Final Report

FHWA/IN/JTRP-2005/18

FLOOD FREQUENCY RELATIONSHIPS FOR INDIANA

by

A. Ramachandra Rao
Professor Emeritus
Principal Investigator
School of Civil Engineering
Purdue University

Joint Transportation Research Program
Project No. C-36-62O
File No. 9-8-15
SPR-2858

Prepared in Cooperation with the
Indiana Department of Transportation and the
Federal Highway Administration
U.S. Department of Transportation

School of Civil Engineering
Purdue University
March 2006

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA/IN/JTRP-2005/18		

4. Title and Subtitle	5. Report Date
Flood Frequency Relationships for Indiana	March 2006
	6. Performing Organization Code
7. Author(s) A. Ramachandra Rao	8. Performing Organization Report No. FHWA/IN/JTRP-2005/18
9. Performing Organization Name and Address Joint Transportation Research Program Purdue University West Lafayette, IN 47907-1284	10. Work Unit No.
	11. Contract or Grant No. SPR-2858
12. Sponsoring Agency Name and Address Indiana Department of Transportation, State Office Bldg, 100 N Senate Ave., Indianapolis, IN 46204	13. Type of Report and Period Covered Final Report
	14. Sponsoring Agency Code

15. Supplementary Notes

Prepared in cooperation with the Indiana Department of Transportation and Federal Highway Administration.

16. Abstract

The objective of the research presented in this report is the development of relationships to estimate flood magnitudes for Indiana streams. In order to achieve this goal several probability distributions were evaluated. The Pearson (3) (LP(3)) and the Generalized Extreme Value (GEV) distributions were found to be the best distributions for Indiana data. Because of the requirement that Log Pearson (3) (LP(3)) distribution must be used in federally-funded projects it was retained in the study.

Relationships were developed for the flood frequencies to be estimated by the LP(3) distributions. The State of Indiana has been divided into regions, seven of which are homogeneous and one heterogeneous. The floods of specific return periods were related to watershed characteristics which are relatively easy to measure by the generalized least squares (GLS) method.

The regional flood estimates based on L-moments have been developed and presented for all the eight regions. These are based on P(3), GEV and LP(3) distributions. The GLS based regional regression analysis was used to relate the flood magnitudes based on these distributions and watershed parameters. The L-moment based methods and the regional regression relationships are compared to each other by split sample tests.

The following conclusions are presented based on this study. (1) Identifying homogeneous regions prior to development of flood frequency relationships substantially reduce the prediction errors. (2) The L-moment based flood frequency relationships are more accurate than those developed by regional regression analysis (3) The Pearson (3) and GEV distributions give more accurate flood flow estimates than the LP(3) distribution.

17. Key Words	18. Distribution Statement
flood frequency analysis, watersheds, generalized least squares	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	138	

Form DOT F 1700.7 (8-69)

TECHNICAL *Summary*

INDOT Research

Technology Transfer and Project Implementation Information

TRB Subject Code: 22-7 River and Stream Mechanics
Publication No.: FHWA/IN/JTRP-2005/18, SPR-2858

March 2006
Final Report

Flood Frequency Relationships for Indiana

Introduction

The objective of the research was the development of relationships to estimate flood magnitudes for Indiana streams. In order to achieve this goal several probability distributions were evaluated. The Pearson (3) (LP(3)) and the Generalized Extreme Value (GEV) distributions were found to be the best distributions for Indiana data. Because of the requirement that Log Pearson (3) (LP(3)) distribution must be used in federally-funded projects, it was retained in the study.

Findings

Relationships were developed for the flood frequencies to be estimated by the LP(3) distributions. The State of Indiana has been divided into regions, seven of which are homogeneous and one heterogeneous. The floods of specific return periods were related to watershed characteristics which are relatively easy to measure by the generalized least squares (GLS) method.

The regional flood estimates based on L-moments have been developed and presented for all the eight regions. These are based on P(3), GEV and LP(3) distributions. The GLS based regional regression analysis was used to relate the flood magnitudes based on these distributions and watershed parameters. The L-moment based methods and the regional regression relationships are compared to each other by split sample tests.

Following are the conclusions of this study.

1. The prediction errors were smallest for homogeneous watersheds and highest for heterogeneous watersheds.
2. The L-moment based method is more accurate than the GLS method.
3. The Pearson (3) and generalized extreme value distributions give more accurate predictions than the log Pearson (3) distribution.

Implementation

A proposal for an implementation project will be developed by the Principal Investigator, which will include a manual and a CD-ROM to use the relationships discussed in the final report. A workshop to train interested engineers in using these relationships will be presented.

Contacts

For more information:

Prof. A. Ramanchandra Rao
Principal Investigator
School of Civil Engineering
Purdue University
West Lafayette IN 47907
Phone: (765) 494-2176
Fax: (765) 496-1988
E-mail: rao@purdue.edu

Indiana Department of Transportation
Division of Research
1205 Montgomery Street
P.O. Box 2279
West Lafayette, IN 47906
Phone: (765) 463-1521
Fax: (765) 497-1665

Purdue University
Joint Transportation Research Program
School of Civil Engineering
West Lafayette, IN 47907-1284
Phone: (765) 494-9310
Fax: (765) 496-7996
E-mail: jtrp@ecn.purdue.edu
http://www.purdue.edu/jtrp

INDOT Research

TECHNICAL *Summary*

Technology Transfer and Project Implementation Information

TRB Subject Code: 22-7 River and Stream Mechanics
Publication No.: FHWA/IN/JTRP-2005/18, SPR-2858
March 2006
Final Report

Flood Frequency Relationships for Indiana

Introduction

The objective of the research was the development of relationships to estimate flood magnitudes for Indiana streams. In order to achieve this goal several probability distributions were evaluated. The Pearson (3) (LP(3)) and the Generalized Extreme Value (GEV) distributions were found to be the best distributions for Indiana data. Because of the requirement that Log Pearson (3) (LP(3)) distribution must be used in federally-funded projects, it was retained in the study.

Findings

Relationships were developed for the flood frequencies to be estimated by the LP(3) distributions. The State of Indiana has been divided into regions, seven of which are homogeneous and one heterogeneous. The floods of specific return periods were related to watershed characteristics which are relatively easy to measure by the generalized least squares (GLS) method.

The regional flood estimates based on L-moments have been developed and presented for all the eight regions. These are based on P(3), GEV and LP(3) distributions. The GLS based regional regression analysis was used to relate the flood magnitudes based on these distributions and watershed parameters. The L-moment based methods and the regional regression relationships are compared to each other by split sample tests.

Following are the conclusions of this study.

1. The prediction errors were smallest for homogeneous watersheds and highest for heterogeneous watersheds.
2. The L-moment based method is more accurate than the GLS method.
3. The Pearson (3) and generalized extreme value distributions give more accurate predictions than the log Pearson (3) distribution.

Implementation

A proposal for an implementation project will be developed by the Principal Investigator, which will include a manual and a CD-ROM to use the relationships discussed in the final report. A workshop to train interested engineers in using these relationships will be presented.

Contacts

For more information:

Prof. A. Ramanchandra Rao
Principal Investigator
School of Civil Engineering
Purdue University
West Lafayette IN 47907
Phone: (765) 494-2176
Fax: (765) 496-1988
E-mail: rao@purdue.edu

Indiana Department of Transportation
Division of Research
1205 Montgomery Street
P.O. Box 2279
West Lafayette, IN 47906
Phone: (765) 463-1521
Fax: (765) 497-1665

Purdue University
Joint Transportation Research Program
School of Civil Engineering
West Lafayette, IN 47907-1284
Phone: (765) 494-9310
Fax: (765) 496-7996
E-mail: jtrp@ecn.purdue.edu
http://www.purdue.edu/jtrp

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. Introduction

The basic objective of the research reported herein is to analyze Indiana flood data and to develop regional equations to estimate magnitudes of floods corresponding to specified recurrence intervals. The commonly used recurrence intervals of 2, 5, 10, 25, 50 and 100 years are used in this research. In Indiana, the equations which are being used presently to estimate floods were developed by Glatfelter (1984) using data available up to or a few years before 1982. More than 20 years of additional data are available since Glatfelter's work. The additional data offers an incentive to develop more accurate relationships to estimate flood magnitudes.

In addition to the improvements which can be brought about by using the additional data, there are two strong reasons to develop new flood frequency relationships. Both of these are related to the drawbacks in Glatfelter's work. The first of these is that the ordinary least squares (OLS) method was used by him to develop these relationships. This was the standard practice in U.S.G.S. at that time. In fact, all the states followed the same procedure. However, the nature of the flood data is such that, later, the Generalized Least Squares (GLS) method was shown to be better suited for the problem. The GLS was shown to reduce the rather substantial errors in these relationships (Fig. 1.1.1).

The second major drawback in Glatfelter's work is that he used the data from the major river basins in Indiana. These river basins are shown in Fig. 1.1.2. However, as demonstrated by Rao and Hamed (1997), these river basins are not homogeneous in their flood characteristics. Consequently the flood frequency relationships developed by using these data can be improved substantially.

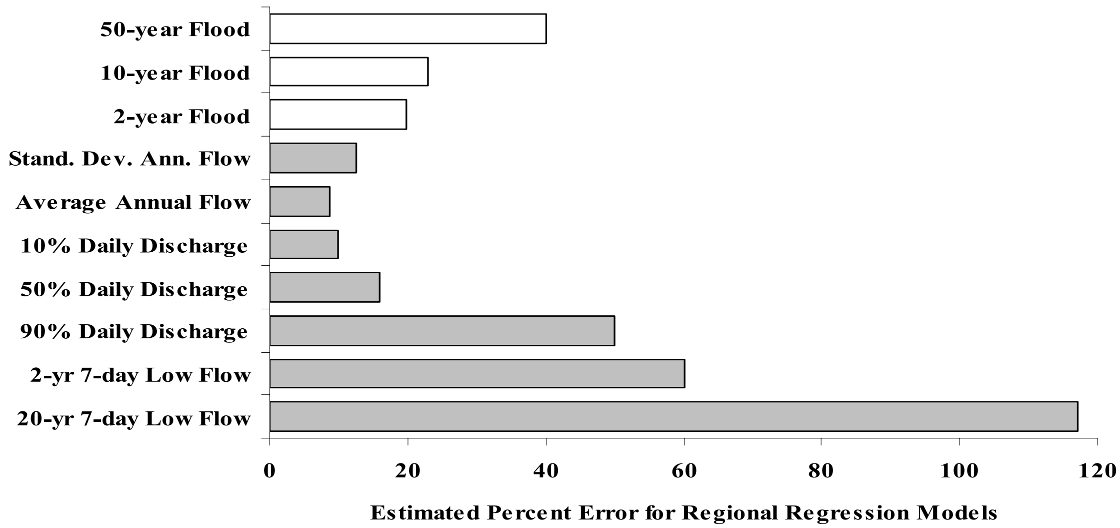

Figure 1.1.1. Percentage error for regional regression estimators of different statistics in the Potomac River basin (after Thomas and Benson (1970))

Figure 1.1.2. Regions by Glatfelter (1984)

In order to identify homogeneous regions in Indiana, a JTRP study was conducted at Purdue University. Different methods based on trial and error, clustering algorithms, fuzzy algorithms and neural networks were used to identify homogeneous regions in Indiana [(Rao et al. (2002), Srinivas and Rao (2002), Iblings and Rao (2003), Srinivas and Rao (2003)]. The homogeneity of these regions was tested by using the statistics developed by Hosking and Wallis (1993, 1997). Three of these results are shown in figures 1.1.3-1.1.5. A comparison of these regions in Fig. 1.1.2 and Figs. 1.1.3-1.1.5 demonstrates the fact that flood homogeneous regions in Indiana do not correspond to river basin boundaries. The annual maximum flood data from the regions in Figs. 1.1.4 and 1.1.5 or slight modifications are used in the present study. Separate flood frequency relationships are developed for each region.

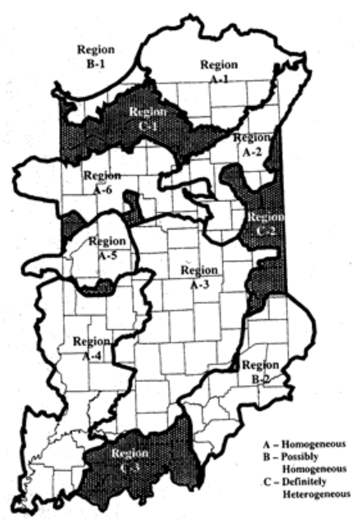

Figure 1.1.3. Flood Homogeneous regions of Indiana (Rao et al. (2002))

The U.S. Water Resources Council mandated the use of log-Pearson (III) (LP(III)) distribution for estimating floods in the U.S whenever Federal funds are used. The LP(III) distribution is very sensitive to skewness coefficients of the annual maximum flood data. These skewness coefficients vary considerably in any given region and hence the flood estimates based on them also vary (McCormick and Rao (1995)). Also, the LP(III) distribution may not be the best distribution to describe the flood data (Wallis and Wood (1985), Rao and Hamed (2000),

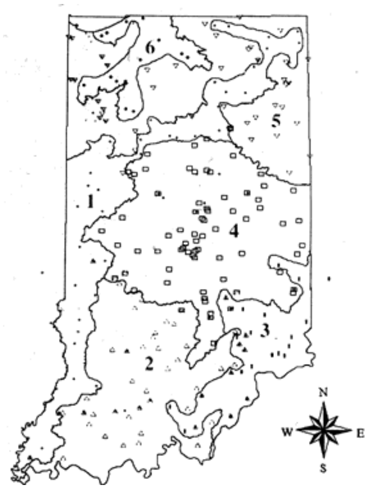

Figure 1.1.4. Flood Homogeneous regions by hybrid cluster method (Srinivas and Rao (2002)).

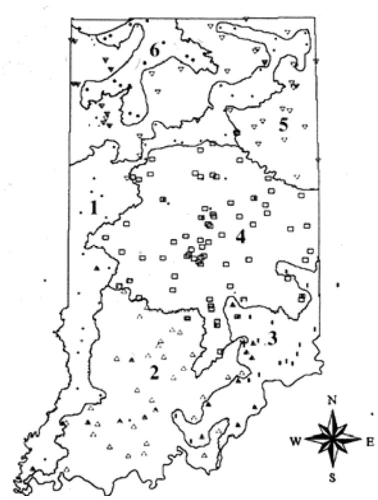

Figure 1.1.5. Flood Homogeneous Regions by Fuzzy Cluster Analysis (Srinivas and Rao (2003))

Rao et al. (2003)). Consequently, two sets of relationships, one based on LP(III) and another one based on a better distribution for a given region are developed for each region. The relationships based on LP(III) distribution are used where it is required. The other set of relationships are designed to be used where the relationships based on LP(III) distributions are not required.

1.1. Objectives of the Study

The optimal statistical distribution which may be used may vary from one region to another. Consequently, data from each region are analyzed to determine the best distribution for each region. The commonly used distributions such as Generalized Extreme Value (GEV), Generalized Logistic, LP(III) and other distributions are selected for this analysis. The tests designed by Hosking and Wallis (1993, 1997) as well as other standard tests such as χ^2_- or Kolmogorov-Smirnov tests are used to select the best distribution for each region. This is discussed in chapter 2.

The flood frequency relationships based on LP (III) distribution must be used whenever federal funds are used in any project. Consequently, the flood magnitudes corresponding to the specified frequencies are estimated by using all the available data and the Water Resources Council (WRC) method. These flood values are related to the physiographic and meteorologic variables so that they may be used to estimate the flood magnitudes at locations where flood data are not available. This aspect of the study is discussed in chapter 3.

In developing flood frequency relationships, regression-based relationships are commonly used. However, recent research based on L-moments has demonstrated that the results based on L-moments are as good as or better than those based on other regression relationships. Consequently, L-moment based relationships are developed by using Indiana data. The accuracy of the L-moment based method is tested by using split sample tests. The development of L-Moment based relationships is discussed in chapter 4.

In order to use the L-moment based approach for ungaged watersheds, the average annual flood or a similar statistic must be estimated from easily measured watershed and meteorological characteristics. These relationships are developed by using GLS techniques. The accuracy of these relationships is tested by using split sample tests. Consequently, development of relationships for estimating the average annual maximum flow and testing them are discussed in chapter 5.

Although the L-moment based methods are supposed to be better than those based purely on regression relationships, the universality of this assertion has not been established. The claims of superiority of the L-moment method compared to the regression relationships are investigated by using a comparison of these methods. Consequently the regression relationships for floods of different return periods are developed for each region. The GLS method is used for developing these relationships. The correlation between the dependent variables are tested and only one of the variables of a pair tested is retained in order to eliminate spurious correlations. Development of these flood frequency regression relationships is discussed in chapter 6.

In selecting these procedures for flood frequency analysis the accuracies of these methods must be established. The accuracies of L-moment and regression analysis methods are established by using the split sample technique. Part of the data from a region is used to establish these relationships. The remaining part of the data is used to test the accuracy of these relationships. Thus the errors of estimation are determined. This aspect of the study discussed in chapter 7.

A summary and a set of conclusions are presented in chapter 8.

The details of much of the work reported herein are found in three reports:

1. Estimation of Peak Discharges of Indiana Streams by Using log Pearson (III) Distribution, Interim Report No. 1, by David Knipe and A.R. Rao, May, 2005.
   (Knipe and Rao (2005))

2. Indiana Flood Data Analysis, Interim Report No. 2, by Shalini Kedia and A. R. Rao, July, 2005. *(Kedia and Rao (2005))*
3. Flood Estimates for Indiana Steams, Interim Report No. 3, by En-Ching Hsu and A. R. Rao, August, 2005. *(Hsu and Rao (2005)).*

In order to keep the length of this report within reasonable limits the readers are referred to these reports. They are available from Purdue University libraries.

II. Selection of Distributions

An important problem in hydrology is the estimation of flood magnitudes, especially because planning and design of water resource projects and flood plain management depend on the frequency and magnitude of peak discharges. A flood event can be described as a multivariate event whose main characteristics can be summarized by its peak, volume, and duration, which may be correlated. However, flood frequency analysis has often concentrated on the analysis of flood peaks. Several summaries, discussions and extensive reviews of the field of flood frequency analysis are given by Chow (1964), Yevjevich (1972), Kite (1977), Singh (1987), Potter (1987), Bobee and Ashkar (1991), McCuen (1993), Stedinger et al. (1993), and Rao and Hamed (2000).

In the statistical analysis of floods extreme value probability distributions are fitted to measured peak flows. This method is data intensive and is applicable only to gauged watersheds. Selection of probability distribution is generally arbitrary, as no physical basis is available to rationalize the use of any particular distribution. Several distributions, Log-Normal, Pearson type III, Wiebull, log Pearson Type III, Generalized Extreme Value, to name a few, have been used and these may seem appropriate for a given sample of data. To check the validity of accepting a distribution, goodness-of-fit tests are used.

The U.S. Water Resources Council recommends the use of log-Pearson (III) (LP (III)) distribution for estimating floods in the U.S. Studies by Wallis and Wood (1985), Rao and Hamed (2000), and Rao et al. (2003) show that LP (III) distribution may not be the best distribution for the flood data in U.S. Therefore it is useful to test the adequacy of the distributions to determine the best distribution for a given region.

Generalized Extreme Value (GEV) distribution is considered to be an appropriate choice for annual peak floods. Stedinger and Lu (1991) developed critical values and formulas for goodness-of-fit-tests for the GEV distribution. In the past, three tests, namely, Kolmogorov-Smirnov test, the probability plot correlation test, and sample L moment ratio tests have been investigated. These tests are used to check if data available for a site are consistent with a regional GEV distribution. Zempleni (1991) proposed a test based on the stability property of GEV distributions. It provides a tool for testing the hypothesis of a sample having GEV distribution against any other probability distribution.

To identify the homogeneous regions in Indiana, different methods based on trial and error, clustering algorithms, fuzzy algorithms and neural networks have been used [(Rao et al. (2002), Srinivas and Rao (2002), Iblings and Rao (2003), Srinivas and Rao (2003)]. The homogeneity of these regions is tested by using the statistics developed by Hosking and Wallis (1993, 1997). The study by Srinivas and Rao (2002) yielded six regions shown in Figure 2.1.1. In the present study, the annual maximum flood data from the regions in Figure 2.1.1 are used. Regions 1-5 are found to be homogeneous and region six in Figure 2.1.1, containing the Kankakee River basin, is heterogeneous.

The objective of the research discussed in this chapter is to use Indiana data for a comparative analysis, and determine the best distribution for each region. These distributions include Log Pearson Type III (LP III), Generalized Extreme Value (GEV), Pearson Type III, Log Normal (III), Gamma, Generalized Pareto and Logistic distributions. The method of moments, maximum Likelihood and probability weighted moments are used for parameter estimation. The distributions fitted by using these methods are tested by using the Chi-Square and Kolmogorov-Smirnov tests. The results of these goodness-of-fit tests are used to select a distribution for a region.

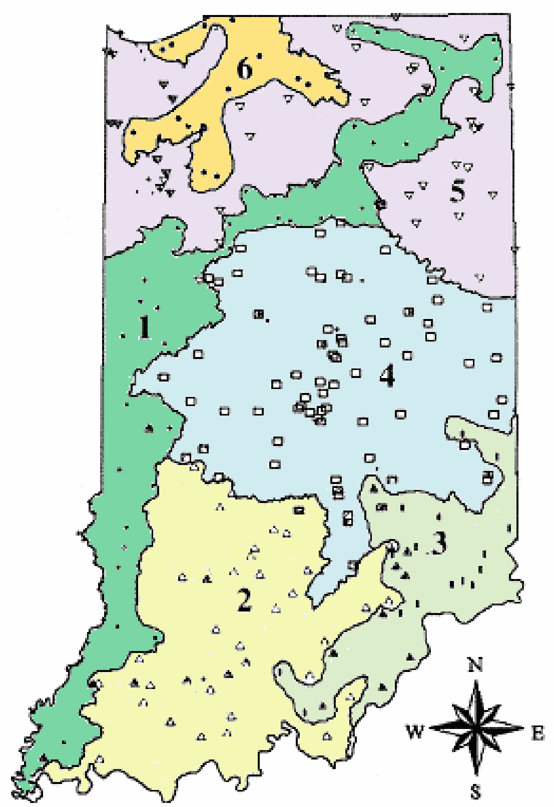

Figure 2.1.1. Homogeneous regions developed by Srinivas and Rao (2002)

2.1. Parameter and Quantile Estimation

In flood frequency analysis, an assumed probability distribution is fitted to the available data to estimate the flood magnitude for a specified return period. The choice of an appropriate probability distribution is quite arbitrary, as no physical basis is available to rationalize the use of any particular distribution. The first type of error which is associated with wrong assumption of a particular distribution for the given data can be checked to a certain extent by using goodness-of-fit tests. These are statistical tests which provide a probabilistic framework to evaluate the adequacy of a distribution.

Even if an acceptable distribution is selected, proper estimation of parameters is important. Some of the parameter estimation methods may not yield good estimates, or even converge. Therefore, some guidance is needed about the parameter estimation methods.

2.2. Parameter Estimation

Several methods can be used for parameter estimation. In this study, the method of moments (MOM), the maximum likelihood method (MLM) and the probability weighted moment method (PWM) are used for parameter estimation.

The maximum likelihood method (MLM) is considered to be the most accurate method, especially for large data sets since it leads to efficient parameter estimators with Gaussian asymptotic distributions. It provides the smallest variance of the estimated parameters, and hence of the estimated quantiles, compared to other methods. However, with small samples the results may not converge.

The method of moments (MOM) is relatively easy and is more commonly used. It can also be used to obtain starting values for numerical procedures involved in ML estimation. However, MOM estimates are generally not as efficient as the ML estimates, especially for distributions with large number of parameters, because higher order moments are more likely to be highly biased for relatively small samples.

The PWM method gives parameter estimates comparable to the ML estimates. Yet, in some cases the estimation procedures are not as complicated as in other methods and the computations are simpler. Parameter estimates from samples using PWM are sometimes more accurate than the ML estimates. Further details on this topic are found in Rao and Hamed (2000).

2.3. Quantile Estimation

After the parameters of a distribution are estimated, quantile estimates (x_T) which correspond to different return periods T may be computed. The return period is related to the probability of non-exceedence (F) by the relation,

$$F = 1 - \frac{1}{T} \tag{2.3.1}$$

where $F = F(x_T)$ is the probability of having a flood of magnitude x_T or smaller. The problem then reduces to evaluating x_T for a given value of F. In practice, two types of distribution functions are encountered. The first type is that which can be expressed in the inverse form $x_T = \phi(F)$. In this case, x_T is evaluated by replacing $\phi(F)$ by its value from equation 2.3.1. In the second type the distribution cannot be expressed directly in the inverse form $x_T = \phi(F)$. In this case numerical methods are used to evaluate x_T corresponding to a given value of $\phi(F)$.

2.4. Selection of Probability Distributions

There are many distributions which are used in flood frequency analysis. A few distributions which are commonly used in modeling flood data, are listed below and are used in the present study.

a. Three–Parameter Lognormal (LN (3)) Distribution

b. Pearson (3) Distribution

c. Log Pearson (3) Distribution

d. Generalized Extreme Value (GEV) Distribution

The choice of distributions to be used in flood frequency analysis has been a topic of interest for a long time. The best probability distribution to be used to fit the observed data cannot be determined analytically. Often, the selection of the distribution is based on an understanding of the underlying physical process. For example, the extreme value distribution might be an appropriate choice for annual peak floods. Many times, the range of the variable in the distribution function, the general shape of the distribution, and descriptors like skewness and kurtosis indicate whether a particular distribution is appropriate to a given situation. If the sample data are insufficient, the reliability in estimating more than two or three parameters may be quite low. So, a compromise has to be made between flexibility of the distribution and reliability of the parameters.

To assess the reasonability of the selected distribution, statistical tests like Chi-Square test, Kolmogrov-Smirnov tests and Akaike's Information Criterion are used. The Chi square test and Kolmogrov-Smirnov tests are discussed below.

2.4.1. Chi-Square Test

In the chi-square test, data are first divided into k class intervals. The statistic χ^2 in equation 2.4.1 is distributed as chi-square with k-1 degrees of freedom.

$$\chi^2 = \sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j} \tag{2.4.1}$$

In equation 2.4.1, O_j is the observed number of events in the class interval j, E_j is the number of events that would be expected from the theoretical distribution, and k is the number of classes to which the observed data are sorted. If the class intervals are chosen such that each interval corresponds to an equal probability, then $E_j = n/k$ where n is the sample size and k is the number of class intervals, and equation 2.4.1 reduces to equation 2.4.2.

$$\chi^2 = \frac{k}{n}\sum_{j=1}^{k} O_j^2 - n \tag{2.4.2}$$

Class intervals can be computed by using the inverse of the distribution function corresponding to different values of probability F, similar to estimating quantiles.

2.4.2. Kolmogrov-Smirnov Test

A statistic based on the deviations of the sample distribution function $F_N(x)$ from the completely specified continuous hypothetical distribution function $F_0(x)$ is used in this test. The test statistic D_N is defined in equation 2.4.3.

$$D_N = \max\left|F_N(x) - F_0(x)\right| \tag{2.4.3}$$

The values of $F_N(x)$ are estimated as N_j / N where N_j is the cumulative number of sample events in class j. $F_0(x)$ is then 1/k, 2/k, …etc., similar to the chi-square test. The value of D_N must be less than a tabulated value of D_N at the specified confidence level for the distribution to be accepted.

2.5. Procedure to Select the Distributions

The selection of probability distributions by using data from Indiana Watersheds is discussed in this section. The probability distributions included in this study are: Log Pearson Type III (LP III), Generalized Extreme Value (GEV), Pearson Type III, Log Normal (III), Gamma, Generalized Pareto and Logistic distributions. The method of moments, maximum Likelihood and probability weighted moments are used for parameter estimation. The distributions fitted by using above mentioned methods of parameter estimation are tested by using the Chi-Square and Kolmogorov-Smirnov tests for goodness-of-fit. Conclusions from these goodness-of-fit tests are used to select the distributions.

The annual peak flows from 279 gaging stations are used in this study. The annual peak flow data, as well as attributes for each gage, are found at the USGS website http://water.usgs.gov/nwis/peak. The USGS site numbers of these gaging stations are included in tables in Kedia and Rao (2005). More information can be found about these sites by using the USGS site number as an input in the USGS website. These gaging stations are divided into 6 regions by Srinivas and Rao (2002) as shown in Figure 2.1.1.

A software package in MATLAB was developed by Khaled Hamed (2001). This package has been used in this research for selecting the best distribution for each region in Indiana.

The following nine distributions are selected as candidates for the best distribution suitable to each region in Indiana: Pearson Type III, Log Pearson Type III, Generalized Extreme Value, Log Normal III, Gamma, Generalized Pareto, Logistic, Gamma and Weibull distribution. Pearson Type I, Extreme Value Type II, and Log Normal II distributions are not considered because the same distributions with three parameters are selected. Some data sets from region 1 were selected to evaluate the nine distributions. The plots of goodness of fit obtained for many of the stations, for Gamma, Generalized Pareto, Logistic and Weibull Distribution showed a very poor fit. Consequently, four distributions (log Normal III, Log Pearson III, Pearson Type III and GEV) are chosen for further investigation.

Method of moments, maximum likelihood and probability weighted moments were used to estimate the parameters. These parameters are used to calculate the quantiles corresponding to return periods of 10, 20, 50 and 100 years. Standard errors corresponding to the observed values are also obtained. Results of goodness of fit at 95% confidence limit are tabulated for each gage station in a region corresponding to each distribution and method of parameter estimation. As an example of the results for Log Pearson III distribution fitted to the data from Region 3 are shown in Table 2.5.1.

Table 2.5.1. Results for Log Pearson III distribution for Region 3.

1	2	3	4	5	6	7	8	9	10	11	12	13
station No.	USGS No.	no. of Obs.	Std Table Chi Square	Std Table K-Smirnov	ML		MOM		PWM		**Ranks of each Distribution**	**Best Method of Parameter estimation**
					actual chi square	actual k smirnov	actual chi square	actual k smirnov	actual chi square	actual k smirnov		
1	3242100	16	9.49	0.34	3	0.14	6	0.11	inv	inv	3131	MOM
2	3262750	17	9.49	0.33	9.82	0.2	9.82	0.16	inv	inv	3412	ML
3	3272900	17	9.49	0.33	inv	inv	2.29	0.08	inv	inv	2431	MOM
4	3274950	23	9.49	0.28	inv	inv	12.83	0.32	11.43	0.2	1324	MOM
5	3275900	10	9.49	0.41	inv	inv	2.8	0.15	inv	inv	3421	MOM
6	3276000	47	9.49	0.20	inv	inv	26.36	0.15	25.68	3.45	4312	ML
7	3276640	17	9.49	0.33	6.06	0.16	8.88	0.14	inv	inv	3231	MOM
8	3276700	33	9.49	0.24	4.58	0.1	6.52	0.16	inv	inv	4311	ML
9	3276770	10	9.49	0.41	6	0.15	2.8	0.12	4.4	0.17	3321	MOM
10	3276950	10	9.49	0.41	9.2	0.18	10.8	0.17	inv	inv	3421	MOM
11	3277000	41	9.49	0.21	inv	inv	4.07	0.09	inv	inv	1422	MOM
12	3277250	10	9.49	0.41	6	0.19	9.2	0.18	inv	inv	2221	MOM
13	3291780	33	9.49	0.24	inv	inv	3.61	0.07	inv	inv	1112	ML
14	3292350	17	9.49	0.33	4.67	0.08	3	0.08	3	0.08	2121	MOM
15	3294000	48	9.49	0.20	8.67	0.06	3.33	0.08	inv	inv	3311	ML
16	3302500	50	9.49	0.19	5.36	0.09	12.72	0.15	27.44	3.01	2421	ML
17	3302690	10	9.49	0.41	4.4	0.2	2.8	0.14	inv	inv	4312	ML
18	3303000	77	9.49	0.15	5.6	0.06	5.6	0.06	inv	inv	1111	ML
19	3364100	17	9.49	0.33	4.18	0.14	2.29	0.13	inv	inv	1411	MOM
20	3364500	56	9.49	0.18	3.71	0.08	4.57	0.08	inv	inv	2122	MOM
21	3364570	10	9.49	0.41	inv	inv	inv	inv	inv	inv	1110	MOM
22	3366000	20	9.49	0.29	1.6	0.07	5.6	0.13	inv	inv	4321	ML
23	3366200	34	9.49	0.23	inv	inv	6	0.09	inv	inv	1111	MOM
24	3366400	10	9.49	0.41	9.2	0.25	9.2	0.14	inv	inv	1023	ML
25	3367600	10	9.49	0.41	6	0.3	6	0.22	inv	inv	4123	ML
26	3368000	47	9.49	0.20	3.55	0.07	10.02	0.11	inv	inv	4111	ML
27	3369000	61	9.49	0.17	1.82	0.06	2.08	0.06	inv	inv	1222	ML
28	3369500	62	9.49	0.17	5.87	0.07	4.58	0.06	inv	inv	1431	MOM
29	3369700	10	9.49	0.41	inv	inv	inv	inv	inv	inv	1233	PWM
30	3374455	33	9.49	0.24	8.45	0.13	12.33	0.19	inv	inv	2412	ML

ML: Maximum Likelihhood Method
MOM: Method of Moments
PWM: Probability Weighted Moment
Actual K – Smirnov: Computed value using Kolmogorov Smirnov Test.
Actual Chi-Square: Computed value using Chi-Square Test.

An explanation to each column of Table 2.5.1 is given below.

Column 1: Station Number

Column 2: USGS Site number

Column 3: Number of observations each gauging station

Column 4: Chi Square value using the Standard Tables

Column 5: Kolmogorov Smirnov (K-S) value using Standard Tables

Column 6: Actual Chi Square Value for the data set of the particular gauging station using ML method of parameter estimation.

Column 7: Actual K-S Value for the data set with ML method.

Column 8: Actual Chi Square Value for the data set with MOM method.

Column 9: Actual K-S Value for the data set with MOM method.

Column 10: Actual Chi Square Value for the data set with PWM method.

Column 11: Actual K-S Value for the data set with PWM method.

Column 12: Ranks of each distribution (starting from GEV, followed by Pearson III, Log Normal III and Log Pearson III) for the data set. (Highest Rank 1 to Lowest Rank 4).

Column 13: Best method of parameter estimation.

Note: 'inv' in the table denotes that the results for that particular method of parameter estimation did not converge.

A larger deviation of theoretical quantile estimates from regional quantile estimates is observed for Region 6. After tabulating the results for all the regions, the best distribution is selected by comparing the results from Chi-Square and Kolmogorov-Smirnov tests with the values from standard tables at 95% confidence limits. For each region and gauging station, all four distributions are ranked in order. The distribution with lowest Chi-Square test value and Kolmogorov-Smirnov test are assigned the highest rank, Rank 1. A histogram is plotted to exhibit the frequency of Rank 1 for each frequency distribution (Figure 2.5.1 – Figure 2.5.6). The distribution with highest frequency is selected as the best distribution for that particular region. These rankings are shown in Table 2.5.1 (column 12) for region 3. For other regions the results are included Kedia and Rao (2005).

To select the best method of parameter estimation, the Chi-Square and the Kolmogorov-Smirnov test values for each distribution and gauging station, are compared with values obtained for the three methods of parameter estimation. The method with the lowest value is given the highest rank, Rank 1. Same procedure is followed for each distribution and gauging station. The method having highest frequency of Rank 1 within each station is selected as the best method of parameter estimation for that particular gauging station. The selected method of parameter estimation for each gauging site in region 3 is shown in Col. 13 of Table 2.5.1. For other regions, the results are found in Kedia and Rao (2005). In most cases, maximum likelihood method is the best one. The final results are tabulated in Table 2.5.2.

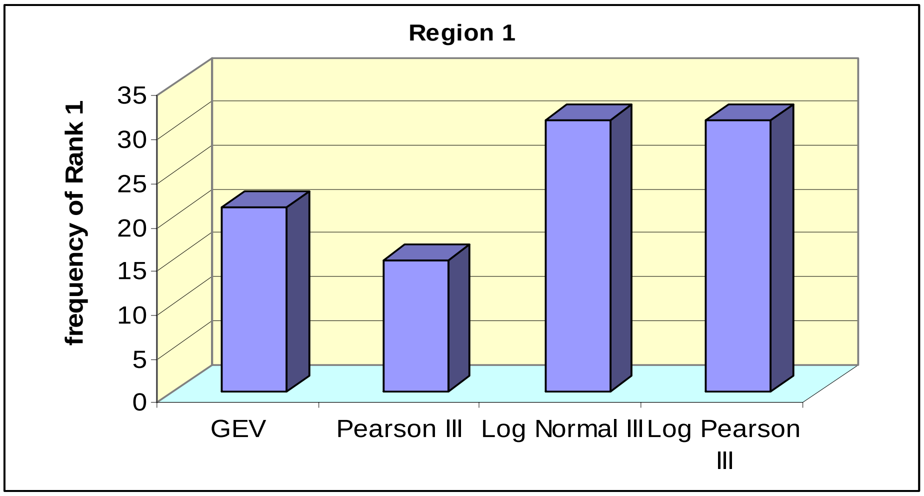

Figure 2.5.1. Region 1- Frequency of Rank 1 for selecting the best Distribution

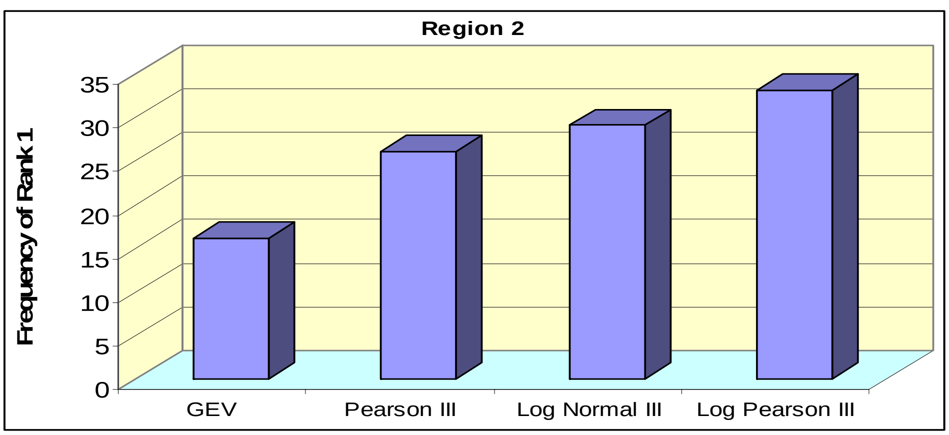

Figure 2.5.2. Region 2- Frequency of Rank 1 for selecting the best Distribution

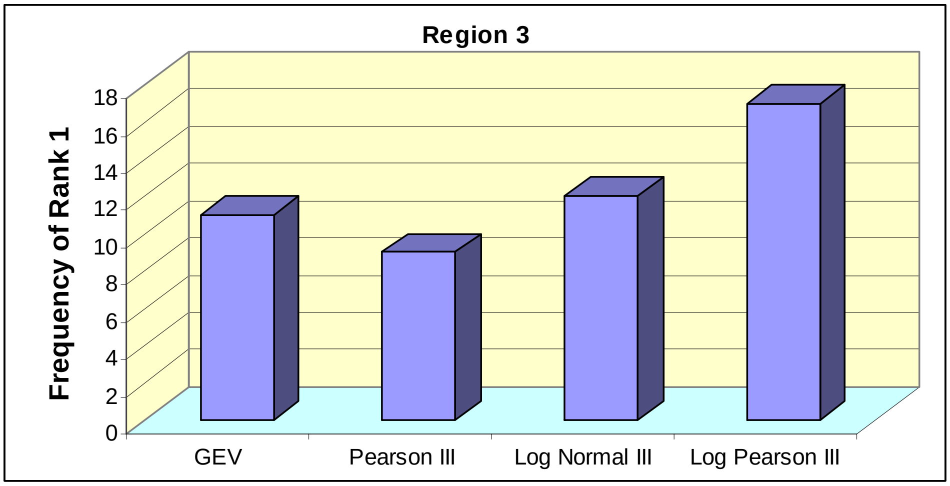

Figure 2.5.3. Region 3- Frequency of Rank 1 for selecting the best Distribution

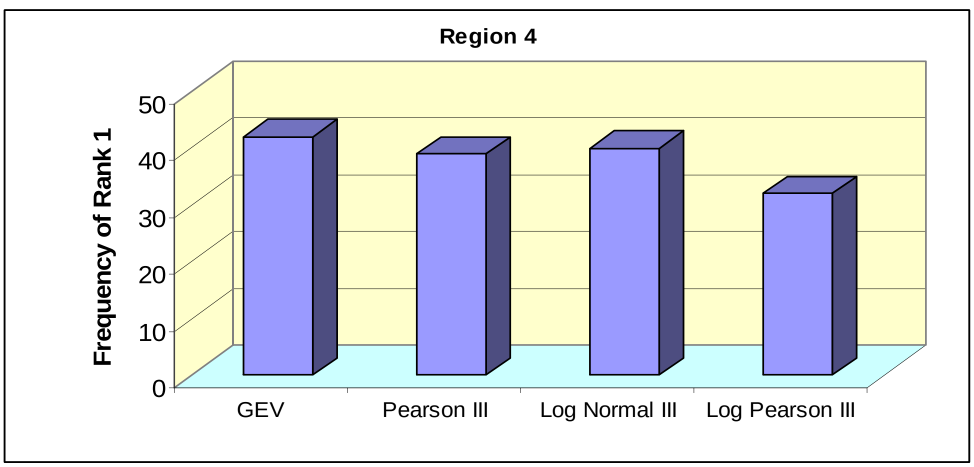

Figure 2.5.4. Region 4- Frequency of Rank 1 for selecting the best Distribution

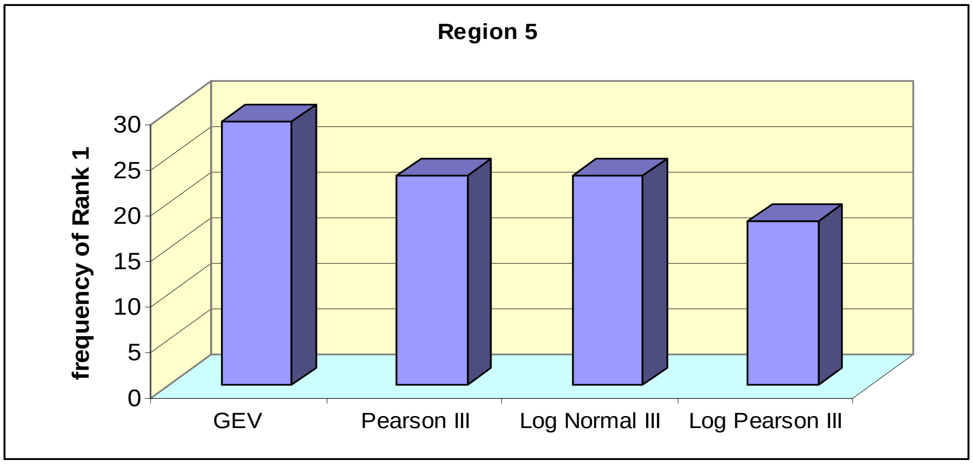

Figure 2.5.5. Region 5- Frequency of Rank 1 for selecting the best Distribution

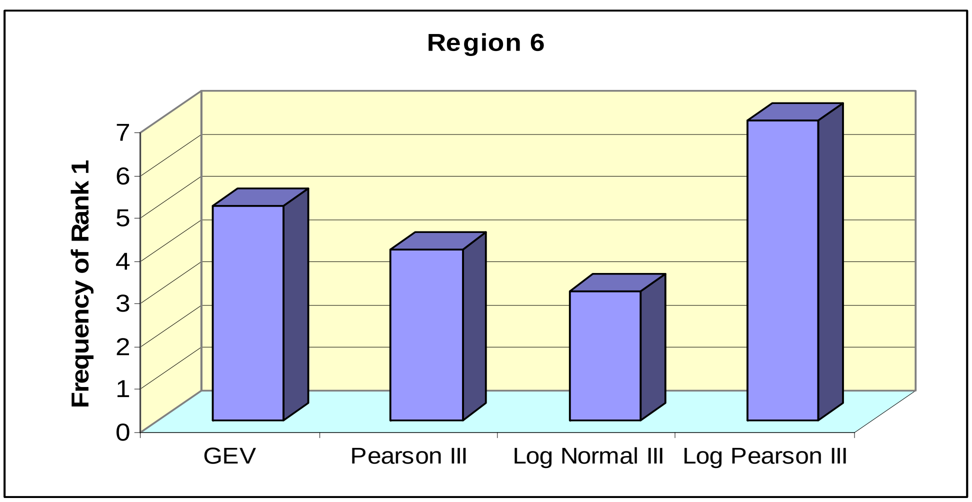

Figure 2.5.6. Region 6- Frequency of Rank 1 for selecting the best Distribution

Table 2.5.2. Selection of Best Distribution in each Region

Region Number	Number of Stations	Rank 1	Rank2	Rank3	Rank4	Best method of Parameter estimation
1	62	LP III LGN III		GEV	P III	ML
2	58	LP III	LGN III	PIII	GEV	ML
3	30	LPIII	LGN III	GEV	P III	MOM
4	73	GEV	LGN III	P III	LP III	ML
5	42	GEV	LGN III	P III	LP III	ML
6	14	LPIII	GEV	P III	LGN III	ML

The results given in Table 2.5.2 are obtained by using observations from all of Indiana watersheds. In many of these watersheds the data are quite short. For example, in Table 2.5.1, the number of observations is less than 20 in 15 out of 30 sites. The goodness-of-fit tests are not reliable for such a small number of observations. Therefore, only those data sets which have more than 30 observations are considered. Same procedures for ranking the four distributions and the method of parameter estimation are adopted and the results are shown in Figure 2.5.7 – 2.5.12. The new rankings given to the distributions and method of parameter estimation for each region of Indiana watersheds are given in Table 2.5.3.

