NCHRP

Web-Only Document 199:

Estimating Joint Probabilities of Design Coincident Flows at Stream Confluences

Roger T. Kilgore
David B. Thompson
David T. Ford
Kilgore Consulting and Management
Denver, CO

Contractor's Final Report for NCHRP Project 15-36
Submitted March 2013

National Cooperative Highway Research Program
TRANSPORTATION RESEARCH BOARD
OF THE NATIONAL ACADEMIES

ACKNOWLEDGMENT

This work was sponsored by the American Association of State Highway and Transportation Officials (AASHTO), in cooperation with the Federal Highway Administration, and was conducted in the National Cooperative Highway Research Program (NCHRP), which is administered by the Transportation Research Board (TRB) of the National Academies.

COPYRIGHT INFORMATION

Authors herein are responsible for the authenticity of their materials and for obtaining written permissions from publishers or persons who own the copyright to any previously published or copyrighted material used herein.

Cooperative Research Programs (CRP) grants permission to reproduce material in this publication for classroom and not-for-profit purposes. Permission is given with the understanding that none of the material will be used to imply TRB, AASHTO, FAA, FHWA, FMCSA, FTA, Transit Development Corporation, or AOC endorsement of a particular product, method, or practice. It is expected that those reproducing the material in this document for educational and not-for-profit uses will give appropriate acknowledgment of the source of any reprinted or reproduced material. For other uses of the material, request permission from CRP.

DISCLAIMER

The opinions and conclusions expressed or implied in this report are those of the researchers who performed the research. They are not necessarily those of the Transportation Research Board, the National Research Council, or the program sponsors.

The information contained in this document was taken directly from the submission of the author(s). This material has not been edited by TRB.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. On the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, on its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

The **Transportation Research Board** is one of six major divisions of the National Research Council. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal. The Board's varied activities annually engage about 7,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation. **www.TRB.org**

www.national-academies.org

AUTHOR ACKNOWLEDGMENTS

The research reported herein was performed under NCHRP Project 15-36 by Kilgore Consulting and Management. Mr. Roger T. Kilgore was the Project Director and co-Principal Investigator. The co-authors for this report are Dr. David B. Thompson with R.O. Anderson, Inc. and Dr. David Ford of Ford Consulting Engineers, Inc. Additional support was provided by Brad Moore, P.E. and Nathan Pingel of David Ford Consulting Engineers, Inc.

Mark Heggli of Innovative Hydrology and John Henz of HDR, Inc. provided literature review and analysis of the nature of storm cells and radar as potential tools for assessment.

ABSTRACT

The results of a research study to develop practical procedures for estimating joint probabilities of coincident flows at stream confluences and guidelines for applying the procedures to design problems are documented in this report. The study was based on a database of 85 gage pairs located throughout the coterminous United States. Two practical design methods emerged from the study. The first method for determining a set of exceedance probability combinations associated with the desired joint probability for design is based on copulas. The Gumbel-Hougaard copula was selected as the most appropriate for this application. Use of the copula requires estimation of the Kendall's τ rank correlation for an ungaged watershed pair. Tools for estimating this parameter based on watershed characteristics are also a product of the research. For a given Kendall's τ, the appropriate combinations of events on the main and tributary streams for a desired joint exceedance probability are provided in tabular form. The second method is the use of the total probability method. Although the total probability method, based on the total probability theorem, is a well-documented procedure, this study provided recommended conditional probability matrices needed to apply the theorem. Both methods result in the design conditions representing the joint probability desired by the designer for analysis.

Table of Contents

List of Tables iv
List of Figures vi
Executive Summary vii
1. Background 1
1.1 Problem Statement 1
1.2 Literature Review 3
1.3 Research Objectives and Scope 4
2. Research Method 5
2.1 Work Plan 5
2.1.1 Phase I 5
2.1.1.1 Task 1. Critically Review Literature/Summarize Procedures 5
2.1.1.2 Task 2. Critically Review Procedures 5
2.1.1.3 Task 3. Identify Confluent Stream Sites 5
2.1.1.4 Task 4. Propose Procedures 5
2.1.1.5 Task 5. Submit Interim Report 6
2.1.2 Phase II 6
2.1.2.1 Task 6. Collect/Compile Data 6
2.1.2.2 Task 7. Develop Practical Procedures 6
2.1.2.3 Task 8. Validate Procedures 6
2.1.2.4 Task 9. Final Report 6
2.2 Potential Strategies 6
2.2.1 Bivariate Probability Distributions 7
2.2.2 Univariate Probability Distributions with Linking Copulas 7
2.2.3 Total Probability Method 8
2.2.4 Regression Relations 9
2.2.5 Marginal Analysis 9
2.2.6 Synthetic Storm Cell/Runoff Modeling 10
2.2.7 Tabular Summaries 10
2.3 Database 11
2.3.1 Gage Pairs 11
2.3.2 Instantaneous Data 13

2.3.3 Watershed and Meteorological Data 13
2.3.3.1 Watershed Data 13
2.3.3.2 Meteorological Data 14
3. Data Evaluation 15
3.1 Annual versus Partial Duration Series 18
3.2 Definition of Coincident Flow 19
3.3 Mean Daily versus Instantaneous Data 27
3.3.1 Mean Daily Flow/Instantaneous Peak Correlation 27
3.3.2 Annual Series Correlation 28
3.3.3 Paired Dataset Correlation 30
4. Bivariate Distributions and Univariate Distributions with Copulas 34
4.1 Fitting the Distributions 34
4.1.1 Graphical Assessment and Application 34
4.1.2 Best-fit Distribution 38
4.2 Correlation Parameter Estimation 39
4.2.1 Objective 39
4.2.2 Data 39
4.2.2.1 Watershed, Meteorological, and Hydrological Data 39
4.2.2.2 Estimated Correlation Parameters 40
4.2.2.3 Correlation Matrix 41
4.2.3 Correlation Parameter Regressions 41
4.2.4 Correlation Parameter Groupings 46
4.2.5 Recommended Method 48
4.3 Event Combination Development 49
4.4 Complementary Data in Paired Datasets 53
5. Total Probability Method 56
5.1 Pair Count Matrices 56
5.2 Conditional Probability Matrices 58
6. Conclusions and Recommendations 62
6.1 Factors Influencing Joint Probability 62
6.2 Recommended Tools 62
6.2.1 Copulas 63
6.2.2 Total Probability Method 63

6.3 Suggested Research 64
7. References 66
Appendix A. Literature Review A-1
Appendix B. Databases B-1
Appendix C. Bivariate Distributions and Copulas C-1
Appendix D. Total Probability Method D-1
Appendix E. Other Strategies E-1
Appendix F. Data Analysis F-1
Appendix G. Application Guidelines G-1
Appendix H. Step by Step Application Guide H-1

List of Tables

Table 2.1. Tabular Summary of Joint Probabilities of Coincident Flow 11
Table 2.2. Gage Pair Drainage Area Distribution 12
Table 2.3. Gage Pair Drainage Area Ratio Distribution 13
Table 2.4. Sources for Watershed Characteristics 13
Table 3.1. Prototype Gage Drainage Areas 17
Table 3.2. Instantaneous Annual and Partial Duration Series Availability 18
Table 3.3. Instantaneous Peak Data for Selected Events 20
Table 3.4. Correlation Between Instantaneous Annual Peaks and Same-day Mean Daily Flow 28
Table 3.5. Daily Data Correlation with Instantaneous Data 30
Table 3.6. Correlation Between Main and Tributary (Gage Pair 05)..... 33
Table 4.1. Annual Peak Summary Statistics 35
Table 4.2. Summary Statistics for Daily Data Concurrent Pair Datasets 35
Table 4.3. 10-percent Exceedance Combinations for Gage Pair 06 (POM) 38
Table 4.4. Summary of RMSE Computations..... 39
Table 4.5a. Correlation Matrix for the Peaks on Main (POM) Dataset..... 42
Table 4.5b. Correlation Matrix for the Peaks on Tributary (POT) Dataset 43
Table 4.6. Pearson's ρ (POM)..... 47
Table 4.7. Pearson's ρ (POT)..... 47
Table 4.8. Kendall's τ (POM)..... 47
Table 4.9. Kendall's τ (POT) 47
Table 4.10. Correlation Parameters Based on Grouping..... 48
Table 4.11. 10-yr Joint Return Periods..... 50
Table 4.12. 25-yr Joint Return Periods..... 50
Table 4.13. 50-yr Joint Return Periods..... 51
Table 4.14. 100-yr Joint Return Periods..... 51
Table 4.15. 500-yr Joint Return Periods..... 51
Table 4.16. 10-yr Joint Return Periods (Groupings)..... 52
Table 4.17. 25-yr Joint Return Periods (Groupings)..... 52
Table 4.18. 50-yr Joint Return Periods (Groupings)..... 52
Table 4.19. 100-yr Joint Return Periods (Groupings)..... 53

