.0698
2	1.0000	0.3393	0.2078	0.1551	0.0688
1	1.0000	0.3678	0.2177	0.1477	0.0755
Region No:		**Average Simulated L-Moments using GEV**			
4	1.0000	0.2328	0.1748	0.1730	0.0692
5	1.0000	0.2779	0.2175	0.1894	0.0716
3	1.0000	0.3057	0.2609	0.1976	0.1015
2	1.0000	0.3423	0.2673	0.1905	0.0870
1	1.0000	0.3950	0.3424	0.2305	0.1295

* Regions arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK)

TABLE A4. Comparison of Regional Average Historical and Simulated L-Moments: Clustering on LCV, LSKEW and QSP. *

Region No:			Historic L-Moments		
4	1.0000	0.2229	0.0530	0.1324	0.0519
3	1.0000	0.2613	0.2005	0.1713	0.0784
5	1.0000	0.3224	0.2420	0.1853	0.0809
1	1.0000	0.3383	0.3187	0.2178	0.1176
2	1.0000	0.4615	0.4672	0.3275	0.1847
Region No:			**Average Simulated L-Moments using EV1**		
4	1.0000	0.2263	0.1785	0.1751	0.0615
3	1.0000	0.2612	0.1791	0.1618	0.0679
5	1.0000	0.3109	0.1991	0.1604	0.0627
1	1.0000	0.3240	0.1962	0.1545	0.0730
2	1.0000	0.3825	0.2274	0.1480	0.0817
Region No:			**Average Simulated L-Moments using GEV**		
4	1.0000	0.2241	0.0886	0.1432	0.0391
3	1.0000	0.2604	0.2050	0.1748	0.0782
1	1.0000	0.3296	0.3083	0.2198	0.1200
5	1.0000	0.3124	0.2465	0.1915	0.0847
2	1.0000	0.4221	0.4047	0.2669	0.1533

* Regions arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK)

TABLE A5. Comparison of Regional Average Historical and Simulated L-Moments: Clustering on GEV Parameters and QSP. *

Region No:			Historic L-Moments		
5	1.0000	0.2379	0.1054	0.1379	0.0527
4	1.0000	0.2735	0.0767	0.1239	0.0644
1	1.0000	0.2827	0.2877	0.2218	0.1145
2	1.0000	0.3330	0.3218	0.2319	0.0952
3	1.0000	0.3566	0.2335	0.1385	0.0700
6	1.0000	0.4233	0.4544	0.3236	0.1868
Region No:			**Average Simulated L-Moments using EV1**		
5	1.0000	0.2399	0.1783	0.1685	0.0648
4	1.0000	0.2731	0.1889	0.1787	0.0543
1	1.0000	0.2802	0.1841	0.1593	0.0687
2	1.0000	0.3186	0.1961	0.1559	0.0660
3	1.0000	0.3363	0.2029	0.1535	0.0743
6	1.0000	0.3671	0.2185	0.1494	0.0766
Region No:			**Average Simulated L-Moments using GEV**		
5	1.0000	0.2386	0.1284	0.1498	0.0516
4	1.0000	0.2701	0.1240	0.1476	0.0451
1	1.0000	0.2800	0.2804	0.2113	0.1086
3	1.0000	0.3419	0.2429	0.1782	0.0921
2	1.0000	0.3229	0.3022	0.2191	0.1034
6	1.0000	0.3983	0.4013	0.2682	0.1547

* Regions arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK)

TABLE A6. Comparison of Regional Average Historical and Simulated L-Moments: USGS Regions. *

Region No:					
			Historic L-Moments		
6	1.0000	0.2698	0.2230	0.1867	0.0743
1	1.0000	0.2781	0.2728	0.2139	0.0989
4	1.0000	0.2830	0.2037	0.1582	0.0801
2	1.0000	0.2839	0.2265	0.1994	0.1044
7	1.0000	0.3034	0.2691	0.2028	0.1061
3	1.0000	0.3115	0.2695	0.1830	0.0887
5	1.0000	0.3185	0.2852	0.2061	0.1042
Region No:			Average Simulated L-Moments using EV1		
6	1.0000	0.2622	0.1850	0.1648	0.0648
1	1.0000	0.2679	0.1887	0.1596	0.0669
4	1.0000	0.2770	0.1872	0.1626	0.0702
2	1.0000	0.2815	0.1881	0.1600	0.0682
7	1.0000	0.2959	0.1853	0.1564	0.0672
3	1.0000	0.3022	0.1909	0.1576	0.0689
5	1.0000	0.3040	0.1937	0.1577	0.0724
Region No:			Average Simulated L-Moments using GEV		
6	1.0000	0.2656	0.2198	0.1904	0.0835
7	1.0000	0.2983	0.2666	0.2031	0.1026
4	1.0000	0.2821	0.2182	0.1804	0.0865
2	1.0000	0.2844	0.2254	0.1845	0.0843
1	1.0000	0.2729	0.2597	0.2037	0.0986
3	1.0000	0.3028	0.2586	0.1934	0.0942
5	1.0000	0.3083	0.2727	0.2030	0.1041

* Regions arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK)

TABLE A.7. Comparison of Regional Average Historic and Simulated Parameters: Clustering on LCV and QSP. *

Region No. *	EV1 MEVL	EV1 AEVL	GEV MGVL	GEV AGVL	GEV KGVL #
			HISTORIC PARAMETERS		
3	0.80	0.34	0.80	0.34	-0.01
4	0.76	0.41	0.76	0.39	-0.04
1	0.73	0.47	0.70	0.40	-0.16
2	0.68	0.56	0.63	0.44	-0.20
5	0.63	0.64	0.55	0.42	-0.33
			AVERAGE SIMULATED PARAMETERS		
3	0.80	0.35	0.80	0.34	-0.03
4	0.77	0.41	0.76	0.38	-0.07
1	0.74	0.45	0.71	0.39	-0.16
2	0.71	0.51	0.66	0.42	-0.19
5	0.68	0.55	0.59	0.42	-0.28

* Regions arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK)
\# MEVL, MGVL = location parameters; AEVL, AGVL = scale parameters; and KGVL = shape parameter.