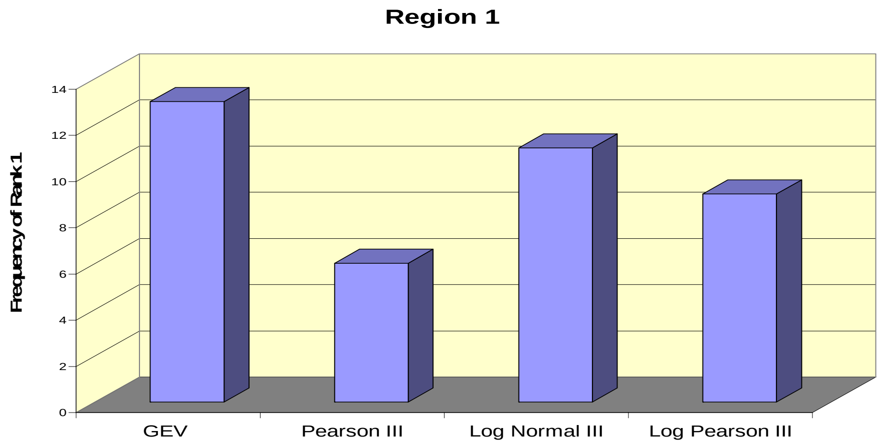

Figure 2.5.7. Region 1- Frequency of Rank 1 for selecting the best Distribution with more than 30 observations at each site.

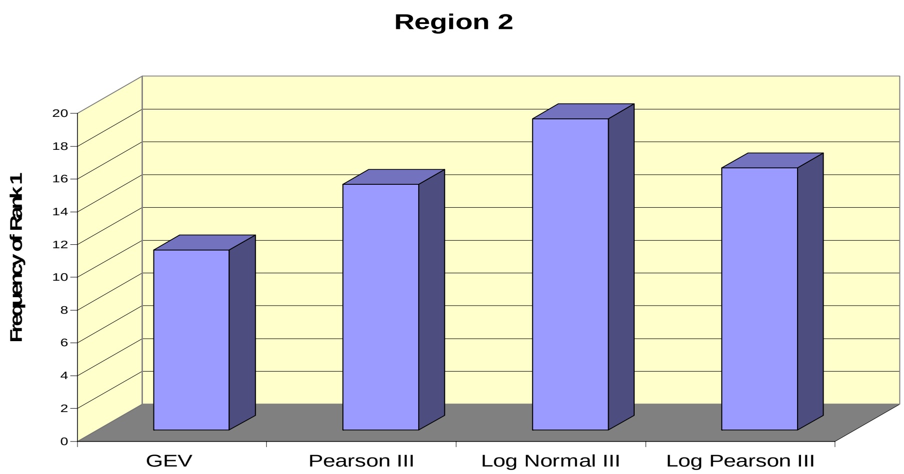

Figure 2.5.8. Region 2- Frequency of Rank 1 for selecting the best Distribution with more than 30 observations at each site.

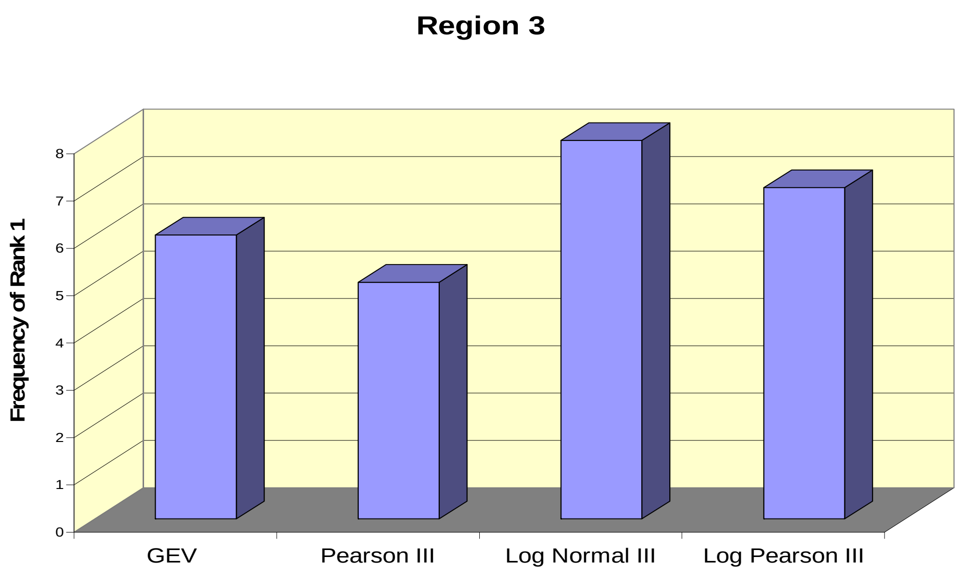

Figure 2.5.9. Region 3- Frequency of Rank 1 for selecting the best Distribution with more than 30 observations at each site.

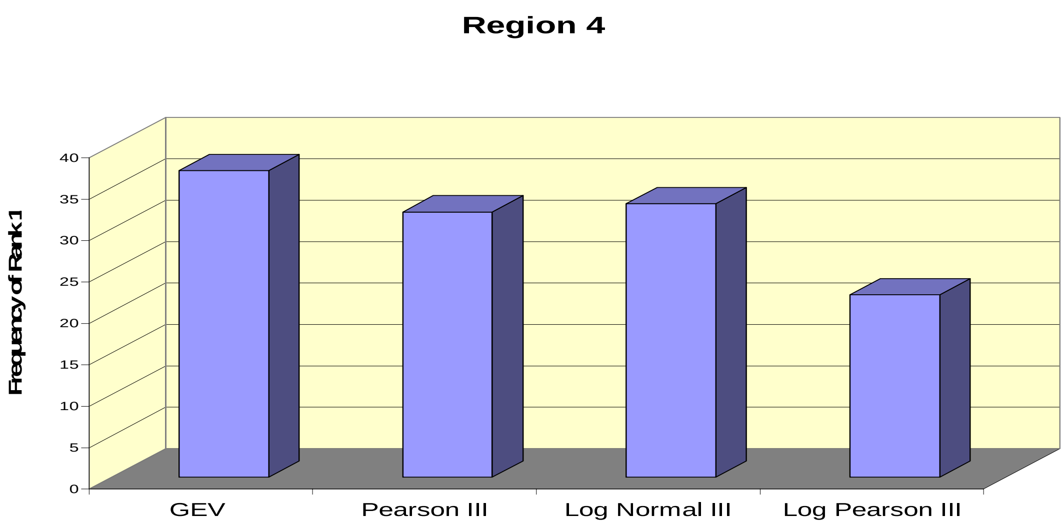

Figure 2.5.10. Region 4- Frequency of Rank 1 for selecting the best Distribution with more than 30 observations at each site.

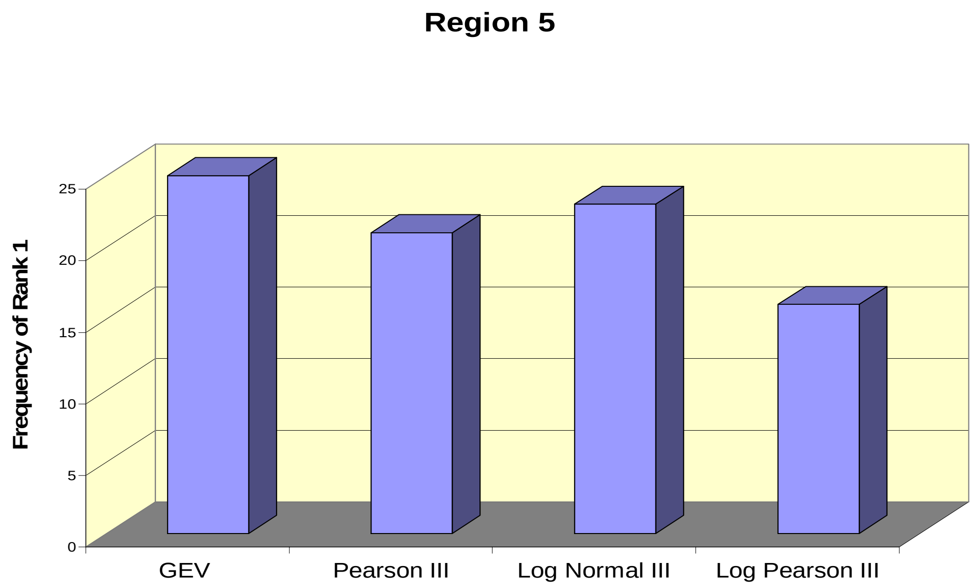

Figure 2.5.11. Region 5- Frequency of Rank 1 for selecting the best Distribution with more than 30 observations at each site.

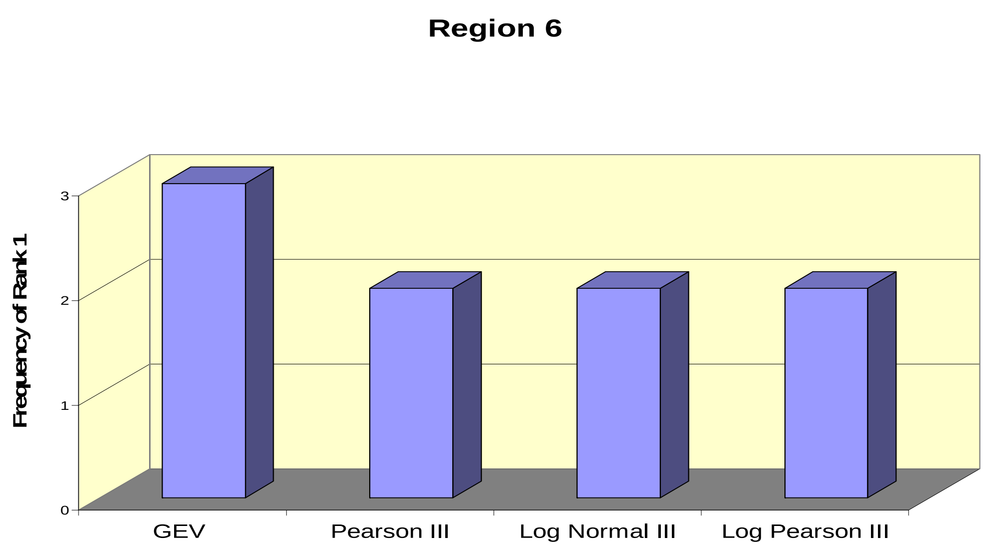

Figure 2.5.12. Region 6- Frequency of Rank 1 for selecting the best Distribution with more than 30 observations at each site.

Table 2.5.3. Selection of Best Distribution in each Region using stations with more than 30 observations.

Region Number	Number of Stations	Rank 1	Rank2	Rank3	Rank4	Best method of Parameter estimation
1	21	GEV	LGN III	LP III	P III	ML
2	30	LGN III	LP III P III		GEV	ML
3	13	LGN III	LP III	GEV	P III	MOM
4	55	GEV	LGN III P III		LP III	ML
5	36	GEV	LGN III P III		LP III	ML
6	07	GEV	LGN III LP III P III			ML

The importance of having longer data sequences in goodness-of-fit tests is clearly brought out by the results in Table 2.5.3. The GEV distribution is the best distribution with larger data sets, followed by Log Normal (III) distribution. Log Pearson (III) distribution which was selected as the best distribution in Table 2.5.2 is no longer in the Rank 1 for any region.

III. Estimation of Peak Discharges by LP(III) Method

3.1. Introduction

For the regression analysis discussed in this chapter, the regions defined by Srinivas and Rao (2004) are used. However, two of the regions were split into two distinct regions. Region 1 and Region 5 were split based on the presence of a significant amount of natural storage in the northern part of Region 1 and the eastern part of Region 5. These regions are identified as Regions 7 and 8, respectively. The Generalized Least Squares method, which is the regression methodology used here, utilizes the distance between stations as a feature of the algorithm. Regions 1 and 5, as previously defined, extended across the state, resulting in long distances between stations. The regression errors were reduced by splitting these two regions, because of the reduction in the distance between stations and incorporation of the percentage of the basin covered by water or wetlands as a regression parameter.

A minor difficulty in regionalization is that the actual region determinations are often based on large scale maps of the state or region examined. In the regions defined by Srinivas and Rao, the regions were delineated based on the gaging stations only, and followed major basin divides only where it was appropriate to do so (Fig. 3.1.1). However, the scale of the map and ignoring drainage divides make the map difficult to apply in practice, since a site for investigation might lie close to a boundary and determination of the proper region may not be accurate. To eliminate any ambiguity in applying the appropriate equations, the regionalization for this chapter was done by fitting the 14-digit Hydrologic Unit Code (HUC) watersheds for Indiana, as described in DeBroka (1999). The 14-digit HUC watersheds are a nomenclature developed and accepted by state and federal water resource agencies for characterizing watersheds.

For the purposes of application, separation of gages into regions as originally determined by Srinivas and Rao has been preserved, but the actual boundaries are modified slightly to follow the 14-digit HUC boundaries whenever possible. This results in a method that is easy to use, since all that is needed to know about a site is the 14-digit HUC basin in which it is located, which is fairly easy to determine. A few 14-digit HUC basins had to be split between regions, but these were kept to a minimum. A CDROM containing a comprehensive listing of the 14-digit HUC basins for Indiana, with an indication of the region(s) for each basin is found in Knipe and Rao (2005). The final map of the regions is shown in Figure 3.1.2.

3.2. Development of Flood Prediction Equations

The annual peak discharges for each of the gages in the study were reviewed for data consistency and possible errors. The original IDNR peak discharge file used in previous studies was compared with peak flow files obtained from the USGS NWIS website (http://waterdata.usgs.gov/nwis). Staff of the USGS and the IDNR researched the discrepancies between the two data sources and corrected the data where necessary. Many of the differences between the two data sources were due to changes in rating curves developed by the USGS after the initial publication of the discharge in the annual Water Resources Data compilations. Corrections have been made to the USGS peak flow files, which are now the definitive source for peak flow information.

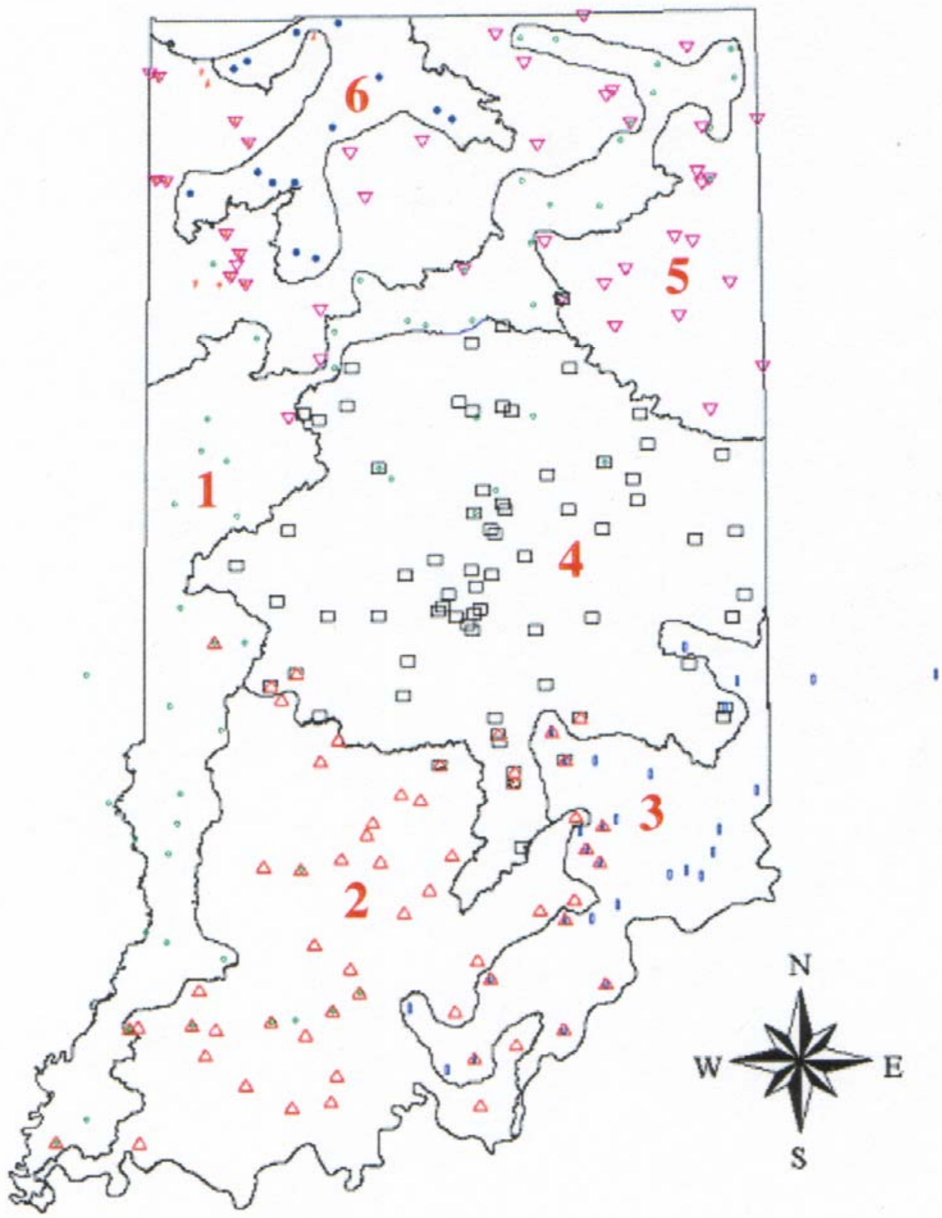

Figure 3.1.1. Regions for Indiana as defined by Srinivas and Rao (2003)

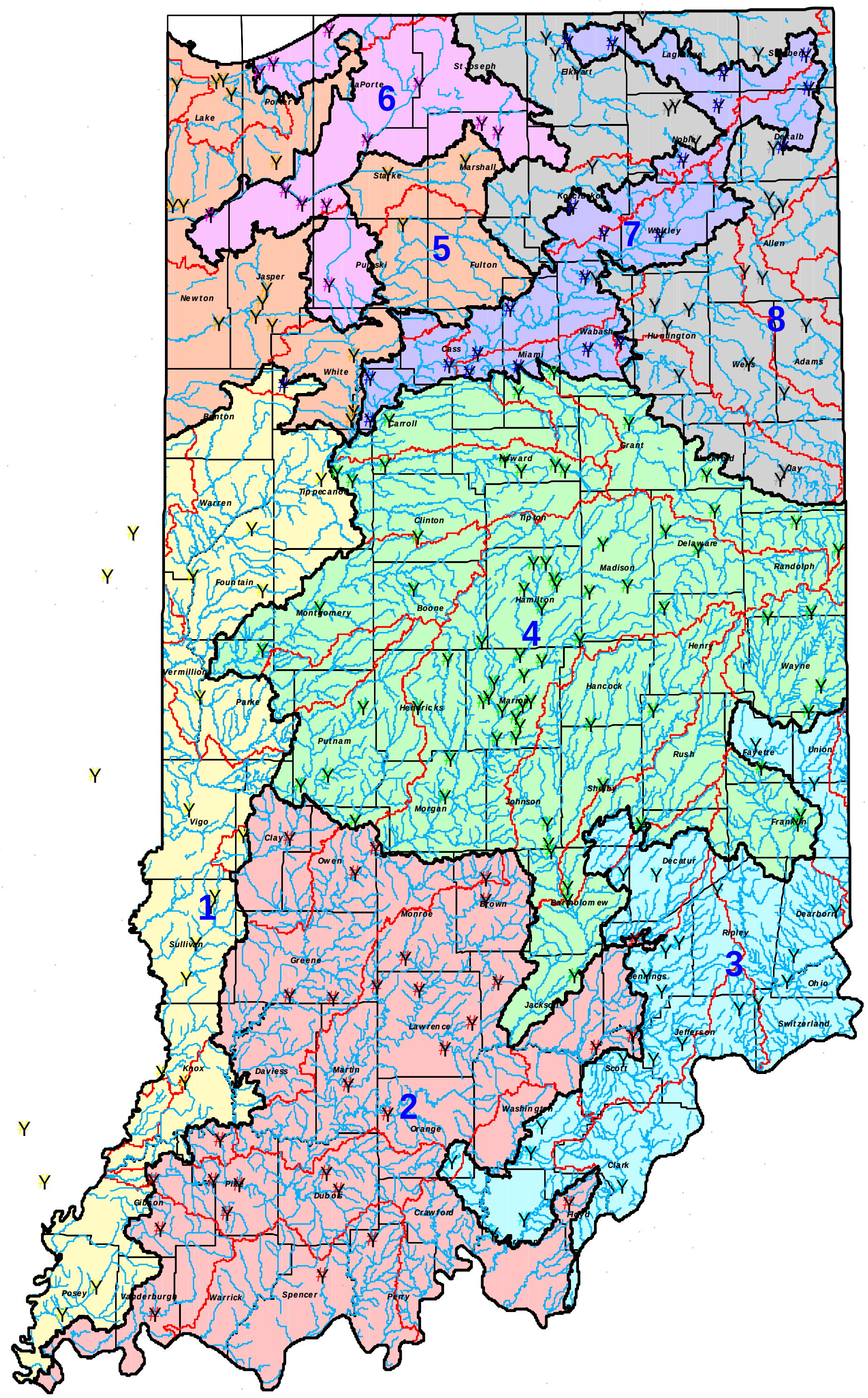

Figure 3.1.2. Regions as defined for the present analysis

Flood frequency curves for each gaging station were calculated by using standard techniques of the U. S. Water Resources Council (USWRC 1982). The USWRC technique is to fit the annual peak flow data from a station using the log-Pearson III distribution. The discharge values are first transformed by computing the logarithm of each value. The mean, standard deviation (S), and skew coefficient (G) for the logarithmic series are computed by using the following equations, where X is the logarithmic of the flow and N is the number of years of record in the annual peak data series:

$$\overline{X} = \frac{\Sigma X}{N} \tag{3.2.1}$$

$$S = \left[\frac{\Sigma(X-\overline{X})^2}{N-1}\right]^{0.5} \tag{3.2.2}$$

$$G = \frac{N\,\Sigma(X-\overline{X})^3}{(N-1)(N-2)S^3} \tag{3.2.3}$$

The skew coefficient is then weighted by using a regional generalized skew coefficient, in order to eliminate local anomalies that may exist for a particular site. The regional skew coefficient used in this study is -0.2. This value is the standard value used by the IDNR and has been agreed to by the other federal agencies (USGS, USACE, NRCS).

$$G_w = \frac{MSE_{\overline{G}}(G) + MSE_G(\overline{G})}{MSE_{\overline{G}} + MSE_G} \tag{3.2.4}$$

The mean square error of the regional skew coefficient is taken from USWRC (1982) to be 0.55. The mean square error of the station skew coefficient is approximated by

$$MGE_G \cong 10^{[A-B[\log_{10}(N/10]]} \tag{3.2.5}$$

$$A = -0.33 + 0.08\,|G| \quad \text{if} \quad |G \leq 0.90|$$

$$-0.52 + 0.30|G| \quad \text{if} \quad |G > 0.90|$$

$$B = 0.94 - 0.26|G| \quad \text{if} \quad |G| \leq 1.50$$

$$0.55 \quad \text{if} \quad |G| > 1.50$$

The flood frequency values for each return period are then computed by using the following equation:

$$\log Q = \overline{X} + KS \tag{3.2.6}$$

where K is based on the log-Pearson III distribution and is a function of the weighted skew coefficient and the return interval. K is normally determined from tables published in USWRC (1982). These calculations are performed by using the USGS computer program PEAKFQ.

The gaging stations used for this study and the respective calculated flood frequency discharges are found in Knipe and Rao (2005). The 10, 25, 50, 100, 200 and 500 year return periods are used for data from each station.

3.3. Basin Characteristics

Determination of basin characteristics for each of the gaged watersheds is a critical step in a hydrologic regression study. The successful application of the final regression equations will depend on the accurate determination of the basin characteristics by the user. Seven basin characteristics are used in this study.

1. The drainage area of a stream.

2. Slope of a stream is computed by the "10-85" method using the following equation:

$$Slope = \frac{E_{85\%} - E_{10\%}}{0.75 * L}$$

E_{10} and E_{85} are the elevations, in feet, of the thalweg of the stream at 10% and 85% of the total length (L, in miles) of the stream upstream from the determination point, respectively.

3. The 2 year, 24 hour rainfall intensity is taken from TP-40 (NWS, 1960).

4. A runoff coefficient was defined using the STATSGO GIS coverage provided by the NRCS (see http://www.ncgc.nrcs.usda.gov/branch/ssb/products/statsgo/index.html).

   The overall soil runoff coefficient is computed by a weighted average of the soil runoff coefficients found in a watershed, based on the aerial extent of each soil complex in a region. Two different soil runoff coefficients were computed, since some soil complexes are defined differently depending on whether the soil is drained or undrained. Accordingly, a drained and undrained soil runoff coefficient is computed.

5-7. The last three variables that are computed are the percentage of the watershed that is covered by water or wetlands (%W), by urbanized areas (%U), and by forested areas (%F). These data are derived from the National Land Cover Dataset (NLCD) compiled by the USGS EROS data center. The data were compiled from satellite imagery and has a spatial resolution of 30 meters. This information is based on ground information from the early 1990s.

The NLCD is a raster grid with each grid cell coded with land use classification. The land use classes were taken from a modified Anderson Land Use classification, a standard nomenclature for describing different land use types. The possible values from the NLCD system are listed in Table 3.3.1.

For use in this study, the grid data were converted from a raster dataset to polygons in a ARC shapefile. These polygons were then clipped by using the watershed area polygons for each gaging station. From these shapefiles, the area of the watershed classified by each code can be determined. A percentage of the watershed covered by each class is then computed by dividing the incremental areas by the total drainage area. %W is then calculated by adding the percentages for codes 11, 12, 91, and 92. %U is the sum of the percentages for codes 21, 22, and 23. %F is the sum of codes 41, 42 and 43.

Calculation of these percentages are the most difficult aspect of the application of the final equations. The values could be estimated from a USGS 7 ½ minute quadrangle map, but practical experience shows that these estimates can vary widely from user to user, and proper application of the method demands that basin characteristics be computed in a similar manner to the methods used to derive the regressed data. Knipe and Rao (2005) include a table of pre-computed values of %W and %U for each 14-digit HUC watershed in Indiana.

Table 3.3.1. NCLD Land Cover Class Definitions.

NLCD Code	Description
11	Open Water
12	Perennial Ice/Snow
21	Low Intensity Residential
22	High Intensity Residential
23	Commercial/Industrial/Transportation
31	Bare Rock/Sand/Clay
32	Quarries/Strip Mines/Gravel Pits
33	Transitional
41	Deciduous Forest
42	Evergreen Forest
43	Mixed Forest
51	Shrubland
61	Orchards/Vineyards/Other
71	Grasslands/Herbaceous
81	Pasture/Hay
82	Row Crops
83	Small Grains
84	Fallow
85	Urban/Recreational Grasses
91	Woody Wetlands
92	Emergent Herbaceous Wetlands

3.4. Generalized Least Squares Regression

Historically, two types of regression analysis have been used for flood frequency analysis. In ordinary least squares the parameters **B** = (b_1, b_2, … b_n) for a model of the response variable Y_n (in this case, the log of the discharge for the given return period), given in equation 3.4.1 are estimated,

$$Y_n = b_0 + b_1 x_1 + b_2 x_2 + ... b_n x_n + \varepsilon \tag{3.4.1}$$

where (x_1, x_2 … x_n) are the various predictor or regressor variables (drainage area, slope, etc.), n is the number of regressor variables in the model and ε represents the error in the model. The regressor variables may be converted to logarithms, and the prediction equation is expressed as a complex power equation. The scheme for ordinary least squares is to estimate the parameters **B** to minimize the sum of the squares of the error term.

While ordinary least squares is a valid model, improvements have been made in the scheme to utilize the unique properties of hydrologic annual maximum flow data. Stedinger and Tasker (1989) have developed and extensively tested a model they have termed generalized least squares (GLS). GLS is an extension of ordinary least squares that incorporates the length of record at each gaging station, differences in the variance at different sites, and any possible cross correlation in the data between stations. The model equation is the same as for ordinary least squares, represented in vector form in equation 3.4.2

$$\hat{\boldsymbol{Y}} = \boldsymbol{X\beta} + \boldsymbol{e} \tag{3.4.2}$$

where **Y** is a (n × 1) vector of flow characteristics at n sites (and $\hat{\mathbf{Y}}$ is an estimate of **Y**), **X** is an (n × p) matrix of (p – 1) basin characteristics augmented by a column of one's, **β** is a (p × 1) vector of regression parameters and ***e*** is an (n × 1) vector of random errors. The GLS estimate of **β** is given by Stedinger and Tasker as

$$\boldsymbol{\beta} = (\mathbf{X}^T \boldsymbol{\Lambda}^{-1}\mathbf{X})^{-1}\mathbf{X}^T \boldsymbol{\Lambda}^{-1}\mathbf{Y} \qquad (3.4.3)$$

where $\boldsymbol{\Lambda}$ is the covariance of the model. In the GLS model $\boldsymbol{\Lambda}$ is estimated by

$$\hat{\boldsymbol{\Lambda}} = \hat{\gamma}^2\mathbf{I} + \hat{\boldsymbol{\Sigma}} \qquad (3.4.4)$$

where $\hat{\gamma}^2$ is an estimate of the model error variance and $\hat{\boldsymbol{\Sigma}}$ is an (n × n) matrix of sampling covariances with elements:

$$\Sigma_{ij} = \hat{\sigma}_i^2\left[1 + K_T^2\left(\frac{\kappa - 1}{4}\right)\right] \Big/ n_i \qquad \text{for } i = j \qquad (3.4.5)$$

$$\Sigma_{ij} = \hat{\rho}_{ij}\frac{m_{ij}\hat{\sigma}_i\hat{\sigma}_i}{n_i n_i}\left[1 + \hat{\rho}_{ij}K_T^2\left(\frac{\kappa - 1}{4}\right)\right] \qquad \text{for } i \neq j \qquad (3.4.6)$$

where:

$\hat{\sigma}_i$ is an estimate of the standard deviation of flows at site i

K_T is the T-year frequency factor for the distribution used

κ is the kurtosis of the distribution used

n_i is the record length at site i

m_{ij} is the concurrent record length of sites i and j

$\hat{\rho}_{ij}$ is an estimate of the lag zero correlation of flows between sites i and j

There are a number of additional steps that can be applied to improve the estimate of these variables, which are detailed in Stedinger and Tasker's various reports. One is the estimate of the lag-zero cross correlation coefficient, ρ_{ij}. To eliminate data problems and increase the robustness of the overall solution, a non-linear regression model is used to smooth out data problems by relating the cross correlation coefficient to the distance between gaging stations. This regression model is of the form:

$$\rho_{ij} = \exp\left\{\left[\frac{d_{ij}}{\alpha d_{ij}+1}\right]\ln\theta\right\} \tag{3.4.7}$$

where d_{ij} is the distance between stations i an j, and α and θ are model parameters.

The GLS regression scheme is implemented in the USGS computer program GLSNET. This program requires input of the annual maximum flood series for each station, including the adjustments for low and high outliers and historic discharges as appropriate. Each station is also required to have latitude and longitude to compute the cross correlation of each station pair in the regression region. The PEAKFQ program needs to be run on the dataset before GLSNET can be run, since the mean, standard deviation and generalized skew from the flood frequency curve computation and estimation of the flood frequency are part of the GLS method. Basin characteristics are also incorporated into the WDM file as user defined variables, for use as the regressor variables.

3.5. Regression Results

The original data set of gaging stations included 439 gages located in Indiana and in the surrounding states of Illinois, Kentucky, Michigan, and Ohio. Through a process of trial and error, this initial set of stations was reduced to 223 based on the homogeneity of certain stations as computed using previous techniques detailed by Srinivas and Rao (2003). The total homogeneity measure of each of the regions with the final station selection is given in Table 3.5.1.

As shown in Table 3.5.1, Regions 1, 3, and 4 are homogeneous, Regions 2, 5, 7, and 8 are possibly homogeneous, and Region 6 is heterogeneous. Region 6 is not a surprise, since all of the previous studies in regionalization had identified that region as heterogeneous. The four regions that are possibly homogeneous are a result of the effort to balance the station selection between homogeneity and the regression diagnostics. The selected stations are a compromise

Table 3.5.1. Homogeneity measures for defined regions.

Region No.	*# of gages*	H_1	H_2	H_3	*Region type*
1	*21*	0.66	-1.83	-2.40	Homogeneous
2	*30*	1.17	-1.18	-2.00	Possible homogeneous
3	*24*	0.26	0.53	0.12	Homogeneous
4	*72*	0.79	-0.97	-1.45	Homogeneous
5	*18*	1.18	-0.30	-0.09	Possible homogeneous
6	*12*	14.68	5.42	2.47	Heterogeneous
7	*22*	1.56	0.04	-0.24	Possible homogeneous
8	*25*	1.07	-0.59	-0.96	Possible homogeneous

between these two goals. It should be noted, however, that 3 of the regions have H_1 values less than 1.2, meaning that they are fairly close to being considered homogeneous by the common standard. These homogeneity measures do not match the results from previous data sets exactly due to the refinement of the peak flow file performed as a part of this study, and the addition of the 2003 water year data.

The final station selection has 223 stations selected for the 8 regions. The location of the gaging sites is shown in Figure 3.1.2.

The return periods chosen for evaluation in this study are the 10, 25, 50, 100, 200 and 500 year frequency flood discharges. The 100-year flood is the basis for most of the regulatory programs in the State of Indiana regarding water resources, while the lower return periods provide information regarding more frequent events that are also helpful in design. The 500-year flood is estimated here even though the length of the period of record for most gages does not support the estimation of the discharge for such a large return period. However, the 500-year flood discharge is a parameter in some of the equations for estimating depth of scour at bridge piers and abutments, and therefore it is useful to have an estimate of this discharge. This estimate should be used with extreme caution.

The regression variables for each of the regions were chosen from evaluating the regression results using trial and error. Runoff coefficient, $I_{2,24}$ and %F did not contribute positively to any of the regional regression models and therefore were not considered in any of the equations. Runoff coefficient, in particular, varied from region to region, but did not vary greatly within a region, meaning that it was of little use in a regression analysis. Given that the regionalization was found to follow geologic and soil type regions throughout the state, this conclusion is not surprising.

All regions have effective drainage area (DA) as factor in the regression, which is expected. Slope is a factor in all regions except Region 8. This may be due to the nature of the stations chosen in those regions, but Glatfelter's study found that slope was not a significant variable in the corresponding region in that area. In this case %W is an indirect measure of the slope of the watershed, since higher water storage in a watershed is an indication of gentler slopes. %W is a factor in Regions 7 and 8 (the lake country) and %U is a factor only in Region 4, which is the only region where urban gages (in the Indianapolis metropolitan area) are present in significant numbers. For purposes of the regression analysis, %W and %U are expressed as percentages, not decimals, and that a value of one is added to each variable. This was to eliminate %W and %U values of zero, which resulted in matrices that could not be inverted.

The average model error is the main regression output used to evaluate the quality of the regression. It is calculated from equation 3.5.1. The percent error is given by Tasker (1995) as in Equation 3.5.1.

$$\%Error = 100\left[\exp(\gamma^2 * 5.3019) - 1\right]^{1/2} \qquad (3.5.1)$$

Average equivalent years of record is a measure developed to express the accuracy of prediction as an equivalent number of years of record required to achieve results of comparable accuracy. It is calculated by equation 3.5.2.

$$EqYOR = \frac{\hat{s}_i^2\left[1 + k_i g_i + \frac{k_i^2}{2}(1 + 0.75 g_i^2)\right]}{\hat{\gamma}_i^2 + \hat{\Sigma}_i} \tag{3.5.2}$$

The form of the prediction equations for Regions 1, 2, and 3 include the effective drainage area and slope as the regressed variables. Table 3.5.2 – 3.5.4 list the values of the regression constant C, and the exponents a_1 and a_2 for use in determining peak discharges using equation 3.5.3 respectively for regions 1, 2 and 3.

$$Q_{RetPer} = (C)(DA)^{a_1}(Slope)^{a_2} \tag{3.5.3}$$

Table 3.5.2. Regression results for Region 1.

Return Period	Constant (C)	DA (a_1)	Slope (a_2)	Avg Model Error	Avg Eq YOR	%Error
10	47.8	0.802	0.535	0.013	4.24	27.1%
25	55.3	0.805	0.561	0.014	5.46	27.8%
50	61.4	0.805	0.573	0.015	6.62	28.3%
100	67.5	0.805	0.585	0.016	6.90	29.5%
200	74.3	0.803	0.592	0.017	7.36	30.6%
500	83.9	0.800	0.599	0.019	7.82	32.2%

Table 3.5.3. Regression results for Region 2

Return Period	Constant (C)	DA (a_1)	Slope (a_2)	Avg Model Error	Avg Eq YOR	%Error
10	69.6	0.798	0.473	0.022	3.12	35.5%
25	102.4	0.777	0.441	0.023	4.23	35.6%
50	133.1	0.762	0.417	0.023	5.01	36.0%
100	169.5	0.748	0.394	0.024	5.70	36.8%
200	213.3	0.734	0.371	0.025	6.24	37.7%
500	283.3	0.716	0.341	0.027	6.80	39.4%

Table 3.5.4. Regression results for Region 3

Return Period	Constant (C)	DA (a_1)	Slope (a_2)	Avg Model Error	Avg Eq YOR	%Error
10	74.6	0.889	0.416	0.008	8.92	20.9%
25	91.5	0.891	0.425	0.007	13.53	19.7%
50	104.5	0.894	0.430	0.007	16.16	19.9%
100	116.8	0.898	0.434	0.008	17.93	20.4%
200	132.5	0.898	0.434	0.009	18.06	22.1%
500	152.1	0.902	0.437	0.011	17.53	24.8%

For Region 4, the urbanization factor %U + 1, is added to the equation for the previous regions.

$$Q_{\mathrm{Re}\,tPer} = (C)(DA)^{a_1}(Slope)^{a_2}(\%U+1)^{a_3} \tag{3.5.4}$$

Table 3.5.5. Regression results for Region 4

Return Period	Constant (C)	DA (a_1)	Slope (a_2)	%U+1 (a_3)	Avg Model Error	Avg Eq YOR	%Error
10	31.1	0.820	0.681	0.080	0.010	7.67	23.1%
25	37.7	0.820	0.698	0.079	0.009	10.64	22.5%
50	42.9	0.819	0.707	0.077	0.009	12.90	22.4%
100	48.4	0.816	0.712	0.075	0.009	15.13	22.4%
200	52.7	0.816	0.722	0.074	0.010	16.59	22.7%
500	58.7	0.815	0.731	0.073	0.010	18.17	23.5%

Equations for Region 5 and 6 are similar to the equations for Regions 1, 2 and 3.

$$Q_{\mathrm{Re}\,tPer} = (C)(DA)^{a_1}(Slope)^{a_2} \tag{3.5.5}$$

Table 3.5.6. Regression results for Region 5

Return Period	Constant (C)	DA (a_1)	Slope (a_2)	Avg Model Error	Avg Eq YOR	%Error
10	35.8	0.776	0.368	0.013	2.96	26.7%
25	45.6	0.764	0.356	0.014	3.70	27.7%
50	53.1	0.756	0.347	0.015	4.24	28.3%
100	60.8	0.748	0.338	0.015	4.75	28.8%
200	68.7	0.742	0.330	0.020	5.23	33.5%
500	79.5	0.734	0.319	0.016	5.79	30.0%

Table 3.5.7. Regression results for Region 6

Return Period	Constant (C)	DA (a_1)	Slope (a_2)	Avg Model Error	Avg Eq YOR	%Error
10	22.4	0.732	0.776	0.025	2.17	37.8%
25	27.9	0.709	0.858	0.026	2.77	38.7%
50	31.5	0.696	0.917	0.027	3.21	39.4%
100	34.6	0.687	0.974	0.028	3.62	40.1%
200	37.3	0.681	1.029	0.029	4.01	40.8%
500	40.3	0.675	1.098	0.030	4.47	41.7%

For Region 7, the factor %W + 1 is added to the equation

$$Q_{\mathrm{Re}\,tPer} = (C)(DA)^{a_1}(Slope)^{a_2}(\%W+1)^{a_3} \qquad (3.5.6)$$

Table 3.5.8. Regression results for Region 7

Return Period	Constant (C)	DA (a_1)	Slope (a_2)	%W+1 (a_3)	Avg Model Error	Avg Eq YOR	%Error
10	65.0	0.873	0.372	-0.795	0.030	2.36	41.7%
25	89.0	0.858	0.361	-0.801	0.034	2.84	44.4%
50	108.4	0.849	0.354	-0.803	0.037	3.19	46.2%
100	129.3	0.839	0.347	-0.803	0.034	3.53	44.3%
200	151.1	0.831	0.343	-0.802	0.041	3.82	49.4%
500	182.2	0.821	0.336	-0.800	0.044	4.18	51.3%

Region 8 is different from the other equations in that the slope is not a factor in the equation. %W + 1 is reflected in the final equation.

$$Q_{\mathrm{Re}\,tPer} = (C)(DA)^{a_1}(\%W+1)^{a_2} \qquad (3.5.7)$$

Table 3.5.9. Regression results for Region 8

Return Period	Constant (C)	DA (a_1)	%W+1 (a_2)	Avg Model Error	Avg Eq YOR	%Error
10	106.0	0.835	-0.733	0.029	1.20	41.0%
25	118.2	0.839	-0.719	0.029	1.66	40.4%
50	126.5	0.842	-0.707	0.028	2.04	39.9%
100	134.2	0.843	-0.695	0.027	2.44	39.5%
200	141.1	0.845	-0.683	0.027	2.84	39.1%
500	149.8	0.846	-0.667	0.026	3.40	38.6%

The ranges of values for each of the watershed parameters in these equations are given in Table 3.5.9. Applying these equations in circumstances where the values of the watershed parameters are outside of the ranges of the data used in the regression study is not recommended, and should be done with caution. The effect of outlier values of the basin characteristics cannot be determined with any certainty, since the data are non existent, and the response of a particular watershed could vary greatly outside the bounds of the variable ranges.