Table 4.20. 500-yr Joint Return Periods (Groupings) 53
Table 5.1. AEP Bins 56
Table 5.2a. Pair Counts for Gage Pair 06 (POM) 57
Table 5.2b. Pair Counts for Gage Pair 06 (POT) 57
Table 5.3. CPM Selection Table – Degree of Correlation 58
Table 5.4a. CPM for Highly Correlated Flows (POM) 59
Table 5.4b. CPM for Highly Correlated Flows (POT) 60
Table 5.5a. CPM for Moderately Correlated Flows (POM) 60
Table 5.5b. CPM for Moderately Correlated Flows (POT) 60
Table 5.6a. CPM for Mixed Correlated Flows (POM) 60
Table 5.6b. CPM for Mixed Correlated Flows (POT) 61
Table 5.7a. CPM for Poorly Correlated Flows (POM) 61
Table 5.7b. CPM for Poorly Correlated Flows (POT) 61

List of Figures

Figure 1.1. Problem Definition Schematic 2
Figure 1.2. Isoline for Annual Bivariate Exceedance Probability 3
Figure 2.1. Gage Pairs and Hydrologic Units 12
Figure 3.1. Gage Pair 06 (New Jersey) 16
Figure 3.2. Gage Pair 43 (Washington) 16
Figure 3.3. Gage Pair 76 (Texas) 17
Figure 3.4a. Gage Pair 06 June 20-28, 2003 21
Figure 3.4b. Gage Pair 06 September 18-24, 2004 21
Figure 3.4c. Gage Pair 06 April 1-11, 2005 22
Figure 3.4d. Gage Pair 06 October 8-20, 2005 22
Figure 3.5a. Gage Pair 43 December 9-16, 1987 23
Figure 3.5b. Gage Pair 43 October 15-22, 1988 23
Figure 3.5c. Gage Pair 43 January 8-13, 1990 24
Figure 3.5d. Gage Pair 43 November 23-28, 1990 24
Figure 3.6a. Gage Pair 76 October 12-17, 1997 25
Figure 3.6b. Gage Pair 76 October 18-24, 1998 25
Figure 3.6c. Gage Pair 76 May 2-8, 2000 26
Figure 3.6d. Gage Pair 76 August 31-September 5, 2001 26
Figure 3.7. Daily and Instantaneous Peaks for 01403060 (Gage Pair 05) 29
Figure 3.8. Daily and Instantaneous Peaks for 01403150 (Gage Pair 05) 29
Figure 3.9. Daily–Instantaneous Correlation versus Drainage Area 30
Figure 3.10. Peak on Main Dataset (Instantaneous Data, Pair 05) 31
Figure 3.11. Peak on Tributary Dataset (Instantaneous Data, Pair 05) 32
Figure 3.12. Peak on Main Dataset (Daily Data, Pair 05) 32
Figure 3.13. Peak on Tributary Dataset (Daily Data, Pair 05) 33
Figure 4.1. Fit for Gage Pair 06 (POM) 36
Figure 4.2. Joint Distribution for Gage Pair 06 (POM) 37
Figure 4.3. 10-percent Exceedance Probability for Gage Pair 06 (POM) 38
Figure 4.4. Flood Frequency Curves for Gage 01403060 (GP05) 54
Figure 4.5. Flood Frequency Curves for Gage 01446000 (GP06) 55

Executive Summary

Highway drainage structures are often located near the confluence of two streams where they might be subject to inundation by relatively rare hydrologic events from either stream. These structures must be designed to meet specified performance objectives for floods. Because the flooding of structures on one stream can be affected by high flows on the other stream, it is important to know the relation between the exceedance probabilities for flows emanating from the main and tributary streams in a confluent stream pair, that is, the joint probability of the coincident flows.

The joint probability question arises where a structure is located at a site where flows emanating from the main (larger) watershed and tributary (smaller) watershed might result in critical hydraulic design conditions. A portion of the tributary stream is influenced by both the discharge of the tributary stream and the backwater caused by flows in the main stream. This portion of the tributary stream is referred to as the influence reach. The location of the structure within the influence reach, as well as the joint hydrologic behavior of the confluent streams, will determine the importance of the joint behavior of the confluent streams on the appropriate design conditions for the structure.

The objective of this research was to develop practical procedures for estimating joint probabilities of design coincident flows at stream confluences and to provide guidelines for applying the procedures. The scope was limited to riverine areas and did not include consideration of coastal areas.

Several general types of strategies were explored for their suitability in addressing the research objective. They were: 1) bivariate probability distributions, 2) univariate probability distributions with linking copulas, 3) the total probability method, 4) regression analyses, 5) marginal analysis, 6) synthetic design storm cells with rainfall/runoff modeling, and 7) tabular summaries. Preliminary qualitative evaluations eliminated all but the first three strategies.

Three databases were developed to quantitatively evaluate the remaining strategies. The first was a selection of confluent gage pairs from stream gages located in the coterminous United States. Composed of 85 gage pairs, this was the primary database for developing and testing the strategies. A second database contained instantaneous flow records from the United States Geological Survey (USGS) Instantaneous Data Archive. The third database included watershed and meteorological data associated with the confluent gage pairs.

Two practical methods emerged from the analyses. The first method involves determining a set of exceedance probability combinations associated with the desired joint probability for design based on copulas, specifically the Gumbel-Hougaard copula. The second method requires a series of conditional probability matrices for application within the context of the total probability method.

With respect to the first method, the selection of the Gumbel-Hougaard copula was based on the evaluation of four bivariate distribution or copula methods: 1) bivariate Gumbel distribution, 2) bivariate normal distribution (log-transformed), 3) Gumbel-Hougaard copula, and 4) Frank copula.

The Gumbel-Hougaard and Frank copulas require estimation of a single parameter for application to an ungaged watershed pair to characterize the dependence between the main and tributary streams. This dependence parameter is a function of Kendall's τ. This research resulted in methods to estimate Kendall's τ for ungaged watershed pairs based on watershed and meteorological characteristics.

The bivariate normal and Gumbel (mixed model) distributions require the estimation of five parameters: a location and scale parameter for each marginal distribution (main and tributary streams) and a dependence parameter. The dependence parameter is calculated as a function of Pearson's ρ. Because of the need for additional parameters, the bivariate normal and Gumbel (mixed model) distributions are not as easy to use as the copula methods.

The two copula methods offer advantages over the bivariate normal and mixed model Gumbel distributions. First, the mixed model bivariate Gumbel (mixed model) distribution may only be applied when the estimated Pearson's ρ correlation between the paired watersheds is less than two-thirds. Therefore, it cannot be applied to watersheds when the correlation coefficient exceeds two-thirds. Second, a fundamental advantage of the copula methods over both bivariate distributions is that underlying marginal distributions are not assumed. Therefore, a designer may choose any appropriate marginal distribution for both the main and tributary streams or two entirely different marginal distributions for the main and tributary streams. However, given its frequent application for annual peak flow, the Log Pearson III distribution for both the main and tributary streams would likely be a common selection. Both the normal and Gumbel bivariate distributions require that the marginal distributions are normal and Gumbel, respectively.

Considering the fit of the observed gage pair data to the bivariate distributions and copulas, the Gumbel-Hougaard method performed better than the normal and Gumbel bivariate distributions and the Frank copula. Considering the fit with observed data, simplicity (need to estimate a single dependence parameter), and flexibility (no assumption of the marginal distribution), the Gumbel-Hougaard copula was recommended for use. However, application of this method does not require a background in statistics because guidance developed as part of this research transformed the method to a series of tables. These tables provide the designer with a set of combinations of main and tributary stream conditions, each having a specified joint probability. Each combination is analyzed hydraulically to determine the critical design condition.

The second method is the total probability method, which is based on the total probability theorem. The key element of this part of the research was the development of a series of conditional probability matrices for application within the context of the total probability method. Four sets of matrices for high, moderate, mixed, and poor correlation are presented. Full implementation of the total probability method involves evaluating several hydraulic scenarios to generate an exceedance probability curve for stage at the design location. This method may be preferred over the copula method if the objective is to evaluate several alternatives over a range of possible occurrences in the context of a cost-benefit analysis.