TABLE A8. Comparison of Regional Average Historic and Simulated Parameters: Clustering on EV1 Parameters and QSP. *

Region No.*	EV1 MEVL	EV1 AEVL	GEV MGVL	GEV AGVL	GEV # KGVL
			HISTORIC PARAMETERS		
4	0.81	0.33	0.81	0.34	0.01
5	0.77	0.40	0.76	0.39	-0.05
3	0.74	0.45	0.71	0.39	-0.14
2	0.70	0.52	0.66	0.43	-0.17
1	0.65	0.61	0.58	0.43	-0.29
			AVERAGE SIMULATED PARAMETERS		
4	0.81	0.34	0.81	0.33	-0.01
5	0.77	0.40	0.76	0.37	-0.07
3	0.75	0.44	0.72	0.38	-0.14
2	0.72	0.49	0.69	0.42	-0.14
1	0.69	0.53	0.62	0.42	-0.25

* Regions arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK)
# MEVL, MGVL = location parameters; AEVL, AGVL = scale parameters; and KGVL = shape parameter.

TABLE A9. Comparison of Regional Average Historics and Simulated Parameters: Clustering on LCV, LSKEW and QSP. *

Region No. *	EV1 MEVL	EV1 AEVL	GEV MGVL	GEV AGVL	GEV # KGVL
			HISTORIC PARAMETERS		
4	0.81	0.32	0.85	0.37	0.19
3	0.78	0.38	0.77	0.36	-0.05
5	0.73	0.47	0.71	0.42	-0.11
1	0.72	0.49	0.68	0.38	-0.22
2	0.62	0.67	0.52	0.38	-0.42
			AVERAGE SIMULATED PARAMETERS		
4	0.81	0.33	0.83	0.36	0.13
3	0.78	0.38	0.77	0.36	-0.05
5	0.74	0.45	0.72	0.40	-0.12
1	0.73	0.47	0.69	0.38	-0.20
2	0.68	0.55	0.58	0.40	-0.34

* Regions arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK)
# MEVL, MGVL = location parameters; AEVL, AGVL = scale parameters; and KGVL = shape parameter.

TABLE A10. Comparison of Regional Average Historic and Simulated Parameters: Clustering on GEV Parameters and QSP. *

Region No.	EV1 MEVL	EV1 AEVL	GEV MGVL	GEV AGVL	GEV # KGVL
			HISTORIC PARAMETERS		
5	0.80	0.34	0.82	0.37	0.10
4	0.77	0.39	0.80	0.45	0.15
1	0.76	0.41	0.74	0.34	-0.18
2	0.72	0.48	0.68	0.37	-0.22
3	0.70	0.51	0.68	0.47	-0.10
6	0.65	0.61	0.56	0.36	-0.40
			AVERAGE SIMULATED PARAMETERS		
5	0.80	0.35	0.81	0.36	0.07
4	0.77	0.39	0.79	0.41	0.07
1	0.77	0.40	0.74	0.34	-0.16
2	0.73	0.46	0.69	0.37	-0.20
3	0.72	0.49	0.69	0.44	-0.11
6	0.69	0.53	0.60	0.38	-0.33

* Regions arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK)
\# MEVL, MGVL = location parameters; AEVL, AGVL = scale parameters; and KGVL = shape parameter.

TABLE A11. Comparison of Regional Average Historic and Simulated Parameters: Clustering on USGS Regions. *

Region No. *	EV1 MEVL	EV1 AEVL	GEV MGVL	GEV AGVL	GEV # KGVL
			HISTORIC PARAMETERS		
6	0.78	0.39	0.76	0.36	-0.08
1	0.77	0.40	0.74	0.34	-0.15
4	0.76	0.41	0.75	0.39	-0.05
2	0.76	0.42	0.74	0.38	-0.09
7	0.75	0.44	0.72	0.37	-0.15
3	0.74	0.45	0.71	0.38	-0.15
5	0.73	0.46	0.70	0.38	-0.17
			AVERAGE SIMULATED PARAMETERS		
6	0.78	0.38	0.77	0.36	-0.08
1	0.78	0.39	0.75	0.34	-0.13
4	0.77	0.40	0.75	0.38	-0.07
2	0.77	0.41	0.75	0.38	-0.08
7	0.75	0.43	0.73	0.37	-0.14
3	0.75	0.44	0.72	0.38	-0.13
5	0.75	0.44	0.71	0.38	-0.15

* Regions arranged in increasing steepness of the corresponding flood frequency growth curves (i.e. increasing LCV or LSK)
\# MEVL, MGVL = location parameters; AEVL, AGVL = scale parameters; and KGVL = shape parameter.