Table 3.5.10. Ranges for various watershed characteristics

Region	DA (sq mi)	Slope (ft/mi)	%W (%)	%U (%)
1	0.27-13,706	1.4-79		
2	0.15-11,125	1.2-267		
3	0.07-284	3.8-253		
4	0.31-2,444	2.7-48.7		0-83.9
5	5.82-1,869	1.6-8.6		
6	1.5-1,779	0.9-15.8		
7	0.17-4,072	2.4-43.7	0-7.2	
8	0.45-3,370		0-12.1	

Examining the error results, regions 3 and 4 have the smallest percentage errors and the largest equivalent years of record. This corresponds to the heterogeneity measures, which identified these regions as homogeneous. Region 1, the other homogeneous region, has error values slightly higher than Regions 3 and 4, but still better than four of the other five regions. Errors for the other four regions compare to the errors found in Glatfelter's study. Region 5 has results that are comparable to the three homogenous regions.

Equations for computing confidence limits for each of the predictive equations have also been derived as part of the GLS methodology. A 100(1-α) prediction interval is given in Equations 3.5.8 and 3.5.9 for a logarithmic transformation of the prediction variable q_0

$$10^{\hat{y}_0 - T} \le q_0 \le 10^{\hat{y}_0 + T} \quad (3.5.8)$$

where

$$T = t_{\frac{\alpha}{2}, n-p'} \sqrt{\hat{\gamma}_0^2 + x_0 (X' \hat{\Lambda}^{-1} X)^{-1} x_0'} \quad (3.5.9)$$

where $t_{\alpha/2,n-p'}$ is the critical value for a t distribution for n-p' degrees of freedom (Tasker, 1995).

3.6 Evaluation of the Prediction Equations

With any study, testing the results with independent methodologies is an important aspect of determining the reliability of the study. The nature of the input data for any hydrologic study is imprecise, and therefore various means of evaluating the study results are warranted. For this study, the results have been tested by using a split sample test, with a comparison to previously determined discharges, and by examining the fit of the regression to the input data points.

As a general examination of the regression results, Figures 3.6.1 through 3.6.8 are plots of the peak 100-year flood frequency discharges for gaging stations in each region (calculated using the USWRC methodology) plotted versus the 100-year frequency flood discharge predicted by the respective regional equation. Given a perfect relationship, these discharges would be equal to each other, and therefore would plot on a straight line at a 45 degree angle. By examining the deviation of the plotted points to this line, the relative strength of the predictive equations can be evaluated.

For these plots, the best fit equations are for Regions 3 and 4, which have the smallest errors from the GLS analysis, and have the lowest homogeneity measures. Other regions do not demonstrate as strong a relationship, but generally show an acceptable relationship between calculated and predicted values for the 100-year discharge.

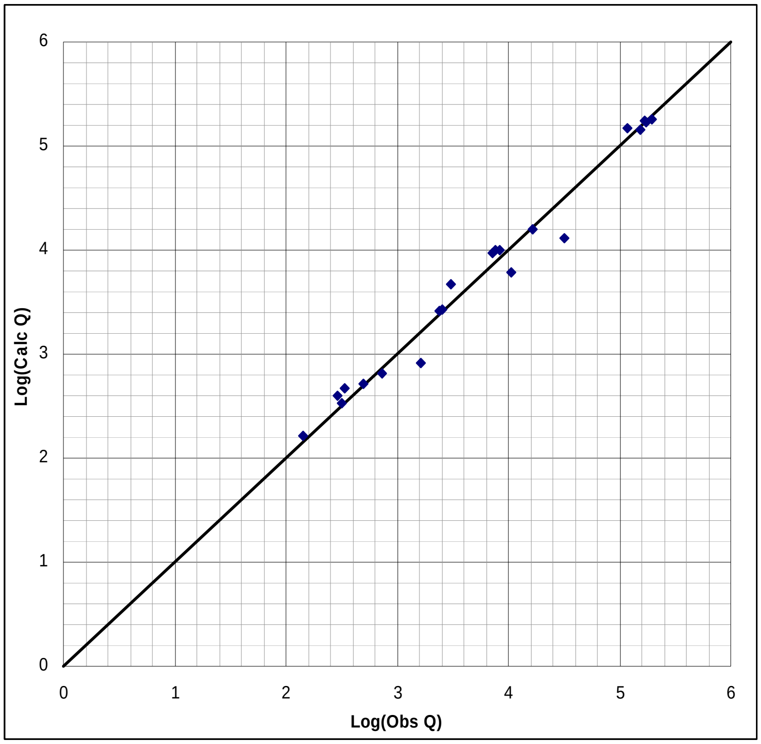

Figure 3.6.1. Comparison of 100 year observed discharges and regression model discharges for Region 1.

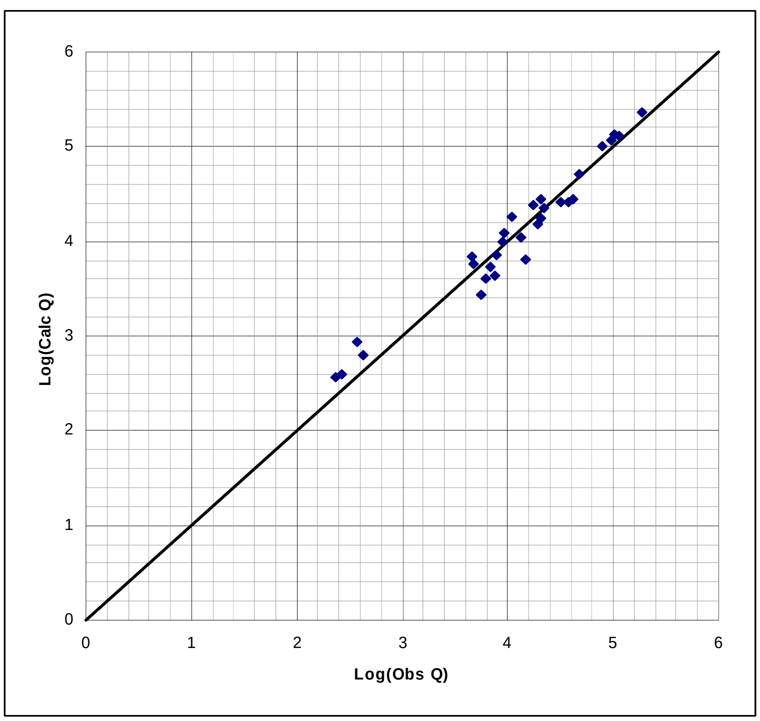

Figure 3.6.2. Comparison of 100 year observed discharges and regression model discharges for Region 2

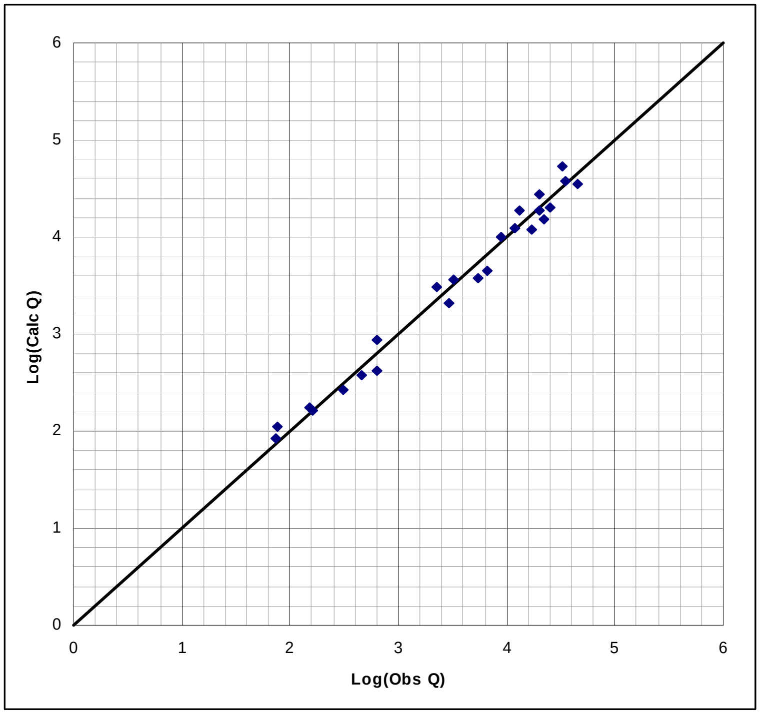

Figure 3.6.3. Comparison of 100 year observed discharges and regression model discharges for Region 3

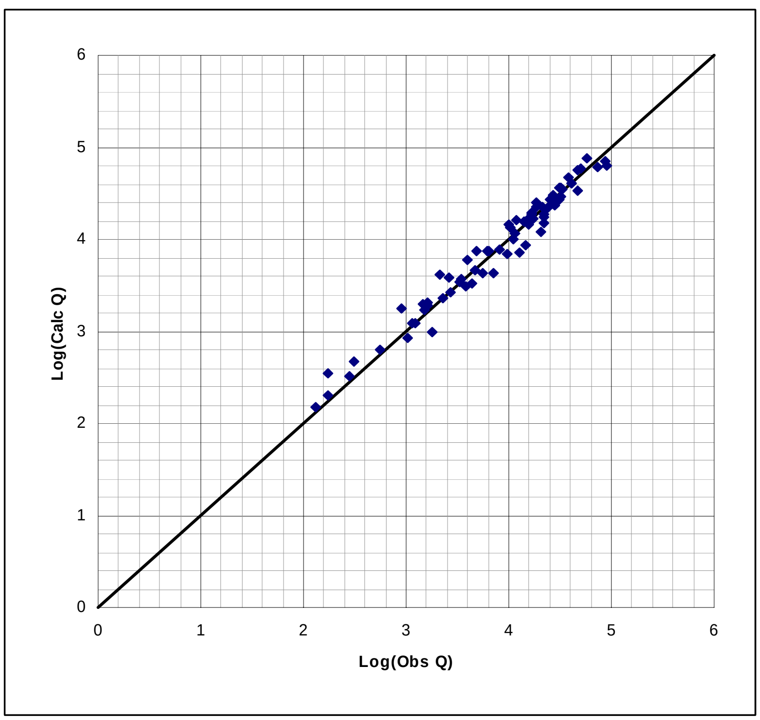

Figure 3.6.4. Comparison of 100 year observed discharges and regression model discharges for Region 4

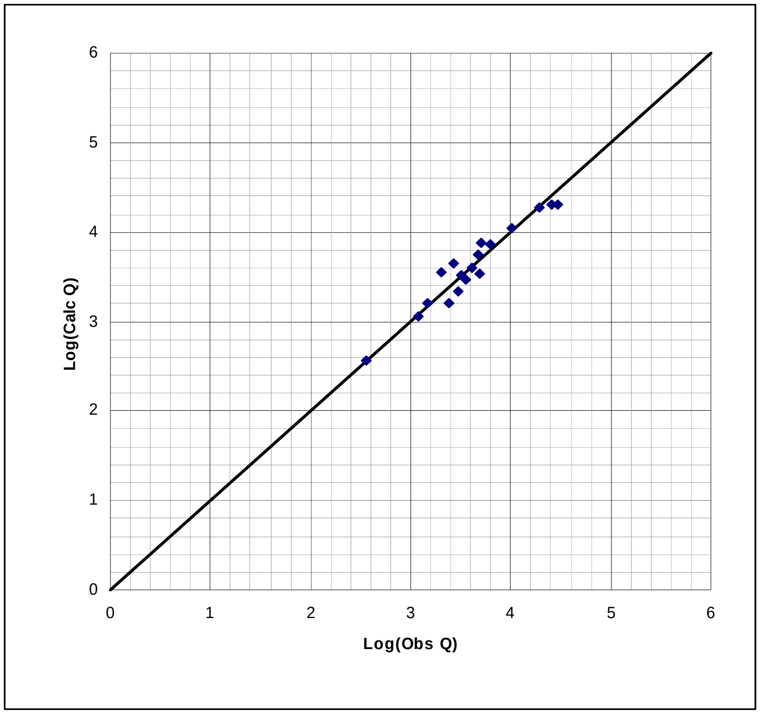

Figure 3.6.5. Comparison of 100 year observed discharges and regression model discharges for Region 5

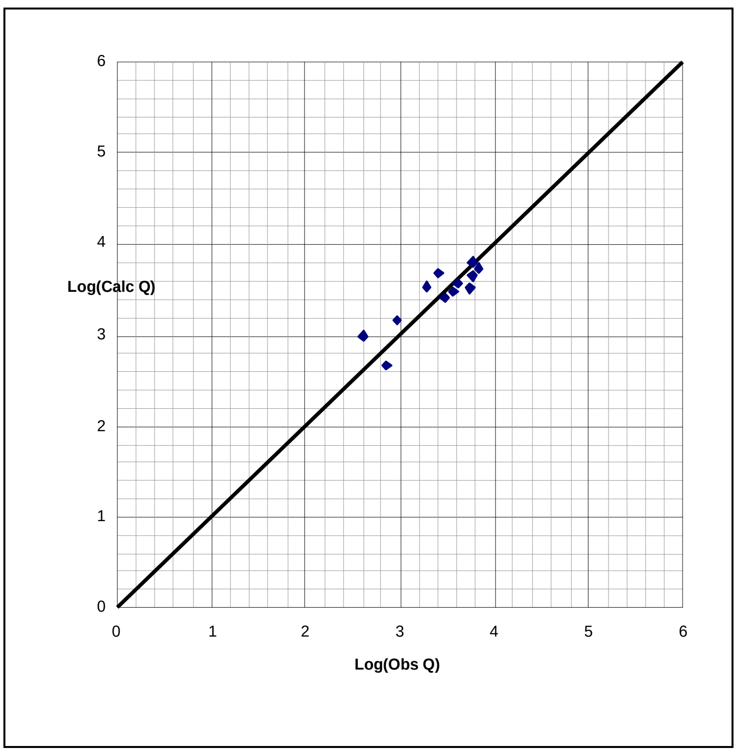

Figure 3.6.6. Comparison of 100 year observed discharges and regression model discharges for Region 6

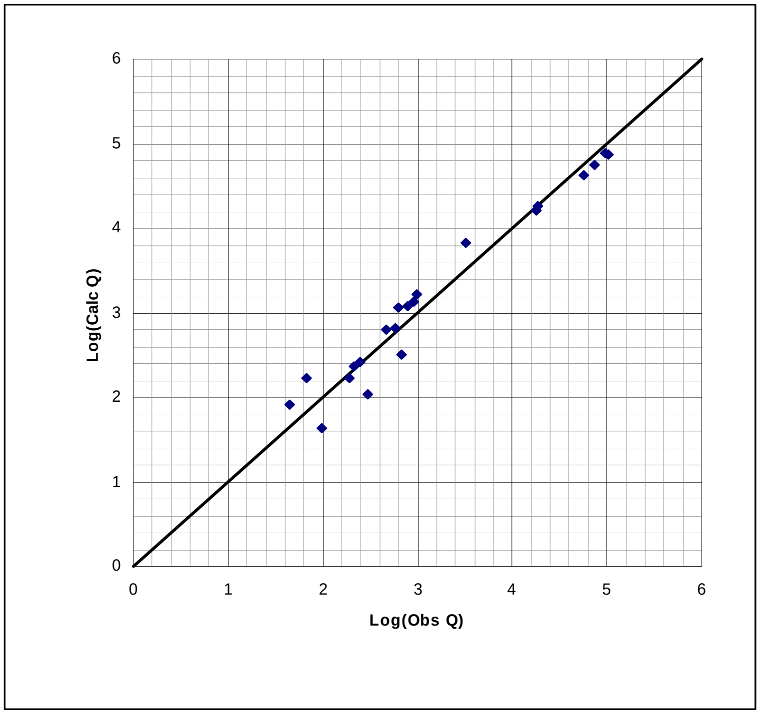

Figure 3.6.7. Comparison of 100 year observed discharges and regression model discharges for Region 7

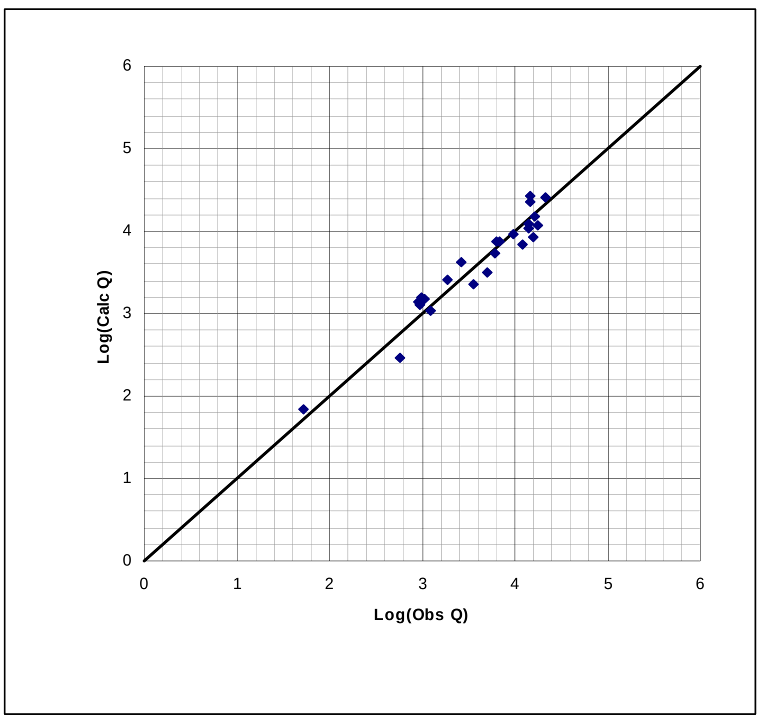

Figure 3.6.8. Comparison of 100 year observed discharges and regression model discharges for Region 8

3.7. Split Sample Test

A split sample test is useful in identifying how stable and reliable a dataset may be. In a truly homogeneous data sample, a regression model on a significant part of the data set should be comparable to a regression model of the entire data set. For the split sample test in this study, the following methodology was used:

- A random number (between 0 and 1) was assigned to each gaging station, using the Microsoft EXCEL rand() function.
- The stations in each region were then sorted using the random number as the sorting key.
- 20% of the stations in each of the regions were then chosen as the "split" sample, based on the lowest random number generated.
- The GLS regression method was then run using the remaining 80% of the sample set. The regression variables were kept the same as for the original regression analysis. Only the 100-year flood was used for this test.
- The split sample regression equation was then used to predict the flood flows at the stations removed from the test.
- The percent error of the predicted peak discharge was computed based on the peak discharge computed using flood frequency analysis. This percent error was then compared to the overall percent error in the model (as computed in the GLS methodology), and compared to the percent error at the removed stations in the full regression model.

Table 3.7.1 shows the stations removed for the split sample test with the percent errors noted, and Table 3.7.2 is a summary of the results of the test by region.

Table 3.7.1. Stations removed from regression for Spilt Sample test

StatNo	2004Regions	Q100(calc)	Q100 (ss)	%diff	%diff (entire sample)
03335500	1	119,359	164,786	38.1%	24.5%
03336000	1	155,856	155,410	0.3%	6.9%
03360100	1	142	140	1.3%	16.3%
03378550	1	10,666	5,326	50.1%	42.2%
03302300	2	7,489	4,143	44.7%	42.2%
03322100	2	11,092	18,224	64.3%	62.6%
03360000	2	48,371	51,158	5.8%	7.0%
03366500	2	37,426	25,925	30.7%	29.3%
03373700	2	17,716	11,009	37.9%	34.8%
03374000	2	185,277	236,515	27.7%	25.6%
03276640	3	462	292	36.8%	19.4%
03291780	3	8,825	10,208	15.7%	14.0%
03302690	3	75	113	50.1%	49.6%
03302730	3	11,916	12,709	6.7%	4.9%
03369000	3	19,954	19,176	3.9%	5.8%
03274880	4	555	633	14.1%	13.4%
03275500	4	21,766	16,757	23.0%	21.0%
03325500	4	11,548	11,510	0.3%	2.0%
03326000	4	20,639	11,888	42.4%	41.6%
03333600	4	1,596	2,031	27.3%	25.9%
03334500	4	16,635	18,503	11.2%	14.2%
03348020	4	1,633	1,952	19.5%	18.5%
03348350	4	6,401	7,376	15.2%	16.3%
03348700	4	130	153	17.6%	14.0%
03349500	4	4,859	7,400	52.3%	53.5%
03358000	4	13,904	15,301	10.0%	12.3%
03361500	4	18,305	22,121	20.9%	23.8%
03364000	4	73,957	58,382	21.1%	18.3%
03365500	4	89,484	60,543	32.3%	30.1%
03332500	5	19,452	17,480	10.1%	3.1%
03333000	5	25,553	18,919	26.0%	19.7%
04093500	5	4,147	3,909	5.7%	3.2%
05523000	5	1,201	1,156	3.8%	4.7%
03332400	6	2,963	2,725	8.0%	13.2%
05515500	6	1,925	3,687	91.5%	74.5%
03324500	7	17,952	18,948	5.6%	10.4%
03327930	7	666	282	57.7%	52.9%
03328430	7	633	1,451	129.4%	80.9%
03329400	7	794	1,667	110.0%	50.2%
03324300	8	14,066	12,770	9.2%	12.3%
04099750	8	2,648	3,974	50.1%	60.0%
04100220	8	905	1,513	67.2%	54.1%
04180000	8	6,025	5,568	7.6%	12.2%
04181500	8	14,822	23,385	57.8%	54.9%

Table 3.7.2. Split Sample error percentages

Region	(1)	(2)	(3)	(4)
1	22.4%	22.4%	21.8%	29.5%
2	35.2%	33.6%	33.8%	36.8%
3	22.6%	18.7%	25.0%	20.4%
4	22.0%	21.8%	23.0%	22.4%
5	11.4%	7.7%	23.5%	28.8%
6	49.8%	43.8%	43.7%	40.1%
7	75.7%	48.6%	45.0%	44.3%
8	38.4%	38.7%	34.7%	39.5%
Total	30.9%	27.2%	29.1%	---

In Table 3.7.2, the columns are as follows:

(1) is the average percent error of the calculated discharge for the split sample using the censored regression equation, compared to the calculated peak discharge using flood frequency analysis

(2) is the average percent error of the calculated peak discharge for the split sample using the full regression equation, compared to the calculated peak discharge using flood frequency analysis.

(3) is the average percent error of the calculated peak discharge for the entire sample using the full regression equation, compared to the calculated peak discharge using flood frequency analysis.

(4) is the average model error as calculated from the GLS regression diagnostics, using equation 3.5.1.

For most regions the percent error as calculated by these various methods are comparable to each other. This is to be expected, since the regions are mainly homogeneous or possibly homogeneous, and therefore errors inherent within the analysis should be consistent for subsets of the data. The exceptions to this are Region 5, where the split sample errors are much less then the errors for the entire data set, and Region 7, where they are much greater.

The anomalies for these two regions could be due to a number of reasons. The difference for Region 5 is most likely due to a fortunate selection of stations that fit the data unusually well. Note, for example, that station 03333000, Tippecanoe River near Delphi, is in the split sample, while station 03333050, also named Tippecanoe River near Delphi, is not. The second station is actually a replacement of the first located slightly downstream of the original station, and therefore has similar basin characteristics and a similar flood frequency curve. The reduction in the error for the split sample could be a reason for reevaluating the stations for Region 5 and attempting to further reduce the error for the entire sample. However, since there are only 18 stations in Region 5, eliminating further stations would reduce the diversity of basin characteristics at each of the stations in the region, reducing the predictive qualities of the resulting equation. A balance must be struck between having too many stations in a region; resulting in a heterogeneous region, and too few stations; resulting in equations that are not useful for predicting flood frequency flows for basins that have basin characteristics outside of the range of characteristics in the study.

While the split sample for Region 5 had a lower average error than the entire study, Region 7 had a much higher average error for the split sample than for the entire sample. This may be due to the random nature of the stations chosen for the split sample. Three of the four stations removed from the analysis have drainage areas less than 10 square miles, while 10 of the remaining 16 gages have drainage areas greater than 10 square miles (and mostly much greater than 10 square miles). Also, two of the split sample gages (Weesau Creek near Deadsville and Rattlesnake Creek near Patton) are stations with small drainage areas, but fairly long periods of record (31 and 25 years, respectively). This influences the split sample regression to a degree that it is not predicting the peak discharges for the smaller discharges as well as the general model. One of the main advantages of GLS regression over other types of analysis is that the record length is a factor in determining the influence of a station on the model. The nature of the gaging program is such that gaging stations for smaller streams typically do not have as long record lengths as do the stations on larger streams. Therefore, stations such as those two randomly removed from this analysis have a great bearing on defining the lower end of the model, causing the split sample equation to err unacceptably in predicting the peak flows for these stations.

Whether LP (III) distribution gives results with smaller errors than other distributions is not addressed in this chapter. It is considered in the next chapter.

IV. Regional Flood Estimation Based on L-Moments

Two sets of data are used in this and the following chapters. The first set is that used in Chapter 2. The second set is the data used in chapter 3. The reason for using the first set is that the division of Indiana to eight regions was rather arbitrary. The effect of this division on flood prediction equations is investigated in this and following chapters.

The objective of the research reported in this chapter is to investigate the L-moment method to obtain the regional normalized flood quantiles. Basic descriptions of L-moments, parameter estimation and probability distribution are introduced first. The regional flood estimation method is discussed later.

4.1. L-moments and Parameter Estimation

4.1.1. L-Moment

The *r*th L-moments of a random variable x have been defined as in equation 4.1.1 (Hosking, 1990):

$$\lambda_r = \int_0^1 x(F) P^*_{r-1}(F) dF \tag{4.1.1}$$

in which

$$P^*_r = \sum_{k=0}^{r} (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} F^k \quad , \tag{4.1.2}$$

where $F(x)$ is the cumulative distribution of x, and $x(F)$ is the quantile function for the distribution. The first L-moment, λ_1, is the arithmetic mean, while the second L-moment, λ_2, is a measure of dispersion analogous to the standard deviation. Hosking found it convenient to standardize the higher L-moments such that the *r*th L-moment ratio is given by:

$$\tau_r = \lambda_r / \lambda_2 \qquad (4.1.3)$$

where τ_3 is a measure of the symmetry of the sample and is referred to as L-skewness. τ_4 is referred to as L-kurtosis, and $\tau = \lambda_2 / \lambda_1$ is analogous to the conventional coefficient of variation of central moments.

For an ordered random sample, $x_1 \le x_2 \le \cdots \le x_n$, the *r*th sample L-moment, l_r can be estimated by using equation 4.1.4:

$$l_r = \frac{r!(n-r)!}{n!} \sum_{1 \le i_1} \sum_{< i_2 \le} \cdots \sum_{< i_r \le n} \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} X_{i_{r-k}} \,, \; r = 1, \cdots, n \;. \qquad (4.1.4)$$

Hosking (1990) points out that it is not necessary to iterate over all subsamples of size *r*, as l_r can be written as a linear combination of order statistics.

$$l_r = \sum_{k=0}^{r-1} p^*_{r-1,k} b_k \qquad (4.1.5)$$

where

$$b_k = \frac{1}{n^{k+1}} \sum_{i=1}^{n} (i-1)^k x_i \;\text{ and }\; p^*_{rk} = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} \qquad (4.1.6)$$

4.1.2. Moments and Parameter Estimation

For a distribution with a probability density function *f(x)*, the r^{th} theoretical moments μ'_r and r^{th} sample moments m'_r about the origin are given by eq. 4.1.7 and eq. 4.1.8, respectively.

$$\mu'_r = \int_{-\infty}^{\infty} x^r f(x)dx \tag{4.1.7}$$

$$m'_r = \frac{1}{n}\sum_{i=1}^{n} x_i^{\ r} \tag{4.1.8}$$

Parameters of a probability distribution function are estimated by the method of moments (MOM) by equating the moments of samples to the moments of the probability distribution function. The method of moments is a commonly used parameter estimation method. For a distribution with k parameters, $\alpha_1, \alpha_2, \cdots, \alpha_k$ which are to be estimated, the first *k* sample moments are set equal to the corresponding population moments that are given in terms of unknown parameters. These *k* equations are solved simultaneously for the unknown parameters.

For a given probability distribution, its probability weighted moments $M_{p,r,s}$ are defined as

$$M_{p,r,s} = E[X^p F^r (1-F)^s] = \int_0^1 [x(F)]^p F^r (1-F)^s \, dF \tag{4.1.9}$$

where $F = F(x, \phi_1, \phi_2, \cdots, \phi_k) = P(X \leq x)$ is the cumulative distribution function, *x(F)* is the inverse cumulative function, and *p, r* and *s* are integers. Two particular sets of PWMs α_r and β_r are usually considered:

$$\alpha_s \equiv M_{1,0,s} = \int_0^1 x(F)(1-F)^s \, dF = \int x(1-F)^s f(x)dx \tag{4.1.10}$$

$$\beta_s \equiv M_{1,r,0} = \int_0^1 x(F)F^r dF = \int xF^r f(x)dx \tag{4.1.11}$$

where $f(x) = f(x, \phi_1, \phi_2, \cdots, \phi_k)$ is the probability density function. It can be shown that the set of α_s and β_s are linearly dependent, implying that either definition of PWM may be used for parameter estimation without loss of generality (Hosking, 1986).

To estimate the parameters of a distribution, PWM estimators of the ordered sample $x = \{x_1 \le x_2 \le \cdots \le x_{n-1} \le x_n\}$ are defined as follows, using Haktanir's notation (1997):

$$\alpha'_s = \frac{1}{n}\sum_{i=1}^{n}(1 - P_{nex,i})^s x_i \quad \text{for } s = 0,1, 2,\ldots, n\text{-}1 \qquad (4.1.12)$$

$$\beta'_s = \frac{1}{n}\sum_{i=1}^{n} P_{nex,i}\ x_i \quad \text{for } r = 0,1, 2,\ldots, n\text{-}1 \qquad (4.1.13)$$

where $P_{nex,i}$ is an estimate for the non-exceedance probability of the ith event. Using the unbiased estimators suggested by Landwehr et al. (1979), $P^j_{nex,i}$ is taken as

$$P^j_{nex,i} = \frac{\binom{i-1}{j}}{\binom{n-1}{j}} \quad \text{for } j = 0,1, 2,\ldots, n\text{-}1 \qquad (4.1.14)$$

The PWM parameter estimates $\phi_1', \phi_2', \cdots, \phi_k'$ are defined as those values that make the first k theoretical PWMs equal to the first k sample PWM estimators; i.e. $\phi_1', \phi_2', \cdots, \phi_k'$ are those values such that

$$\alpha'_s = \alpha_s(\phi'_1, \phi'_2, \cdots, \phi'_k) \quad \text{for } s = 0,1, 2,\ldots, k\text{-}1 \qquad (4.1.15)$$

$$\beta'_r = \beta_r(\phi'_1, \phi'_2, \cdots, \phi'_k) \quad \text{for } r = 0,1, 2,\ldots, k\text{-}1 \qquad (4.1.16)$$

The mathematical solutions or numerical approximations for parameters of different probability distributions by the method of moments and probability weighted moments have been derived by Rao and Hamed (2000). Mainly, the lognormal III, Pearson III and log Pearson III, and the GEV distributions are used in the study.

To select the candidate probability distributions, the statistical characteristics of the data can be investigated by using the L-moment ratio diagram. An important application of summary statistics calculated from an observed random sample is identification of the distribution from which the sample is drawn. This is achieved, particularly for skewed distributions, by using L-moments. The statistics of L-skewness and L-kurtosis of all sites in each region is shown in Fig. 4.1.1, along with the theoretical lines for some distributions.

The mean square error from the L-moment ratio diagram is the measure used to evaluate the candidate probability distributions. It is calculated by using equation 4.1.17.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_{LCK}^{Obs} - x_{LCK}^{Th})^2}{n}} \tag{4.1.17}$$

where x_{LCK}^{Obs} refers the fourth L-moment ratio of the observed values and x_{LCK}^{Th} is the theoretical fourth L-moment ratio of a specific distribution corresponding to the same third L-moment of the observations. From the results in Figure 4.1.2, it can be concluded that normal distribution and extreme value type I distribution are inappropriate for these data; the other six distributions, which do not show significant differences in the graph, will be evaluated as candidate distributions in the regional index flood analysis. The candidate distributions are lognormal (LNIII), two-parameter Gamma (GMII), Pearson type 3 (PTIII), log-Pearson type 3 (LPIII), generalized extreme value (GEV), and generalized logistic (GLO) distributions. The method used to select the better distribution for regional analysis is to evaluate the distribution which offers better and stable regional estimates than the others.

A split data sample in which the sample is split into 75% for parameter estimation and 25% for parameter validation is used. We use the 75% of data to build the L-moment ratio diagram. Again, the RMSE of L-moment ratio is calculated for each distribution, and the result is shown in Figure 3.1.3. A similar conclusion is indicated as before, in that the sampling statistics of normal distribution (NOR) and extreme value type I (EVI) are not acceptable, so that both of them are not appropriate for further analysis.

Kedia and Rao (2005) performed the goodness-of-fit of the commonly used probability distributions, LP3, GEV, LN3, and PT3, to the regions defined by Sirinivas and Rao (2003). This was discussed in Chapter 2. Their results indicated that GEV and LN3 are better than others in fitting the probability distributions. At a certain significance level, many distributions satisfy the hypothesis and it is not easy to pick up a unique "best" distribution. The results are similar to those in L-moment ratio diagrams. Figure 4.1.1 shows that the data are scattered around theoretical curves, none of them falling on a single curve. The root-mean-square-error (RMSE) values from the L-moment ratio diagram shown in Figures 4.1.2 and Figure 4.1.3 lead to a similar conclusion.

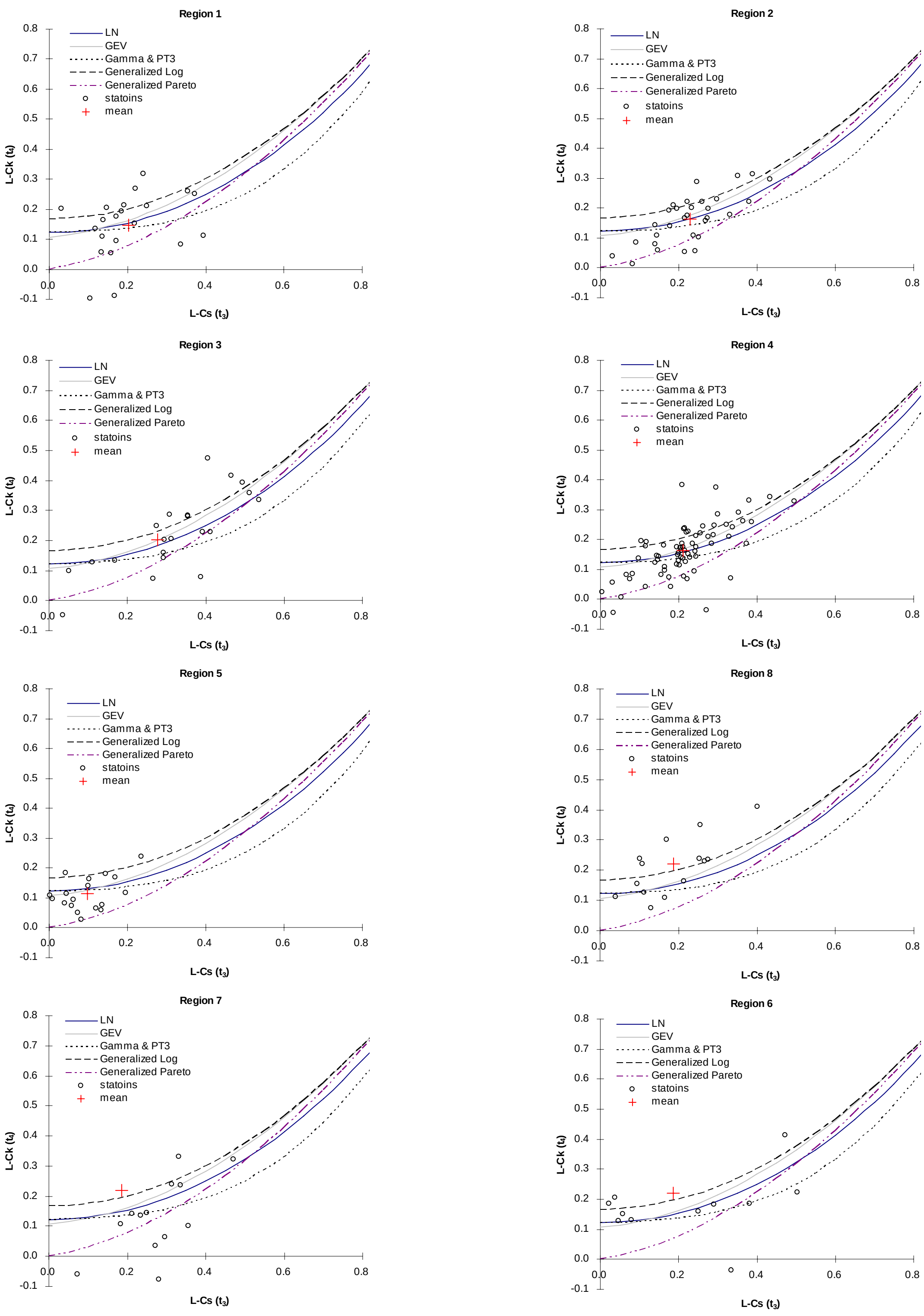

Figure 4.1.1. LCs-LCK moment ratio diagram for the study regions.

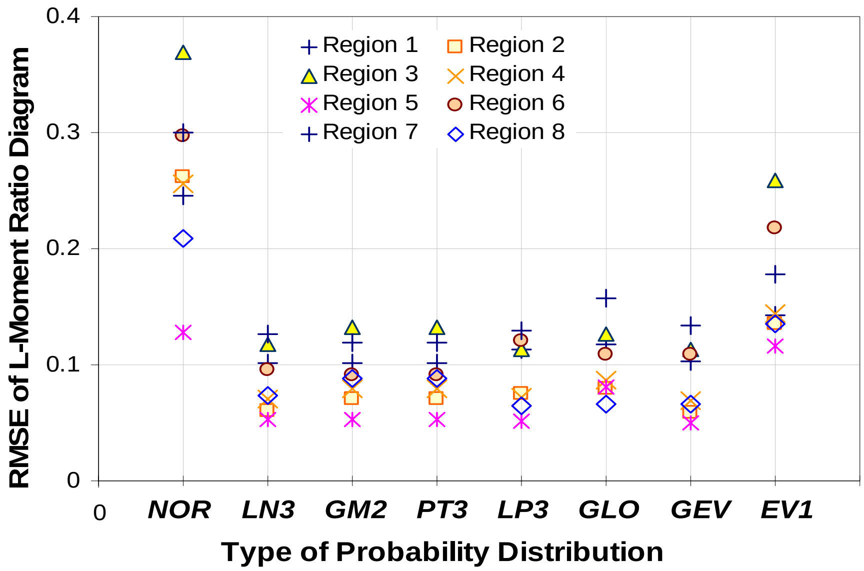

Figure 4.1.2. RMSE of L-Moment ratio diagram comparison for different distributions.

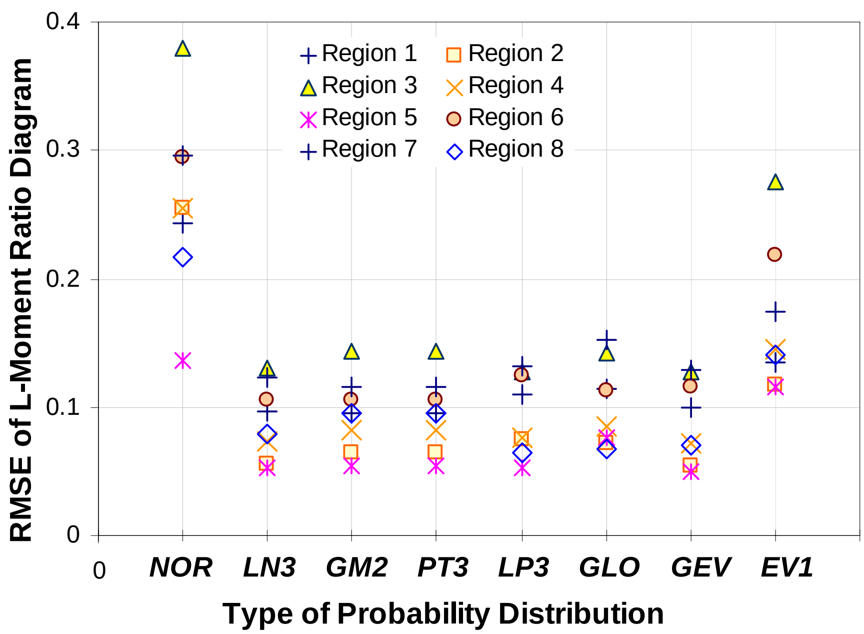

Figure 4.1.3. RMSE of L-Moment ratio diagram comparison of the 75% of data for different distributions.

4.2. Regional Index Flood Method Based on L-Moments

4.2.1. Introduction

The basic idea behind the index flood method, which has been in use for a long time (Dalrymple, (1960)), is that the distributions of floods at different sites in a homogeneous region are similar except for an index-flood parameter. This index flood parameter reflects the important physiographic and meteorologic characteristics of a watershed. The L-moment based index flood method was proposed by Landwehr, Matalas and Wallis and popularized by Wallis and others (Hosking et al. (1985), Wallis (1980), Wallis and Wood (1985)). An important factor in the success of the index flood method is that data from hydrologically similar basins are used (Lettenmaier et al., (1987)).