The preliminary steps in the application of either the copula or total probability methods are common to both methods:

1. Specify the design objective. The designer might be interested in determining stage and/or velocity at the design location for a given joint annual exceedance probability (AEP) or return period. The desired design event or range of events is specified by the designer.
2. Determine if the project site is within the influence reach. If not, joint probability design techniques are not required.
3. Select the joint probability design method: copula-based or total probability.

If the copula-based method is selected, the design process continues with the following steps:

1. Assess site-specific dependence (correlation) between the main and tributary streams as measured by Kendall's τ. Tools for estimating Kendall's τ have been developed as part of this research.
2. Determine possible flow combinations on the main and tributary streams for a selected joint exceedance probability and estimated dependence parameter.
3. Estimate the discharges for the identified flow combinations.
4. Analyze the hydraulics of the system using the flow combinations to determine the critical combination, that is, the one with the most severe (conservative) conditions for the site. This is the appropriate design condition for the design joint annual exceedance probability/return period.

If the total probability method is selected, the design process continues with the following steps:

1. Construct the marginal exceedance frequency curve for the main stream. Index values of the main stream discharge, $Q_{M,I}$, may be used, that is discreet values of Q_M that represent a range of Q_M and the corresponding probability of occurrence for each range, $P(Q_{M,i})$.
2. Select the conditional probability matrix for the tributary stream stage, y_T, for each value of $Q_{Mi,i}$. The result is the conditional probability of the tributary stage given a main stream discharge, $P[y_T|Q_{Mi,i}]$. These matrices have been developed as part of this research along with guidance for the designer on the selection of the appropriate matrix for a given site.
3. Develop the linking relationship between main and tributary stream flows and the resulting stage. The linking relationship may take the form of a table of stages computed for various combinations of Q_T and Q_M. A HEC-RAS model of the confluence and design location is a likely tool.
4. Compute the total probability of occurrence $P[y_T]$ for each stage computed in the linking relationship.
5. Construct the total probability curve. Select the stage corresponding to the selected design event.

Supporting data and analyses are presented in several appendices. The appendices include a literature review summary (Appendix A) and a description of the databases (Appendix B). Background on bivariate distributions and copulas (Appendix C), the total probability method (Appendix D), and other strategies (Appendix E) are provided. Appendix F provides descriptions of the supporting data analyses.

Appendix G contains an applications guide to demonstrate specific design steps for the copula-based and total probability methods. Application of the methods is generally limited to watershed pairs satisfying three conditions: 1) neither watershed in the pair is substantially affected by regulation, 2) the sum of the watershed areas must be less than 9,000 mi^2, and 3) the smaller watershed of the pair must have a drainage area greater than 1 mi^2. In some cases, the methods developed in this research might be applied beyond these limits if the designer considers the risks and benefits of such application in light of the alternative tools that might be available.

Appendix H provides a step-by-step applications guide in a user friendly format. It is intended to be applied to most design situations and objectives, except for situations of high sensitivity or vulnerability either for the infrastructure itself or for the traveling public.

1. Background

Highway drainage structures are often located near the confluence of two streams where they may be subject to inundation by high flows from either stream. These structures are designed to meet specified performance objectives for floods of a specified return period, e.g., the 100-year flood. Because the flooding of structures on one stream can be affected by high flows on the other stream, it is important to know the relationship between the joint exceedance probabilities on the confluent stream pair, i.e., the joint probability of the coincident flows.

Accurate estimates of the joint probability of coincident flows at stream confluences are a crucial element in the design of efficient and effective highway drainage structures. No accurate generally accepted estimation procedure for determining coincident flows exists for use in the design of highway structures. A practical procedure for the determination of joint probabilities of design flows at stream confluences is needed and is the subject of this report.

1.1 Problem Statement

The joint probability question arises when a structure is located where the hydraulic behavior of some combination of the main and tributary streams may result in critical hydraulic design conditions. Specifically, there is a portion of the tributary stream that is influenced by both the discharge of the tributary stream and the backwater caused by the main stream. This is referred to as the influence reach. The location of the structure within the influence reach, as well as the joint hydrologic behavior of the confluent streams, will determine the importance of the confluent streams on the appropriate design conditions for the structure.

Three variations of relative importance of the design location are illustrated in Figure 1.1: A) structure very close to the confluence, B) structure some distance, x, from the confluence, and C) structure beyond the influence reach. The relative influence of the main and tributary streams varies depending on the design location and the watershed characteristics. Although both velocity and stage may be of interest to the designer, the stage is the focus of this research.

For a given distance from the confluence, x, there is a unique maximum for stage, y, for a given probability of exceedance at the design location. Defining the following variables:

x	=	distance from confluence
X_{max}	=	maximum range of the zone of influence (influence reach)
y	=	stage
Q_M	=	discharge on the main stream
Q_T	=	discharge on the tributary

For the three cases, the stage may be stated as a function of the following:

Case A) x/X_{max} is close to zero,

$$y_A = f(Q_M+Q_T) \tag{1.1}$$

Case B) $0 < x/X_{max} < 1$,

$$y_B = f(Q_M+Q_T, Q_T) \tag{1.2}$$

Case C) $x/X_{max} > 1$,

$$y_C = f(Q_T) \tag{1.3}$$

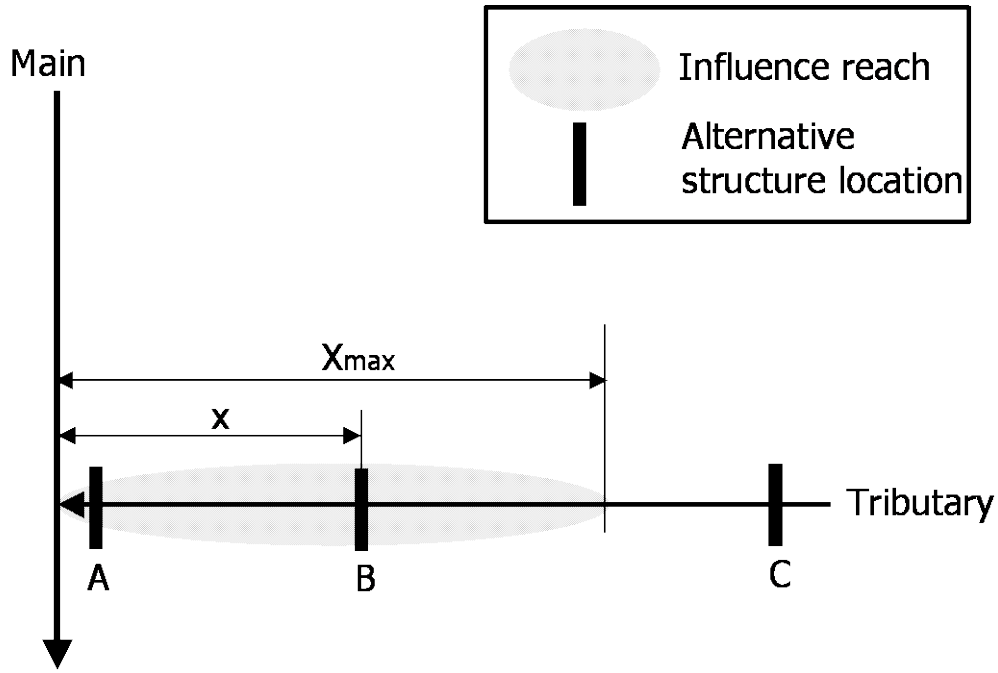

Figure 1.1. Problem Definition Schematic

For Case A, the structure is sufficiently close to the confluence such that the stage is determined by the sum of the flows from the main and tributary streams. This case essentially reduces to an analysis of the flow just downstream of the confluence; therefore, this case is not of great interest for this research.

Case B represents the joint probability problem and the primary focus of this research. The objective in this case is to find the stage, y_B, which corresponds to the joint exceedance probability appropriate for the design problem. The stage at location B at any time is a function of the flow at the confluence, which establishes the downstream control elevation, and the flow in the tributary, which determines the water surface profile from the downstream control to the design location. Alternatively, the stage at location B can be considered a function of the main and tributary stream discharges.

For Case C, the structure is outside of the influence reach and is not representative of the joint probability problem so there will be no further discussion of this case. Case C represents the typical univariate riverine hydraulic design condition.

It is assumed in the characterization of the problem that the design location and, therefore, the influence reach are on the tributary (smaller) watershed. However, the analysis is analogous if the design location is on the main stream.