Regional index flood methods based on probability weighted moments and L-moments have been studied, generally with Generalized Extreme Value (GEV) or Wakeby distributions (Hosking and Wallis (1988), Jin and Stedinger (1989), Landwehr et al. (1987), Potter and Lattenmaeir (1990), Wallis and Wood (1985)). These results, especially with GEV distribution have been demonstrated to be robust. They have been claimed to be more accurate than other procedures based on two or more parameters and short records. This assertion will be tested in this chapter.

4.2.2. Regional L-moment Method

Assume that there are K sites in a region with annual maximum flow records [$x_t(k)$, t=1,2,…,n_k] and k=1,2,…,K. The first three L-moment estimators $\hat{\lambda}_1(k)$, $\hat{\lambda}_2(k)$, and $\hat{\lambda}_3(k)$ are computed by using the unbiased probability weighted moment (PWM) estimators. The regional average of the normalized L-moments of orders 2 and 3 are computed by using equation 4.2.1 and the 1st order normalized L-moment is 1.0.

$$\lambda_r^R = \frac{\sum_{k=1}^{K} w_k \left[\hat{\lambda}_r(k) / \hat{\lambda}_1(k)\right]}{\sum_{k=1}^{K} w_k}, \qquad r = 2, 3 \tag{4.2.1}$$

In equation 4.2.1, w_k are the weights. A simple choice for w_k is N_k where N_k is the number of observations at site *k*. The weighting parameter w_k may depend on the heterogeneity of a region (Tasker and Stedinger (1986, 1989)) and some modification might be required. The normalized parameters for different probability distributions can be calculated by probability weighted moment method based on the first three normalized L-moments. For various recurrence intervals, the quantiles $\hat{x}_p^R$ of the normalized regional distribution are estimated. The mean of annual maximum flood series is generally used as the index flood. Hence, the estimator of the 100p percentile of the flood distribution $\hat{x}_p(k)$ at any site *k* is given by equation 4.2.2.

$$\hat{x}_p(k) = \hat{\lambda}_1^k \, \hat{x}_p^R \tag{4.2.2}$$

where $\hat{\lambda}_1^k$ is the mean for site *k*.

Since $\hat{\lambda}_1^k$ is the regressor, the confidence limit for the regional L-moment quantile estimate can be calculated by equation 4.2.3.

$$CL = \hat{x}_i(k) \pm t_{\alpha/2, N-2} \sqrt{MSE\left(\frac{1}{N} + \frac{(\lambda_i - \overline{\lambda})^2}{S_{\lambda\lambda}}\right)} \tag{4.2.3}$$

Where N is the total number of observations of the annual peak flow, $\overline{\lambda}$ is the average of λ_i^k values, $S_{\lambda\lambda}$ is the sum of square of λ_i^k, MSE is the mean square of the residuals, λ_i is any possible value of λ and $\hat{x}_i(k)$ is the predicted value of x at λ_i. $t_{\alpha/2,N-2}$ is the value of the student's t-distribution for a $100(1-\alpha)$ percent of confidence interval with N-2 degrees of freedom.

An advantage of using the L-moment method is that the $\hat{x}_p^R$ is estimated directly by using the best distribution for a region. The results about the best distribution for a region are used to estimate $\hat{x}_p^R$. The importance of using data from a homogeneous region for this analysis is stressed by Lattenmaier et al. (1987). One of the important variables which must be estimated for ungaged locations is $\hat{\lambda}_1^k$. The usual practice is to estimate this variable by relating it to other variables which are easily available.

4.2.3. At-site and regional parameter estimation

The mathematical derivations or numerical approximations for six candidate distributions based on method of moments and probability weighted moments (Rao, A.R. and K.H. Hamed, 2000) are used for parameter estimation. First, for the annual maximum streamflow data at each site, the conventional moments (mean, standard deviation, skewness and kurtosis) and L-moments (l_1, l_2, l_3, l_4, $t= l_2/ l_1$, $t_3= l_3/ l_2$, $t_4= l_4/ l_2$) are obtained. These are key statistics for parameter estimation. Once these parameters are calculated, the quantile estimates are calculated for recurrence intervals: 2, 5, 10, 20, 25, 50, 100, and 200 years.

Following the study approach described previously, the regional average normalized L-moments can be calculated from the estimates at all sites. A region yields only one set of normalized L-moments, and hence a unique set of parameters is produced. The normalized quantiles for seven distributions with eight different recurrence intervals are summarized in

Table 4.2.1. Hence, the flood estimate of each site is calculated by multiplying the normalized regional quantile with the at-site first L-moment (l_1).

An estimate of the precision of regional flood quantiles is of interest. It can be evaluated by the variance v^2, which is the difference between the actual normalized quantiles x_p^s for different sites in a region and the average regional estimator x_p^R. It is a measure of the heterogeneity of a region. The variance of $\hat{x}_p$ is given by equation (4.2.4).

$$v^2 = \text{var}(\hat{x}_p) = E\left[\hat{\lambda}_1 \hat{x}_p^R - \lambda_1 \tilde{x}_p^s\right]^2 \tag{4.2.4}$$

Higher variance v^2 refers to high variability within the region and smaller variance indicates strong homogeneity within a region.

The results for Regions 1~8 defined by Knipe and Rao (2004) are given in Table 4.2.1. When GLS regression is used, the regressions in Region 7 and Region 8 do not yield good results. Hence, the region refined by Srinivas and Rao (2003) was considered. Region 1 and Region 7 (by Knipe and Rao, 2004) are merged as one region, which is the Region 1 defined by Srinivas and Rao (2003). Region 5 and Region 8 are merged as one region, which is the Region 5 defined by Srinivas and Rao. The normalized regional quantile estimates of these two regions are listed in Table 4.2.3.

An example of the at-site quantile estimates against the regional quantiles estimates is shown in Fig. 4.2.1. The goodness-of-fit can be observed and the results of GEV, PTIII and LNIII are closely approaching 45 degree line. Log-Pearson type three (LPIII) and two-parameter Gamma distribution are the worst two distributions from this analysis.

Table 4.2.1. Normalized regional quantile estimates.

	T (year)	LN3	G2	PT3	LP3	GEV	GLO	LOG
Region 1	2	0.9098	0.8478	0.9088	1.0036	0.6319	0.9209	1.0000
	5	1.3367	1.7612	1.3469	1.0482	1.0522	1.3046	1.3566
	10	1.6257	2.3350	1.6353	1.0696	1.3418	1.5809	1.5652
	20	1.9066	2.8622	1.9069	1.0863	1.6286	1.8761	1.7574
	25	1.9966	3.0252	1.9920	1.0910	1.7214	1.9774	1.8175
	50	2.2767	3.5163	2.2513	1.1041	2.0131	2.3162	2.0011
	100	2.5599	3.9897	2.5049	1.1153	2.3114	2.6984	2.1821
	200	2.8481	4.4501	2.7545	1.1252	2.6177	3.1320	2.3617
Region 2	2	0.8761	0.8298	0.8739	1.0023	0.5716	0.8916	1.0000
	5	1.3556	1.7468	1.3737	1.0531	1.0362	1.3200	1.3979
	10	1.7013	2.3349	1.7194	1.0785	1.3788	1.6438	1.6306
	20	2.0512	2.8814	2.0538	1.0988	1.7363	2.0024	1.8451
	25	2.1658	3.0514	2.1600	1.1046	1.8561	2.1280	1.9121
	50	2.5305	3.5663	2.4871	1.1209	2.2452	2.5580	2.1169
	100	2.9100	4.0658	2.8116	1.1353	2.6638	3.0587	2.3188
	200	3.3066	4.5544	3.1350	1.1481	3.1156	3.6453	2.5192
Region 3	2	0.8456	0.8257	0.8409	0.9989	0.4719	0.8656	1.0000
	5	1.3352	1.7433	1.3635	1.0614	0.9369	1.2999	1.4074
	10	1.7163	2.3347	1.7477	1.0947	1.3075	1.6479	1.6458
	20	2.1217	2.8857	2.1306	1.1226	1.7187	2.0497	1.8654
	25	2.2585	3.0573	2.2541	1.1307	1.8619	2.1942	1.9340
	50	2.7046	3.5777	2.6393	1.1543	2.3458	2.7019	2.1438
	100	3.1859	4.0835	3.0276	1.1756	2.8976	3.3163	2.3505
	200	3.7055	4.5787	3.4195	1.1954	3.5287	4.0638	2.5557
Region 4	2	0.8824	0.8250	0.8808	1.0040	0.6083	0.8964	1.0000
	5	1.3748	1.7426	1.3903	1.0593	1.0885	1.3376	1.4092
	10	1.7204	2.3346	1.7355	1.0859	1.4328	1.6648	1.6486
	20	2.0642	2.8865	2.0657	1.1069	1.7843	2.0222	1.8691
	25	2.1758	3.0584	2.1700	1.1128	1.9004	2.1464	1.9381
	50	2.5276	3.5798	2.4898	1.1293	2.2724	2.5678	2.1488
	100	2.8893	4.0867	2.8052	1.1435	2.6643	3.0525	2.3564
	200	3.2630	4.5832	3.1179	1.1561	3.0785	3.6133	2.5624

Table 4.2.1. Normalized regional quantile estimates (Cont.)

	T (year)	LN3	G2	PT3	LP3	GEV	GLO	LOG
Region 5	2	0.9578	0.8657	0.9576	1.0061	0.7422	0.9623	1.0000
	5	1.3213	1.7747	1.3235	1.0510	1.1097	1.2938	1.3144
	10	1.5387	2.3336	1.5406	1.0713	1.3294	1.5106	1.4983
	20	1.7340	2.8411	1.7337	1.0865	1.5242	1.7268	1.6678
	25	1.7938	2.9971	1.7925	1.0907	1.5829	1.7979	1.7208
	50	1.9725	3.4644	1.9668	1.1019	1.7550	2.0257	1.8827
	100	2.1433	3.9118	2.1317	1.1113	1.9135	2.2670	2.0422
	200	2.3084	4.3443	2.2894	1.1193	2.0599	2.5242	2.2006
Region 6	2	0.9275	0.8831	0.9267	1.0017	0.5879	0.9538	1.0000
	5	1.2558	1.7867	1.2645	1.0456	0.9097	1.2459	1.2737
	10	1.4809	2.3309	1.4891	1.0677	1.1348	1.4458	1.4338
	20	1.7014	2.8196	1.7018	1.0854	1.3600	1.6519	1.5813
	25	1.7723	2.9688	1.7686	1.0904	1.4334	1.7210	1.6274
	50	1.9940	3.4136	1.9727	1.1048	1.6657	1.9472	1.7684
	100	2.2195	3.8366	2.1728	1.1174	1.9057	2.1940	1.9072
	200	2.4502	4.2430	2.3703	1.1288	2.1545	2.4650	2.0450
Region 7	2	0.8562	0.8262	0.8527	0.9966	0.5123	0.8713	1.0000
	5	1.3460	1.7436	1.3704	1.0711	0.9811	1.3060	1.4065
	10	1.7163	2.3347	1.7422	1.1121	1.3440	1.6496	1.6444
	20	2.1027	2.8853	2.1087	1.1470	1.7376	2.0425	1.8635
	25	2.2316	3.0568	2.2262	1.1573	1.8726	2.1830	1.9320
	50	2.6482	3.5767	2.5911	1.1874	2.3224	2.6733	2.1413
	100	3.0915	4.0818	2.9567	1.2150	2.8243	3.2614	2.3476
	200	3.5640	4.5765	3.3240	1.2407	3.3859	3.9705	2.5523
Region 8	2	0.9396	0.8722	0.9392	1.0031	0.6683	0.9477	1.0000
	5	1.2937	1.7793	1.2989	1.0449	1.0212	1.2678	1.2992
	10	1.5207	2.3328	1.5253	1.0651	1.2500	1.4880	1.4742
	20	1.7339	2.8332	1.7336	1.0808	1.4662	1.7159	1.6355
	25	1.8008	2.9866	1.7980	1.0853	1.5340	1.7925	1.6859
	50	2.0055	3.4455	1.9923	1.0977	1.7410	2.0438	1.8399
	100	2.2073	3.8837	2.1798	1.1084	1.9435	2.3189	1.9917
	200	2.4082	4.3064	2.3622	1.1178	2.1422	2.6221	2.1424

An example of the variance of estimation errors in eight regions is shown in Fig. 4.2.2. Pearson Type III (PTIII) has smaller variance than others, especially for longer recurrence intervals. Generalized extreme value (GEV) and three-parameter log-normal distribution (LNIII) yield results which are close to the variance from the PTIII distribution. Overall, GEV, PTIII, and LNIII have good estimates for all regions. Sometimes LNIII cannot yield convergent parameter estimates. The other issue is that although LPIII is not a good candidate for regional index flood estimation, it may have to be used in engineering design. As a result, we will use PTIII, GEV and LPIII for the following analysis.

The 95% confidence intervals for the regional L-moment flood quantile estimates are shown in Fig. 4.2.3. This is calculated from equation 4.2.3 based on regression of the mean annual peak discharge, which is the first L-moment. It is plotted in a log-log axis, hence the smaller mean flows look better. A conclusion from Figures 4.2.1 and Figure 4.2.2 is that the LPIII has regional L-moment estimates which are inferior to PTIII and GEV. The confidence intervals for LPIII is much wider than PTIII and GEV from the results in Fig. 4.2.3.

In Table 4.2.3, candidate probability distributions are listed based on the mean-square-error of L-moment ratio diagram. The order in Table 4.2.3 begins with the one having the minimum MSE. Optimal distributions for regional L-moment estimates are obtained from the variances of regional estimates. It turns out that PTIII, GEV and LNIII are good probability distributions for regional L-moment flood estimates.

Table 4.2.2. Normalized regional quantile estimates for Region 1 and Region 5 defined by Srinivas and Rao (2003).

	T (year)	LN3	G2	PT3	LP3	GEV	GLO	LOG
Region 1 + Region 7 = Region 1	2	0.8854	0.8387	0.8836	1.0012	0.5751	0.9529	1.0000
	5	1.3401	1.7541	1.3563	1.0595	1.0168	1.3921	1.3775
	10	1.6651	2.3351	1.6812	1.0893	1.3395	1.7252	1.5983
	20	1.9922	2.8721	1.9943	1.1136	1.6738	2.0948	1.8017
	25	2.0991	3.0386	2.0935	1.1206	1.7853	2.2246	1.8653
	50	2.4379	3.5416	2.3988	1.1404	2.1459	2.6690	2.0597
	100	2.7892	4.0282	2.7012	1.1581	2.5312	3.1876	2.2512
	200	3.1549	4.5027	3.0019	1.1741	2.9442	3.7963	2.4413
Region 5 + Region 8 = Region 5	2	0.9473	0.8695	0.9470	1.0041	0.7009	0.9619	1.0000
	5	1.3054	1.7775	1.3092	1.0468	1.0601	1.2912	1.3054
	10	1.5283	2.3331	1.5316	1.0670	1.2853	1.5140	1.4841
	20	1.7337	2.8365	1.7333	1.0825	1.4922	1.7418	1.6487
	25	1.7974	2.9909	1.7953	1.0870	1.5560	1.8178	1.7002
	50	1.9905	3.4533	1.9809	1.0990	1.7475	2.0653	1.8574
	100	2.1783	3.8952	2.1584	1.1092	1.9300	2.3333	2.0124
	200	2.3629	4.3219	2.3301	1.1182	2.1046	2.6254	2.1662

Table 4.2.3. Determine the optimal probability distributions for regional L-moment flood estimates of the entire series of data.

Region No.	Candidate Probability Distributions	Optimal Distributions for Regional Estimates
1	PT3, GM2, LN3, GEV, LP3	PT3, LN3, GEV
2	GEV, LN3, PT3, GM2, GLO	GEV, PT3, LN3
3	LP3, GEV, LN3, GLO, PT3	PT3, LN3, GEV
4	GEV, LN3, LP3, PT3, GM2	PT3, LN3, GEV
5	GEV, LP3, LN3, PT3, GM2	GEV, PT3, LN3
6	LN3, PT3, GM2, GLO, GEV	PT3, GEV, LN3
7	PT3, GM2, LN3, LP3, GEV	PT3, LN3, GEV
8	LP3, GLO, GEV, LN3, PT3	PT3, LN3, GLO

Note: 1). Candidate probability distributions are determined from the mean-square-error of L-moment ratio diagram, and the order is beginning with the one having the minimum MSE. 2). Optimal distributions for regional L-moment flood estimates are obtained from the variances of regional estimates.

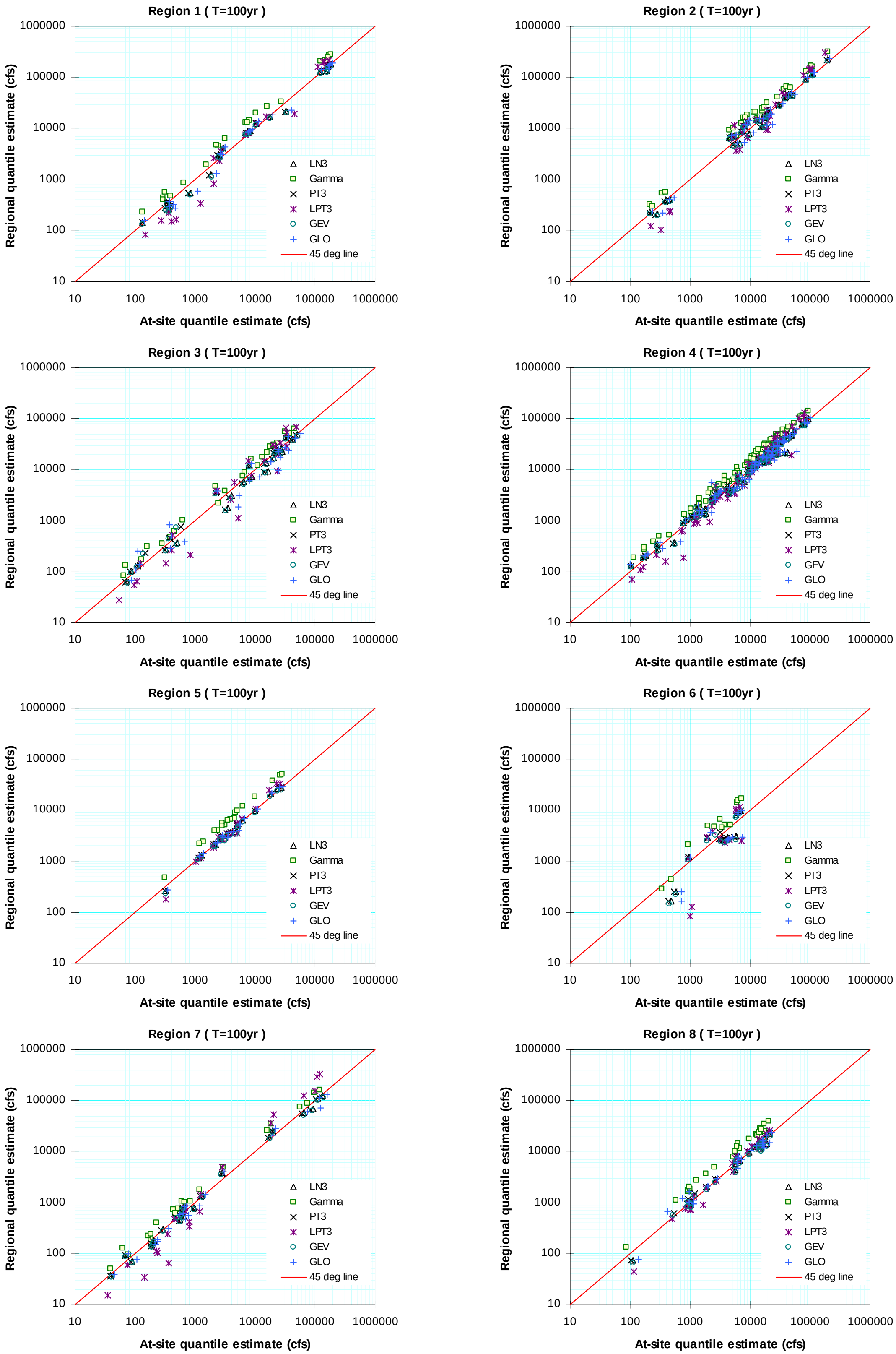

Figure 4.2.1. At-site and regional quantile flood estimates (T = 100 year).

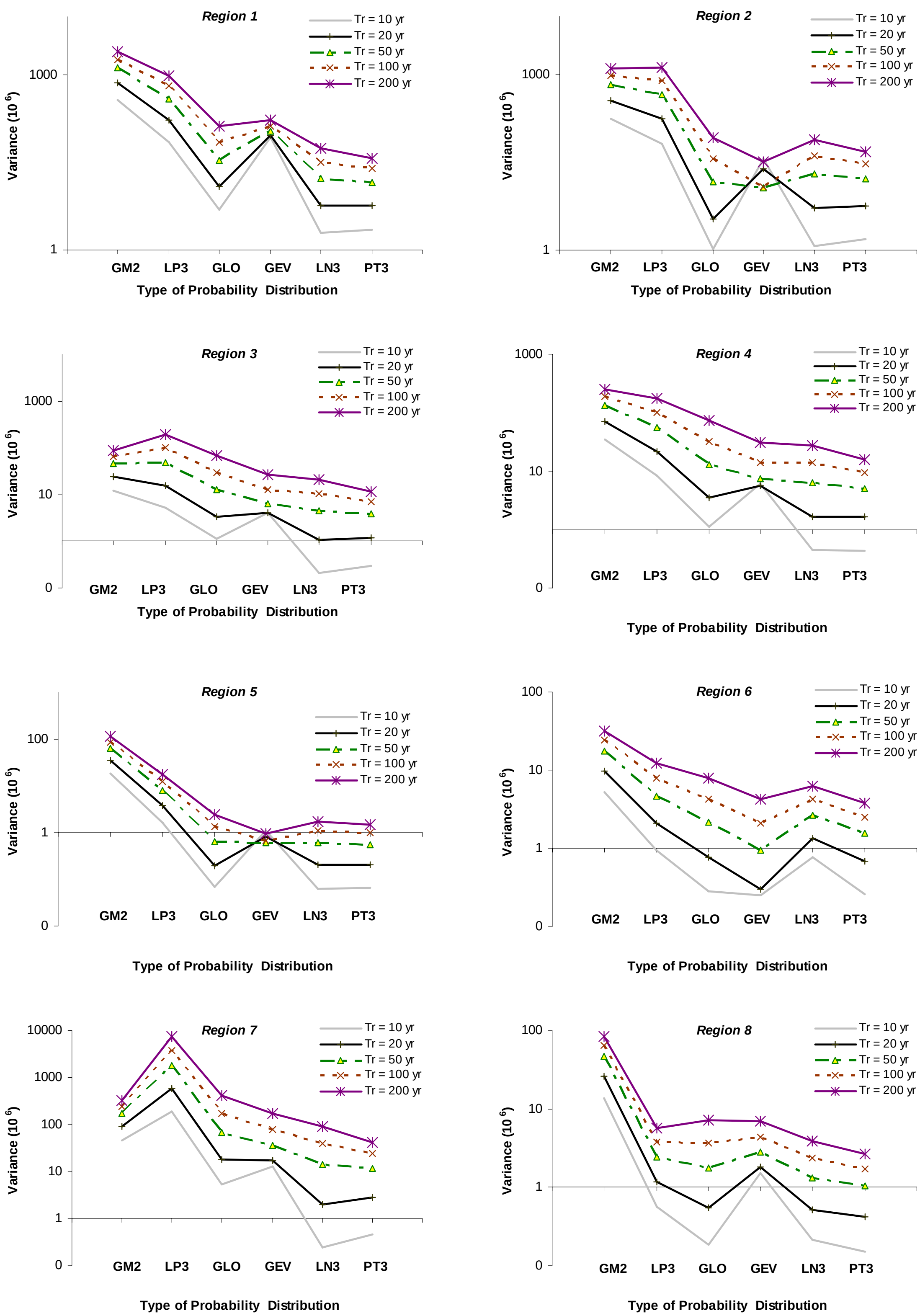

Figure 4.2.2. Variance of the difference between at-site and regional estimates.

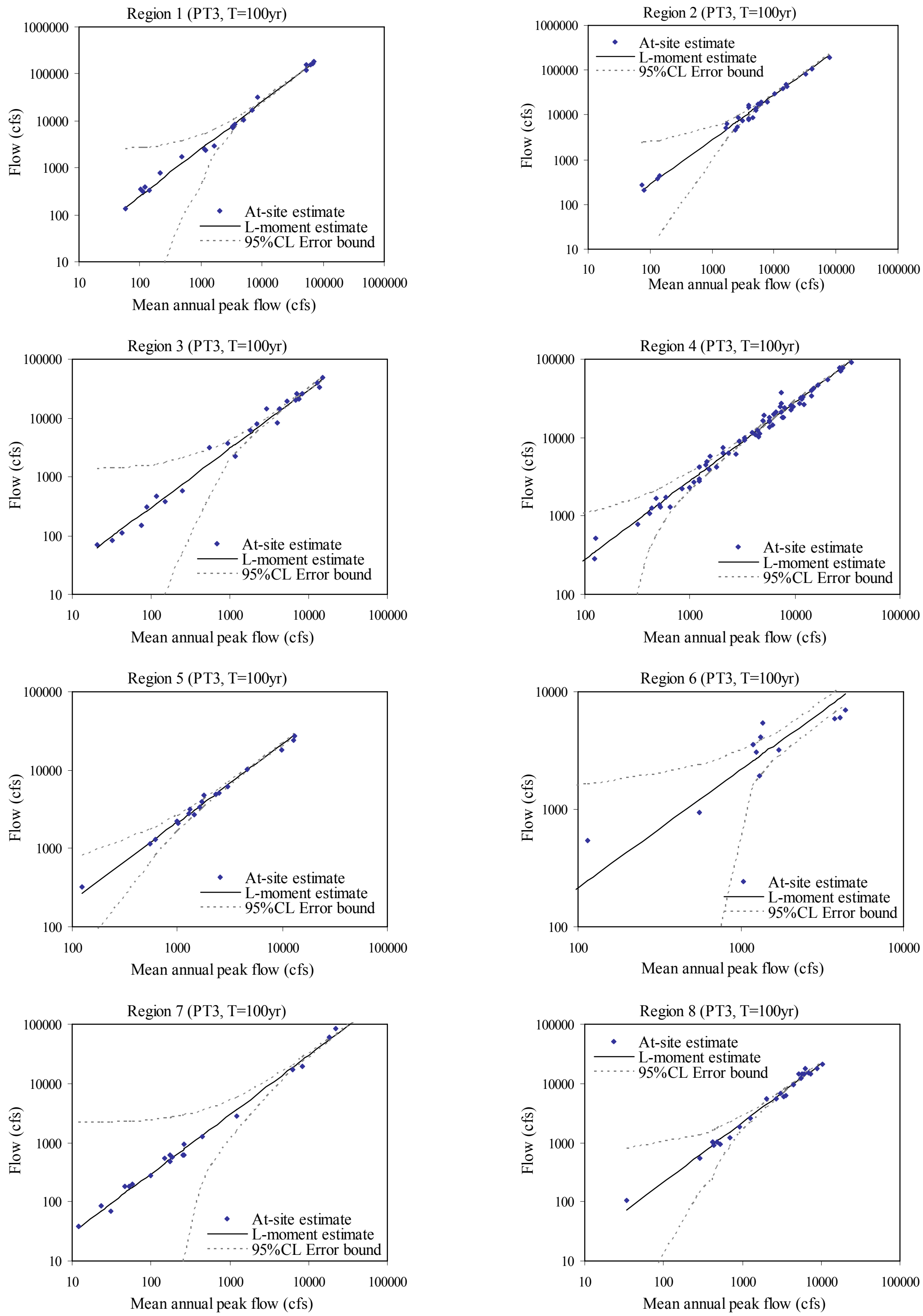

Figure 4.2.3. 95% confidence intervals for regional PTIII L-moment estimates.

V. Regional Regression Analysis

5.1. Introduction

In performing flood frequency analysis for a location of interest, the information about flow measurements is gathered first. However, data may not be available for the location of interest. Regional regression is an idea in which the flood characteristics are related to the geographical or hydrological attributes, which are measurable for any location in a watershed. Generalized least square regression (GLS) discussed in Chapter III is introduced by Stedinger and Tasker (1985) to develop these relationships. It takes the data consistency (lengths of record and correlation) and geographical distance into account. Consequently, the method is physically based. The results from GLS method are presented next.

5.2. GLS regional regression results

To investigate the governing hydrological attributes to estimate peak flows the square of the correlation coefficient between each hydrological feature and the at-site quantile estimates are calculated. The result is shown in Table 5.2.1 for PTIII distribution and Table 5.2.2 for GEV distribution. The drainage area (A) and slope (S) are the primary factors. For the secondary factor, wet area (%W) is considered. Urbanization factor (%U) shows smaller correlation and hence it is not taken into further consideration. Therefore, three different models are set up for GLS regional regression. Equation 5.2.1, equation 5.2.2 and equation 5.2.3 are used to fit the quantile floods and hydrological features. The regression parameters, a, b, c, d and squares of the correlation coefficients values are estimated.

Table 5.2.1. R^2 values for the relationship between the individual hydrological attributes and PTIII flood quantile estimates.

Region No.	Attriibute	T=10yr	T=20yr	T=50yr	T=100yr	T=200yr
1	Drainage area (mi^2)	0.978	0.975	0.970	0.967	0.9642
	Slope	0.856	0.850	0.841	0.836	0.8311
	%W	0.380	0.306	0.260	0.232	0.2108
	%U	0.280	0.223	0.191	0.172	0.158
2	Drainage area (mi^2)	0.935	0.935	0.933	0.931	0.9289
	Slope	0.729	0.731	0.732	0.733	0.7331
	%W	0.183	0.166	0.160	0.156	0.1541
	%U	0.288	0.230	0.196	0.175	0.160
3	Drainage area (mi^2)	0.978	0.976	0.973	0.969	0.9662
	Slope	0.753	0.749	0.744	0.739	0.7355
	%W	0.096	0.067	0.054	0.046	0.0409
	%U	0.042	0.029	0.023	0.020	0.018
4	Drainage area (mi^2)	0.935	0.934	0.932	0.931	0.9287
	Slope	0.397	0.394	0.390	0.388	0.3858
	%W	0.234	0.213	0.203	0.196	0.1923
	%U	0.008	0.001	0.000	0.001	0.002
5	Drainage area (mi^2)	0.939	0.938	0.937	0.936	0.9351
	Slope	0.453	0.456	0.458	0.460	0.4609
	%W	0.028	0.037	0.043	0.048	0.0526
	%U	0.150	0.165	0.170	0.173	0.175
6	Drainage area (mi^2)	0.797	0.750	0.689	0.644	0.6005
	Slope	0.466	0.405	0.333	0.285	0.2433
	%W	0.074	0.091	0.112	0.127	0.141
	%U	0.228	0.193	0.188	0.186	0.187
7	Drainage area (mi^2)	0.907	0.903	0.899	0.896	0.8936
	Slope	0.534	0.529	0.523	0.520	0.5166
	%W	0.001	0.007	0.011	0.014	0.017
	%U	0.377	0.379	0.380	0.380	0.381
8	Drainage area (mi^2)	0.795	0.792	0.788	0.785	0.7814
	Slope	0.674	0.677	0.680	0.682	0.6826
	%W	0.000	0.003	0.007	0.009	0.0115
	%U	0.005	0.000	0.002	0.006	0.010

Table 5.2.2. R^2 values for the relationship between the individual hydrological attributes and GEV flood quantile estimates.

Region No.	Attriibute	T=10yr	T=20yr	T=50yr	T=100yr	T=200yr
1	Drainage area (mi^2)	0.982	0.978	0.971	0.964	0.956
	Slope	0.865	0.856	0.842	0.831	0.8175
	%W	0.389	0.311	0.260	0.227	0.2007
	%U	0.284	0.226	0.191	0.169	0.153
2	Drainage area (mi^2)	0.935	0.935	0.933	0.929	0.923
	Slope	0.723	0.728	0.732	0.733	0.7328
	%W	0.177	0.162	0.159	0.159	0.1602
	%U	0.296	0.235	0.197	0.172	0.153
3	Drainage area (mi^2)	0.975	0.978	0.974	0.967	0.9563
	Slope	0.753	0.753	0.748	0.740	0.7298
	%W	0.101	0.072	0.056	0.046	0.039
	%U	0.044	0.031	0.025	0.021	0.018
4	Drainage area (mi^2)	0.933	0.934	0.932	0.928	0.9221
	Slope	0.398	0.393	0.387	0.382	0.3763
	%W	0.226	0.207	0.201	0.198	0.1961
	%U	0.008	0.001	0.000	0.001	0.003
5	Drainage area (mi^2)	0.939	0.939	0.938	0.936	0.9344
	Slope	0.453	0.456	0.459	0.462	0.4641
	%W	0.029	0.037	0.042	0.047	0.0498
	%U	0.152	0.166	0.168	0.169	0.168
6	Drainage area (mi^2)	0.816	0.770	0.689	0.609	0.5126
	Slope	0.511	0.441	0.336	0.250	0.165
	%W	0.069	0.093	0.130	0.161	0.1931
	%U	0.228	0.199	0.207	0.216	0.227
7	Drainage area (mi^2)	0.911	0.906	0.899	0.894	0.8888
	Slope	0.541	0.534	0.524	0.517	0.5089
	%W	0.001	0.006	0.011	0.014	0.0162
	%U	0.375	0.377	0.379	0.381	0.383
8	Drainage area (mi^2)	0.799	0.796	0.789	0.781	0.771
	Slope	0.673	0.677	0.681	0.682	0.6816
	%W	0.000	0.003	0.006	0.009	0.0119
	%U	0.006	0.000	0.002	0.007	0.013

$$\text{Model I: } Q_T = aA^b \quad (5.2.1)$$

$$\text{Model II: } Q_T = aA^b S^c \quad (5.2.2)$$

$$\text{Model III: } Q_T = aA^b S^c (1+\%W)^d \quad (5.2.3)$$

The probability distributions used for regional regression are generalized extreme value (GEV), Pearson type III (PTIII) and log-Pearson type III (LPIII). GEV and PTIII distributions fit the observed data well and also provide stable results in regional flood index evaluation. Hence they are used further in the analysis. LPIII distribution is used because it may be required to be used in engineering design. However, from the previous analysis, it is not a good distribution to estimate regional flood values and this aspect should be kept in mind.

The coefficients calculated by the GLS method are summarized in Table 5.2.3 for PTIII distribution, Table 5.2.4 for GEV distribution and Table 5.2.5 for LPIII distribution. There are eight sub-tables in each table and they refer to different recurrence intervals of 2, 5, 10, 20, 25, 50, 100 and 200 years. In each sub-table, the coefficients, a, b, c, d and R^2, are given for each model and region. The unit for drainage area is square miles, slope is in percentage, wet area is in percentage and the regressed quantile flow is in cubic feet per second (cfs).

Examples of goodness-to-fit of the GLS regression results compared to the at-site quantile estimates are shown in Figure 5.2.1 (for PTIII), Figure 5.2.2 (for GEV), and Figure 5.2.3 (for LPIII) distributions. The model used for each plot is based on the maximum R-square value of the three regression models. They show the best-fitting GLS regional regression results for each distribution for eight hydrological regions. The discharges are plotted against the drainage areas.

We also use the ordinary least square (OLS) regression scheme to fit these data. The results from OLS are shown as dashed lines in Figures 5.2.1 to Figure 5.2.3. Graphically they are very close to the solid lines which are the GLS regression result; however, the result from GLS is slightly better than OLS in goodness-to-fit. Also, the most important part of GLS and its benefit is GLS contains more physical information than OLS, which is simple curve fitting.

Table 5.2.3. GLS Regression coefficients for the drainage areas and PTIII flood quantile estimates.

Regional regression for PT3 at T=2yr

	Parameter	a	b	c	d	R^2
	Model I	149.945	0.638			0.986
Region1	Model II	43.057	0.757	0.375		0.993
	Model III	43.529	0.756	0.375	-0.001	0.993
	Model I	256.676	0.594			0.960
Region2	Model II	23.476	0.847	0.569		0.985
	Model III	71.702	0.796	0.392	-0.416	0.979
	Model I	189.646	0.736			0.896
Region3	Model II	40.159	0.884	0.388		0.921
	Model III	35.647	0.883	0.392	0.114	0.918
	Model I	113.509	0.690			0.902
Region4	Model II	14.067	0.851	0.702		0.919
	Model III	20.346	0.868	0.677	-0.361	0.936
	Model I	46.832	0.706			0.955
Region5	Model II	17.384	0.814	0.409		0.971
	Model III	24.869	0.805	0.390	-0.169	0.971
	Model I	55.649	0.594			0.930
Region6	Model II	11.294	0.805	0.553		0.951
	Model III	10.119	0.787	0.502	0.129	0.950
	Model I	46.041	0.734			0.975
Region7	Model II	7.827	0.882	0.595		0.967
	Model III	50.878	0.885	0.505	-1.163	0.980
	Model I	45.274	0.704			0.714
Region8	Model II	76.124	0.649	-0.173		0.746
	Model III	113.421	0.829	0.075	-0.981	0.754

Regional regression for PT3 at T=5yr

	Parameter	a	b	c	d	R^2
	Model I	236.255	0.629			0.986
Region1	Model II	50.233	0.777	0.466		0.993
	Model III	64.427	0.773	0.428	-0.116	0.994
	Model I	414.755	0.587			0.960
Region2	Model II	44.143	0.824	0.534		0.984
	Model III	109.237	0.782	0.391	-0.336	0.978
	Model I	314.877	0.736			0.892
Region3	Model II	55.282	0.903	0.432		0.934
	Model III	54.276	0.904	0.435	0.005	0.934
	Model I	176.838	0.692			0.907
Region4	Model II	19.259	0.864	0.744		0.929
	Model III	22.998	0.873	0.734	-0.182	0.937
	Model I	71.136	0.689			0.949
Region5	Model II	27.561	0.792	0.388		0.965
	Model III	34.284	0.787	0.375	-0.103	0.965
	Model I	146.734	0.476			0.865
Region6	Model II	18.900	0.748	0.696		0.902
	Model III	15.591	0.720	0.613	0.215	0.911
	Model I	77.389	0.724			0.980
Region7	Model II	12.343	0.877	0.616		0.975
	Model III	81.087	0.882	0.526	-1.173	0.987
	Model I	63.815	0.701			0.705
Region8	Model II	135.992	0.619	-0.252		0.758
	Model III	202.458	0.798	-0.005	-0.976	0.750

Table 5.2.3. GLS Regression coefficients for the drainage areas and PTIII flood quantile estimates (Cont.)

Regional regression for PT3 at T=10yr

	Parameter	a	b	c	d	R^2
	Model I	296.745	0.624			0.986
Region1	Model II	52.398	0.789	0.522		0.992
	Model III	75.826	0.783	0.465	-0.171	0.993
	Model I	527.720	0.582			0.960
Region2	Model II	61.815	0.809	0.512		0.982
	Model III	135.469	0.772	0.388	-0.290	0.978
	Model I	405.407	0.739			0.884
Region3	Model II	63.246	0.917	0.459		0.935
	Model III	65.980	0.920	0.462	-0.059	0.935
	Model I	217.520	0.695			0.908
Region4	Model II	22.431	0.871	0.761		0.936
	Model III	24.246	0.876	0.758	-0.087	0.939
	Model I	86.460	0.680			0.946
Region5	Model II	34.119	0.782	0.377		0.962
	Model III	39.692	0.778	0.367	-0.070	0.962
	Model I	235.108	0.420			0.805
Region6	Model II	21.986	0.735	0.795		0.860
	Model III	17.452	0.706	0.705	0.239	0.881
	Model I	97.937	0.725			0.983
Region7	Model II	14.852	0.882	0.634		0.979
	Model III	96.520	0.888	0.544	-1.170	0.991
	Model I	76.143	0.698			0.700
Region8	Model II	185.371	0.602	-0.297		0.766
	Model III	275.664	0.780	-0.050	-0.972	0.741

Regional regression for PT3 at T=20yr

	Parameter	a	b	c	d	R^2
Region1	Model I	355.567	0.620			0.986
	Model II	54.017	0.799	0.568		0.991
	Model III	85.855	0.792	0.496	-0.213	0.993
Region2	Model I	638.916	0.579			0.960
	Model II	81.073	0.797	0.493		0.981
	Model III	161.046	0.764	0.385	-0.253	0.977
Region3	Model I	494.365	0.741			0.874
	Model II	69.992	0.930	0.482		0.932
	Model III	76.711	0.933	0.483	-0.108	0.931
Region4	Model I	255.045	0.697			0.908
	Model II	25.322	0.876	0.772		0.940
	Model III	25.383	0.878	0.775	-0.014	0.941
Region5	Model I	100.560	0.674			0.944
	Model II	40.279	0.775	0.368		0.960
	Model III	44.488	0.771	0.360	-0.043	0.960
Region6	Model I	339.097	0.377			0.737
	Model II	23.976	0.730	0.880		0.819
	Model III	18.500	0.702	0.787	0.252	0.852
Region7	Model I	117.108	0.727			0.985
	Model II	17.006	0.888	0.649		0.981
	Model III	108.747	0.894	0.560	-1.164	0.992
Region8	Model I	87.609	0.695			0.694
	Model II	237.709	0.588	-0.333		0.772
	Model III	352.797	0.765	-0.087	-0.967	0.730

Table 5.2.3. GLS Regression coefficients for the drainage areas and PTIII flood quantile estimates (Cont.)

Regional regression for PT3 at T=25yr

	Parameter	a	b	c	d	R^2
	Model I	374.290	0.619			0.985
Region1	Model II	54.493	0.802	0.581		0.991
	Model III	88.905	0.794	0.505	-0.225	0.992
	Model I	674.570	0.578			0.960
Region2	Model II	87.600	0.793	0.488		0.981
	Model III	169.212	0.762	0.385	-0.242	0.977
	Model I	522.800	0.742			0.871
Region3	Model II	72.001	0.933	0.488		0.931
	Model III	80.031	0.937	0.489	-0.122	0.929
	Model I	266.643	0.698			0.907
Region4	Model II	26.210	0.878	0.775		0.941
	Model III	25.735	0.879	0.780	0.006	0.941
	Model I	104.922	0.672			0.943
Region5	Model II	42.213	0.772	0.365		0.959
	Model III	45.949	0.769	0.357	-0.035	0.959
	Model I	375.840	0.365			0.714
Region6	Model II	24.475	0.729	0.905		0.805
	Model III	18.743	0.701	0.812	0.255	0.842
	Model I	123.072	0.727			0.985
Region7	Model II	17.648	0.890	0.653		0.981
	Model III	112.207	0.896	0.565	-1.161	0.992
	Model I	91.157	0.694			0.692
Region8	Model II	255.114	0.584	-0.343		0.774
	Model III	378.326	0.761	-0.097	-0.966	0.727

Regional regression for PT3 at T=50yr

	Parameter	a	b	c	d	R^2
Region1	Model I	432.021	0.615			0.985
	Model II	55.905	0.810	0.617		0.989
	Model III	97.995	0.800	0.528	-0.256	0.991
Region2	Model I	785.272	0.575			0.960
	Model II	108.870	0.783	0.473		0.980
	Model III	194.494	0.755	0.382	-0.213	0.976
Region3	Model I	610.741	0.744			0.860
	Model II	77.883	0.943	0.505		0.925
	Model III	90.045	0.947	0.505	-0.160	0.922
Region4	Model I	301.516	0.700			0.905
	Model II	28.869	0.882	0.783		0.943
	Model III	26.797	0.880	0.791	0.062	0.941
Region5	Model I	118.068	0.668			0.941
	Model II	48.131	0.766	0.358		0.957
	Model III	50.312	0.764	0.350	-0.012	0.956
Region6	Model I	500.040	0.332			0.642
	Model II	25.739	0.728	0.975		0.767
	Model III	19.328	0.702	0.881	0.261	0.815
Region7	Model I	141.101	0.730			0.985
	Model II	19.528	0.895	0.665		0.981
	Model III	121.854	0.903	0.579	-1.154	0.992
Region8	Model I	101.811	0.693			0.686
	Model II	310.715	0.573	-0.372		0.777
	Model III	459.468	0.749	-0.126	-0.961	0.717

Table 5.2.3. GLS Regression coefficients for the drainage areas and PTIII flood quantile estimates (Cont.)