Considering Case B in more detail, each possible combination of main and tributary stream discharge has a joint annual exceedance probability (AEP). In standard univariate analysis (considering a single stream) a more extreme event is simply an event that is larger than the event associated with the given AEP. In a bivariate analysis a more extreme event is one where the main and tributary stream discharges are both larger (one can be equal) than given values of the main and tributary stream discharges, respectively. If Q_{M1} and Q_{T1} represent one combination of main and tributary stream discharges and Q_{M2} and Q_{T2} represent a second combination, the first combination is considered more extreme if:

$$Q_{M1} \geq Q_{M2} \text{ and } Q_{T1} \geq Q_{T2} \tag{1.4}$$

where,

Q_M = Discharge on the main stream

Q_T = Discharge on the tributary stream

Given that each combination of main and tributary stream discharges has a joint AEP, there is a series of combinations that form an annual exceedance probability isoline as shown in Figure 1.2. Characteristic of the exceedance probability isoline is that all points on the line have the same joint exceedance probability. A designer will apply estimated discharge pairs from various points on the isoline to determine an appropriate water surface elevation or velocity to consider in the design. Although several combinations of discharge pairs may have the same exceedance probability, the one that should be used for design will depend on the site location within the influence reach. Finding that critical combination for a given site may involve selecting two or more potential combinations from the appropriate isoline and selecting the one that exhibits the most severe consequences at the site. Developing the tools to accomplish this was the objective of this research.

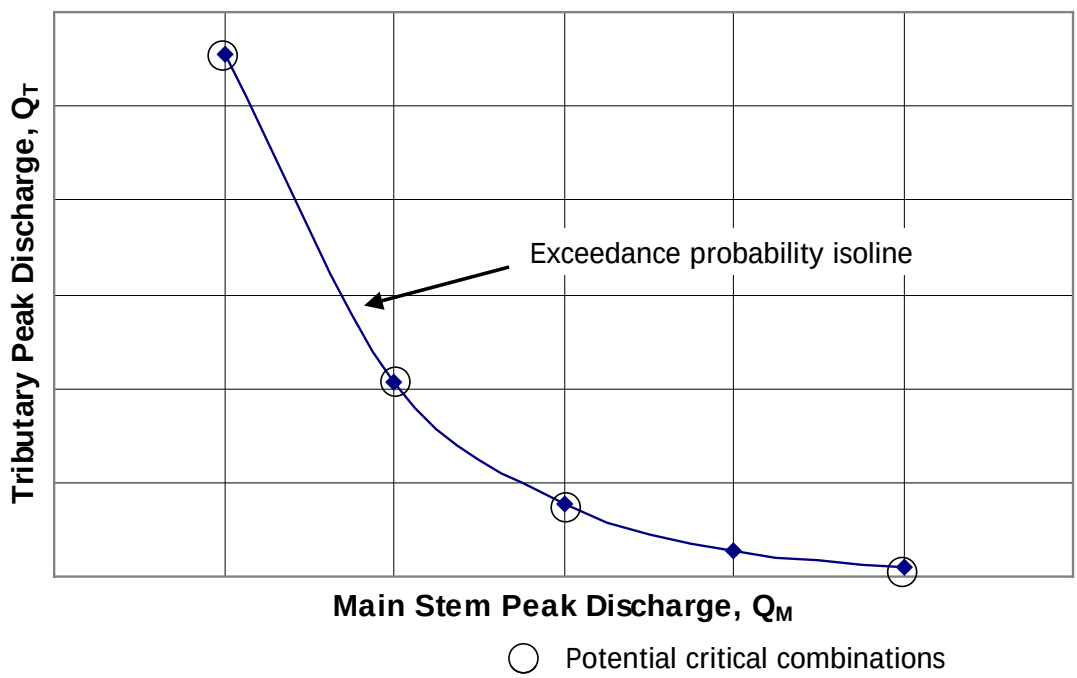

Figure 1.2. Isoline for Annual Bivariate Exceedance Probability

1.2 Literature Review

The literature was reviewed to identify potential strategies for addressing the research objective and to find applications in other fields, which might be transferable to hydrology. The review was focused in four areas: 1) bivariate distributions, 2) total probability method, 3) related joint probability topics, and 4) storm cell characterization. Appendix A provides a summary of the literature review.

1.3 Research Objectives and Scope

The objective of this research was to develop practical procedures for estimating joint probabilities of design coincident flows at stream confluences and guidelines for applying the procedures. The scope of this research was limited to riverine areas and did not include consideration of coastal areas.

2. Research Method

The research objectives were pursued through the identification of potential strategies and tools for application to the joint probability problem, development of a gage pair database to test various methods, and preparation of recommendations for application. In this section, the work plan is described along with the potential strategies that were investigated. The development of the gage pair database is also described.

2.1 Work Plan

The research work plan was divided into two phases and nine tasks. At the conclusion of Phase I, an interim report and revised Phase II work plan was prepared. This report marks the conclusion of Phase II.

2.1.1 Phase I

2.1.1.1 Task 1. Critically Review Literature/Summarize Procedures

The first task was to critically review the literature pertaining to joint probabilities for flows at stream confluences. The main purpose of this review was to identify and summarize procedures used to estimate joint probabilities of coincident events. The review included the hydrologic literature as well as literature that might address coincident events, for example, traffic-flow, seismic analyses, and other subject areas that might be characterized by stochastic processes.

2.1.1.2 Task 2. Critically Review Procedures

Next was to critically review the procedures identified in Task 1 and evaluate the suitability of promising procedures for use in hydrologic modeling by highway agencies. Following this, the data requirements, data availability, and level of effort needed to apply each procedure were determined.

2.1.1.3 Task 3. Identify Confluent Stream Sites

Using available stream gage and other hydrologic data, confluent stream sites with concurrent data of sufficient record length to estimate joint probabilities were identified. It was intended that the sites be regionally diverse throughout the United States and include a wide range of basin characteristics and hydrologic conditions.

Based on the hypotheses postulated in Tasks 1 and 2, along with the findings of those two tasks, appropriate stream sites, and types of precipitation data, were to be identified for potential use. Depending on the potential utility of radar data, sites with useful radar data might be identified.

2.1.1.4 Task 4. Propose Procedures

Task 4 was to propose the development of one or more innovative procedures or refinements of existing procedures for estimating joint probabilities of design coincident flows. For each proposed procedure the availability of data, ease of use, statistical robustness, and accuracy of the procedure were to be addressed. For the proposed procedures identified, the intent was to seek balance between data availability, ease of use, statistical robustness, and accuracy.

2.1.1.5 Task 5. Submit Interim Report

An interim report documenting the work completed in Tasks 1 through 4 was prepared. The report included an updated, detailed work plan for completing Phase II of the research as a separate appendix.

2.1.2 Phase II

2.1.2.1 Task 6. Collect/Compile Data

Stream gage and other hydrologic data from the sites identified during Task 3 were collected and compiled. Quality control for the data was developed to ensure the integrity of the data and the reproducibility of results.

2.1.2.2 Task 7. Develop Practical Procedures

Using the Task 6 data, Task 7 was to develop the practical procedures proposed in Phase I for estimating joint probabilities of design coincident flows at stream confluences for riverine areas. The model types that seemed likely for investigation prior to conducting Phase I included the following:

1. Bivariate probability distributions
2. Tabular summaries
3. Regression analyses
4. Adaptations to the Total Probability Method

Other model types revealed in the literature search and that survived evaluation in Task 4 would be evaluated in this task.

2.1.2.3 Task 8. Validate Procedures

Task 8 was to validate and refine the Task 7 procedures. Development of application guidelines was also a part of this task. The guidelines are intended to address the regional applicability of the procedures.

2.1.2.4 Task 9. Final Report

A final report documenting the entire research effort is prepared in this task. The final report should include illustrative examples showing the application of the procedures and the ranges of variables. This document is the final report.

2.2 Potential Strategies

Although four potential strategies were listed in the original work plan in Task 7, the literature search and subsequent evaluation resulted in seven strategies that exhibited potential utility. They are:

1. Bivariate probability distributions
2. Univariate probability distributions with linking copulas
3. Total probability method
4. Regression analyses
5. Marginal analysis

6. Synthetic storm cell/runoff modeling
7. Tabular summaries

Each is described briefly in the following sections.

2.2.1 Bivariate Probability Distributions

The use of bivariate probability distributions for analyzing the joint probability of coincident flows is an extension of the established procedure of using univariate distributions to analyze gage data for a single stream. Probably the most well known example of this is the use of the Pearson III distribution when analyzing log-transformed annual peaks, that is, Log Pearson III analysis. In that case, discharge data are used to develop the distribution parameters and then the fitted distribution is used to define events with a given exceedance probability. The univariate distribution for the main or tributary streams is referred to as the marginal distribution in the context of bivariate distributions. A description of selected univariate distributions is provided in Appendix C.1.

A key element in the evaluation of bivariate distributions is a determination of which bivariate distributions, if any, fit coincident stream flows. Bivariate distributions contain the complete structure of the marginal distributions and the correlation between the marginal distributions. Although their distribution parameters may differ, the marginal distributions that form the bivariate distribution are the same.

Direct application of bivariate distributions to confluent stream flows was not found in the literature review, but applications to flood peaks and volume as well as other paired data were identified. Examples of the Normal (transformed by logarithms or Box-Cox), Gumbel, and generalized extreme value (GEV) bivariate distributions (Bogardi, et al., 1975; Sackl and Bergmann, 1987; Shiau, 2003; Morris and Calise, 1987; Raynal and Salas, 1987; Loganathan, et al., 1987) were published in the literature. Weibull and exponential bivariate distributions have been applied, but generally not to flow data. See, for example, Prior-Jones and Beiboer (1990). Based on the literature review, the bivariate Gumbel and normal (log) distributions were further investigated. Detailed analysis and application of the bivariate distributions to the confluent flooding problem is presented in Appendix C.2.

2.2.2 Univariate Probability Distributions with Linking Copulas

An alternative method is to apply univariate marginals with the dependence structure between the two marginal distributions expressed through a copula. A useful feature of copulas is that they are independent of the marginal distributions. Although not applied to flows at confluent streams, water resources applications of this method are illustrated by De Michele, et al. (2005), Favre, et al. (2004), and Zhang and Singh (2006).

The Archimedean copula family is desirable for hydrologic analyses because it is easily constructed, includes a large family of copula types, and applies whether the correlation is positive or negative (Nelsen, 2006). The Archimedean family includes the Gumbel-Hougaard, Ali-Mikhail-Haq, Frank, Cook-Johnson copulas, which are assessed in this research, and other copula families. Shiau, et al. (2006) notes the common use of the Clayton copula in hydrology. One application was the relation between flood peak and flood volume. Mathematical descriptions of selected Archimedean copulas are presented in Appendix C.3.

Based on the literature, several univariate marginal distributions have been used with copulas including the Normal (transformed by logarithms or Box-Cox), Gumbel, and generalized extreme value (GEV). However, because copulas can be used with any marginal distribution and the

Pearson Type III (log transformed) distribution is commonly applied to annual peak flow data, the Log Pearson III method was also considered for this research.

2.2.3 Total Probability Method

The total probability method, based on the total probability theorem, has been applied at river confluences (e.g. Dyhouse, 1985; Pingel and Ford, 2004). For application of the total probability theorem for joint probability, one can use the procedure described in EM 1110-2-1415 (USACE, 1993).

The procedure is applied for two conditions: main stream flow as the dominant variable and tributary flow as the dominant variable. The condition yielding the higher stage for the probability of interest is taken as the design condition. In the development that follows, the dominant variable is assumed to be the main stream flow, Q_M. The procedure also requires that there is some function, referred to as a linking relationship, that is used to calculate the tributary stage, y_T, such that:

$$y_T = f(Q_T, Q_M) \tag{2.1}$$

where,

y_T	=	Tributary stage
Q_T	=	Tributary flow
Q_M	=	Main stream flow

Typically, this linking relationship is provided with a HEC-RAS model of the confluence with Q_T and Q_M as inputs. Using these definitions, the total probability theorem may be stated as follows:

$$P[y_T] = \sum_{i=1}^{n} P[y_T \mid Q_{M,i}] P[Q_{M,i}] \tag{2.2}$$

where,

$P[y_T]$	=	probability that y_T will occur
$P[y_T \mid Q_{M,i}]$	=	probability that y_T will occur given $Q_{M,i}$
$P[Q_{M,i}]$	=	probability that $Q_{M,i}$ will occur
i	=	index

The values for $P[Q_{M,i}]$ may be developed from a flow-duration curve, but for the ungaged situations anticipated by this research, it would be developed from the marginal distribution for the main stream annual peak flow.

The total probability method is potentially more computationally intensive than the methods using the bivariate distributions or the univariate distributions with copulas. However, the total probability method produces a complete probability distribution function for tributary stage, y_T, as described in Equation 2.2. The design exceedance probability for the site is used to determine the design stage. By contrast, the bivariate distribution and copula methods assume that a given joint exceedance probability for flow yields (by evaluation of several critical combinations) the equivalent exceedence probability for stage.

2.2.4 Regression Relations

This section addresses application of regression relations to directly determine appropriate design conditions for the joint probability at stream confluences. A method for regression equations might be to calculate the desired probability of exceedance on the main stream, given the joint probability of exceedance, from the probability of exceedance on the tributary and appropriate watershed and meteorological variables. For example:

$$P_{e,M} = f(P_{e,T}, \text{watershed variables, meteorological variables}) \quad (2.3)$$

where,

$P_{e,M}$ = probability of exceedance to be used on the main stream (M)

$P_{e,T}$ = probability of exceedance desired for the tributary (T)

A series of critical combinations, with the same joint exceedance probability, would be derived from Equation 2.3 by supplying the needed watershed and meteorological variables along with a series of values for $P_{e,T}$. This would result in a series of $P_{e,M}$ values.

Research efforts are needed to develop the type of equations represented by Equation 2.3. However, the research team was not optimistic that this strategy would be successful and focused project resources in other areas. No further attention was given to this strategy.

2.2.5 Marginal Analysis

Marginal analysis is a strategy based on considering the probability of exceedance of the combined tributary and main stream flow at the confluence and then allocating this flow, in various combinations, between the confluent streams. The concept is represented by the following equation:

$$Q_M = Q_{M+T,Pe} - Q_T \quad (2.4)$$

where,

Q_M = residual flow on the main stream

Q_T = flow on the tributary

$Q_{M+T,Pe}$ = flow below the confluence of the two streams for exceedance probability P_e

$Q_{M+T,Pe}$ is calculated based on any applicable hydrological technique using the combined (main and tributary) watershed characteristics. Equation 2.4 is then used successively with a series of Q_T values of known exceedance probability to calculate the corresponding values of Q_M. For a given P_e, the values of Q_M would be a maximum of $Q_{M,Pe}$ and range downward and the values of Q_T would be a maximum of $Q_{T,Pe}$ and range downward. These paired values are then evaluated to determine which results in the most severe circumstances.

Application of this strategy requires no further research. It is also not useful for addressing the joint probability problem as described by Case B (Equation 1.2 in Section 1.1). By focusing attention on discharge probabilities at the confluence, it does not consider other possibilities on the influence reach other than those that sum to a selected quantity at the confluence.

Furthermore, for Case A (Equation 1.1 in Section 1.1) it is immaterial to what extent the tributary and main stream contributes to a given return period flow at the confluence. In this case, all that matters is the flow, and therefore the stage, at the confluence. Therefore, the marginal analysis procedure adds no new insight to that assessment and no further consideration will be made of this method in addressing coincident flooding. An application of the method is illustrated on the

prototype watersheds as summarized in Appendix E.1. The technique may have utility for other applications.

2.2.6 Synthetic Storm Cell/Runoff Modeling

This strategy employs traditional rainfall/runoff modeling tools based on the selection of an appropriate storm cell to drive the modeling process. The strategy involves selecting a storm or storms appropriate for the location and exceedance probability of interest. The storm cell characteristics would include the storm size, spatial distribution, and temporal distribution. For practicality, it may be necessary to assume the storm is stationary.

Next, the design storm cell would be applied as an input to a rainfall/runoff model. Any suitable runoff model capable of accepting spatially and temporally varied rainfall could be used. HEC-HMS is an example of a widely available model meeting this criterion. Assuming the storm is stationary, an assumption regarding the location of the storm cell with respect to the watershed is required. A possible assumption is that the centroid of the storm cell is located above the centroid of the combined watershed areas.

Research is needed to establish the appropriate storm cell characteristics, size, duration, intensity, etc., for given probabilities of exceedance and geographic region. Analysis of radar data may be a tool for achieving this. Research is also needed to determine if more than one cell type, for a given location and probability of exceedance, could result in the desired runoff conditions. Working with multiple storm types would represent a significant investment of time. Even with multiple storm types, the designer must accept the assumption embedded in all rainfall/runoff modeling methods: the T-year precipitation results in the T-year runoff.