Regional regression for PT3 at T=100yr

	Parameter	a	b	c	d	R^2
Region1	Model I	489.317	0.613			0.984
	Model II	57.265	0.817	0.648		0.988
	Model III	106.649	0.806	0.549	-0.282	0.990
Region2	Model I	896.268	0.573			0.959
	Model II	131.618	0.774	0.460		0.978
	Model III	219.760	0.750	0.380	-0.187	0.975
Region3	Model I	698.344	0.746			0.849
	Model II	83.356	0.952	0.521		0.919
	Model III	99.717	0.955	0.519	-0.191	0.913
Region4	Model I	334.991	0.702			0.901
	Model II	31.407	0.886	0.789		0.943
	Model III	27.817	0.882	0.800	0.110	0.940
Region5	Model I	130.743	0.663			0.939
	Model II	53.972	0.761	0.351		0.954
	Model III	54.480	0.758	0.343	0.008	0.954
Region6	Model I	639.416	0.303			0.569
	Model II	26.717	0.729	1.036		0.732
	Model III	19.759	0.704	0.942	0.264	0.790
Region7	Model I	158.534	0.733			0.985
	Model II	21.273	0.901	0.676		0.981
	Model III	130.208	0.909	0.591	-1.146	0.991
Region8	Model I	111.992	0.692			0.679
	Model II	368.409	0.564	-0.397		0.778
	Model III	543.001	0.739	-0.152	-0.956	0.707

Regional regression for PT3 at T=200yr

	Parameter	a	b	c	d	R^2
Region1	Model I	546.361	0.610			0.983
	Model II	58.599	0.823	0.675		0.986
	Model III	114.985	0.810	0.566	-0.304	0.989
Region2	Model I	1007.864	0.570			0.959
	Model II	155.814	0.766	0.448		0.977
	Model III	245.104	0.744	0.378	-0.165	0.974
Region3	Model I	785.830	0.747			0.838
	Model II	88.529	0.959	0.534		0.912
	Model III	109.137	0.962	0.531	-0.218	0.905
Region4	Model I	367.385	0.704			0.897
	Model II	33.848	0.889	0.794		0.942
	Model III	28.799	0.883	0.808	0.151	0.939
Region5	Model I	143.074	0.660			0.937
	Model II	59.790	0.756	0.344		0.952
	Model III	58.519	0.753	0.336	0.027	0.952
Region6	Model I	793.937	0.278			0.500
	Model II	27.519	0.730	1.090		0.702
	Model III	20.111	0.706	0.997	0.266	0.768
Region7	Model I	175.509	0.735			0.984
	Model II	22.918	0.905	0.686		0.980
	Model III	137.604	0.915	0.602	-1.137	0.989
Region8	Model I	121.778	0.691			0.672
	Model II	427.940	0.556	-0.419		0.779
	Model III	628.520	0.730	-0.174	-0.952	0.697

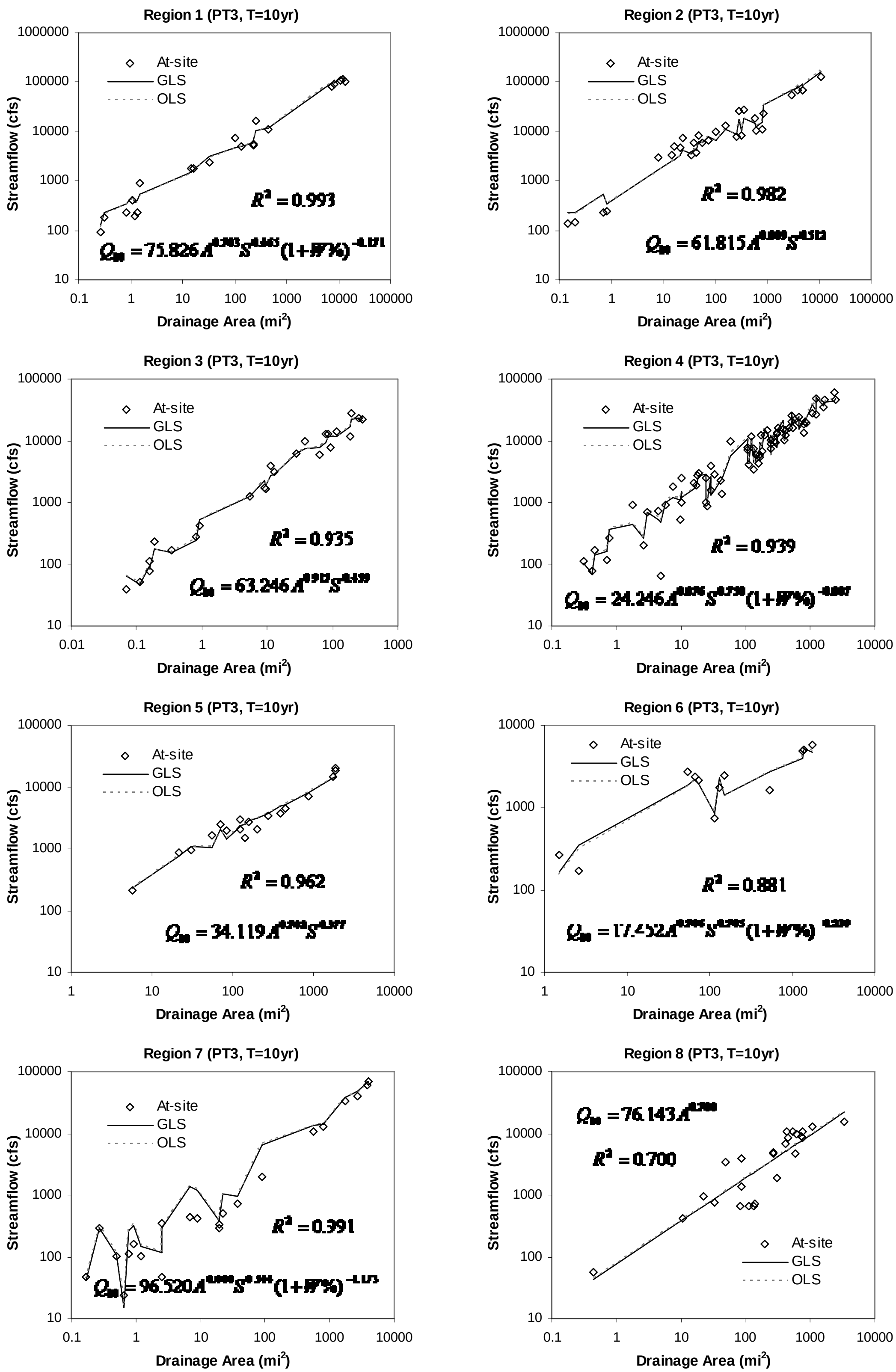

Figure 5.2.1(a). GLS regional regression for PTIII (T = 10 years)

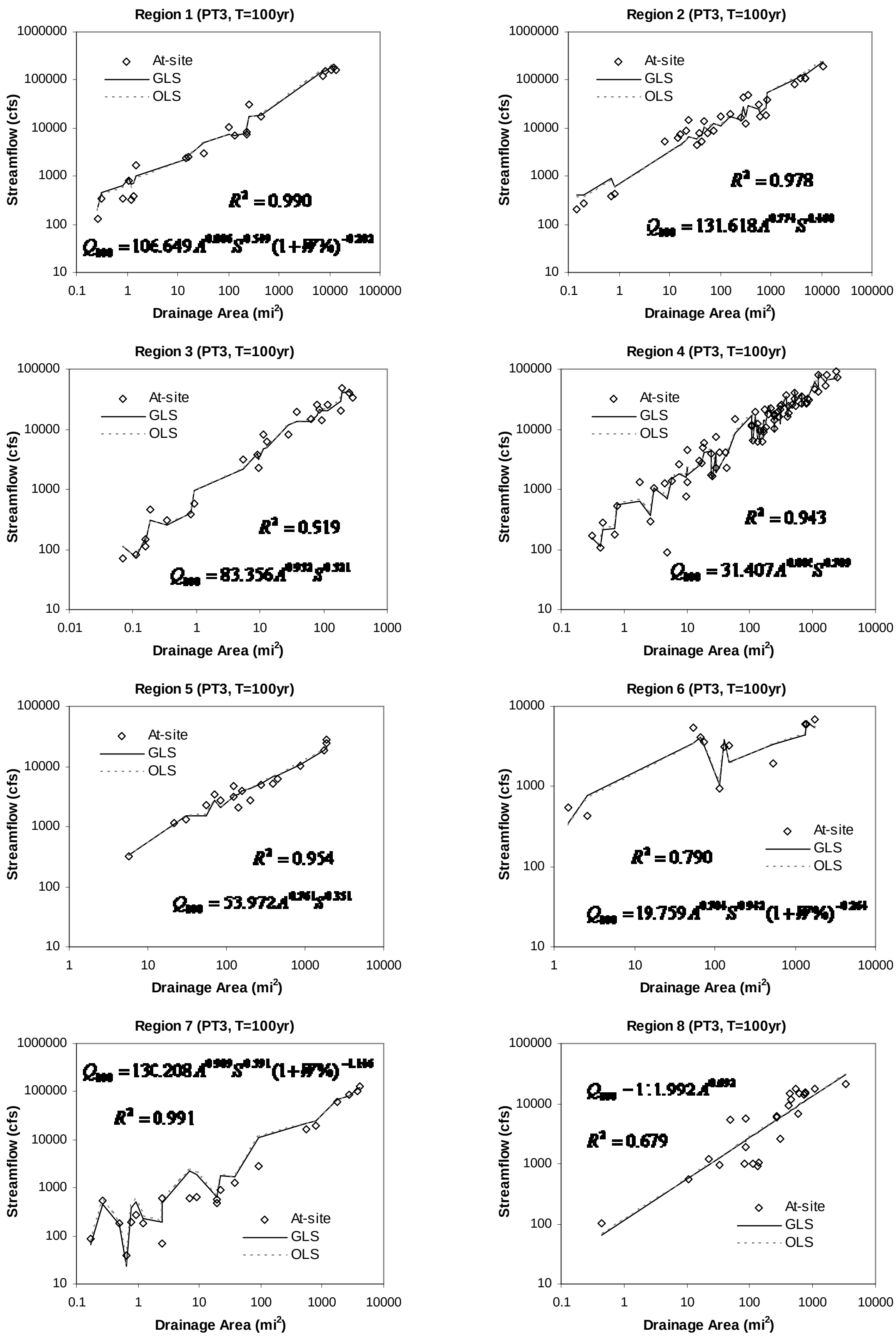

Figure 5.2.1(b). GLS regional regression for PTIII (T = 100 years)

Table 5.2.4. GLS Regression coefficients for the drainage areas and GEV flood quantile estimates.

Regional regression for GEV at T=2yr

	Parameter	***a***	***b***	***c***	***d***	***R^2***
Region1	Model I	140.925	0.646			0.986
	Model II	47.621	0.750	0.325		0.992
	Model III	45.221	0.750	0.335	0.028	0.992
Region2	Model I	243.479	0.599			0.960
	Model II	18.719	0.871	0.610		0.985
	Model III	65.081	0.814	0.413	-0.464	0.978
Region3	Model I	167.474	0.728			0.882
	Model II	32.504	0.884	0.410		0.905
	Model III	25.534	0.880	0.415	0.252	0.896
Region4	Model I	111.202	0.687			0.891
	Model II	15.383	0.840	0.666		0.903
	Model III	26.699	0.863	0.628	-0.530	0.929
Region5	Model I	46.595	0.706			0.955
	Model II	17.303	0.814	0.409		0.971
	Model III	25.007	0.805	0.389	-0.174	0.971
Region6	Model I	33.428	0.676			0.931
	Model II	9.398	0.840	0.480		0.947
	Model III	7.998	0.814	0.410	0.187	0.944
Region7	Model I	43.693	0.734			0.955
	Model II	8.200	0.874	0.561		0.949
	Model III	51.651	0.876	0.474	-1.144	0.962
Region8	Model I	43.710	0.710			0.714
	Model II	69.677	0.660	-0.156		0.742
	Model III	103.789	0.839	0.092	-0.979	0.757

Regional regression for GEV at T=5yr

	Parameter	***a***	***b***	***c***	***d***	***R^2***
Region1	Model I	219.915	0.637			0.986
	Model II	57.278	0.766	0.404		0.993
	Model III	67.874	0.763	0.379	-0.080	0.993
Region2	Model I	387.126	0.592			0.960
	Model II	35.122	0.846	0.571		0.984
	Model III	98.483	0.798	0.408	-0.383	0.978
Region3	Model I	280.467	0.730			0.887
	Model II	53.527	0.888	0.413		0.924
	Model III	48.053	0.887	0.416	0.107	0.922
Region4	Model I	171.392	0.690			0.898
	Model II	19.863	0.856	0.725		0.913
	Model III	27.043	0.870	0.705	-0.304	0.928
Region5	Model I	70.720	0.689			0.948
	Model II	27.182	0.794	0.391		0.965
	Model III	34.452	0.788	0.378	-0.112	0.965
Region6	Model I	103.740	0.529			0.887
	Model II	18.653	0.755	0.603		0.912
	Model III	14.859	0.721	0.509	0.255	0.918
Region7	Model I	73.201	0.721			0.964
	Model II	12.901	0.866	0.582		0.961
	Model III	83.719	0.870	0.494	-1.165	0.974
Region8	Model I	60.632	0.708			0.703
	Model II	119.005	0.635	-0.225		0.749
	Model III	177.392	0.814	0.022	-0.976	0.756

Table 5.2.4. GLS Regression coefficients for the drainage areas and GEV flood quantile estimates (Cont.)

Regional regression for GEV at T=10yr

	Parameter	a	b	c	d	R^2
	Model I	278.174	0.631			0.986
Region1	Model II	58.976	0.780	0.466		0.993
	Model III	79.415	0.775	0.421	-0.140	0.993
	Model I	496.428	0.587			0.960
Region2	Model II	50.565	0.829	0.544		0.983
	Model III	123.245	0.787	0.403	-0.330	0.977
	Model I	364.046	0.736			0.886
Region3	Model II	64.232	0.902	0.431		0.932
	Model III	61.594	0.903	0.434	0.033	0.932
	Model I	211.430	0.693			0.903
Region4	Model II	22.507	0.866	0.752		0.922
	Model III	27.167	0.876	0.741	-0.193	0.931
	Model I	86.216	0.681			0.945
Region5	Model II	33.818	0.783	0.379		0.962
	Model III	39.903	0.779	0.369	-0.077	0.962
	Model I	176.101	0.463			0.839
Region6	Model II	22.526	0.735	0.710		0.878
	Model III	16.885	0.698	0.601	0.298	0.897
	Model I	93.841	0.721			0.972
Region7	Model II	15.538	0.871	0.604		0.969
	Model III	100.402	0.875	0.516	-1.164	0.981
	Model I	72.370	0.705			0.699
Region8	Model II	163.334	0.618	-0.272		0.759
	Model III	243.387	0.795	-0.025	-0.972	0.749

Regional regression for GEV at T=20yr

	Parameter	a	b	c	d	R^2
Region1	Model I	339.463	0.626			0.986
	Model II	58.344	0.794	0.530		0.992
	Model III	88.214	0.788	0.466	-0.193	0.993
Region2	Model I	613.925	0.583			0.960
	Model II	70.293	0.812	0.517		0.981
	Model III	149.811	0.776	0.397	-0.279	0.977
Region3	Model I	453.341	0.742			0.879
	Model II	73.361	0.917	0.451		0.934
	Model III	74.639	0.919	0.454	-0.033	0.934
Region4	Model I	250.053	0.697			0.906
	Model II	24.820	0.876	0.774		0.932
	Model III	26.993	0.881	0.771	-0.094	0.935
Region5	Model I	100.618	0.674			0.943
	Model II	40.344	0.775	0.368		0.960
	Model III	44.761	0.771	0.359	-0.045	0.959
Region6	Model I	275.116	0.407			0.770
	Model II	24.486	0.728	0.821		0.835
	Model III	17.334	0.692	0.704	0.330	0.874
Region7	Model I	114.570	0.723			0.978
	Model II	17.687	0.879	0.628		0.976
	Model III	112.653	0.884	0.541	-1.159	0.987
Region8	Model I	84.217	0.701			0.694
	Model II	217.582	0.600	-0.317		0.768
	Model III	323.533	0.776	-0.071	-0.967	0.736

Table 5.2.4. GLS Regression coefficients for the drainage areas and GEV flood quantile estimates (Cont.)

Regional regression for GEV at T=25yr

	Parameter	a	b	c	d	R^2
	Model I	360.139	0.624			0.986
Region1	Model II	57.807	0.799	0.551		0.992
	Model III	90.587	0.792	0.481	-0.209	0.993
	Model I	654.072	0.581			0.960
Region2	Model II	77.818	0.806	0.508		0.981
	Model III	158.931	0.772	0.395	-0.263	0.977
	Model I	483.849	0.744			0.875
Region3	Model II	76.093	0.922	0.458		0.933
	Model III	78.862	0.925	0.461	-0.053	0.933
	Model I	262.353	0.698			0.907
Region4	Model II	25.513	0.879	0.780		0.934
	Model III	26.878	0.883	0.780	-0.063	0.937
	Model I	105.085	0.672			0.943
Region5	Model II	42.458	0.772	0.364		0.959
	Model III	46.246	0.769	0.356	-0.035	0.959
	Model I	314.451	0.390			0.743
Region6	Model II	24.783	0.728	0.858		0.819
	Model III	17.249	0.692	0.738	0.339	0.865
	Model I	121.345	0.724			0.980
Region7	Model II	18.290	0.882	0.636		0.977
	Model III	115.773	0.888	0.549	-1.156	0.989
	Model I	88.116	0.700			0.692
Region8	Model II	237.672	0.594	-0.331		0.770
	Model III	353.014	0.770	-0.085	-0.966	0.731

Regional regression for GEV at T=50yr

	Parameter	a	b	c	d	R^2
Region1	Model I	427.953	0.619			0.985
	Model II	55.415	0.814	0.616		0.990
	Model III	96.752	0.805	0.528	-0.257	0.992
Region2	Model I	787.373	0.577			0.960
	Model II	105.820	0.788	0.480		0.979
	Model III	189.496	0.761	0.388	-0.214	0.976
Region3	Model I	585.435	0.751			0.860
	Model II	84.115	0.938	0.479		0.926
	Model III	92.281	0.942	0.481	-0.114	0.924
Region4	Model I	300.415	0.702			0.906
	Model II	27.538	0.887	0.798		0.941
	Model III	26.356	0.887	0.804	0.032	0.939
Region5	Model I	118.526	0.667			0.940
	Model II	49.123	0.764	0.351		0.956
	Model III	50.688	0.761	0.344	-0.006	0.956
Region6	Model I	466.918	0.340			0.639
	Model II	24.868	0.731	0.975		0.766
	Model III	16.516	0.697	0.852	0.358	0.836
Region7	Model I	142.840	0.729			0.983
	Model II	19.911	0.894	0.662		0.981
	Model III	123.148	0.900	0.578	-1.146	0.991
Region8	Model I	100.607	0.696			0.684
	Model II	309.992	0.575	-0.375		0.776
	Model III	458.217	0.751	-0.130	-0.960	0.712

Table 5.2.4. GLS Regression coefficients for the drainage areas and GEV flood quantile estimates (Cont.)

Regional regression for GEV at T=100yr

	Parameter	a	b	c	d	R^2
Region1	Model I	502.034	0.613			0.983
	Model II	52.261	0.829	0.682		0.987
	Model III	101.327	0.818	0.577	-0.302	0.989
Region2	Model I	935.570	0.572			0.959
	Model II	142.691	0.770	0.450		0.977
	Model III	224.135	0.748	0.380	-0.165	0.974
Region3	Model I	699.260	0.758			0.837
	Model II	91.561	0.955	0.500		0.911
	Model III	106.380	0.959	0.501	-0.174	0.905
Region4	Model I	338.494	0.706			0.899
	Model II	29.401	0.896	0.815		0.942
	Model III	25.616	0.891	0.827	0.127	0.939
Region5	Model I	131.386	0.662			0.938
	Model II	55.997	0.756	0.338		0.953
	Model III	54.942	0.754	0.332	0.023	0.953
Region6	Model I	679.210	0.292			0.508
	Model II	23.975	0.740	1.095		0.712
	Model III	15.343	0.708	0.971	0.367	0.803
Region7	Model I	165.136	0.735			0.983
	Model II	21.175	0.906	0.691		0.980
	Model III	127.332	0.914	0.609	-1.134	0.990
Region8	Model I	113.807	0.691			0.673
	Model II	400.359	0.556	-0.420		0.779
	Model III	587.756	0.732	-0.174	-0.954	0.689

Regional regression for GEV at T=200yr

	Parameter	a	b	c	d	R^2
Region1	Model I	583.377	0.608			0.980
	Model II	48.663	0.844	0.750		0.982
	Model III	104.543	0.831	0.626	-0.346	0.985
Region2	Model I	1101.084	0.567			0.957
	Model II	191.505	0.751	0.419		0.973
	Model III	263.778	0.736	0.370	-0.117	0.971
Region3	Model I	827.871	0.765			0.806
	Model II	98.503	0.971	0.522		0.886
	Model III	121.376	0.976	0.522	-0.233	0.876
Region4	Model I	376.795	0.711			0.883
	Model II	31.130	0.904	0.830		0.935
	Model III	24.698	0.894	0.848	0.223	0.932
Region5	Model I	143.723	0.658			0.936
	Model II	63.126	0.749	0.325		0.950
	Model III	59.066	0.746	0.318	0.051	0.949
Region6	Model I	975.037	0.245			0.362
	Model II	22.440	0.751	1.218		0.667
	Model III	13.982	0.723	1.097	0.370	0.769
Region7	Model I	188.332	0.741			0.979
	Model II	22.107	0.920	0.721		0.975
	Model III	128.715	0.930	0.642	-1.121	0.984
Region8	Model I	127.852	0.686			0.657
	Model II	513.403	0.537	-0.464		0.777
	Model III	747.139	0.712	-0.218	-0.948	0.662

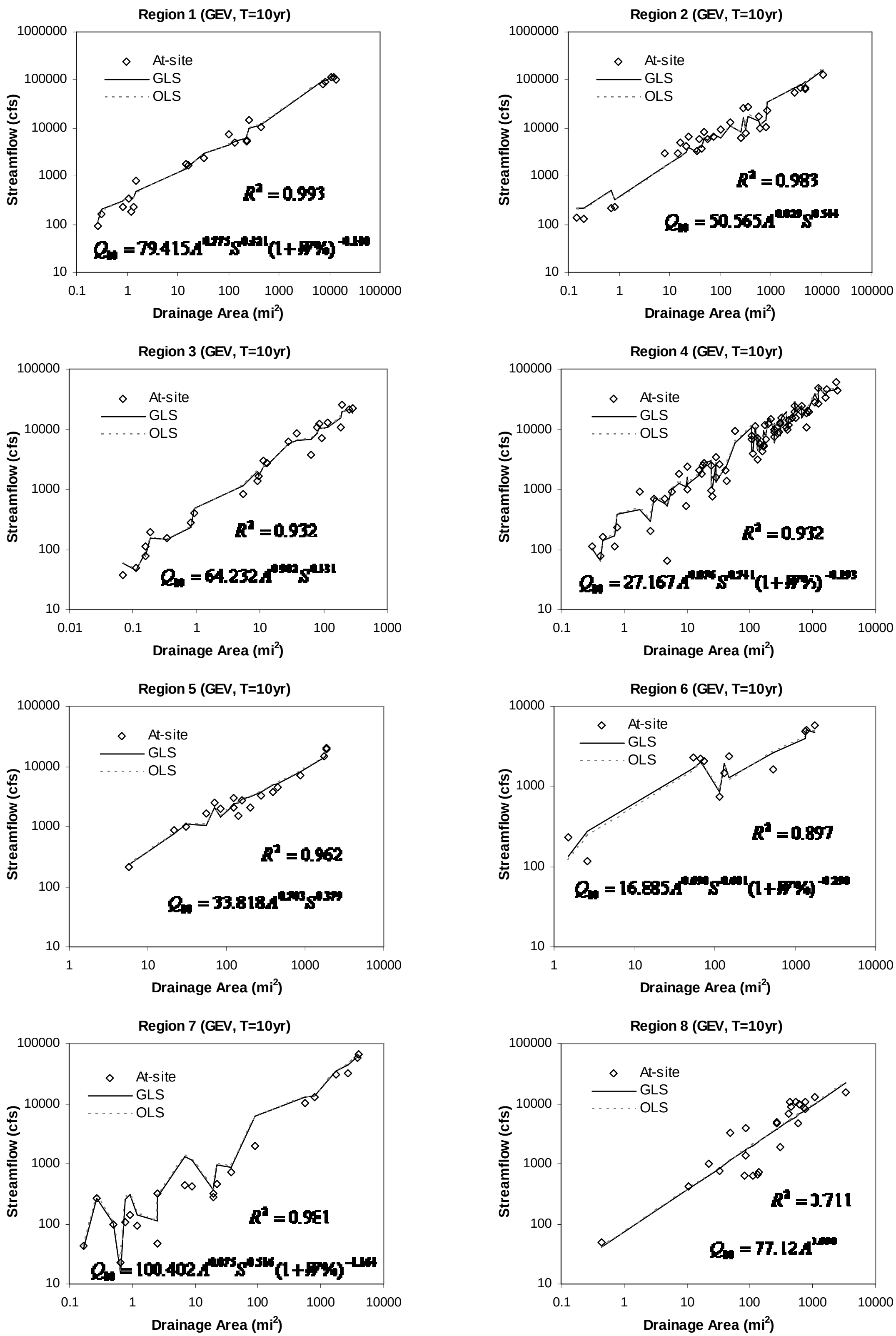

Figure 5.2.2(a). GLS regional regression for GEV (T = 10 years).

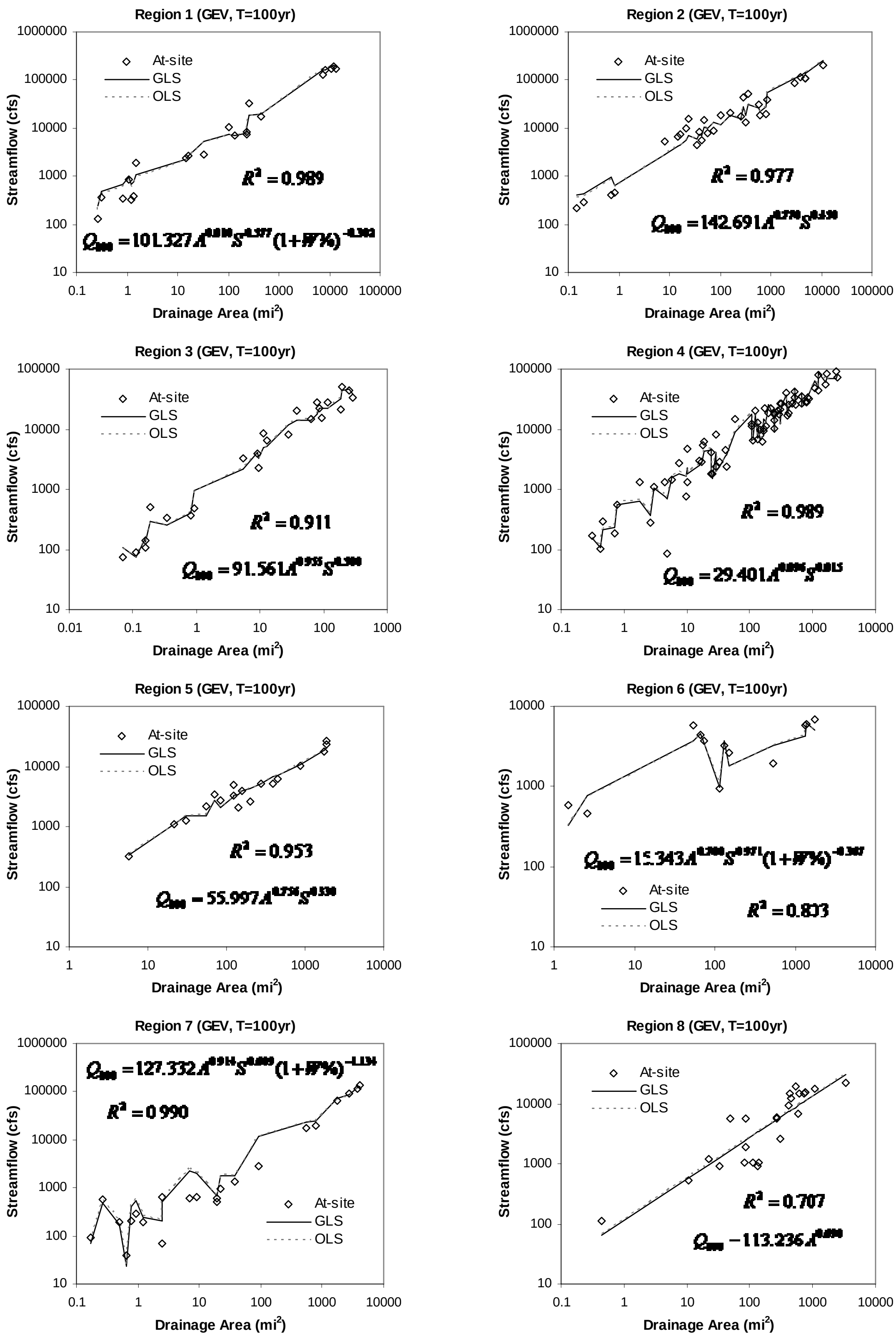

Figure 5.2.2(b). GLS regional regression for GEV (T = 100 years).

Table 5.2.5. GLS Regression coefficients for the drainage areas and LPIII flood quantile estimates.

Regional regression for LP3 at T=2yr

	Parameter	a	b	c	d	R^2
	Model I	146.884	0.640			0.987
Region1	Model II	37.223	0.772	0.410		0.994
	Model III	39.164	0.771	0.403	-0.022	0.994
	Model I	287.294	0.584			0.949
Region2	Model II	28.058	0.830	0.553		0.978
	Model III	67.892	0.789	0.414	-0.329	0.972
	Model I	193.046	0.740			0.896
Region3	Model II	39.770	0.890	0.396		0.926
	Model III	36.516	0.888	0.397	0.092	0.924
	Model I	115.386	0.691			0.899
Region4	Model II	14.011	0.854	0.711		0.919
	Model III	20.845	0.869	0.682	-0.373	0.937
	Model I	46.707	0.707			0.955
Region5	Model II	17.813	0.811	0.399		0.971
	Model III	26.121	0.802	0.379	-0.180	0.971
	Model I	55.398	0.598			0.922
Region6	Model II	10.451	0.818	0.574		0.942
	Model III	9.711	0.804	0.539	0.092	0.941
	Model I	44.146	0.741			0.977
Region7	Model II	7.289	0.891	0.604		0.968
	Model III	48.104	0.894	0.515	-1.175	0.981
	Model I	47.553	0.701			0.724
Region8	Model II	81.781	0.643	-0.181		0.756
	Model III	118.870	0.811	0.052	-0.919	0.767

Regional regression for LP3 at T=5yr

	Parameter	a	b	c	d	R^2
	Model I	230.406	0.630			0.985
Region1	Model II	48.429	0.780	0.466		0.992
	Model III	64.032	0.777	0.423	-0.135	0.993
	Model I	460.726	0.575			0.946
Region2	Model II	49.685	0.811	0.530		0.974
	Model III	105.987	0.776	0.411	-0.281	0.969
	Model I	311.620	0.736			0.889
Region3	Model II	61.061	0.891	0.407		0.931
	Model III	59.102	0.890	0.408	0.032	0.931
	Model I	180.331	0.691			0.899
Region4	Model II	19.209	0.863	0.755		0.922
	Model III	24.870	0.874	0.737	-0.249	0.933
	Model I	72.273	0.688			0.947
Region5	Model II	28.531	0.789	0.381		0.964
	Model III	35.827	0.783	0.369	-0.107	0.964
	Model I	141.277	0.483			0.869
Region6	Model II	19.246	0.747	0.677		0.902
	Model III	16.183	0.720	0.602	0.198	0.910
	Model I	76.881	0.720			0.977
Region7	Model II	12.134	0.875	0.619		0.971
	Model III	84.255	0.879	0.528	-1.209	0.985
	Model I	65.529	0.700			0.712
Region8	Model II	138.783	0.620	-0.250		0.763
	Model III	201.844	0.787	-0.019	-0.914	0.771

Table 5.2.5. GLS Regression coefficients for the drainage areas and LPIII flood quantile estimates (Cont.)

Regional regression for LP3 at T=10yr

	Parameter	a	b	c	d	R^2
	Model I	294.188	0.622			0.982
Region1	Model II	50.433	0.791	0.528		0.990
	Model III	77.185	0.787	0.462	-0.206	0.991
	Model I	597.250	0.567			0.946
Region2	Model II	72.813	0.789	0.502		0.973
	Model III	150.543	0.755	0.387	-0.268	0.968
	Model I	396.151	0.740			0.882
Region3	Model II	75.108	0.898	0.415		0.929
	Model III	73.933	0.899	0.417	0.010	0.929
	Model I	222.117	0.692			0.899
Region4	Model II	21.547	0.872	0.785		0.929
	Model III	25.522	0.880	0.774	-0.169	0.936
	Model I	87.923	0.679			0.944
Region5	Model II	35.139	0.779	0.373		0.960
	Model III	39.813	0.776	0.365	-0.057	0.960
	Model I	249.432	0.409			0.805
Region6	Model II	25.015	0.715	0.768		0.859
	Model III	19.615	0.684	0.676	0.253	0.883
	Model I	102.550	0.712			0.983
Region7	Model II	14.355	0.876	0.659		0.977
	Model III	103.024	0.881	0.567	-1.232	0.990
	Model I	77.120	0.698			0.711
Region8	Model II	184.753	0.605	-0.292		0.774
	Model III	268.128	0.771	-0.061	-0.908	0.769

Regional regression for LP3 at T=20yr

	Parameter	a	b	c	d	R^2
Region1	Model I	362.225	0.614			0.979
	Model II	49.654	0.804	0.595		0.986
	Model III	86.990	0.798	0.508	-0.269	0.988
Region2	Model I	744.503	0.558			0.946
	Model II	104.190	0.765	0.470		0.972
	Model III	212.958	0.732	0.357	-0.263	0.968
Region3	Model I	482.516	0.745			0.867
	Model II	88.017	0.908	0.423		0.918
	Model III	87.416	0.909	0.427	-0.006	0.918
Region4	Model I	260.735	0.694			0.896
	Model II	23.201	0.881	0.812		0.934
	Model III	25.426	0.886	0.807	-0.097	0.938
Region5	Model I	101.687	0.672			0.940
	Model II	41.066	0.772	0.367		0.957
	Model III	42.425	0.770	0.361	-0.009	0.957
Region6	Model I	414.003	0.341			0.709
	Model II	29.747	0.694	0.863		0.797
	Model III	21.750	0.661	0.756	0.299	0.846
Region7	Model I	129.909	0.706			0.987
	Model II	15.710	0.883	0.709		0.981
	Model III	116.037	0.889	0.614	-1.253	0.994
Region8	Model I	88.137	0.696			0.711
	Model II	235.569	0.591	-0.328		0.785
	Model III	340.265	0.756	-0.099	-0.901	0.764

Table 5.2.5. GLS Regression coefficients for the drainage areas and LPIII flood quantile estimates (Cont.)

Regional regression for LP3 at T=25yr

	Parameter	a	b	c	d	R^2
Region1	Model I	385.335	0.611			0.977
	Model II	49.037	0.809	0.618		0.985
	Model III	89.581	0.802	0.523	-0.288	0.987
Region2	Model I	794.626	0.555			0.946
	Model II	116.487	0.758	0.459		0.972
	Model III	237.777	0.724	0.347	-0.263	0.967
Region3	Model I	511.157	0.747			0.859
	Model II	91.988	0.911	0.426		0.912
	Model III	91.518	0.913	0.430	-0.010	0.912
Region4	Model I	272.681	0.694			0.893
	Model II	23.633	0.884	0.820		0.935
	Model III	25.307	0.888	0.817	-0.075	0.938
Region5	Model I	105.803	0.671			0.940
	Model II	42.872	0.770	0.364		0.956
	Model III	43.075	0.768	0.360	0.006	0.955
Region6	Model I	482.603	0.321			0.670
	Model II	31.024	0.689	0.893		0.773
	Model III	22.181	0.656	0.783	0.312	0.831
Region7	Model I	139.131	0.705			0.988
	Model II	16.000	0.886	0.725		0.982
	Model III	119.188	0.892	0.631	-1.260	0.994
Region8	Model I	91.624	0.695			0.711
	Model II	253.144	0.586	-0.339		0.789
	Model III	364.903	0.751	-0.110	-0.899	0.761

Regional regression for LP3 at T=50yr

	Parameter	a	b	c	d	R^2
Region1	Model I	461.583	0.603			0.972
	Model II	46.397	0.823	0.689		0.979
	Model III	96.152	0.813	0.574	-0.346	0.982
Region2	Model I	959.847	0.546			0.945
	Model II	163.181	0.733	0.424		0.970
	Model III	333.100	0.699	0.312	-0.263	0.965
Region3	Model I	603.822	0.753			0.826
	Model II	103.885	0.922	0.437		0.879
	Model III	103.635	0.925	0.442	-0.021	0.879
Region4	Model I	308.622	0.696			0.880
	Model II	24.740	0.892	0.844		0.934
	Model III	24.768	0.893	0.846	-0.010	0.934
Region5	Model I	117.769	0.666			0.937
	Model II	48.243	0.764	0.358		0.952
	Model III	44.677	0.763	0.355	0.050	0.952
Region6	Model I	758.207	0.260			0.525
	Model II	34.208	0.677	0.988		0.685
	Model III	22.836	0.644	0.868	0.349	0.773
Region7	Model I	169.157	0.702			0.989
	Model II	16.530	0.896	0.780		0.981
	Model III	126.182	0.904	0.685	-1.279	0.993
Region8	Model I	102.412	0.693			0.710
	Model II	312.159	0.573	-0.372		0.797
	Model III	446.450	0.737	-0.143	-0.891	0.751

Table 5.2.5. GLS Regression coefficients for the drainage areas and LPIII flood quantile estimates (Cont.)

Regional regression for LP3 at T=100yr

	Parameter	a	b	c	d	R^2
Region1	Model I	545.593	0.594			0.965
	Model II	43.115	0.837	0.762		0.971
	Model III	100.881	0.825	0.627	-0.400	0.976
Region2	Model I	1140.736	0.537			0.942
	Model II	225.895	0.707	0.388		0.966
	Model III	462.830	0.674	0.276	-0.264	0.962
Region3	Model I	703.266	0.760			0.771
	Model II	115.414	0.934	0.447		0.823
	Model III	115.061	0.937	0.455	-0.031	0.822
Region4	Model I	343.147	0.699			0.857
	Model II	25.572	0.900	0.866		0.923
	Model III	24.071	0.899	0.872	0.051	0.923
Region5	Model I	128.672	0.662			0.933
	Model II	53.337	0.759	0.352		0.948
	Model III	45.829	0.759	0.350	0.091	0.948
Region6	Model I	1154.568	0.202			0.358
	Model II	36.271	0.671	1.082		0.588
	Model III	22.666	0.639	0.955	0.381	0.703
Region7	Model I	201.377	0.700			0.983
	Model II	16.612	0.909	0.838		0.973
	Model III	129.619	0.918	0.741	-1.296	0.984
Region8	Model I	113.236	0.690			0.707
	Model II	378.588	0.560	-0.403		0.804
	Model III	536.122	0.723	-0.175	-0.883	0.739

Regional regression for LP3 at T=200yr

	Parameter	a	b	c	d	R^2
Region1	Model I	638.605	0.585			0.954
	Model II	39.569	0.851	0.835		0.959
	Model III	104.145	0.837	0.681	-0.452	0.966
Region2	Model I	1338.893	0.528			0.937
	Model II	309.724	0.682	0.352		0.959
	Model III	637.815	0.648	0.239	-0.267	0.956
Region3	Model I	810.942	0.766			0.690
	Model II	126.844	0.945	0.458		0.737
	Model III	126.029	0.950	0.468	-0.038	0.736
Region4	Model I	376.568	0.701			0.818
	Model II	26.199	0.908	0.886		0.899
	Model III	23.286	0.903	0.896	0.110	0.899
Region5	Model I	138.683	0.659			0.930
	Model II	58.230	0.755	0.345		0.944
	Model III	46.684	0.755	0.344	0.131	0.944
Region6	Model I	1713.317	0.148			0.199
	Model II	37.354	0.668	1.175		0.498
	Model III	21.930	0.639	1.042	0.409	0.629
Region7	Model I	235.898	0.699			0.968
	Model II	16.371	0.922	0.896		0.957
	Model III	130.285	0.933	0.798	-1.312	0.966
Region8	Model I	124.222	0.686			0.701
	Model II	453.553	0.547	-0.432		0.808
	Model III	634.820	0.710	-0.205	-0.874	0.724

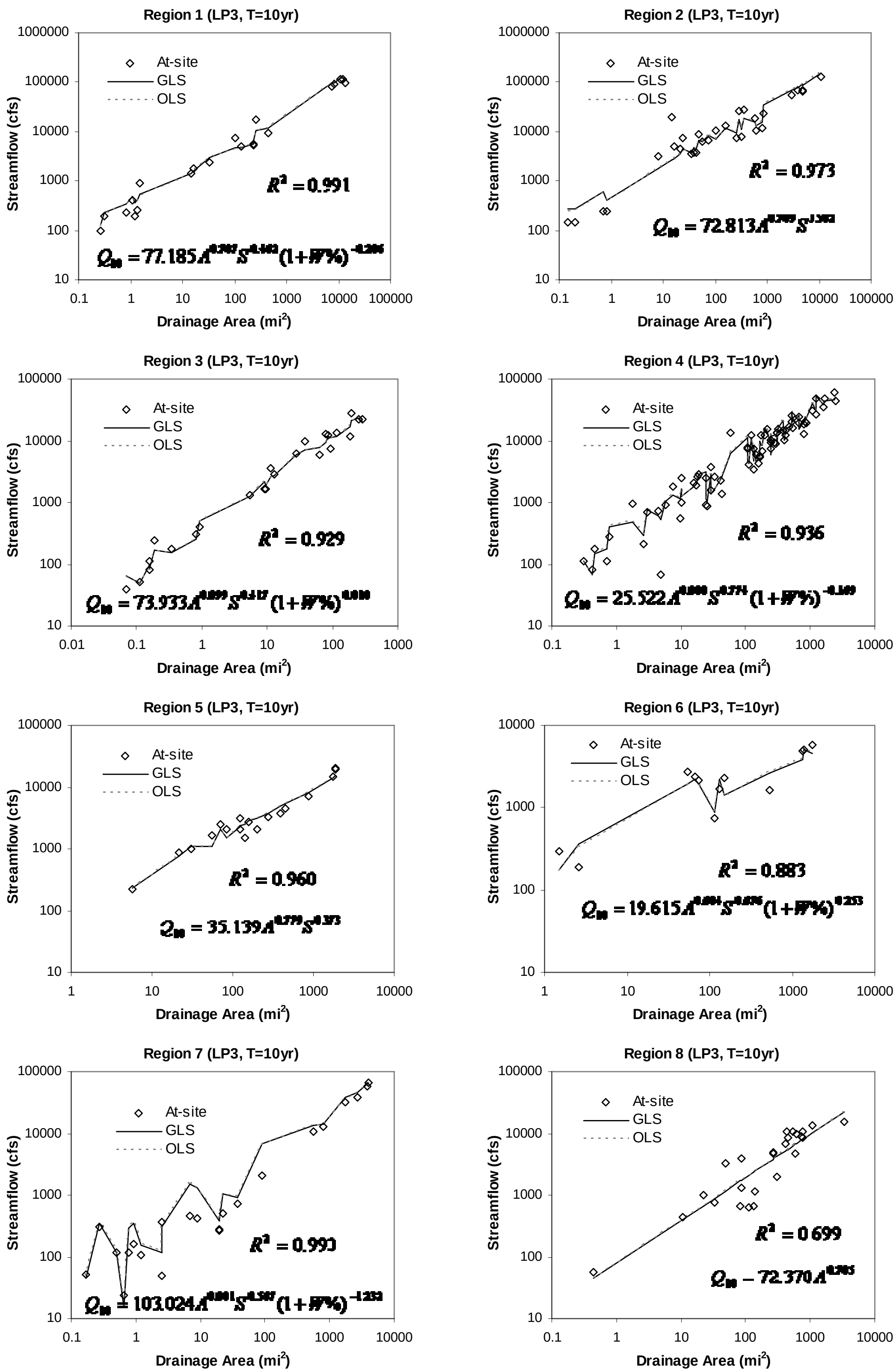

Figure 5.2.3(a). GLS regional regression for LPIII (T = 10 years)

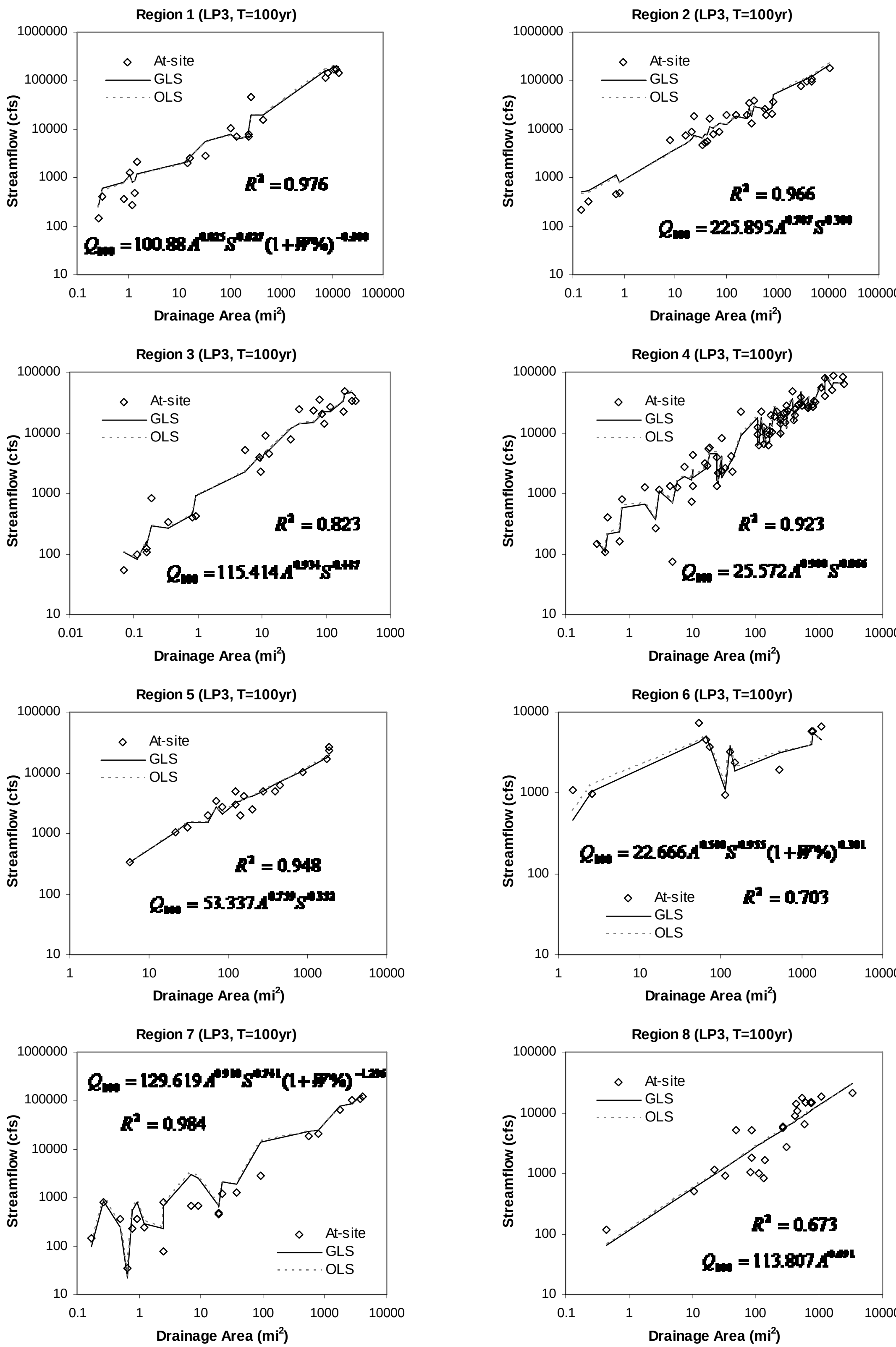

Figure 5.2.3(b). GLS regional regression for LPIII (T = 100 years)

From Table 5.2.3 to Table 5.2.5, we can see that for most of cases, adding W% increases the accuracy of regression; however, it does not yield a significantly better prediction result. Therefore it is easier to only take drainage area and slope into consideration. Also, including only area and slope is better since the more detailed attributes of most ungaged sites are unavailable. Drainage area and stream slope are the basic information which are easily available.

Ideally, the slope should be proportional to the quantile floods. But, region 8 has the opposite behavior with negative coefficients for the slope term. To avoid this unreasonable situation, only drainage area is considered for region 8. As for other regions, the factors contributing to the best fit in region 1, 4, 6 and 7 are area (A), slope (S) and percentage wet area (%W); factors in region 2 and region 3 are area and slope. Region 5 could have area and slope or area, slope and percent wet area.

Additionally, the regional regression is performed for regions 1 and 5 by using the data used by Srinivas and Rao (2003). The results for region 2, 3, 4, and 6 are kept because both Srinivas and Rao (2003, Figure 3.1.1) and Knipe and Rao (2004, Figure 3.1.2) have the same definition for these four regions. Merging region 1 and region 7 in Figure 3.1.2, results in region 1 in Figure 3.1.1, and merging region 5 and region 8 in Figure 3.1.2 yields region 5 in Figure 3.1.1. The coefficients of generalized least square regression for the mean annual peak flow and log-mean peak flow for these six regions are shown in Table 5.2.6 (a) and Table 5.2.6(b), respectively. The GLS regression is also performed for PTIII, GEV, and LPIII distributions for region 1 and 5 in Figure 3.1.1 and the results are listed in Table 5.2.7, Table 5.2.8 and Table 5.2.9, respectively.

Taking region 5 (Figure 3.1.1) for example, the R^2 values in Table 5.2.7 to Table 5.2.8 compared to the R^2 values in Table 5.2.3 to Table 5.2.5, the R^2 is lower than the R^2 in region 5 (Figure 3.1.2) but higher than region 8 (Figure 3.1.2). For instance, in Table 5.2.7 for PTIII distribution with 100-year recurrence interval, the R^2 value is 0.856, which is between 0.707 (Table 5.2.3 for PTIII, 100-year in Region 8) and 0.954 (Table 5.2.3 for PTIII, 100-year in Region 5). For Region 1, it is better to keep using the equation derived from Knipe and Rao (2004) since it yields a better fit than that obtained by considering the merged area. Further analysis of data from region 8 did not yield better results than the equation derived from the merged area. The reason is that several validated stations in region 8 have poor correlation between flood magnitudes and drainage areas. Consequently, it is not possible to derive a reasonable regression among these variables. This leads to the high prediction error in region 8. As for region 1 and region 7, in Table 5.2.7 for PTIII distribution with 100-year recurrence interval, the R^2 value is 0.982, which is lower than both 0.990 (Table 5.2.3 for PTIII, 100-year in Region 1) and 0.991 (Table 5.2.3 for PTIII, 100-year in Region 7). Therefore, for region 1 it is also better to use the equation derived from Knipe and Rao (2004), which contains eight regions.

5.3. Combination of GLS regional regression and L-moment method.

To use the regional L-moment method, the first moment (the mean annual peak flow) is the statistic of interest. However, this is usually not available for an ungauged location. In such situations, GLS regression is one approach to obtain the quantile floods for various recurrence intervals. However, the hydrological or geographical information may be combined with the L-moment method simply by developing equations for the mean (or logarithm of mean annual) flows with the GLS method. At each site, the mean and the mean of logarithms are calculated from the data. The mean of the logarithms of the annual maximum flows are calculated because it is needed in LPIII.

Table 5.2.6. GLS regression coefficients of mean and logmean annual peak flow for Region 1 and Region 5 derived by Srinivas and Rao (2003).

(a) Mean peak flow

	Parameter	a	b	c	d	R^2
Region1 + Region7	Model I	89.4602	0.6951			0.983
	Model II	9.5021	0.8995	0.7122		0.991
	Model III	40.9509	0.8911	0.5377	-0.7661	0.995
Region5 + Region8	Model I	49.2186	0.7021			0.823
	Model II	44.3388	0.7132	0.0387		0.818
	Model III	130.5746	0.7560	0.0413	-0.7623	0.882

(b) Log-mean peak flow

	Parameter	a	b	c	d	R^2
Region1 + Region7	Model I	4.5137	0.0985			0.918
	Model II	3.5134	0.1191	0.0883		0.924
	Model III	4.1319	0.1189	0.0729	-0.0942	0.931
Region5 + Region8	Model I	4.4142	0.0989			0.844
	Model II	3.9897	0.1098	0.0340		0.841
	Model III	4.5339	0.1158	0.0447	-0.0975	0.920

The GLS regression equation is constructed by using the drainage area, slope and wet area percentage. Three models are constructed. Model I is based only on the area, which is $Q_{mean}=aA^b$, where Q_{mean} is the estimated mean flow, A is drainage area, and *a, b* are GLS regression coefficients. Similarly, Model II considers area and slope, which is $Q_{mean}=aA^bS^c$, and Model III considers area, slope and wet area percentage, which is $Q_{mean}=aA^bS^c(1+W\%)^d$. The coefficients for each region and each model are listed in Table 5.3.1. For the logarithms of peak flows, the results are listed in Table 5.3.1. For the logarithms of peak flows, the unit for drainage area is square miles, slope is percentage, wet area is percentage and if the regressed value is *Q*, then the quantile flow is *exp*(*Q*) which is in unit of cubic feet per second (cfs).

Table 5.2.7. GLS Regression coefficients of PTIII flood quantile estimates for merged area.

PT3 GLS regional regression for (Region 1 + Region 7)

T (years)	Parameter	a	b	c	d	R^2
	Model I	77.039	0.699			0.983
2	Model II	9.567	0.889	0.663		0.989
	Model III	39.963	0.880	0.490	-0.744	0.992
	Model I	125.150	0.689			0.982
5	Model II	14.441	0.886	0.686		0.989
	Model III	65.052	0.877	0.505	-0.784	0.994
	Model I	158.127	0.686			0.979
10	Model II	17.127	0.889	0.707		0.987
	Model III	79.761	0.879	0.521	-0.802	0.993
	Model I	189.808	0.684			0.974
20	Model II	19.471	0.891	0.725		0.984
	Model III	92.791	0.882	0.537	-0.815	0.991
	Model I	199.830	0.684			0.972
25	Model II	20.179	0.892	0.730		0.983
	Model III	96.739	0.882	0.541	-0.818	0.990
	Model I	230.586	0.682			0.967
50	Model II	22.275	0.895	0.744		0.978
	Model III	108.428	0.885	0.554	-0.828	0.986
	Model I	260.939	0.681			0.961
100	Model II	24.253	0.897	0.757		0.973
	Model III	119.424	0.888	0.566	-0.835	0.982
	Model I	291.035	0.680			0.954
200	Model II	26.144	0.899	0.768		0.968
	Model III	129.886	0.890	0.577	-0.841	0.977

PT3 GLS regional regression for (Region 5 + Region 8)

T (years)	Parameter	a	b	c	d	R^2
	Model I	45.243	0.707			0.817
2	Model II	39.260	0.723	0.052		0.810
	Model III	117.116	0.765	0.053	-0.767	0.880
	Model I	65.254	0.700			0.813
5	Model II	65.315	0.699	0.001		0.813
	Model III	187.308	0.743	0.008	-0.753	0.879
	Model I	78.017	0.696			0.810
10	Model II	83.033	0.689	-0.022		0.813
	Model III	232.255	0.734	-0.008	-0.747	0.876
	Model I	89.698	0.693			0.807
20	Model II	100.009	0.681	-0.038		0.812
	Model III	273.589	0.727	-0.019	-0.742	0.872
	Model I	93.287	0.692			0.805
25	Model II	105.372	0.678	-0.043		0.812
	Model III	286.342	0.726	-0.022	-0.741	0.870
	Model I	104.015	0.690			0.800
50	Model II	121.813	0.672	-0.056		0.809
	Model III	324.647	0.721	-0.031	-0.736	0.864
	Model I	114.215	0.688			0.795
100	Model II	138.017	0.667	-0.067		0.804
	Model III	361.380	0.717	-0.038	-0.731	0.856
	Model I	123.992	0.687			0.788
200	Model II	154.068	0.663	-0.076		0.799
	Model III	396.924	0.714	-0.045	-0.726	0.849

Table 5.2.8. GLS Regression coefficients of GEV flood quantile estimates for merged area.

GEV GLS regional regression for (Region 1 + Region 7)

T (years)	Parameter	a	b	c	d	R^2
	Model I	72.863	0.704			0.980
2	Model II	10.240	0.882	0.623		0.986
	Model III	41.071	0.874	0.455	-0.723	0.988
	Model I	117.387	0.693			0.982
5	Model II	15.517	0.877	0.643		0.987
	Model III	67.285	0.868	0.466	-0.765	0.991
	Model I	149.787	0.689			0.981
10	Model II	18.356	0.880	0.667		0.988
	Model III	82.611	0.871	0.486	-0.785	0.993
	Model I	183.474	0.686			0.978
20	Model II	20.558	0.885	0.696		0.987
	Model III	95.386	0.876	0.511	-0.802	0.993
	Model I	194.749	0.685			0.977
25	Model II	21.149	0.887	0.706		0.986
	Model III	99.012	0.878	0.521	-0.807	0.992
	Model I	231.436	0.683			0.969
50	Model II	22.647	0.895	0.740		0.980
	Model III	108.840	0.885	0.552	-0.822	0.987
	Model I	271.042	0.681			0.955
100	Model II	23.670	0.903	0.777		0.968
	Model III	116.572	0.894	0.587	-0.837	0.977
	Model I	314.032	0.680			0.933
200	Model II	24.260	0.913	0.816		0.950
	Model III	122.302	0.904	0.625	-0.852	0.960

GEV GLS regional regression for (Region 5 + Region 8)

T (years)	Parameter	a	b	c	d	R^2
	Model I	44.189	0.711			0.817
2	Model II	37.975	0.727	0.056		0.810
	Model III	112.978	0.769	0.056	-0.763	0.880
	Model I	63.126	0.704			0.811
5	Model II	62.460	0.705	0.005		0.811
	Model III	179.037	0.748	0.011	-0.749	0.880
	Model I	75.581	0.700			0.810
10	Model II	79.713	0.694	-0.018		0.813
	Model III	222.955	0.739	-0.006	-0.744	0.878
	Model I	87.595	0.697			0.808
20	Model II	97.253	0.685	-0.037		0.813
	Model III	266.073	0.731	-0.020	-0.739	0.874
	Model I	91.430	0.695			0.806
25	Model II	103.052	0.682	-0.042		0.812
	Model III	280.066	0.729	-0.023	-0.737	0.872
	Model I	103.347	0.692			0.799
50	Model II	121.725	0.674	-0.058		0.808
	Model III	324.438	0.721	-0.035	-0.732	0.861
	Model I	115.359	0.688			0.787
100	Model II	141.619	0.665	-0.072		0.798
	Model III	370.810	0.714	-0.047	-0.726	0.845
	Model I	127.550	0.684			0.770
200	Model II	162.989	0.657	-0.086		0.782
	Model III	419.860	0.706	-0.059	-0.720	0.822

Table 5.2.9. GLS Regression coefficients of LPIII flood quantile estimates for merged area.

LP3 GLS regional regression for (Region 1 + Region 7)

T (years)	Parameter	a	b	c	d	R^2
	Model I	75.280	0.702			0.983
2	Model II	8.581	0.900	0.690		0.990
	Model III	37.428	0.891	0.512	-0.769	0.993
	Model I	123.321	0.688			0.981
5	Model II	13.966	0.887	0.692		0.988
	Model III	66.194	0.878	0.505	-0.814	0.993
	Model I	160.587	0.680			0.978
10	Model II	16.544	0.887	0.722		0.987
	Model III	82.744	0.877	0.529	-0.842	0.993
	Model I	200.511	0.672			0.972
20	Model II	18.207	0.891	0.763		0.982
	Model III	95.758	0.881	0.564	-0.870	0.990
	Model I	214.075	0.670			0.968
25	Model II	18.584	0.892	0.777		0.980
	Model III	99.294	0.883	0.577	-0.879	0.988
	Model I	258.755	0.663			0.955
50	Model II	19.340	0.899	0.825		0.969
	Model III	108.432	0.889	0.620	-0.906	0.979
	Model I	307.767	0.656			0.933
100	Model II	19.582	0.906	0.877		0.950
	Model III	115.116	0.897	0.667	-0.933	0.963
	Model I	361.705	0.649			0.901
200	Model II	19.433	0.915	0.931		0.922
	Model III	119.741	0.905	0.717	-0.961	0.937

LP3 GLS regional regression for (Region 5 + Region 8)

T (years)	Parameter	a	b	c	d	R^2
	Model I	46.618	0.706			0.823
2	Model II	39.250	0.724	0.063		0.815
	Model III	110.705	0.764	0.063	-0.726	0.886
	Model I	66.826	0.699			0.815
5	Model II	66.022	0.700	0.005		0.815
	Model III	177.884	0.741	0.012	-0.708	0.885
	Model I	79.279	0.695			0.815
10	Model II	83.580	0.689	-0.018		0.818
	Model III	218.152	0.731	-0.005	-0.696	0.885
	Model I	90.592	0.692			0.815
20	Model II	100.006	0.681	-0.034		0.820
	Model III	253.678	0.724	-0.017	-0.686	0.884
	Model I	94.065	0.691			0.815
25	Model II	105.134	0.679	-0.039		0.820
	Model III	264.400	0.722	-0.021	-0.682	0.882
	Model I	104.475	0.689			0.812
50	Model II	120.747	0.673	-0.051		0.819
	Model III	296.133	0.717	-0.029	-0.672	0.876
	Model I	114.441	0.686			0.806
100	Model II	136.061	0.667	-0.060		0.814
	Model III	326.092	0.712	-0.037	-0.661	0.866
	Model I	124.093	0.684			0.797
200	Model II	151.280	0.662	-0.069		0.806
	Model III	354.926	0.707	-0.045	-0.650	0.853

The 95% confidence interval are calculated for all the measurements in each region. The results for annual peak flow is shown in Figure 5.3.1. There are four series of data plotted: observed mean annual peak flow, GLS-regressed mean annual peak flow, 95% confidence interval upper limit and 95% confidence limits. Simple linear fitting is applied for each of them in order to show the trend of each data set. The observed and estimated means are very close to each other. Except for Region 4 and Region 6, the two trend lines in the other six regions are almost overlapping. Figure 5.3.2 shows the histograms of the distribution of the drainage area. In Regions 4 and 6 most of the drainage areas are greater than 100 square miles and there are few small drainage areas with high variability flow than in the other regions.

Table 5.3.1. GLS regional regression for mean annual peak flows.

	Parameter	***a***	***b***	***c***	***d***	***R^2***
Region 1	**Model I**	173.4825	0.6309			0.9881
	Model II	37.3649	0.7779	0.4602		0.9951
	Model III	42.5792	0.7763	0.4406	-0.0627	0.9953
Region 2	**Model I**	299.5047	0.5894			0.961
	Model II	32.1958	0.8256	0.5314		0.9846
	Model III	81.8451	0.7822	0.3841	-0.3468	0.9793
Region 3	**Model I**	226.0649	0.7443			0.8972
	Model II	41.8274	0.9055	0.4197		0.9335
	Model III	40.0037	0.9065	0.4231	0.0339	0.9333
Region 4	**Model I**	145.918	0.6712			0.9091
	Model II	21.6708	0.8161	0.6508		0.937
	Model III	34.0161	0.8328	0.6172	-0.4202	0.9524
Region 5	**Model I**	50.2466	0.7005			0.9556
	Model II	19.3921	0.8038	0.3927		0.9709
	Model III	25.7894	0.7971	0.3779	-0.1358	0.971
Region 6	**Model I**	94.6589	0.5154			0.9026
	Model II	12.9361	0.7811	0.6608		0.9352
	Model III	11.5316	0.7621	0.6054	0.1367	0.9378
Region 7	**Model I**	53.8668	0.7342			0.9845
	Model II	8.2252	0.8914	0.6306		0.9765
	Model III	52.7069	0.896	0.539	-1.1561	0.9892
Region 8	**Model I**	49.8341	0.6986			0.7204
	Model II	96.4878	0.6277	-0.2204		0.7637
	Model III	143.2239	0.8076	0.0277	-0.9783	0.7449

Table 5.3.2. GLS regional regression for mean of logarithms of annual peak flows.

	Parameter	***a***	***b***	***c***	***d***	***R^2***
	Model I	5.3345	0.0779			0.9645
Region 1	**Model II**	5.0811	0.0824	0.0152		0.9639
	Model III	4.7038	0.0857	0.033	0.0223	0.9645
	Model I	5.7749	0.0728			0.894
Region 2	**Model II**	4.6272	0.0944	0.0557		0.9063
	Model III	5.896	0.0811	0.0178	-0.0796	0.9101
	Model I	5.2372	0.1085			0.9546
Region 3	**Model II**	3.6401	0.1458	0.087		0.9694
	Model III	3.5076	0.1467	0.0896	0.0287	0.9703
	Model I	5.2189	0.0874			0.9264
Region 4	**Model II**	4.1978	0.1041	0.073		0.9517
	Model III	4.5072	0.1075	0.0679	-0.0693	0.9583
	Model I	4.7561	0.0868			0.9472
Region 5	**Model II**	3.7632	0.1137	0.079		0.9603
	Model III	3.9458	0.1149	0.08	-0.0316	0.9677
	Model I	4.8775	0.0708			0.881
Region 6	**Model II**	3.4816	0.1175	0.0924		0.9253
	Model III	3.4933	0.1174	0.0899	-0.0017	0.9242
	Model I	3.6987	0.1276			0.9461
Region 7	**Model II**	2.9408	0.1448	0.0874		0.9532
	Model III	3.9225	0.1416	0.0891	-0.1828	0.979
	Model I	4.0622	0.1119			0.7936
Region 8	**Model II**	4.2363	0.1072	-0.013		0.795
	Model III	4.6362	0.1228	0.0269	-0.1297	0.9148

In all regression methods high variations in small drainage areas is not reflected. Also, the 95% confidence intervals are added to the logarithms of mean annual peak flow and these are shown in Figure 5.3.3. Regions 3, 5, 7 and 8 have perfect match and Regions 1, 2, 4 and 6 are not as close.

By the using the GLS regression, we can obtain the mean (or logmean) for the location of interest. Furthermore, we need to use the information of regional normalized quantile $\hat{x}_T$, which is presetned in Chapter 4. It is based on regional L-moment method and ends up with a parameter $\hat{x}_T$ for a specific recurrence interval T years. Once the first L-moment λ_k (mean) at site

k is available, it is multiplied by $\hat{x}_T$. In other words, the estimated T-year quantile flood is $\hat{x}_T^k$ (= $\lambda_k \cdot \hat{x}_T$ in unit cfs). For LPIII distribution, the first L-moment should be replaced by the mean of the logarithms of the flows λ_k', and the estimated T-year quantile flood is $\hat{x}_T^k$ ($= 10^{\lambda_1 \hat{x}_T}$) in unit cfs).

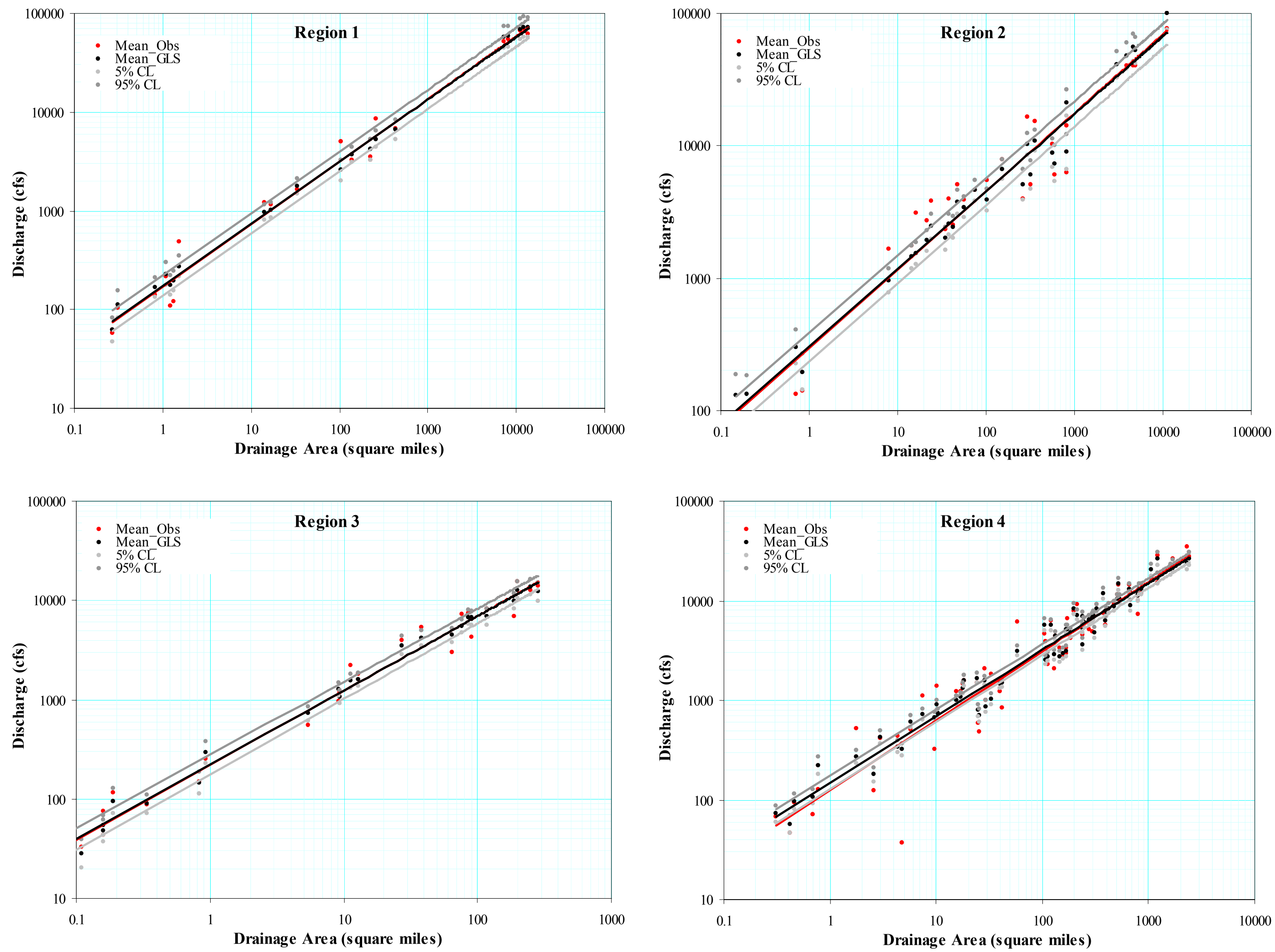

Figure 5.3.1 (a). At-site observed mean annual peak flows compared with GLS regression results and the 95% confidence upper and lower limits for Region 1~4 .

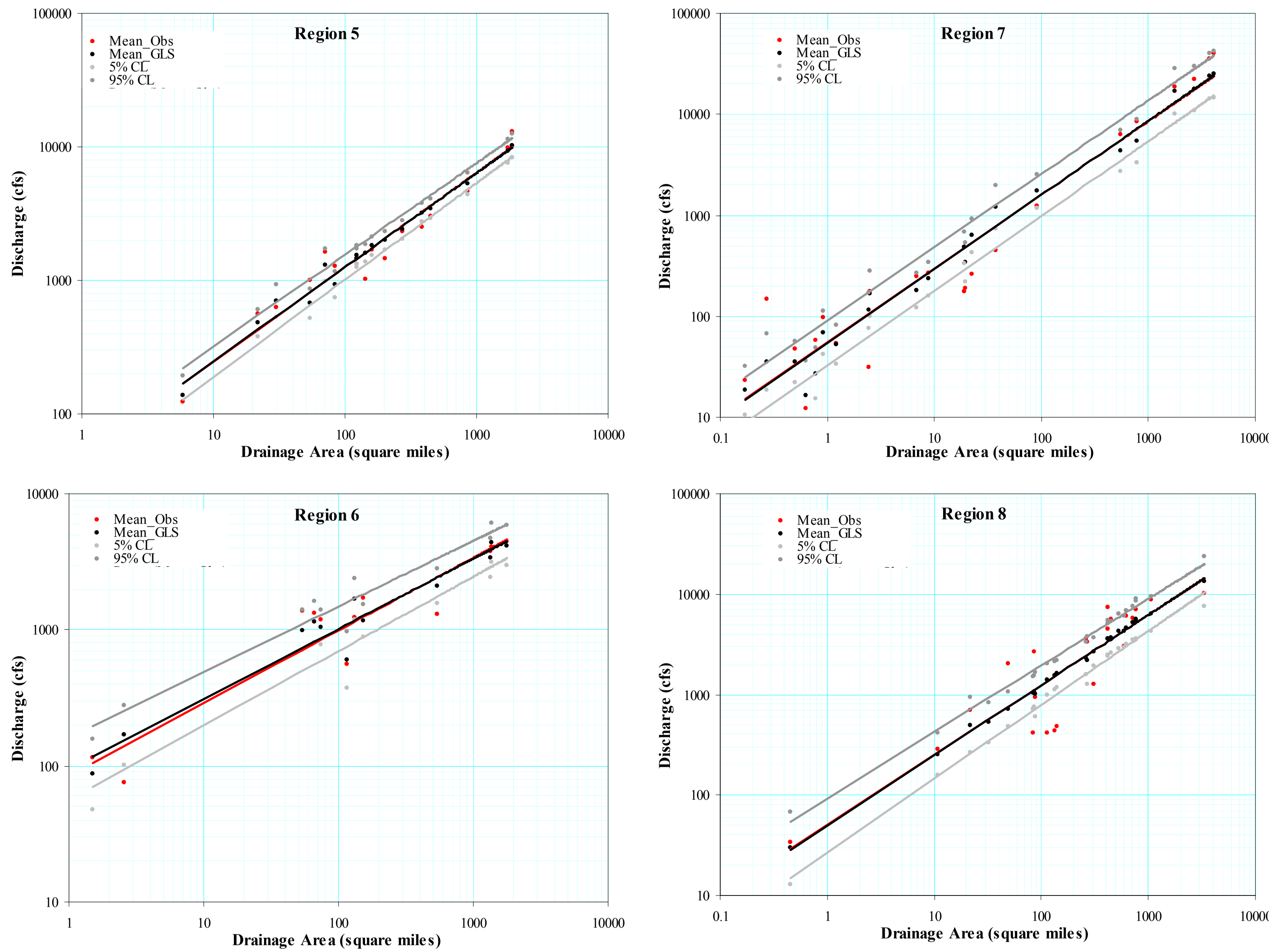

Figure 5.3.1(b). At-site observed mean annual peak flows compared with GLS regression results and the 95% confidence upper and lower limits for Region 5~8 .

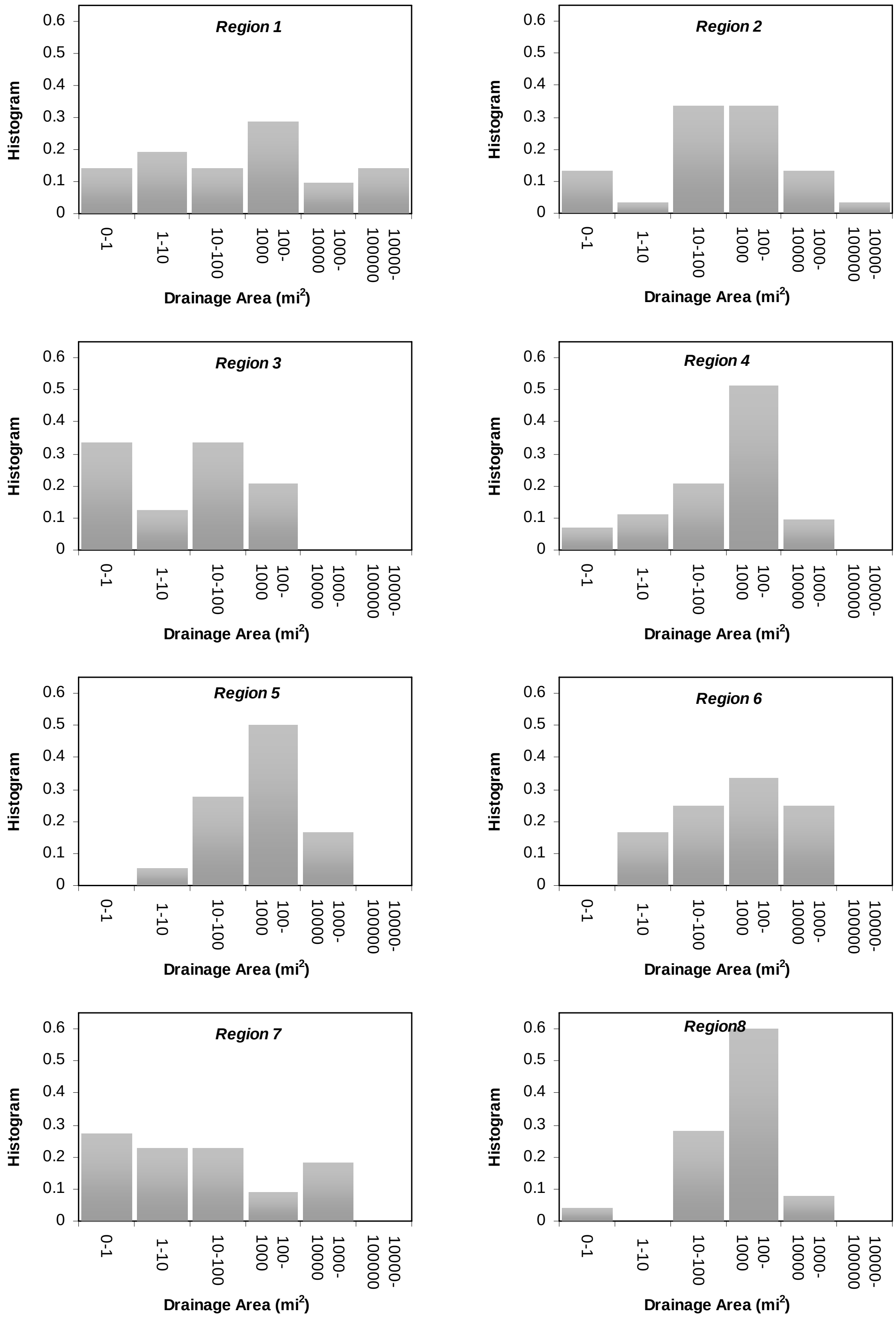

Figure 5.3.2. Histographs of drainage areas for each region.

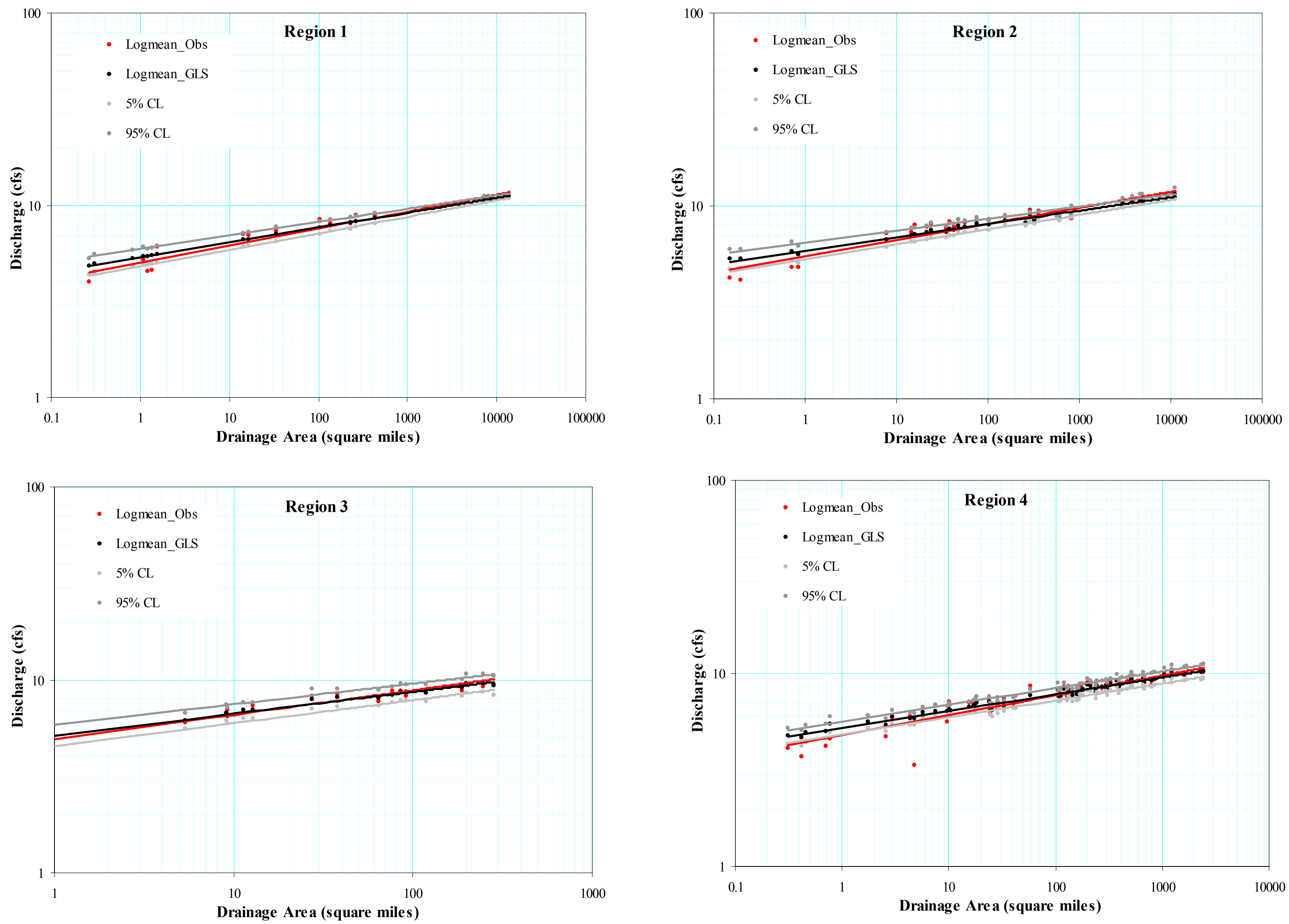

Figure 5.3.3 (a). At-site logarithms of mean annual peak flows compared with GLS regression results and the 95% confidence upper and lower limits for Region 1~4.