This method relies heavily on the success in deriving appropriate storm cell(s) for design. Storm cells differ with respect to the precipitation mechanisms and region of the country although these distinctions are not completely independent. Selected parts of the literature that address this topic are described in Appendix A.4.

Because of the complex nature of storms, the research team was not optimistic that characterization of storm cells for the purpose of analyzing confluent streams was an attainable goal within the context this research effort. Therefore, no further consideration of this method was undertaken.

2.2.7 Tabular Summaries

Tabular summaries are one of the most common tools currently in use for addressing the joint probability of coincident flows. Such tables may simply be a summarization of the results of a more complex methodology and prepared for ease of application or they may represent the culmination of substantial data analyses with no underlying causal linkages.

Table 2.1, taken from AASHTO (2000), is an example of a tabular summary that is widely used. However, no documentation of the source of this table was found, though it is believed to have been developed by the Norfolk District of the U.S. Army Corps of Engineers in the mid 1970s. Because of the uncertainty regarding the source of Table 2.1, it is unclear whether it should have the wide application it does. Application of tabular summaries requires determination of necessary watershed characteristics to enter the table appropriate for the hydrologic region. In the case of Table 2.1, the only required data is the area ratio. The table is presumed to be equally applicable across the country. The joint probability is then determined based on the table values.

Table 2.1. Tabular Summary of Joint Probabilities of Coincident Flow

Area Ratio	Frequencies for Coincident Flow			
	10-year Design		100-year Design	
	Main Stream	Tributary	Main Stream	Tributary
10,000 to 1	1	10	2	100
	10	1	100	2
1,000 to 1	2	10	10	100
	10	2	100	10
100 to 1	5	10	25	100
	10	5	100	25
10 to 1	10	10	50	100
	10	10	100	50
1 to 1	10	10	100	100
	10	10	100	100

To implement a tabular strategy this research must analyze, choose, and develop empirical relations similar to those shown in Table 2.1. A tabular method may be appropriate for regions or situations where statistical relations cannot be quantified, but appear to exist.

2.3 Database

This research included the development of three databases. The first database was a collection of confluent gage pairs from the coterminous United States. A second database included instantaneous flow records from the USGS Instantaneous Data Archive. The third database included watershed and meteorological data associated with the confluent gage pairs.

2.3.1 Gage Pairs

A primary resource for the research was a database of confluent gage pairs from the set of USGS stream gages within the coterminous U.S. The screening criteria to select the gage pairs were as follows:

1. Geographic distribution throughout the coterminous 48 states
2. Area ratios within the range of 1:1 through 1,000:1
3. Drainage areas within the range of 1 to 10,000 mi^2
4. 25-yr minimum record length
5. Gage location captures substantial portion of drainage area to the confluence
6. Gage record does not indicate regulation
7. 10-mile maximum distance between gages in a gage pair

The location and distribution of 93 database gage pairs identified is shown on Figure 2.1 and listed in Table B.1 (Appendix B.1). Information for each gage included the main stream station name, main stream drainage area, tributary stream station name, and tributary stream drainage area, as well as the longitude and latitude of each gage.

Figure 2.1. Gage Pairs and Hydrologic Units

As described in Appendix B.1, eight gage pairs were dropped from the database for various reasons, leaving 85 pairs for this research. Drainage area distributions for the main and tributary streams are summarized in Table 2.2.

Table 2.2. Gage Pair Drainage Area Distribution

Area (mi^2)	Number of Main Stream Gages	Number of Tributary Stream Gages
1-10	2	11
10-100	7	31
100-500	42	34
500-1000	12	5
1000-10000	22	4
Total	85	85

For each gage pair, the ratio of drainage areas and distance between gages were computed. The area ratio distribution is summarized in Table 2.3. Because methods based on drainage area ratio are currently in use, it was desirable to secure a dataset with a wide range of representative values.

Table 2.3. Gage Pair Drainage Area Ratio Distribution

Area Ratio	Number of Gage Pairs
1-2	28
2-5	32
5-10	7
10-100	9
100-500	9
Total	85

2.3.2 Instantaneous Data

Instantaneous data (15 to 60 minute increment) were available for watersheds in 15 states at the time the dataset was assembled. The intent was to collect and review a subset of data from 20 gages in the Instantaneous Data Archive (IDA). The Panel recommended collecting these data from gage pairs rather than individual gages so that pair wise comparisons could be conducted. Data were collected for the 11 pairs (22 gages) summarized in Appendix B.2.

The selected pairs were chosen to emphasize the smaller watersheds in the dataset while maintaining geographic diversity. It was believed that the instantaneous data would be less informative for the larger watersheds because they do not respond as rapidly as smaller watersheds.

2.3.3 Watershed and Meteorological Data

Complementing the discharge data for the gage pairs were various watershed and meteorological data elements that were compiled for this study.

2.3.3.1 Watershed Data

Watershed characteristics sought for the study are listed in Table 2.3. The first column lists the characteristics and the second column lists the sources of data from which these characteristics were computed.

Table 2.4. Sources for Watershed Characteristics

Watershed Characteristic	Source
Drainage area	USGS stream gages dataset
Latitude/longitude of watershed centroid	Computed from NHDPlus dataset
Maximum flow length	Computed from NHDPlus dataset
Watershed channel slope	Computed from NHDPlus dataset
Distance between centroids for each gage pair	Computed from NHDPlus dataset

Gaged drainage area values were obtained from a USGS spatial database (Stewart, et al., 2006) of 23,426 current and historical stream gages. Additional watershed characteristics were derived from NHDPlus (EPA and USGS, 2007). NHDPlus includes stream network features from the National Hydrography Dataset (NHD) and elevation data from the National Elevation Dataset (NED). Stream networks are based on the 1:100,000 scale (medium resolution) NHD and elevation data are based on 30-meter grid resolution NED. A complete description of NHDPlus is provided in the NHDPlus User Guide (EPA and USGS, 2007). NHDPlus data were obtained from http://www.horizonsystems.com/nhdplus/. Additional details on the derivation of watershed variables are provided in Appendix B.3.

A time of concentration for each watershed was estimated using the Kirpich formula. However, many of the watersheds did not exhibit characteristics appropriate for the Kirpich formula. For example, many were much larger than is appropriate for the formula. Nevertheless, it was considered useful to explore measures of watershed "size" other than drainage area to use in the evaluation of correlation structures between gage pairs. Since the Kirpich formula uses a channel length and slope in computing a response time, the Kirpich equation was used for all watersheds, but was renamed "slope-length parameter" so that the quantity would not be confused with time of concentration. As described later, this parameter did exhibit utility in the research. See Appendix B.3 for more information.

2.3.3.2 Meteorological Data

Two meteorological characteristics were compiled: mean annual precipitation at each watershed centroid and the 2-yr 24-h precipitation for each watershed centroid. These data were retrieved from the USGS database containing watershed attributes for approximately 13,000 unregulated gages. Not all attributes are available for all gages, but the attributes of interest are: mean annual precipitation (PRECIP) and 2-yr/24-hour rainfall intensity (I24100).

When unavailable from the USGS database, mean annual precipitation (MAP) was estimated from the precipitation map in NOAA "Climatography of the U.S. No. 81 – Supplement #3" that covers the period from 1961 to 1990.

When unavailable, the 2-yr 24-h precipitation was retrieved from NOAA Atlas 14 (Bonnin, et al., 2006a, 2006b) for areas covered by that tool. For those locations not covered by NOAA Atlas 14, NOAA Atlas 2 (NOAA, 1973) was used in the western U.S. and TP40 (NOAA, 1961) in the eastern U.S.

3. Data Evaluation

Coincident sets of flow data for each gage pair were required for the study. Because flooding data are typically analyzed based on annual rather than partial duration series, the use of the annual series was adopted. This research was primarily based on two sets of annual peak series flow data for each gage pair. These datasets are referred to as the "Peaks on Main" (POM) dataset and the "Peaks on Tributary" (POT) dataset. They were developed for each gage pair by first deriving an annual peak series on one of the gages (the primary gage). For the POM dataset, the main stream (larger drainage area) gage was the primary gage and for the POT dataset the tributary gage was the primary gage. Then, each peak value recorded for the primary gage was paired with the "coincident" value occurring at the same time on the other (complementary) gage in the pair.