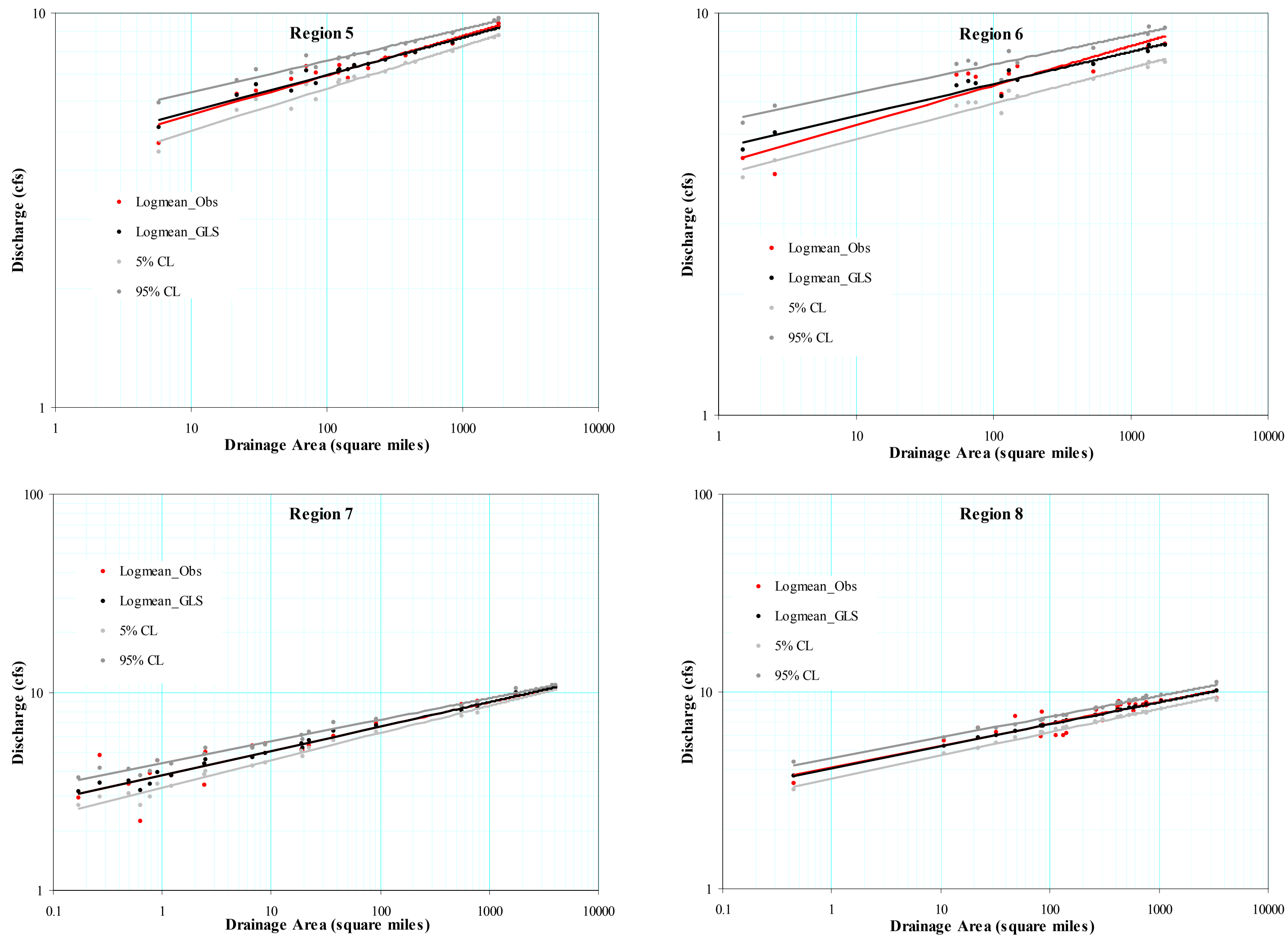

Figure 5.3.3 (b). At-site logarithm of mean annual peak flows compared with GLS regression results and the 95% confidence upper and lower limits for Region 5~8.

VI. Comparative Analysis

There are three methods by which the equations derived in this study can be used. In the first method, the normalized regional quantiles can be used with the observed mean value of annual maximum flows at a site to compute the flood frequencies at specific recurrence intervals. In the second method, the mean values are estimated by using hydrological characteristics at a site, such as the watershed area and stream slope and these mean values are used with the normalized regional quantiles to estimate the flood magnitudes. In the third method, the equations for quantiles derived by the GLS method are used directly to obtain flood magnitudes.

The accuracies of these methods differ. The first method should give the smallest error because only the normalized quantiles are used with the observed means of annual maximum flows. In the second method, the mean peak flows are estimated by using the regression relationships along with the normalized regional quantiles. The third method is based entirely on regression relationships. The errors associated with the second and third methods may depend on the regions to which the equations apply.

Before these procedures, especially the second and third methods, are recommended for use the errors associated with their use must be quantified. This is achieved by using split sample tests. The data from each region is divided into two parts, each containing 75% and 25% of the data. The 25% of the data is selected to reflect the distribution of the watershed areas so that the test involves a range of areas. The 75% of the data is used to estimate the parameters of the equations used for that method. The data in 25% of the sample are used to estimate the floods and these are compared to the observed quantiles. This analysis would enable us to estimate the errors. The procedures used in the three methods are schematically shown in Figure 6.1.1.

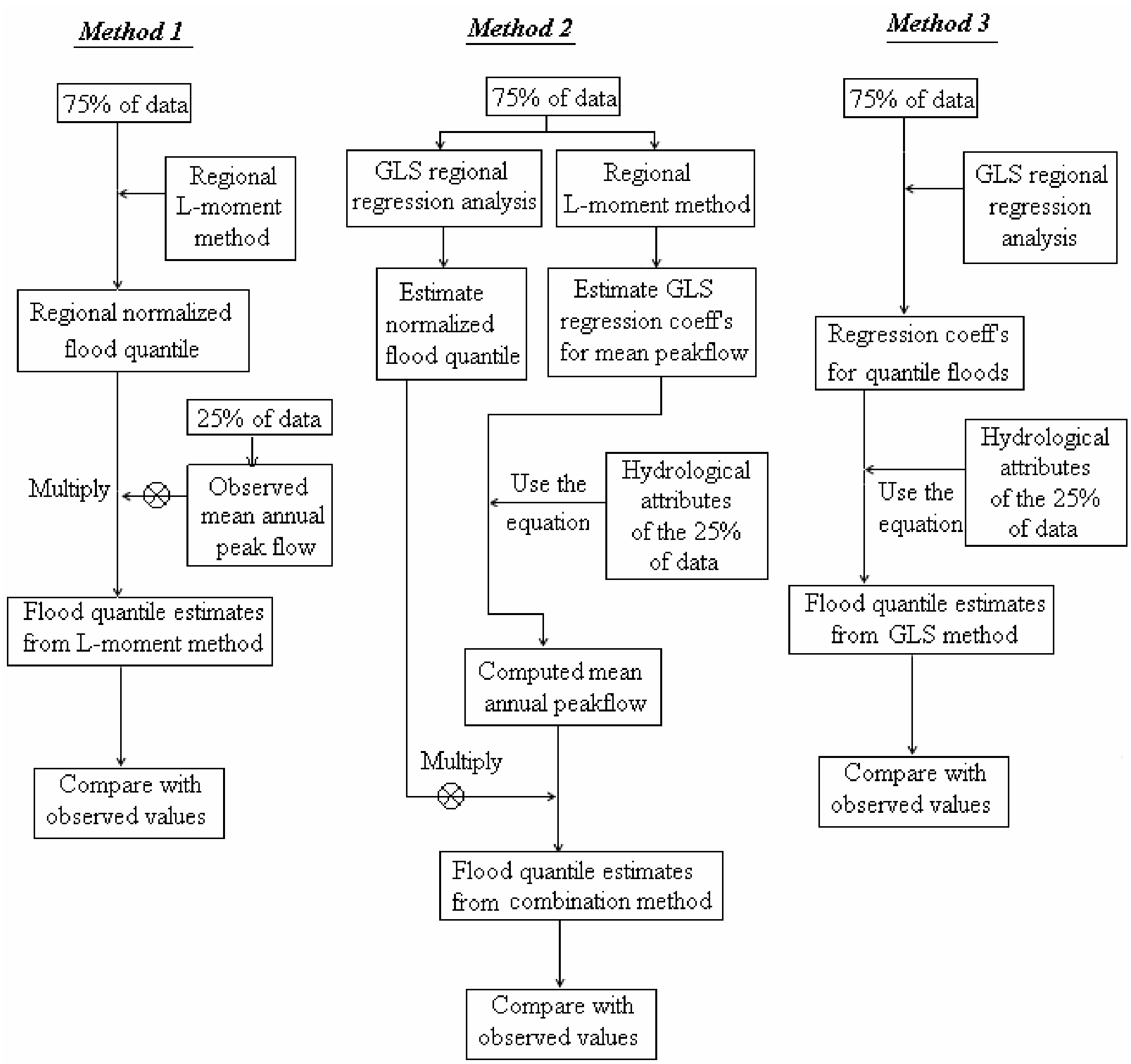

Figure 6.1.1. Flowchart of three comparison methods.

6.1. Split sample test for first method

The mean annual peak flow in this method is calculated from 25% of the data. The observed mean annual peak flow computed from observed data is multiplied by the normalized regional quantiles and compared to at-site quantiles of observed data. The measure to evaluate the error is the variance calculated by Eqn.3.2.4. The total number of stations in each region are

21, 30, 24, 72, 18, 12, 22 and 25; hence, 25% of the observations are 5, 8, 6, 18, 5, 3, 6 and 6, respectively, and they are selected based on their drainage areas. This analysis is valid only in homogeneous or possibly homogeneous regions. The data from region 6, which is heterogeneous, is included only for the sake of completion.

The validated results are plotted in Figure 6.1.2 and the variances are shown in Figure 6.1.3. The recurrence intervals of 50, 100 and 200 years are shown as examples. The *x*-axis has the at-site quantile estimates and *y*-axis is quantile estimates calculated from the regional L-moment method. These results show that Gamma distribution mostly overestimates the floods while LPIII distribution underestimates them. PTIII, LNIII and GEV produce consistent and better estimates. Also, results from Figure 6.1.3 indicate that Gamma and LPIII distribution are not good candidates for regional flood estimation. Results from PTIII are better than the others and hence PTIII is the best distribution according to this test. The optimal probability distributions for regional flood estimates from this test are summarized in Table 6.1.1. It is a two step selection. The results show that PTIII is the favored distribution to Regions 3, 4, 5 and 7, LP3 is preferred for Regions 1 and 8, and GEV is good for Regions 2 and 6. PTIII is acceptable for regions 1, 2, 6 and 8 as the second best distribution.

All these homogeneous or possible homogeneous regions (1, 2, 3, 4, 5, 7, and 8) have estimates close to 45 degree lines; the estimates in region 6, which is the heterogeneous region, is further away the 45 degree line. It indicates that flood estimates from a heterogeneous region are not accurate. Hence, once a region fails homogeneity tests, stable regional estimation is not possible.

As the recurrence interval increases, the estimates deviate more from the 45-degree line, especially for 200 year floods. This behavior is caused by extrapolation errors. Also, we find that the prediction result is less stable for low flows which are mostly from small drainage areas. The hydrological responses from small watersheds are easily affected by local events. Higher value of streamflow corresponds to larger drainage areas which follow the regional properties well and are less influenced by local events. Similar conclusion is seen in Method 2 and Method 3.

Table 6.1.1. Optimal probability distributions for regional flood estimates.

Region No.	Candidate Probability Distributions	Optimal Distributions for Regional Estimates
1	PT3, GM2, LN3, GEV, LP3	LP3, PT3, LN3
2	GEV, LN3, PT3, GM2, GLO	GEV, PT3, LN3
3	GEV, LN3, LP3, GLO, PT3	PT3, LN3, GEV
4	GEV, LN3, LP3, PT3, GM2	PT3, LN3, GEV
5	GEV, LN3, LP3, PT3, GM2	PT3, LN3, GEV
6	LN3, PT3, GM2, GLO, GEV	GEV, PT3, LN3
7	PT3, GM2, LN3, GEV, LP3	PT3, LN3, GEV
8	LP3, GLO, GEV, LN3, PT3	LP3, PT3, LN3

Note: 1). Candidate probability distributions are determined from the mean-square-error of L-moment ratio diagram of the 75% of data, and the order is beginning with the one having the minimum MSE. 2). Optimal distributions for regional estimates are obtained from the variances of L-moment regional estimates from the 25% of data.

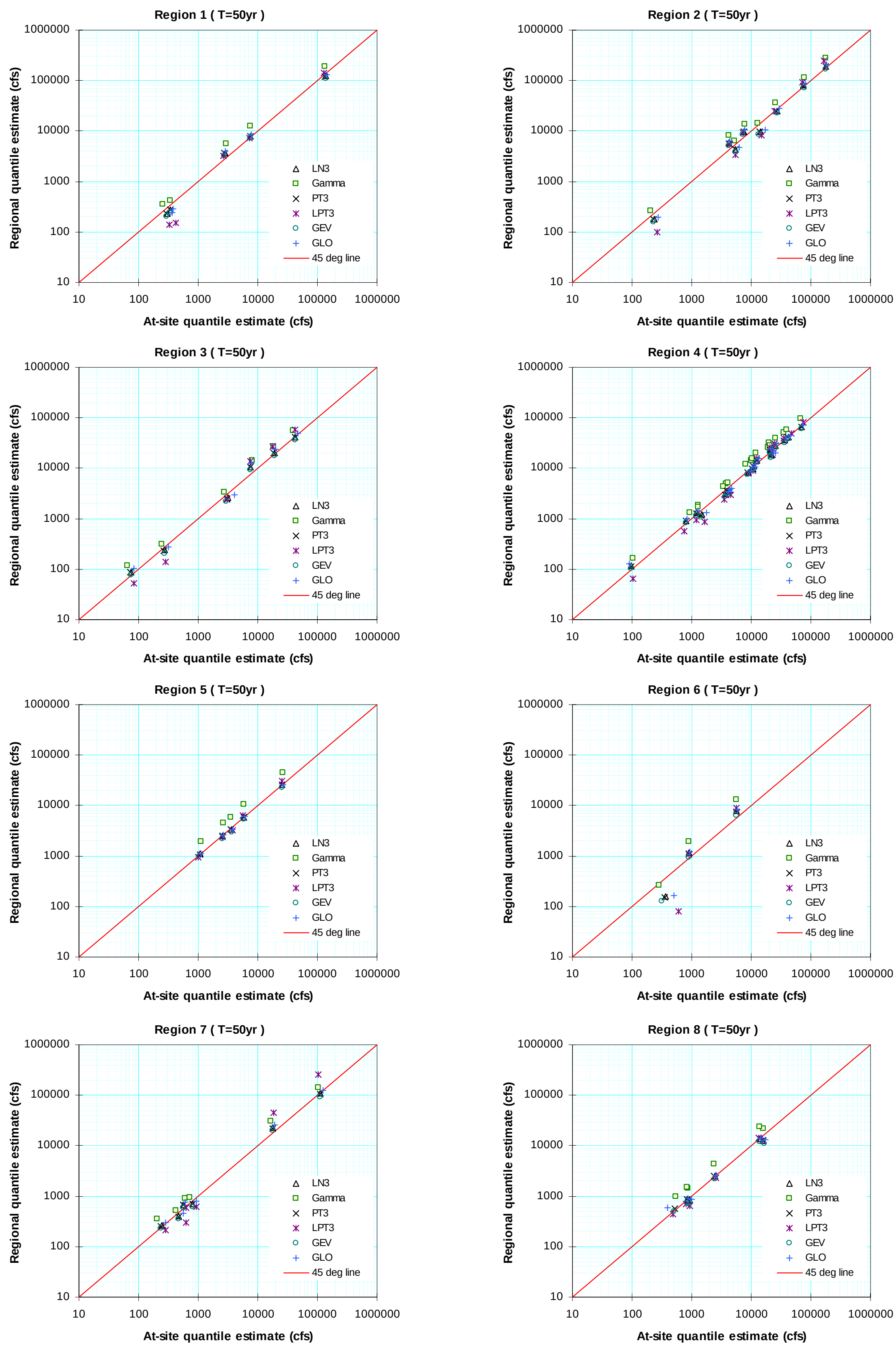

Figure 6.1.2(a) Results of at-site and regional quantile floods from method 1 (T = 50 year).

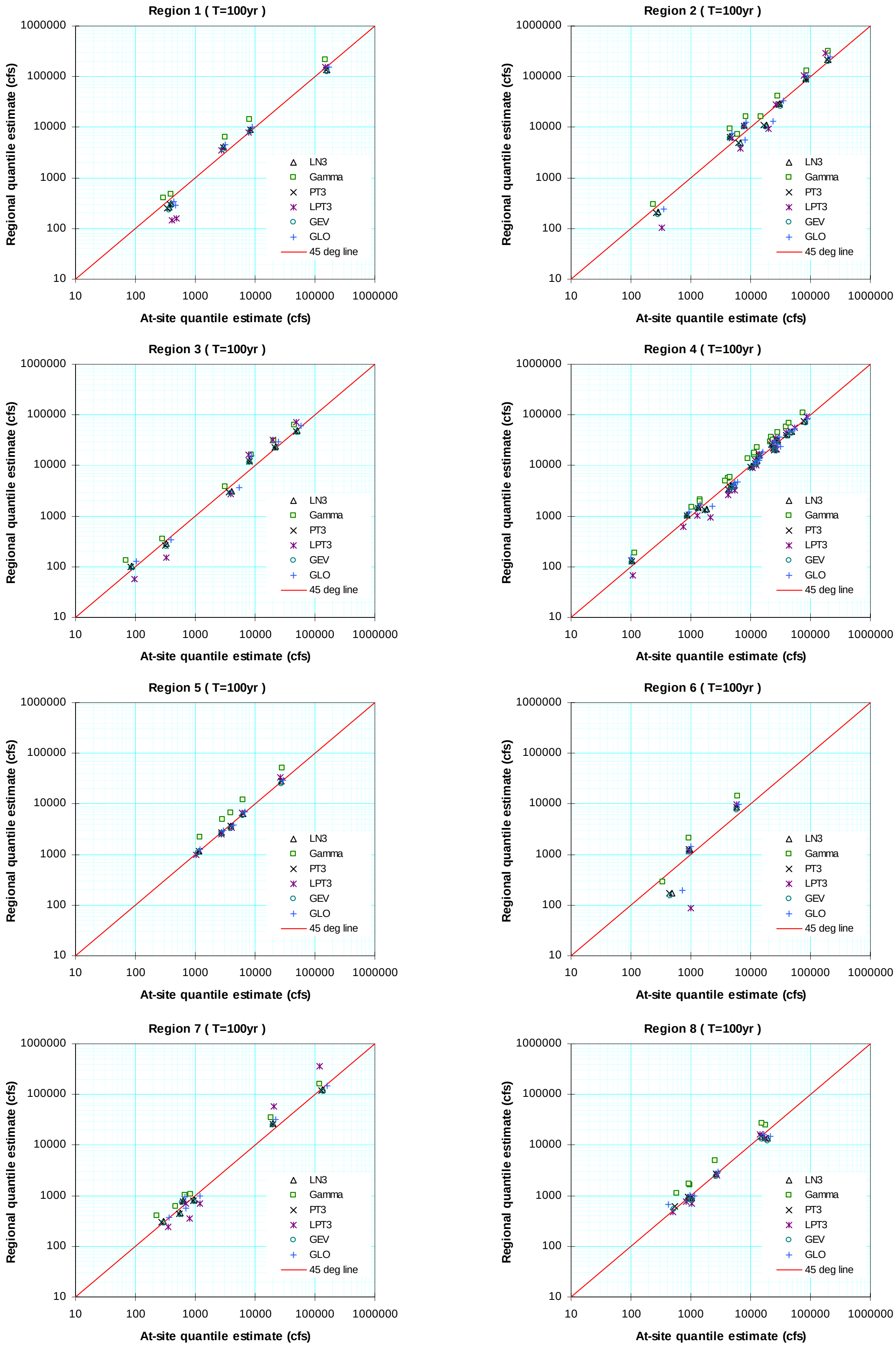

Figure 6.1.2(b) Results of at-site and regional quantile floods from method 1 (T = 100 year).

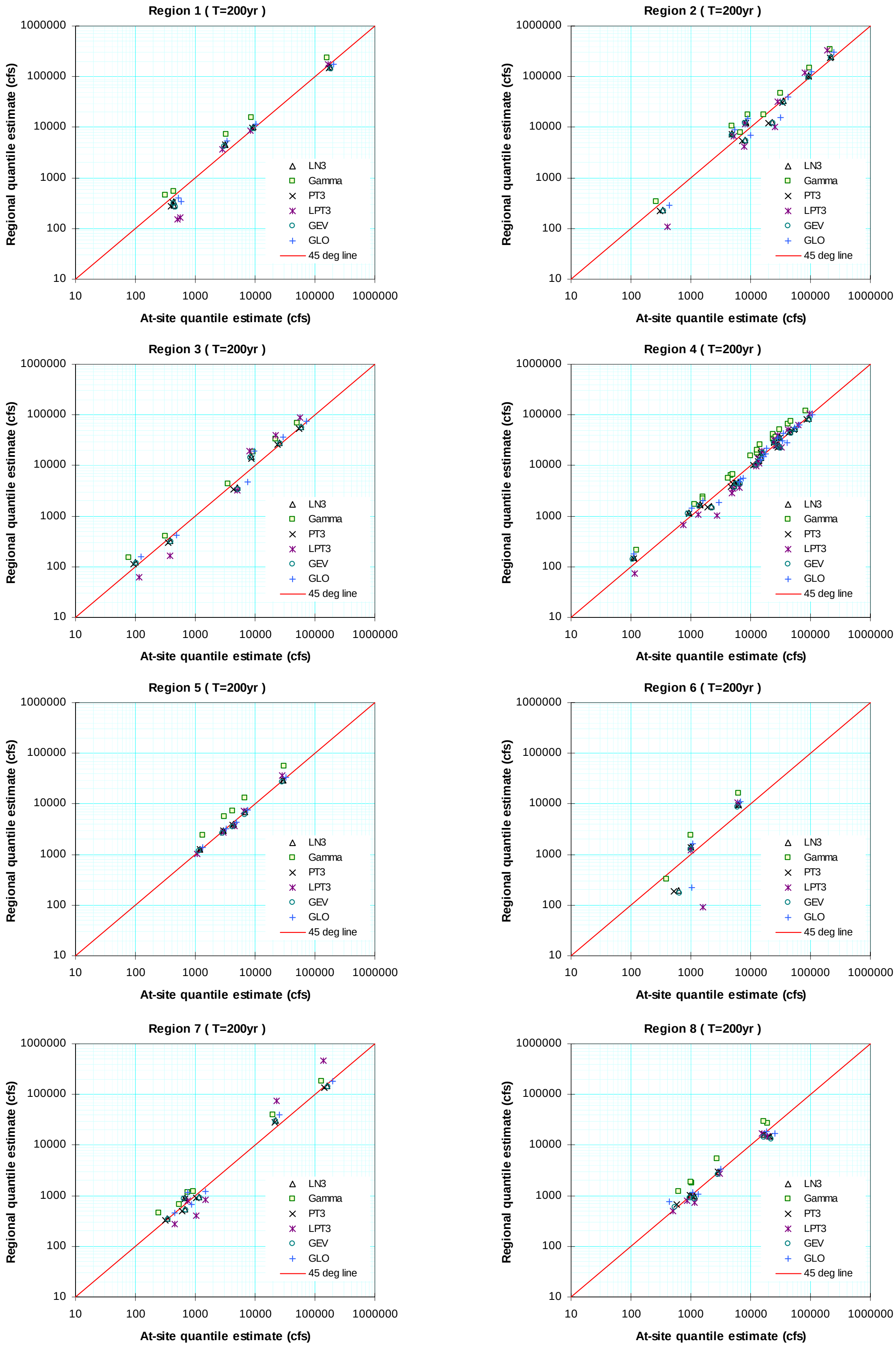

Figure 6.1.2(c) Results of at-site and regional quantile floods from method 1 (T = 200 year).

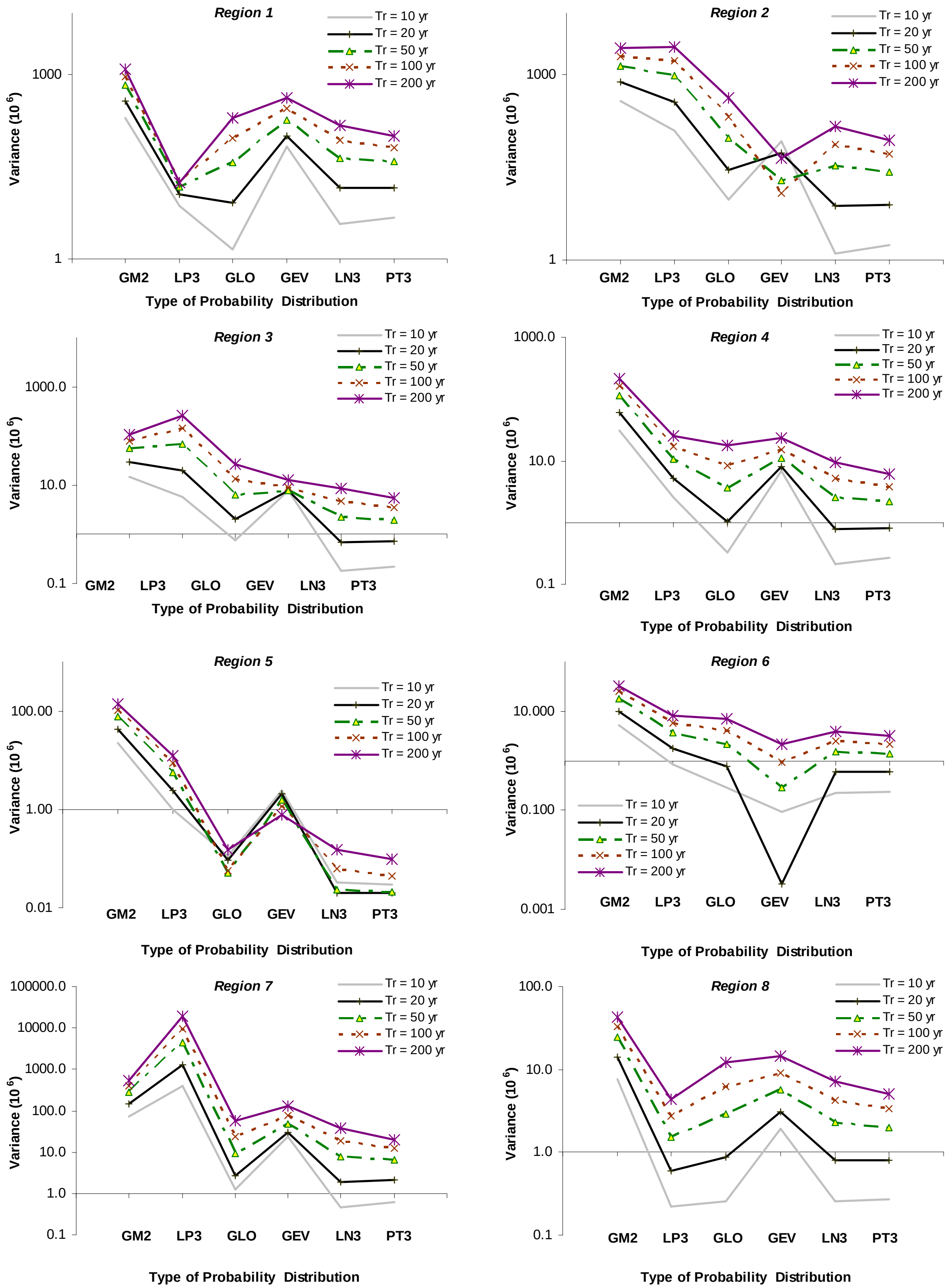

Figure 6.1.3. Variance of the difference between at-site and regional estimates from method 1.

6.2. Split sample test for the second method

The second comparative method is the combination of GLS and L-moment methods. The concepts of regional index flood and GLS regional regression are used to test the prediction accuracy. The normalized regional flood quantile is established from the 75% data. The GLS regional regression is used to estimate the mean annual peakflows from the 75% data.

The at-site and regional flood quantile estimations are calculated by Method 2 by using 25% of data. Results from Figure 6.2.1(a), Region 3 and Region 5 have best accuracy for PTIII distribution. The best estimates for region 1 are from GEV distribution (Figure 6.2.1(b)) and LPIII distribution gives the best estimates for region 3 (Figure 6.2.1(c)). Results from region 6, 7 and 8 are not good; Region 6 and 7 have poor results for small drainage areas. The results in Table 6.2.1 explain these poor results because of the small correlation coefficient between hydrological attributes and qunatile floods. The correlation between drainage area and quantile flood in Region 8 is the poorest one, reflecting the poor results for region 8.

6.3. Split sample test for the third method

To examine the accuracy flood estimated from GLS regional regression, the estimated parameters are directly applied with the hydrological attributes. 75% of data is used to calculate the GLS regional regression in order to obtain the coefficients and exponents (for example, a', b', c', d' from equation $Q_T = a' A^{b'} S^{c'} (1+W\%)^{d'}$), are used to estimate floods from the 25% of data.

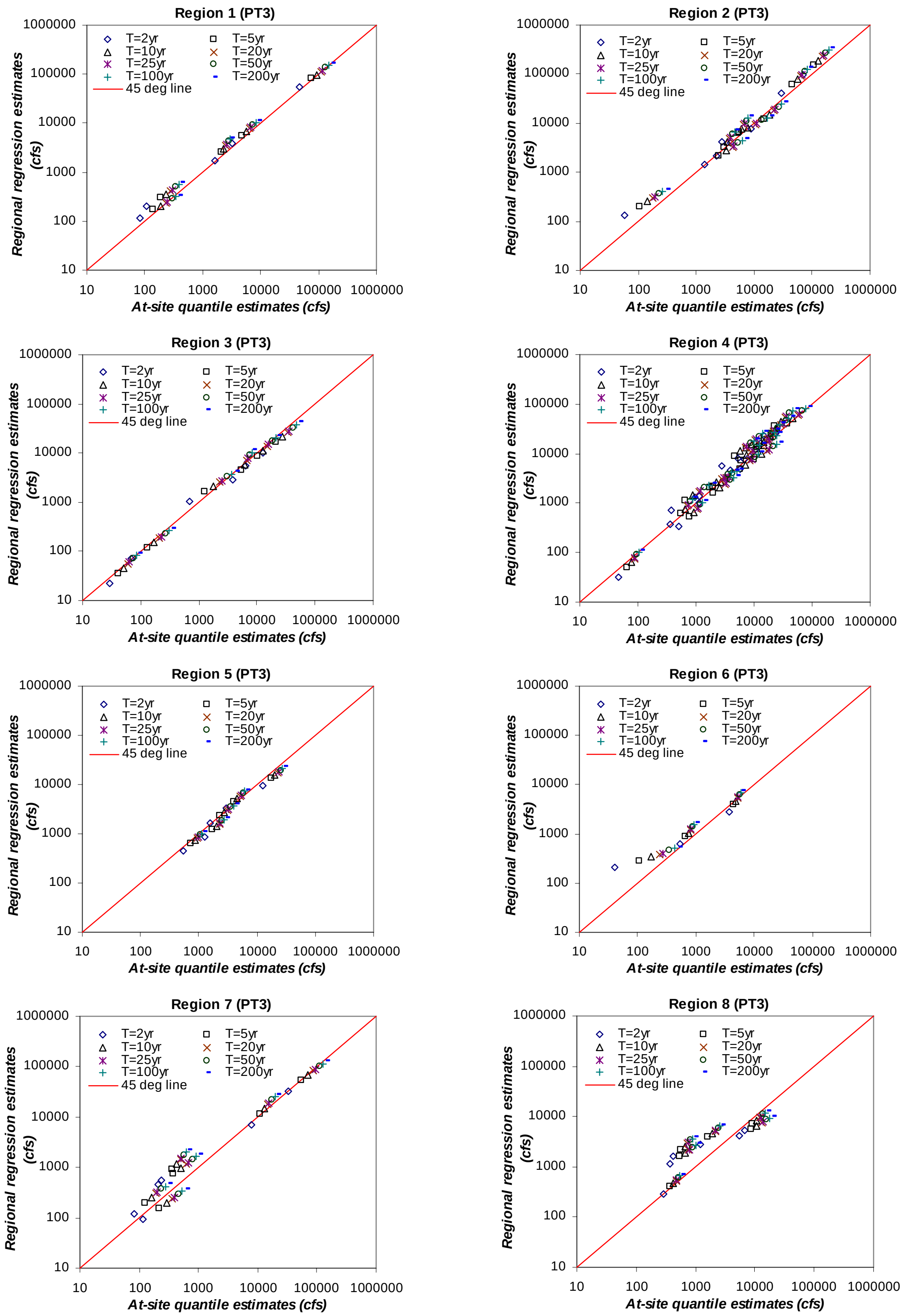

Figure 6.2.1(a). At-site quantile floods and the quantile floods obtained by Method 2 for 25% of the data (PTIII).

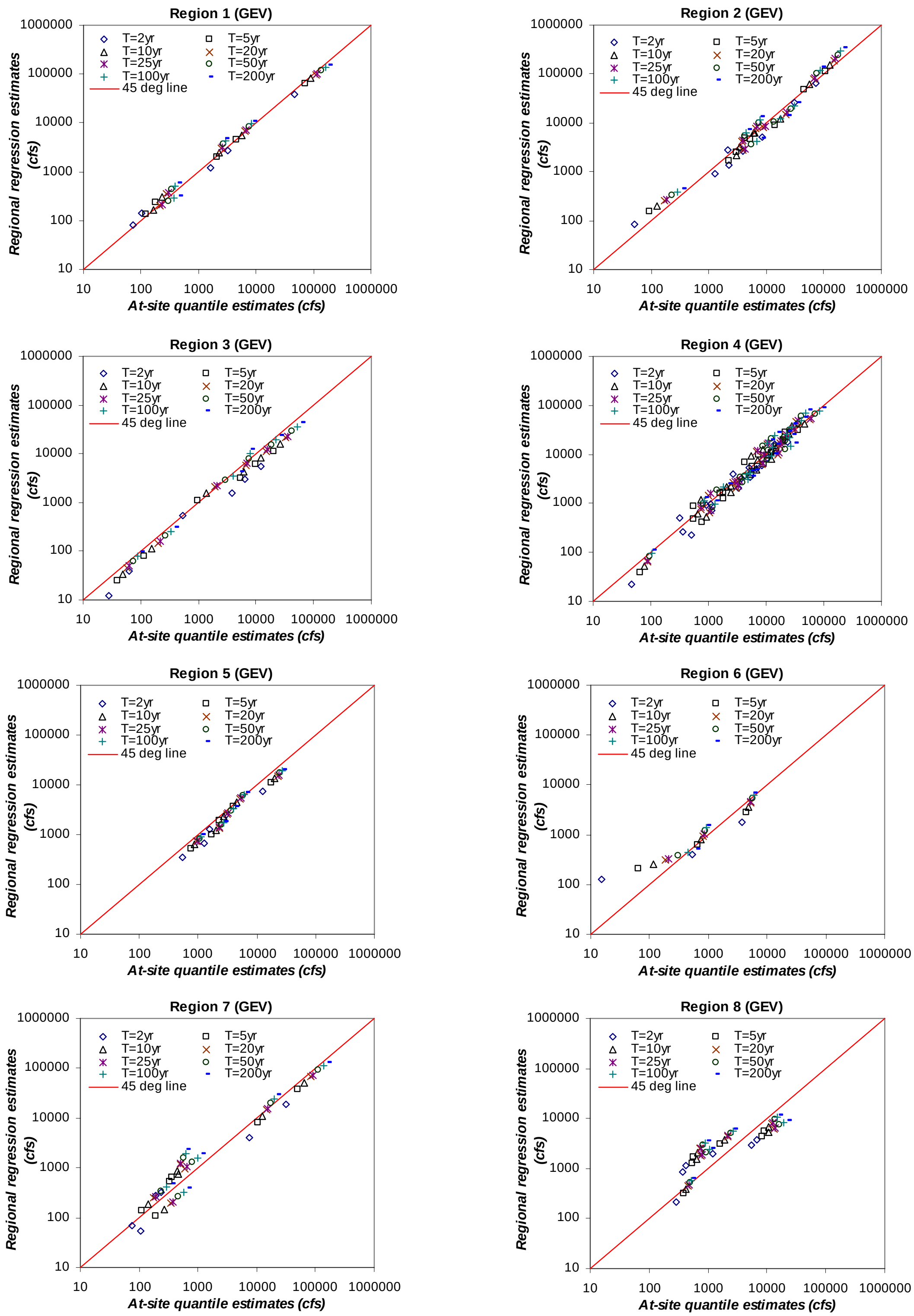

Figure 6.2.1(b). At-site quantile floods and the quantile floods obtained by Method 2 for 25% of the data (GEV).

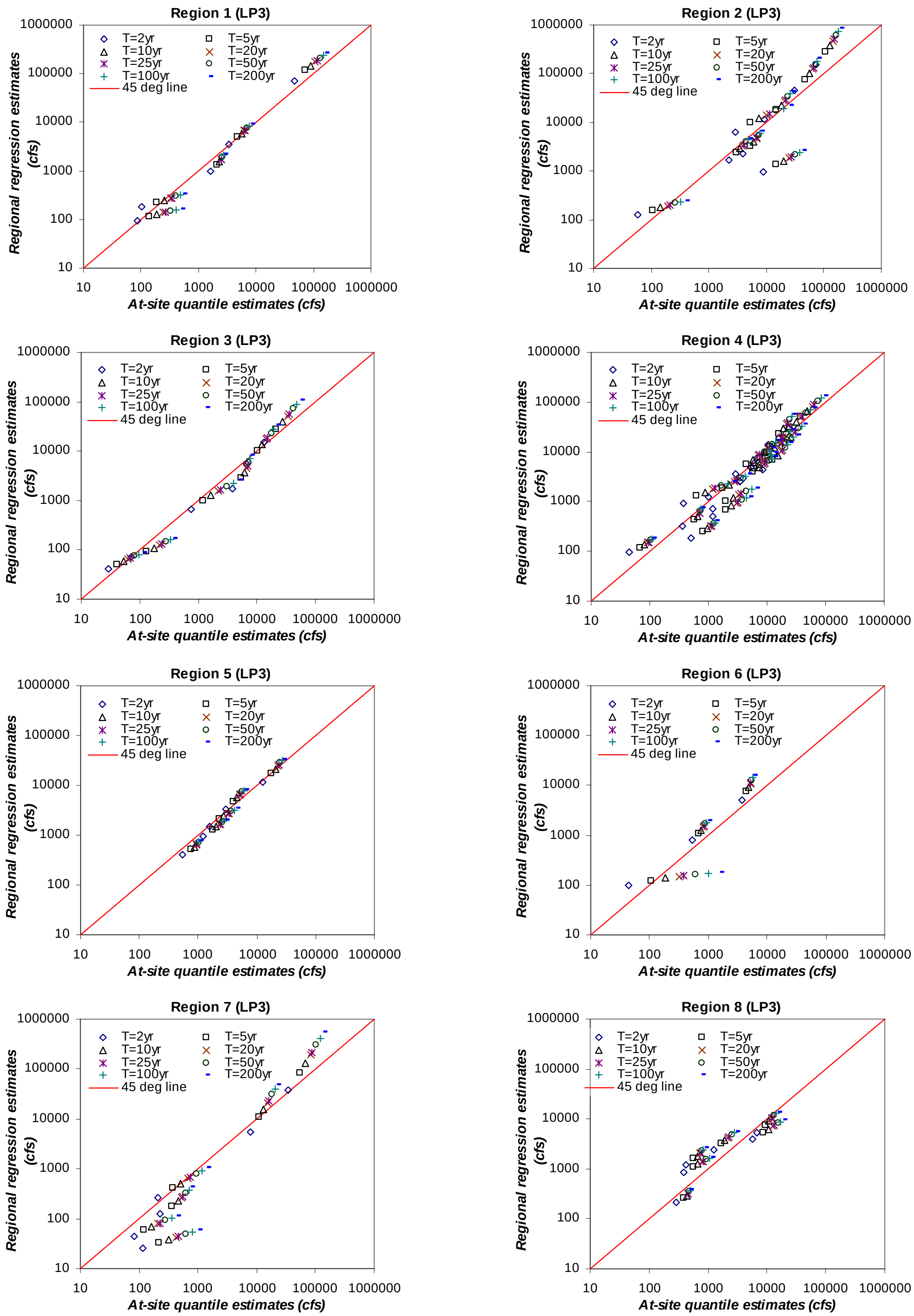

Figure 6.2.1(c). At-site quantile floods and the quantile floods obtained by Method 2 for 25% of the data (LPIII).

The validation results are shown in Figure 6.3.1(a), 6.3.1(b) and 6.3.1(c) for PTIII, GEV, and LPIII distributions, respectively. The y-axis is the value calculated from using the drainage area, slope and percentage wet area of 25% of the data with the GLS regression equation which is built from the 75% data; the x-axis is the at-site quantile flood estimates for these 25% data. Flood estimates for eight recurrence intervals at each site are plotted in the same figure. If they are approaching 45-degree line, it indicates a better capability to predict. Most cases in Figure 6.3.1(a) shows that GLS regression and PTIII quantile floods are in good agreement, except for some outlier points in Region 6, 7 and 8. Similarly, the same situation occurs to the fitting of GEV floods in Figure 6.3.1(b). For LPIII floods (in Figure 6.3.1(c)), besides the outlier in Region 6, 7 and 8, there are more errors in Region 2 than PTIII and GEV.

In summary, Figure 6.3.1 shows the average errors of at-site and regional quantile floods to the third test case (simply considering the GLS regression), and Figure 6.2.1 is to the second test case (combination of regional index flood and GLS regression). The third method does not indicate too many differences among PTIII, GEV and LPIII probability distributions because the result of GLS regression is dominated by the correlation between hydrologic attributes and quantile floods. The second case shows PTIII and GEV having similar response, but LP3 yields worst result. The reason is the same as we have described in Chapter 3 that LPIII is not a good candidate for regional L-moment method of flood estimation. Except for Region 6, the results from either GLS regression or combination method are quite reliable and follow the trend well. Region 7 may have more estimation errors for small drainage areas (less than 1000 square miles) and Region 8 has more error for the drainage areas in the range of 500~5000 square miles.

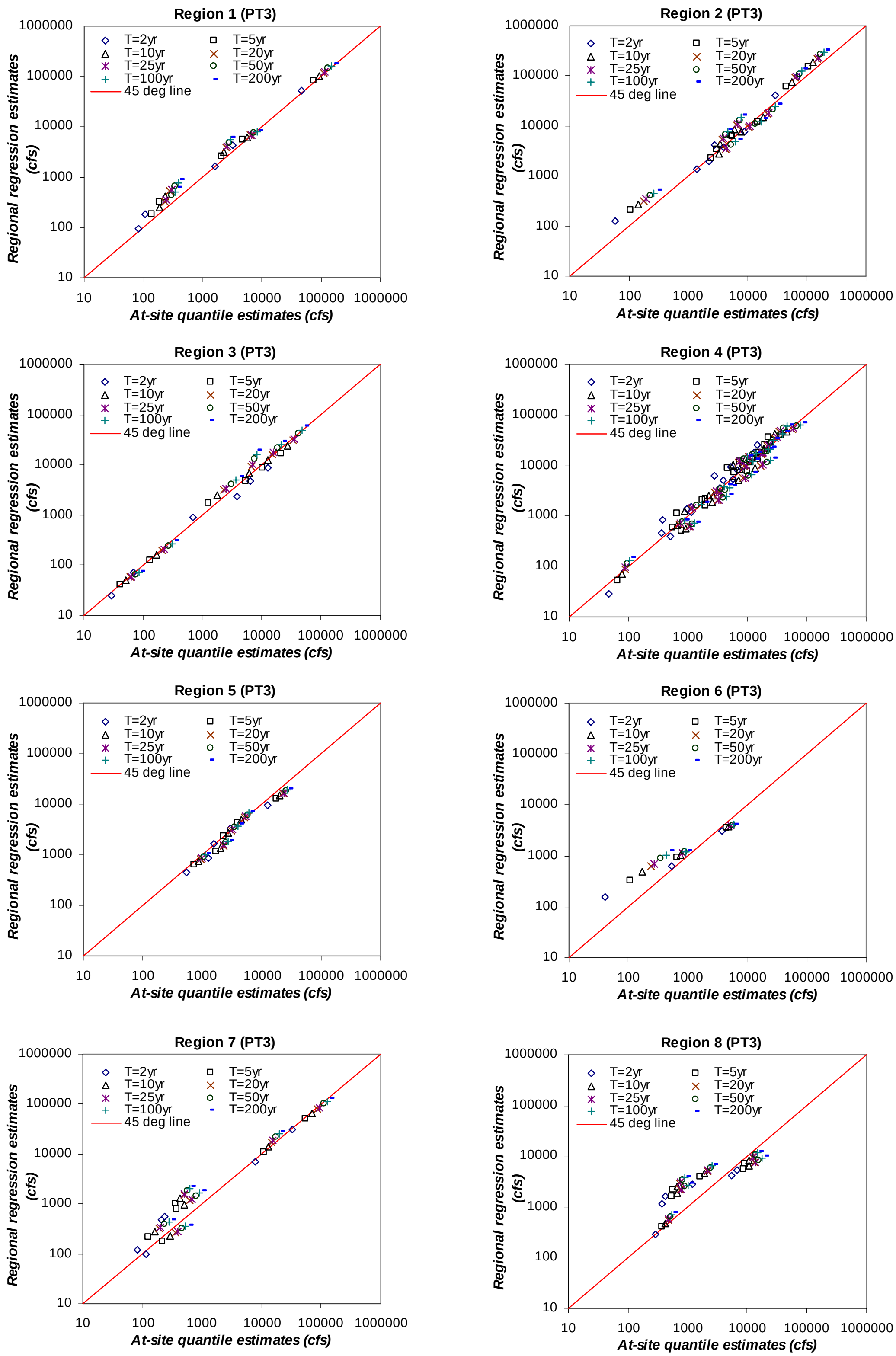

Figure 6.3.1(a). At-site quantile floods and the quantile floods obtained by Method 3 for 25% of the data (PTIII).