Instantaneous peak values at a gaging station are typically used to create annual series of peak flows. However, a potentially significant limitation of creating gage pair datasets based on instantaneous data (readings typically ranging from every fifteen minutes to every hour) is the relative lack of such data at most gaging locations. Another issue was how "coincident" flow should be defined. Should the coincident flow be defined as flows at both gages occurring at the exact same time or should "coincident" be defined to be within some time window in the hope of capturing a local hydrograph peak on both gages in the pair?

The research team decided to use daily flow data because there simply were not enough instantaneous data available from a large enough set of gages over sufficient periods of record to quantitatively evaluate joint probability methods. With the decision to use daily data, "coincident" was defined as the daily flows on both gages occurring on the same day. The justification and potential effects of these decisions are described in this chapter.

The choice to use annual series daily data was driven by the broad availability of daily data and the need for a consistent definition of a coincident occurrence. However, there was concern that the use of daily flow data would not represent the true relation between main and tributary stream flows, especially for smaller watersheds. Three evaluations were conducted to address the following issues:

1. Annual versus partial duration series
2. Definition of a coincident peak
3. Mean daily versus instantaneous data

The assessments were primarily based on prototype gage pairs (a subset of the 85 pair database) that are representative of the database and have sufficient data and data types to support the analyses. Three gage pairs (GP) representing a range of watershed areas and distinct regions of the country were initially selected: New Jersey (GP06), Washington (GP43), and Texas (GP76). These watershed pairs are shown in Figures 3.1, 3.2, and 3.3, respectively.

For some of the analyses, the group of prototype watershed pairs was expanded, primarily to include a greater representation of smaller watersheds. The prototype gage pairs are summarized in Table 3.1. Commentary on each follows:

- Gage Pair 05 (GP05): Tributary gage serves one of the smallest drainage areas in the gage pair database. This pair has one of the longest instantaneous data records available.
- Gage Pair 06 (GP06): Watersheds represent an intermediate size range.

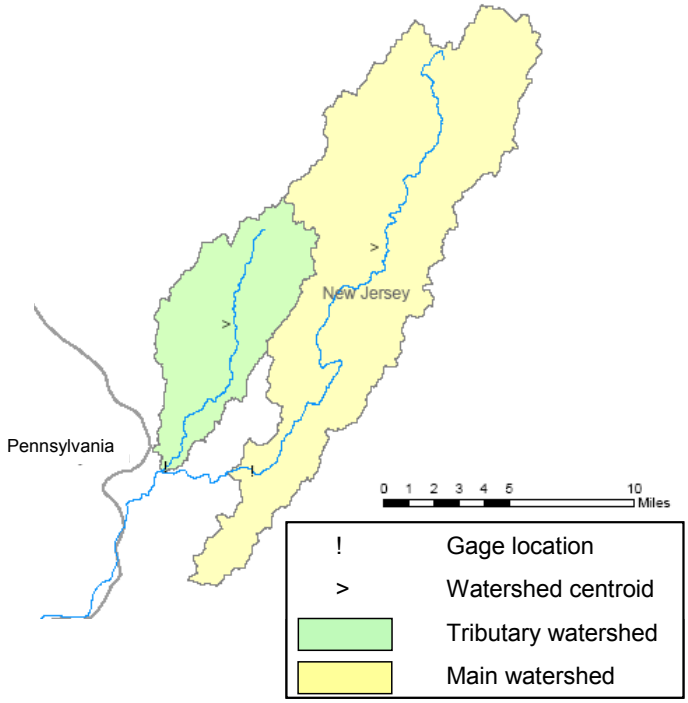

Figure 3.1. Gage Pair 06 (New Jersey)

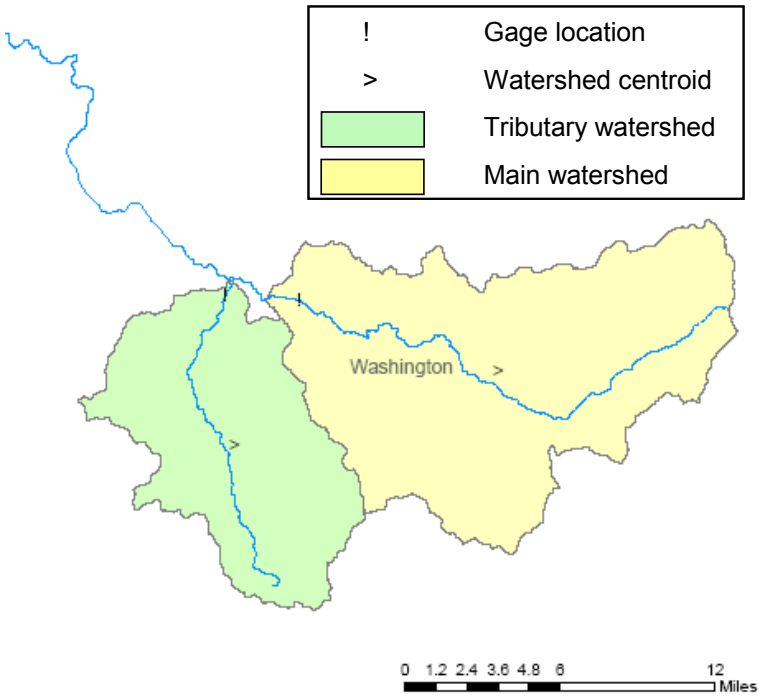

Figure 3.2. Gage Pair 43 (Washington)

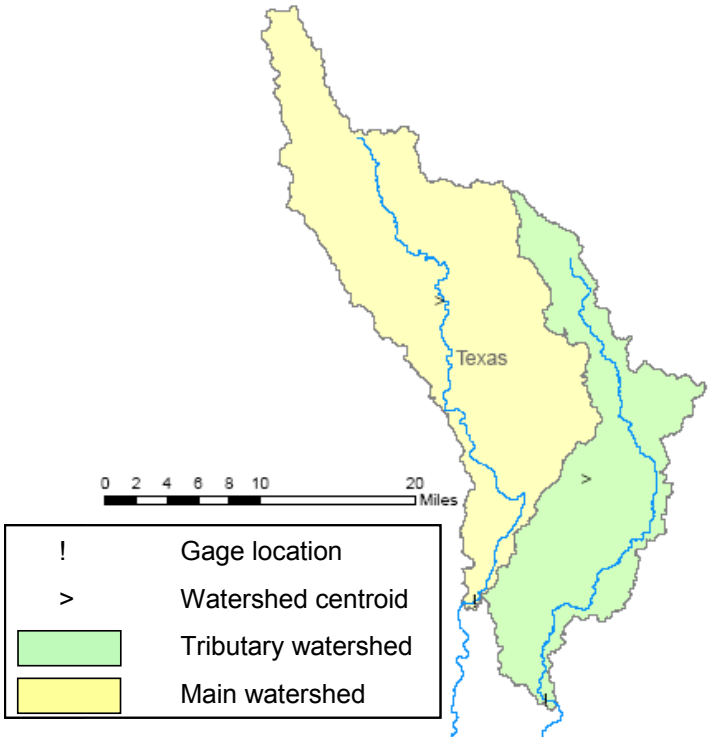

Figure 3.3. Gage Pair 76 (Texas)

Table 3.1. Prototype Gage Drainage Areas

Gage Pair	State	USGS Gage ID	Drainage Area (mi^2)
05	New Jersey	01403060	784
05	New Jersey	01403150	2
06	New Jersey	01445500	106
06	New Jersey	01446000	36.7
08	Maryland	01590000	8.5
08	Maryland	01590500	6.9
21	Michigan	04140500	117
21	Michigan	04141000	1.2
43	Washington	12082500	133
43	Washington	12083000	70.3
76	Texas	08164450	289
76	Texas	08164503	178

- Gage Pair 08 (GP08): Both the main and tributary watersheds are relatively small. There are no instantaneous data available for the main stream and less than a year record on the tributary stream.
- Gage Pair 21 (GP21): Tributary gage serves the smallest drainage area in the gage pair database. There are no instantaneous data available for either the main or tributary streams.

- Gage Pair 43 (GP43): Watersheds represent an intermediate size range.
- Gage Pair 76 (GP76): Watersheds represent an intermediate to large size range.

3.1 Annual Versus Partial Duration Series

The use of annual series data dominates flood frequency analyses in the United States, although partial duration series may also useful. When considering whether to focus on annual or partial duration series for this research, two concerns regarding partial duration series became apparent: 1) such data are less widely available and 2) the data are prepared inconsistently from state to state.