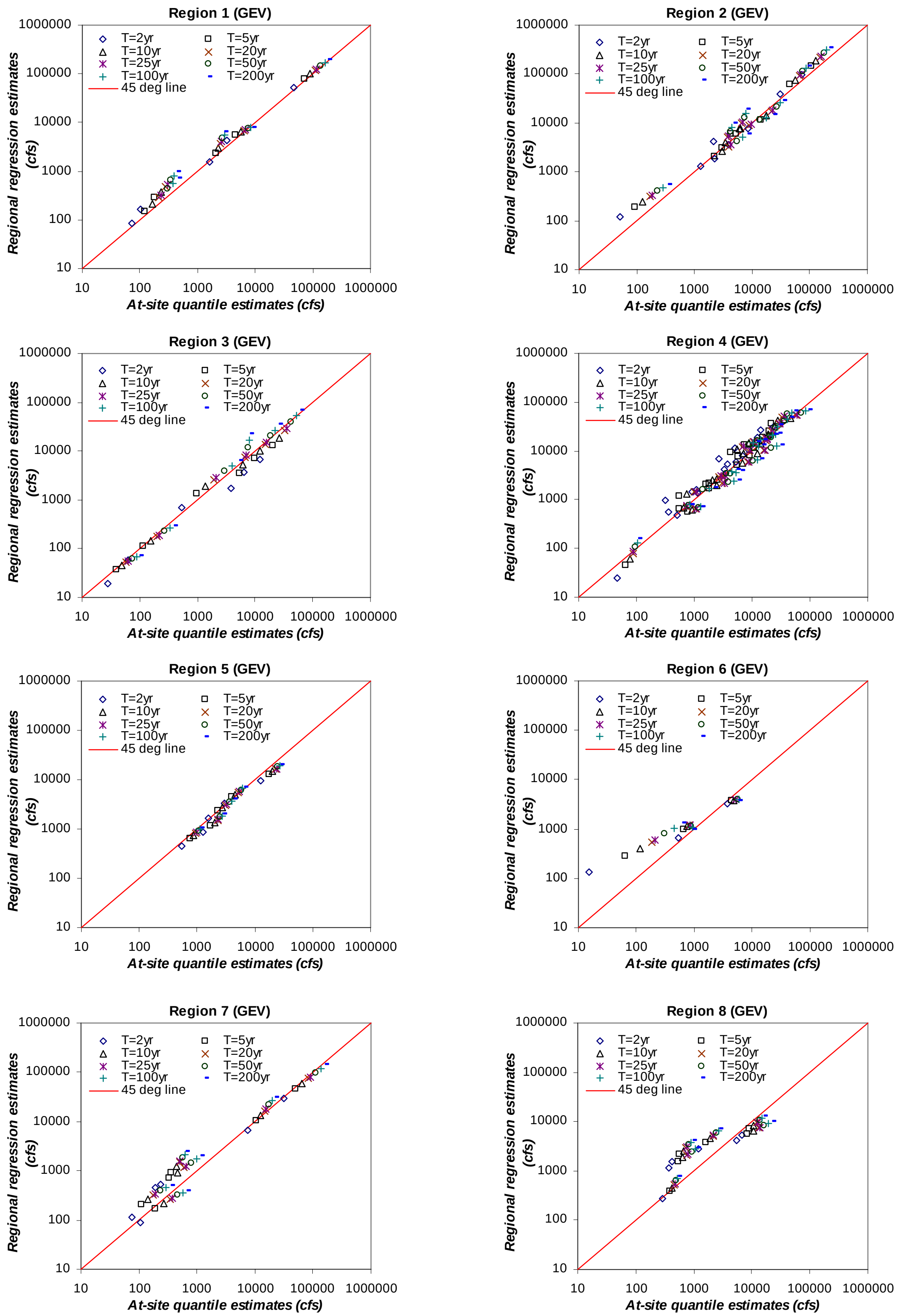

Figure 6.3.1(b). At-site quantile floods and the quantile floods obtained by Method 3 for 25% of the data (GEV).

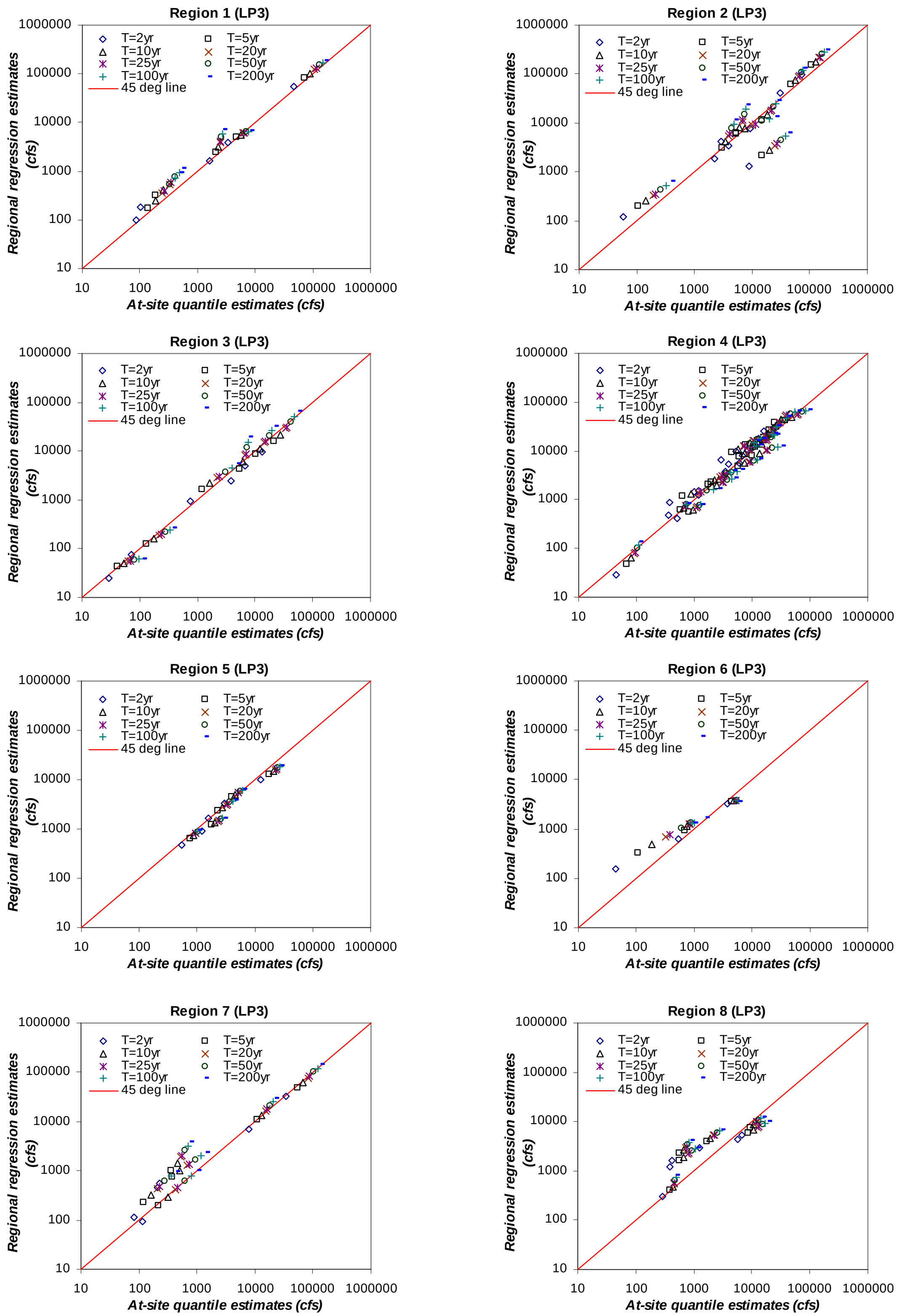

Figure 6.3.1(c). At-site quantile floods and the quantile floods obtained by Method 3 for 25% of the data (LPIII).

6.4. Comparison of the three approaches

To quantify the errors, the estimation errors between at-site quantile estimates and L-moment estimates are calculated for the quantile floods. The percentage error for Region *j* for the L-moment method is calculated by equation (6.4.1).

$$e_{D,T}(\%) = \sum_{i=1}^{N} \frac{| x_{AS}^{D,T}(i) - x_{M1}^{D,T}(i) |}{x_{AS}^{D,T}(i)} \cdot DAR(i) \tag{6.4.1}$$

Where $e_{D,T}$ is the average error percentage for Region *j* with probability distribution *D* (PTIII, GEV or LPIII), and recurrence interval *T* (2, 5, 10, 20, 25, 50, 100 and 200 years). *i*=1,…,N. N is the number of stations in Region *j*. $x_{AS}^{D,T}(i)$ is the quantile flood of distribution *D* and return period *T* at site *i* (AS) $x_{M1}^{D,T}(i)$ is the quantile flood obtained by method *1* (*M1*, L-moment method) of distribution *D* and return period *T* at site *i*. *DAR*(i) is the drainage area ratio, which is ratio of the area at site *i* divided by the sum of drainage areas in the sample.

A similar expression for calculating error percentage is applied to Method 2 (*M2*) and Method 3 (*M3*), with the change that $x_{M1}^{D,T}(i)$ is replaced by $x_{M2}^{D,T}(i)$ and $x_{M3}^{D,T}(i)$, respectively. The average is calculated by weighting by the drainage area instead of simply by the arithmetical average, because it is not reasonable to give same weightings for data from small and large drainage areas and flood magnitudes. The percent errors from small drainage areas are always larger and lead to misinterpretation.

The error percentages are calculated for the 75% of data, which is used for establishing the model parameters, and 25% of data, which is used for validation. The results of the 75% of data for Region 1~ 4 are listed in Table 6.3.1 and of Region 5~8 are listed in Table 6.3.2. Dark shaded box indicates the minimum prediction error in that region. As expected, most of the regions have a more accurate prediction by the L-moment method than GLS and combination method. PTIII has flood better estimates than GEV and LPIII, but PTIII and GEV are close to each other. If we only look at the shaded boxes, except region 6, all the regions have less than 10% average errors. The poor result in region 6 is expected, because it is a heterogeneous region. The error percentage of L-moment method for GEV in most regions is around 15%, and for LPIII distribution is around 20% for most regions. The error percentage of L-moment method for LPIII in region 3 and region 7 is even as high as 30% and 42%, so that LPIII is not preferred for these regions.

As for the GLS regional regression, PTIII distribution is still the preferred distribution and GEV is quite close to it. However, deriving mean flow magnitudes from geographical attributes produce more error than the L-moment method. The best results we see in Table 6.3.1 and Table 6.3.2 are from region 1, 4, 5, and 7, which have error percentage around 10~20%. Region 2 has error as high as 45%, region 3 and region 6 have error around 30% and region 8 has error about 40%. Basically, Table 6.3.1 and Table 6.3.2 give us an idea of residual standard deviation.

The results for 25% of data, which is used for validation, for Region 1~ 4 are listed in Table 6.4.3 and of Region 5~8 are listed in Table 6.4.4. Dark shaded box indicates the minimum prediction error for that region. Most of the regions show more accurate prediction by L-moment method than GLS and combination method. The other comparison of distribution, PTIII has better estimates than GEV and LPIII. If we only look at the shaded boxes, except region 6, all the regions have less than 10% average errors, which is very good prediction for hydrological analysis. The error percentage of L-moment for GEV in most regions is around 10~20%, and for

Table 6.3.1. Estimation errors of 75% split samples obtained from three comparative methods for PTIII, GEV and LPIII distributions (Region 1~4).

Method		Method 1			Method 2			Method 3		
Region	T (yrs)	PT3	LP3	GEV	PT3	LP3	GEV	PT3	LP3	GEV
1	2	1.23	1.8	31.52	11.4	20.31	23.94	8.09	8.33	7.01
	5	1.58	9.4	22.07	11.29	13.85	13.42	12.39	13.06	9.66
	10	1.74	15.1	18.25	11.4	26.33	9.24	14.54	17.19	12.46
	20	2	19.8	15.5	11.54	34.77	6.95	16.14	21.33	14.86
	25	2.22	21.1	14.76	11.59	36.96	6.82	16.46	22.66	15.68
	50	2.83	25.2	12.77	11.73	42.63	6.28	17.56	26.85	18
	100	3.37	28.9	11.15	11.87	47.07	5.6	18.39	30.97	20.19
	200	3.85	32.3	9.8	12.01	50.66	5.12	19.04	35.06	22.49
	Average	2.35	19.2	16.98	11.60	34.07	9.67	15.33	21.93	15.04
	Stdev	0.92	10.2	7.06	0.25	13.01	6.35	3.62	8.99	5.23
2	2	2.02	2.0	35.38	43.58	59.52	9.74	41.79	38.61	38.67
	5	1.24	9.5	23.14	45.81	15.37	15.31	45.59	41.77	42.29
	10	2.24	15.9	17.89	47.27	9.18	21.79	47.49	45.2	44.5
	20	3.14	21.8	13.98	48.5	19.68	27.23	48.76	49.01	46.64
	25	3.49	23.7	12.92	48.86	22.74	28.76	49.13	50.38	47.41
	50	4.45	29.4	9.97	49.88	30.62	33.09	50.09	54.61	49.55
	100	5.29	35.0	7.49	50.78	36.79	37.01	50.81	59.14	51.96
	200	6.03	40.6	5.37	51.59	41.75	40.66	51.47	63.92	54.47
	Average	3.49	22.2	15.77	48.28	29.46	26.70	48.14	50.33	46.94
	Stdev	1.67	12.9	9.75	2.66	16.30	10.61	3.18	8.61	5.14
3	2	6.82	5.4	45.18	27.64	30.94	49.06	27.73	27.16	41.39
	5	1.87	13.1	25.97	22.5	26.5	30.74	21.9	23.41	29.84
	10	5.08	23.1	18.24	19.99	33.22	26.63	20.71	20.62	24.46
	20	8.33	33.2	14.06	18.2	44.27	23.26	23.85	21.93	19.93
	25	9.29	36.5	13	17.7	47.12	22.34	25.13	23.81	19.5
	50	11.95	47.6	14.52	18.12	54.45	19.51	28.8	32.42	25.92
	100	14.23	60.0	18.08	19.92	60.07	19.08	31.96	41.84	33.49
	200	16.22	73.1	22.16	21.53	64.48	23.32	34.79	52.01	42.52
	Average	9.22	36.5	21.40	20.70	45.13	26.74	26.86	30.40	29.63
	Stdev	4.77	23.1	10.55	3.28	14.04	9.77	4.90	11.16	8.92
4	2	4.36	3.4	31.64	25.27	35.01	27.67	30.05	31.34	42.25
	5	1.77	8.9	19.21	23.71	16.21	17.09	22.29	25.86	28.56
	10	3.45	14.5	14.15	22.71	25.4	14.13	18.45	22.37	23.11
	20	5.48	20.0	11.21	22.32	34.84	13.77	15.66	19.22	18.59
	25	6.07	21.9	10.56	22.44	37.4	13.76	14.98	18.35	17.26
	50	7.68	27.5	9.76	22.78	43.94	14.27	13.42	16.58	13.86
	100	9.07	32.9	9.72	23.08	48.94	15.61	12.8	15.8	12.2
	200	10.31	38.1	11.19	23.75	52.91	17.16	12.48	16.48	12.51
	Average	6.02	20.9	14.68	23.26	36.83	16.68	17.52	20.75	21.04
	Stdev	2.88	11.8	7.54	0.97	12.07	4.65	6.04	5.47	10.24

Table 6.4.2. Estimation errors of 75% split samples obtained from three comparative methods for PTIII, GEV and LPIII distributions (Region 5~8).

Method		Method 1			Method 2			Method 3		
Region	T (yrs)	PT3	LP3	GEV	PT3	LP3	GEV	PT3	LP3	GEV
5	2	3.25	2.9	24.81	17.86	17.31	28.54	17.26	17.49	17.4
	5	1.15	8.4	16.4	16.18	26.06	23.12	17.2	18.12	17.73
	10	3.56	14.1	12.22	14.59	32.28	20.5	16.38	16.95	16.8
	20	5.7	19.3	8.74	13.01	36.94	18.2	15.45	15.43	15.54
	25	6.35	20.9	7.71	13.04	38.2	17.5	15.25	15.02	15.15
	50	8.26	25.6	4.81	13.26	41.5	15.46	14.65	13.79	14.28
	100	10.14	30.1	6.99	13.45	44	13.53	14.08	12.58	13.35
	200	11.98	34.5	9.88	13.76	45.96	12.66	13.5	12.97	13.17
	Average	6.30	19.5	11.45	14.39	35.28	18.69	15.47	15.29	15.43
	Stdev	3.68	10.7	6.45	1.76	9.69	5.28	1.39	2.09	1.77
6	2	7.54	1.7	40.14	23.13	20.36	41.69	17.00	15.25	15.80
	5	8.49	20.0	22.13	18.48	25.91	31.39	23.28	22.89	23.35
	10	18.1	33.1	10.57	20.78	28.19	25.58	24.97	26.26	25.76
	20	26.72	45.3	4.67	25.75	30.55	19.61	25.81	28.57	26.95
	25	29.38	49.0	8.21	28.22	31.18	19.37	25.98	29.16	27.19
	50	37.48	60.9	20.36	35.56	32.82	21.23	26.3	30.52	27.43
	100	45.1	72.7	33.04	42.48	34.09	29.11	26.43	31.36	27.09
	200	52.35	84.2	46.32	49.08	35.11	41.19	26.44	31.75	26.29
	Average	28.15	45.9	23.18	30.44	29.78	28.65	24.53	26.97	24.98
	Stdev	16.37	27.3	15.38	10.92	4.86	9.00	3.22	5.56	3.94
7	2	6.18	3.0	31.38	18.56	36.05	25.22	16.18	14.87	23.64
	5	2.67	38.3	17.66	15.46	11.09	17.61	11.53	11.91	17.55
	10	3.21	61.1	14.48	12.37	27.75	12.88	9.27	9.28	13.5
	20	5.93	81.9	13.68	9.88	41.12	9.76	7.76	7.27	9.74
	25	6.66	88.3	13.73	9.2	44.68	8.85	7.37	6.7	8.58
	50	8.61	108.1	14.73	8.37	53.97	10.54	8.01	10.79	5.99
	100	10.17	127.6	17.01	9.76	61.23	13.82	9.65	15.15	9.29
	200	11.45	147.3	19.79	10.91	67.02	17.47	11.03	19.55	13.72
	Average	6.86	81.9	17.81	11.81	42.86	14.52	10.10	11.94	12.75
	Stdev	3.10	47.3	5.89	3.52	18.26	5.43	2.88	4.38	5.70
8	2	2.06	2.2	27.77	39.91	40.12	28.95	40.84	39.02	41.39
	5	2.35	6.4	19.97	39.76	31.07	32.3	39.85	38.5	40.63
	10	4.09	8.8	17.09	39.89	28.7	33.82	39.41	37.38	39.56
	20	5.48	11.2	15.24	40.29	31.74	34.91	39.3	37.02	39.39
	25	5.86	12.0	14.79	40.39	33.66	35.2	39.24	36.85	39.31
	50	6.91	14.2	14.22	40.6	38.77	35.95	39.04	36.25	38.98
	100	7.84	16.2	14.22	40.77	42.91	36.52	38.86	35.71	38.69
	200	8.67	18.0	14.43	41.02	46.35	36.93	38.79	35.14	38.55
	Average	5.41	11.1	17.22	40.33	36.67	34.32	39.42	36.98	39.56
	Stdev	2.43	5.2	4.70	0.45	6.29	2.63	0.67	1.32	0.98

LPIII distribution is as high as 6~114% for all regions. It can be concluded that LPIII is not preferable.

From the results of GLS regional regression, PTIII is the preferred distribution and GEV is quite close to it. Region 1 and region 7 have less than 10% error, region 3, 4, 5, 6 and 7 have error around 16~30%, region 2 has 45% error and region 8 has error as high as 75%. The error from region 8 is from the poor correlation between flood magnitude and drainage area and it makes the regression equation not as reliable as in other regions.

As for Method 2, which is using GLS regression to estimate the mean flow, i.e., the first L-moment, and then applying the L-moment method to calculate the flood magnitude, the error percentages are between the method 1 (L-moment) and method 3 (GLS) or higher than both of them. LPIII is still the less preferred distribution since it embeds the error from both models and makes the result not reliable. Method 2 can be a good substitute for the regions that have higher error in Method 3 but lower error in Method 2.

The same error percentages are calculated for the regions defined by Srinivas and Rao (2003); again, we only put the result for region 1 and region 5 since the other regions are the same as Knipe and Rao (2004). For the 75% of data, which are used to build model parameters, are listed in Table 5.6, and for the 25% of data, which are used to validate the model, are listed in Table 5.7. For the percentage error, we can find out that the values is between the two merged regions. For example, the error for L-moment, PTIII in region 1 is 2.35% and 6.86% in region 7 (Table 6.4.1 and Table 6.4.2), and the error is 5.9% for the merged region. We also have tried to apply the equation for the merged area to region 8 individually, but no significant improvement is got for region 8.

Table 6.4.3. Estimation errors of 25% split samples obtained from three comparative methods for PTIII, GEV and LPIII distributions (Region 1~4).

Method		Method 1			Method 2			Method 3		
Region	T (yrs)	PT3	LP3	GEV	PT3	LP3	GEV	PT3	LP3	GEV
1	2	4.4	2.1	27.2	15.07	25.33	20.61	11.84	12.48	10.63
	5	2.0	5.6	21.1	10.34	13.71	11.74	10.35	11.22	10.11
	10	5.2	6.5	20.0	7.89	28.04	9.69	9.11	11.46	9.32
	20	7.6	6.8	20.1	5.97	37.64	9.31	7.99	12.23	8.21
	25	8.3	6.7	20.3	5.44	40.12	9.36	7.55	12.55	7.88
	50	10.2	6.6	21.2	3.97	46.53	9.93	6.56	13.66	6.73
	100	11.8	6.3	22.6	2.74	51.53	10.94	5.72	14.8	5.62
	200	13.2	5.8	24.1	1.68	55.56	12.26	4.9	15.94	4.51
	Average	7.8	5.8	22.1	6.64	37.31	11.73	8.00	13.04	7.88
	Stdev	3.8	1.6	2.5	4.39	14.22	3.75	2.35	1.64	2.15
2	2	4.3	3.9	36.8	33.42	52.36	14.99	37.03	36.06	36.12
	5	1.0	10.5	24.5	44.24	10.58	11.52	40.15	37.21	37.67
	10	2.7	20.4	18.3	50.89	7.64	23.19	42.49	40.02	40.02
	20	4.6	30.5	13.0	56.64	17.04	33.89	44.71	43.84	42.91
	25	5.4	33.8	11.3	58.38	19.38	37.31	45.37	45.28	44.1
	50	7.4	44.4	6.4	63.34	25.25	47.85	47.2	49.79	47.95
	100	9.3	55.1	2.0	67.74	29.62	58.49	48.74	54.62	52.31
	200	11.0	66.3	6.7	71.71	32.97	69.43	50.19	59.71	56.98
	Average	5.7	33.1	14.9	55.80	24.36	37.08	44.49	45.82	44.76
	Stdev	3.4	21.5	11.4	12.64	14.31	20.58	4.43	8.40	7.25
3	2	2.6	4.8	44.8	17.79	7.83	47.48	29.67	27.69	44.49
	5	1.7	9.2	29.2	17.78	35.6	37.11	19.03	23.28	33.37
	10	3.7	19.9	22.2	17.36	46.8	32.69	12.3	16.59	24.59
	20	5.4	30.9	16.5	17.58	54.53	29.33	11.14	12.48	16.52
	25	5.8	34.5	15.7	17.86	56.52	28.38	11.29	12.55	14.68
	50	7.2	46.3	14.0	18.67	61.65	26.4	11.88	12.81	12.36
	100	8.5	58.7	13.0	19.61	65.63	27.09	16.18	20.56	16.71
	200	9.6	71.7	13.7	20.55	68.81	28.19	22.36	35.5	33.3
	Average	5.6	34.5	21.1	18.40	49.67	32.08	16.73	20.18	24.50
	Stdev	2.8	23.5	11.0	1.13	19.97	7.14	6.64	8.33	11.49
4	2	2.8	2.1	29.2	26.93	37.76	20.45	38.4	38.47	50.96
	5	1.4	5.9	19.2	29.19	17.38	17.94	25.27	28.86	32.75
	10	3.0	9.0	15.6	30.21	25.24	19.3	20.99	22.87	25.21
	20	4.3	11.3	13.2	31	33.95	21.38	20.58	22.11	22.16
	25	4.7	12.0	12.6	31.23	36.51	22.07	20.5	22.05	21.86
	50	5.7	13.7	11.3	31.86	43.57	24.15	20.3	21.95	21.23
	100	6.7	15.0	12.4	32.41	49.02	26.15	20.2	21.96	21.05
	200	7.6	16.1	13.7	32.89	53.37	28.11	20.23	22.06	21.01
	Average	4.5	10.7	15.9	30.72	37.10	22.44	23.31	25.04	27.03
	Stdev	2.1	4.8	5.9	1.93	11.89	3.47	6.33	5.92	10.45

Table 6.4.5. Estimation errors of 25% split samples obtained from three comparative methods for PTIII, GEV and LPIII distributions (Region 5~8).

Method		L-Moment			Method 2			Method 3		
Region	T (yrs)	PT3	LP3	GEV	PT3	LP3	GEV	PT3	LP3	GEV
5	2	1.7	1.4	24.3	20.9	20.12	33.11	20.42	20.7	20.69
	5	2.2	5.5	18.6	21.02	34.58	27.92	22.55	23.48	22.98
	10	2.0	9.6	15.7	20.74	41.51	26.5	23.07	24.08	23.39
	20	1.6	13.3	13.4	20.57	45.76	25.62	23.39	24.41	23.59
	25	1.4	14.3	12.6	20.51	46.79	25.36	23.48	24.44	23.61
	50	1.1	17.7	10.6	20.29	49.29	24.59	23.62	24.4	23.54
	100	1.9	20.8	8.7	20.06	51.03	23.86	23.72	24.59	23.36
	200	2.6	23.7	6.9	19.82	52.25	23.17	23.76	25.02	23.09
	Average	1.8	13.3	13.8	20.49	42.67	26.27	23.00	23.89	23.03
	Stdev	0.5	7.6	5.6	0.41	10.74	3.14	1.12	1.36	0.97
6	2	6.8	0.2	41.2	22.30	19.13	47.86	15.49	13.56	13.06
	5	7.8	19.4	22.5	16.49	26.82	36.78	22.08	21.17	20.01
	10	16.9	31.3	10.9	12.93	30.53	30.43	25.02	26.14	23.91
	20	25.2	42.0	1.3	9.75	33.83	25.21	27.11	30.24	27.07
	25	27.7	45.3	4.4	8.79	34.74	23.72	27.68	31.41	28
	50	35.1	55.3	15.6	5.95	37.13	19.14	29.18	34.54	30.49
	100	42.0	64.9	27.3	3.3	39.05	14.67	30.42	37.15	32.66
	200	48.5	74.2	39.5	5.81	40.63	10.2	31.47	39.28	34.62
	Average	26.2	41.6	20.3	10.67	32.73	26.00	26.06	29.19	26.23
	Stdev	15.3	24.3	15.1	6.31	7.09	12.22	5.22	8.60	7.08
7	2	3.7	2.8	38.0	3.18	11.54	33.35	5.96	7.46	8.9
	5	1.8	38.8	22.6	4.68	23.14	22.7	7.42	8.68	9.36
	10	3.0	70.5	17.0	8.47	36.12	20.06	9.33	8.48	10.02
	20	5.2	104.1	13.5	11.5	45.24	19.26	10.66	7.73	11.64
	25	5.8	115.4	13.6	12.34	47.69	19.25	11.03	7.36	12.14
	50	7.5	152.6	14.7	14.67	54.13	20.84	11.98	5.65	13.59
	100	9.0	193.3	16.5	16.63	59.25	23.7	12.81	5.11	15.04
	200	10.4	237.9	18.7	18.31	63.44	26.82	13.49	9.58	16.45
	Average	5.8	114.4	19.3	11.22	42.57	23.25	10.34	7.51	12.14
	Stdev	3.0	78.5	8.1	5.45	17.98	4.83	2.62	1.51	2.73
8	2	3.3	1.3	26.9	68.64	70.06	61.28	68.22	68.86	67.8
	5	2.4	3.9	22.0	73.1	65.75	66.4	72.13	71.66	70.71
	10	4.6	6.7	20.5	75.43	63.86	69.28	74.28	73.18	72.79
	20	6.4	9.2	19.8	77.29	62.45	71.87	75.97	74.48	74.85
	25	6.8	10.0	19.6	77.81	62.06	72.68	76.44	74.86	75.5
	50	8.2	12.3	19.3	79.29	61.02	75.13	77.72	76.02	77.54
	100	9.2	14.4	19.2	80.58	60.16	77.48	78.8	77.09	79.57
	200	10.2	16.4	19.3	81.73	59.44	79.75	79.72	78.13	81.62
	Average	6.4	9.3	20.8	76.73	63.10	71.73	75.41	74.29	75.05
	Stdev	2.8	5.1	2.6	4.28	3.46	6.02	3.79	3.01	4.58

As a result, the error percentage is obtained by averaging the errors from two regions and by merging the regions still hardly helps to modify the region with bad regression results. Overall, for three methods, the best recommended method is L-moment method. However, if a watershed lacks flow measurements and requires performing regional regression, it is better to compare the accuracy between GLS and combination methods for the region of interest and decide using which model. For three distributions, the order of best-fit distribution is PT3 followed by GEV and finally the LP3 distribution.

.

Table 6.4.6. Comparison the estimation errors of the 75% split samples obtained from three methods for PTIII, GEV and LPIII distributions (Merged regions: 1+7 and 5+8).

Method		Method 1			Method 2			Method 3		
Region	T (yrs)	PT3	LP3	GEV	PT3	LP3	GEV	PT3	LP3	GEV
1 + 7	2	4.8	6.2	32.7	20.5	33.1	25.3	20.5	15.0	24.7
	5	1.2	17.2	22.1	19.1	10.3	13.1	16.6	12.0	19.3
	10	2.5	27.5	17.9	18.6	22.3	7.4	15.7	13.1	16.0
	20	5.0	38.3	15.0	18.4	31.9	4.8	15.6	15.7	13.1
	25	5.7	41.9	14.3	18.5	34.4	5.6	15.8	16.8	12.5
	50	7.7	53.1	12.2	19.1	40.7	8.3	16.4	21.5	11.4
	100	9.4	65.3	11.4	20.0	45.5	12.9	17.4	28.2	11.8
	200	11.0	78.8	13.0	21.3	49.3	17.6	18.6	35.7	14.2
	Average	5.9	41.0	17.3	19.4	33.4	11.9	17.1	19.7	15.4
	Stdev	3.3	24.3	7.1	1.1	12.6	6.9	1.7	8.3	4.6
5 + 8	2	3.1	7.6	27.2	31.1	34.9	22.8	32.6	31.3	32.8
	5	2.2	8.1	18.6	33.4	19.2	18.7	32.3	30.9	32.6
	10	4.2	13.7	14.9	35.0	24.5	20.9	32.0	31.3	31.5
	20	6.2	21.2	12.2	36.7	28.5	22.9	32.2	32.4	30.3
	25	6.8	24.5	11.5	37.2	29.6	23.8	32.4	32.8	30.2
	50	8.6	36.7	10.0	38.8	32.7	27.2	32.8	34.4	29.7
	100	10.2	50.7	11.3	40.3	35.4	30.5	33.2	36.0	30.0
	200	11.6	67.1	13.1	41.6	37.6	33.6	33.8	38.1	30.4
	Average	6.6	28.7	14.8	36.7	30.3	25.1	32.6	33.4	30.9
	Stdev	3.4	21.3	5.7	3.5	6.2	5.0	0.6	2.6	1.2

Table 6.4.7. Comparison the estimation errors of the 25% split samples obtained from three methods for PTIII, GEV and LPIII distributions (Merged regions: 1+7 and 5+8).

Method		Method 1			Method 2			Method 3		
Region	T (yrs)	PT3	LP3	GEV	PT3	LP3	GEV	PT3	LP3	GEV
1 + 7	2	3.6	2.8	38.9	25.2	37.8	21.0	39.7	19.3	25.0
	5	1.8	19.1	24.7	29.1	2.4	5.6	41.5	21.6	24.9
	10	4.7	33.5	18.1	31.0	13.0	6.2	41.7	26.0	24.1
	20	7.1	47.6	13.0	32.5	22.1	11.7	41.6	31.8	23.0
	25	7.8	52.2	11.5	32.9	24.5	13.2	41.5	33.9	22.7
	50	9.6	66.7	7.3	34.1	30.7	17.4	41.1	41.1	21.5
	100	11.3	81.8	6.8	35.1	35.5	21.1	41.5	49.2	20.4
	200	12.7	97.5	6.8	35.9	39.4	24.6	44.1	58.2	19.3
	Average	7.3	50.1	15.9	32.0	25.7	15.1	41.6	35.1	22.6
	Stdev	3.8	31.7	11.2	3.5	13.0	7.1	1.2	13.6	2.1
5 + 8	2	1.5	1.7	23.0	53.1	61.2	25.3	55.3	55.1	53.1
	5	2.0	4.6	17.3	52.6	41.2	26.2	51.3	49.5	49.1
	10	2.6	7.1	15.0	53.3	38.7	27.5	50.0	48.4	46.6
	20	2.9	9.3	13.3	54.3	37.1	28.7	49.4	48.6	44.4
	25	3.0	10.0	12.8	54.6	36.9	29.1	49.2	48.7	43.8
	50	3.1	12.2	11.6	55.7	36.8	30.1	48.8	49.6	42.0
	100	3.5	14.5	10.9	56.8	36.6	31.1	48.6	50.8	40.5
	200	4.1	16.7	10.3	57.8	36.5	31.9	48.5	52.2	39.3
	Average	2.8	9.5	14.3	54.8	40.6	28.7	50.2	50.4	44.8
	Stdev	0.8	5.0	4.2	1.8	8.5	2.3	2.3	2.3	4.6

VII. Conclusions

On the basis of the research presented in this report, the following conclusions are presented.

1. One of the objectives of the study is to select the probability distribution which best fits the data in each of the six regions in Indiana. Based on the results presented in Chapter II, distributions in each region are ranked. In general, for region 4, 5 and 6, Generalized Extreme Value distribution is the best distribution. For regions 2 and 3, Log Normal (III) distribution is the best. Log Pearson (III) distribution is not the best distribution for any region. The Maximum likelihood method is the best parameter estimation method.
2. The equations developed for different regions may be used by the results in Chapter III when LP(III) distributions must be used. If a region is homogeneous the prediction error can be quite small. Otherwise, it may be large. However, the results presented in Chapter II indicate that the LP(III) distribution is inferior to other distributions.
3. The tables needed for the L-moment flood estimates for Indiana watersheds have been presented. The prediction accuracies of three distributions are compared. The LP(III) distribution is the least accurate distribution. If a region is homogeneous, the L-moment method gives quite accurate estimates.
4. The parameters for quantile flood estimation by regression relationships are presented. By using these equations, the flood magnitude is directly calculated. The

GLS regression equations developed for mean and mean of logarithms of annual peak flows are also presented. These equations may be used with L-moment method.

5. Results are presented for the L-moment and GLS methods to estimate flood frequencies. These methods give similar results but the results from L-moment method are slightly superior. Once again the accuracy depends on the homogeneity of regions. The results are quite inferior for heterogeneous watersheds. The LP(III) distribution does not perform as well as other distributions.

Index

Bobee, B., and F. Ashkar (1991). "The Gamma Family and Derived Distributions Applied in Hydrology", *Water Resources Publications*, Littleton, CO.

Chow, V.T. (1964). Editor-in-Chief, Handbook of Applied Hydrology, McGraw-Hill, New York, NY.

Glatfelter, Dale (1984). "Techniques for Estimating Magnitudes and Frequency of Floods of Streams in Indiana", WRI 84-4134, USGS, Indianapolis, IN.

Hosking, J.R.M., and J.R. Wallis (1993). Some Statistics Useful in Regional Frequency Analysis, *Water Resources Research*, 29, 271-281.

Hosking, J.R.M. and J.R. Wallis (1997). Regional Frequency Analysis, Cambridge University Press, Cambridge CB2 2RU, U.K.

Hosking, J.R.M., J. R. Wallis and E.F. Wood (1985). An Appraisal of the Regional Flood Frequency Procedure in the UK Flood Studies Report, *Hydrol. Sci. Jour*., Vol. 30, No. 1, pp. 85-109.

Jin, M., and J.R. Stedinger (1989). Flood Frequency Analysis with Regional and Historical Information", *Water Resources Research*, Vol. 25, No. 5, pp. 925-936.

Iblings, M.L. and Rao, A.R. (2003). Regionalization of Indiana Watersheds for Flood Flow Predictions. III. Use of Precipitation and Flow Data for Regionalization of Watersheds, *Interim Report FHWA/IN/JTRP-2002/3,* No.2, JTRP, School of Civil Engineering, Purdue University, West Lafayette, IN 47907, pp. 98.

Kite, G.W. (1977). Frequency and Risk Analysis in Hydrology, *Water Resources,* Publications, Fort Collins, CO.

Landwehr, J.M., G.D. Tasker and R.D. Jarrett (1987). Discussion of Relative Accuracy of Log Pearson III Procedures, by J.R. Wallis and E.F. Wood, *J. Hydraul. Eng*., Vol. 111, No. 7, pp. 1206-1210.

McCuen, R.H. (1993). *Microcomputer Applications in Statistical Hydrology*, Prentice Hall, Englewood Cliffs, NJ.

McCormick, D.L. and A.R. Rao (1995). Outlier Detection in Indiana Flood Data, Tech. Rept. CE-EHE-95-04, School of Civil Engineering, Purdue University, W. Lafayette, IN 47907, pp. 132.

Potter, K. W. (1987). Research on Flood Frequency Analysis, 1983-1986, *Reviews of Geophysics*, Vol. 25, No. 2, pp. 113-118.

Potter, K.W., and D.P. Lettenmaier, A Comparison of Reginal Flood Frequency Estimation Methods Using a Resampling Method, *Water Resources Research*, Vol. 26, No. 3, pp. 415-424.

Rao, A.R. and K.H. Hamed (1997). Regional Frequency Analysis of Wabash River Flood Data by L-Moments, *ASCE Journal of Hydrologic Engineering*, Vol. 2, No. 4, pp. 169-180.

Rao, A. Ramachandra and Khaled H. Hamed (2000). *Flood Frequency Analysis,* CRC Press LLc, Boca Raton, FL.

Rao, A.R., S. Ernst, and G.D. Jeong, (2002). Regionalization of Indiana Watersheds for Flood Flow Predictions. I. Results from L-Moment Based Method, Interim report FHWA/IN/JTRP-2002/3, JTRP, School of Civil Engineering, Purdue University, W. Lafayette, IN 47907, pp. 98.

Singh, V.P. and K. Singh (1988). Parameter Estimation for TPLN Distribution for Flood frequency Analysis, *Water Resources Bulletin*, Vol. 23, No. 6, pp. 1185-1191.

Srinivas, V.V. and A. R. Rao (2002). Regionalization of Indiana Watersheds for Flood Flow Predictions. II. Regionalization of Indiana Watersheds by Hybrid Cluster Analysis, Interim report FHWA/IN/JTRP-2002/3, JTRP, School of Civil Engineering, Purdue University, W. Lafayette, IN 47907, pp. 98.

Srinivas, V.V. and A.R. Rao (2003). Regionalization of Indiana Watersheds for Flood Flow Predictions. IV. Regionalization of Indiana Watersheds by Fuzzy Cluster Analysis, Interim report FHWA/IN/JTRP-2002/3, JTRP, School of Civil Engineering, Purdue University, W. Lafayette, IN 47907, pp. 98.

Stedinger, J. R. and Tasker, G. D. (1985). Regional Hydrologic Analysis 1.: Ordinary, Weighted, and Generalized Least Squares Compared, *Water Resourc. Res*., v. (22)5: p. 785-793.

Stedinger, J.R., and G.D. Tasker (1985). Regional Hydrologic Regression, 1. Ordinary, Weighted and Generalized Least Squares Compared, *Water Resources Research*, Vol. 21, No. 9, pp. 1421-1432.

Stedinger, J.R., Vogel, R.M. and E. Foufoula-Geogiou, E., (1993). Frequency Analysis of Extreme Events, in “Handbook of Hydrology”, ed. D.R. Maidment, McGraw-Hill, New York, NY, pp. 18.1-18.66.

Tasker, G.D., and J.R. Stedinger (1986). Estimating Generalized Skew with Weighted Least Squares Regression, *J. Water Resour. Plann. Manage.*, Vol. 112, NO. 2, pp. 225-237.

Tasker, G.D., and J.R. Stedinger (1989). An Operational GLS Model for Hydrologic Regression, J. Hydrol., Vol. 111, pp. 361-375.

Thomas, D.M., and M.A. Benson (1970). Generalization of Streamflow Characteristics from Drainage-Basin Characteristics, U.S. Geological Survey, *Water Supply Paper* 1975, U.S. Department of the Interior, Washington D.C.

U.S. Water Resources Council (1982). Guidelines for Determining Flood Flow Frequency, Bulletin 17B (revised), Hydrology Committee, Water Resources Research Council, Washington.

Wallis, J.R. and E.F. Wood (1985). Relative Accuracy of Log Pearson III Procedures, *J. Hydraul. Eng.*, Vol. 111, No. 7, pp. 1043-1056 [with discussion and closure, *J. Hydraul. Eng.*, Vol. 113, No. 7, pp. 1205-1214.]

Yevjevich, V. (1972). Probability and Statistics in Hydrology, *Water Resources* Publications, Fort Collins, CO.

Zempleni, A. (1991). Goodness-of-fit for Generalized Extreme Value Distributions, *Technical Report*, University of Sheffield.