The periods of record for the annual and partial duration series for the prototype gage pairs 06, 43, and 76 and the corresponding number of peak observations are summarized in Table 3.2. In all cases, the partial duration series data available had a shorter period of record. At one gage, 08164503 from Texas, the partial duration series record contained only annual peaks raising doubt as to whether or not this dataset represents a true partial duration series. The other five partial duration series records have between 1.9 to 3.5 observations per year.

Table 3.2. Instantaneous Annual and Partial Duration Series Availability

	Main Gage	Tributary Gage	Coincident[1] Peaks	Notes
Gage Pair 06	01445500	01446000		
Annual Duration Series	WY 1922-2006 (n=85)	WY 1923-1995, 2003-2006 (n=77)	65% (n=50)	
Partial Duration Series	WY 1922-1996 (n=196 over 75 yrs, 2.6/yr)	WY 1923-1995 (n=141 over 73 yrs, 1.9/yr)	36% (n=71)	Some years in both gages only have annual peaks.
Gage Pair 43	12082500	12083000		
Annual Duration Series	WY 1943-2006 (n=64)	WY 1943-2006 (n=64)	53% (n=34)	
Partial Duration Series	WY 1943-1995 (n=187 over 53 yrs, 3.5/yr)	WY 1943-1995 (n=136 over 53 yrs, 2.6/yr)	56% (n=105)	
Gage Pair 76	08164450	08164503		
Annual Duration Series	WY 1978-2006 (n=29)	WY 1978-2006 (n=29)	45% (n=13)	
Partial Duration Series	WY 1980-1991, 1995 (n=26 over 13 yrs, 2/yr)	WY 1980-1991, 1995 (n=13 over 13 yrs, 1/yr)	19% (n=5)	Tributary does not appear to have a true partial duration dataset.

1. For the purposes of this table, coincident peaks are said to occur when the annual peak on tributary stream occurs within plus or minus one day of the annual peak on the main stream.

The frequency of “coincident” peaks in the annual and partial duration series data are also summarized in Table 3.2. The definition of “coincident” is treated later in this chapter, but for the purpose of this discussion, a coincident peak is defined as whether the tributary peak in the same dataset occurs within plus or minus one day of the main stream peak. For the New Jersey pair (Pair 06), 50 of the 77 annual peaks (65 percent) met this definition. In the partial duration series, the frequency of coincidence dropped to 71 of 196 or 36 percent. This drop in

coincidence appears to result because portions of the partial duration series record at both gages appear to only contain annual peaks for several years.

For the Washington pair (Pair 43) 34 of 64 annual peaks (53 percent) were coincident. The percentage of coincidence in the partial duration series data was approximately the same at 56 percent (105 of 187). The partial duration series for both gages in this pair appear more complete than for the New Jersey pair.

For the Texas pair (Pair 76), which has the shortest period of record of the three gage pairs analyzed, 13 of 29 annual peaks (45 percent) were coincident. As with the New Jersey pair, the percentage of coincidence dropped for the partial duration series to 19 percent (5 of 26). As was suspected of the New Jersey gage pair, the partial duration series record for the tributary (gage 08164503) does not appear to be a true partial duration series because it only contains annual peak data for the period of record. Therefore, the frequency of coincidence is lower for the partial duration series, at least in part, because of the data are incomplete.

Typically, partial duration series data are not as widely available and have not received the same scrutiny as have annual series data. In the case of the New Jersey and Texas gage pairs, portions of the partial duration series appear to be populated with only the annual peak data for that year. While that is possible for an isolated year, the pattern seems to hold for many consecutive years suggesting a variation in methodology for preparing the partial duration series data from state to state or even gage to gage. By contrast, the Washington gage pair partial duration series data appear to be more consistently compiled. Consequently, the frequency of coincidence does not substantially drop between the annual series and partial duration series as is observed for the New Jersey and Texas gage pairs.

Annual series data are more consistently scrutinized, more widely distributed, and more frequently applied than partial duration series data. In addition, widely applied hydrological methods for design are based on annual series data. Therefore, annual series data are used in this research project.

3.2 Definition of Coincident Flow

A coincident flow pair is the flow on the main stream and the flow on the tributary stream that occur simultaneously. When considering instantaneous flow records (measurements typically at fifteen to sixty minute intervals), the coincident flow at the complementary gage is the flow that occurs at the exact same time as the instantaneous peak is occurring on the primary gage. Similarly, when considering daily average flow records, the coincident flow at the complementary gage is the flow that occurs at the same day as the daily average peak is occurring on the primary gage.

Because stream flow hydrographs may maintain flows close to peak values over a period of minutes, hours, or days, depending on the size and other characteristics of the watershed, as well as the intensity, duration, and spatial extent of the precipitation event, an overly strict definition of “coincident” might miss the true correlative relation between adjacent watersheds. Conversely, an overly loose definition might also result in a misinterpretation of correlation.

To examine the potential implications of alternative definitions of coincident flow, twelve events, four from each prototype gage pair, with both instantaneous and mean daily data were selected for analysis. The instantaneous peak flows are summarized in Table 3.3. For each case, the main stream flow is the annual peak for that water year. In most cases, the tributary flow is also an annual peak, though the peaks do not occur at exactly at the same time. With the exception of the June 23, 2003 event for Gage Pair 06, the peaks occurred on the same calendar day.

Based on a fitted Log Pearson III distribution, the estimated return period for each peak flow for each gage is also summarized in Table 3.3. In some cases, such as the October 9, 2005 event for Gage Pair 06, both flows exhibit an equivalent return period. However, in other cases, such as the October 19, 1998 event for Gage Pair 76, substantially different return periods are observed. The hydrographs for the events listed in Table 3.3 are shown in Figures 3.4a through 3.6d.

Table 3.3. Instantaneous Peak Data for Selected Events

	Main stream			Tributary			
Pair	Date	Peak Flow (ft^3/s)	Return Period (yrs)	Date	Peak Flow (ft^3/s)	Return Period (yrs)	Notes
06	6/22/2003	1040	3	6/23/2003	530	2	Annual peak for both gages
06	9/18/2004	1490	10	9/18/2004	1330	38	Annual peak for both gages
06	4/3/2005	1520	11	4/3/2005	1080	16	Annual peak for both gages
06	10/9/2005	1280	6	10/9/2005	798	6	Annual peak for both gages
43	12/9/1987	9200	4	12/9/1987	7220	7	Annual peak for both gages
43	10/16/1988	4130	<2	10/16/1988	389	<2	Annual peak for main
43	1/9/1990	14500	19	1/9/1990	9440	26	Annual peak for both gages. IDA incomplete for tributary
43	11/24/1990	11000	7	11/24/1990	9030	20	Annual peak for main, partial peak for tributary (partial peak dataset shows 10000)
76	10/14/1997	8760	4	10/14/1997	5860	3	Annual peak for main
76	10/19/1998	63400	>500	10/19/1998	7110	3	Annual peak for both gages (IDA incomplete for main)
76	5/3/2000	2610	2	5/3/2000	1740	<2	Annual peak for main (daily data missing for main)
76	9/1/2001	10000	4	9/1/2001	10600	7	Annual peak for both gages (tributary peak greater than main peak)

The June 2003 event for Gage Pair 06 is illustrated in Figure 3.4a. The lines represent the instantaneous hydrograph and the individual points represent the mean daily flow values. Because of the watershed characteristics (drainage areas for the main and tributary gages are 106 and 36.7 mi^2) and storm characteristics, the instantaneous hydrograph and daily mean values are reasonably consistent. However, neither the instantaneous nor daily peaks are "coincident" in the sense that they did not occur simultaneously even though the hydrographs appear to be generally coincident. Both the instantaneous and mean daily peak for the main stream occurred on June 22, 2003 while both the instantaneous and mean daily peak for the tributary stream occurred the following day.

An October 1988 event for Gage Pair 43 is depicted in Figure 3.5b. In this case, the instantaneous and mean daily peaks for both the main and tributary gages all occur on the same day (October 16). However, the instantaneous peak for gage 1208300 occurs several hours after the instantaneous peak on gage 12082500 with the value at that time at approximately 60 percent of the peak.

Variations in the relative consistency between peaks in both the mean daily data and instantaneous data are shown in Figures 3.4a through 3.6d. In addition, gaps in the instantaneous records are revealed in Figures 3.5c (Pair 43), 3.6a (Pair 76), and 3.6b (Pair 76). Data gaps and short records are two limitations on the use of instantaneous data.

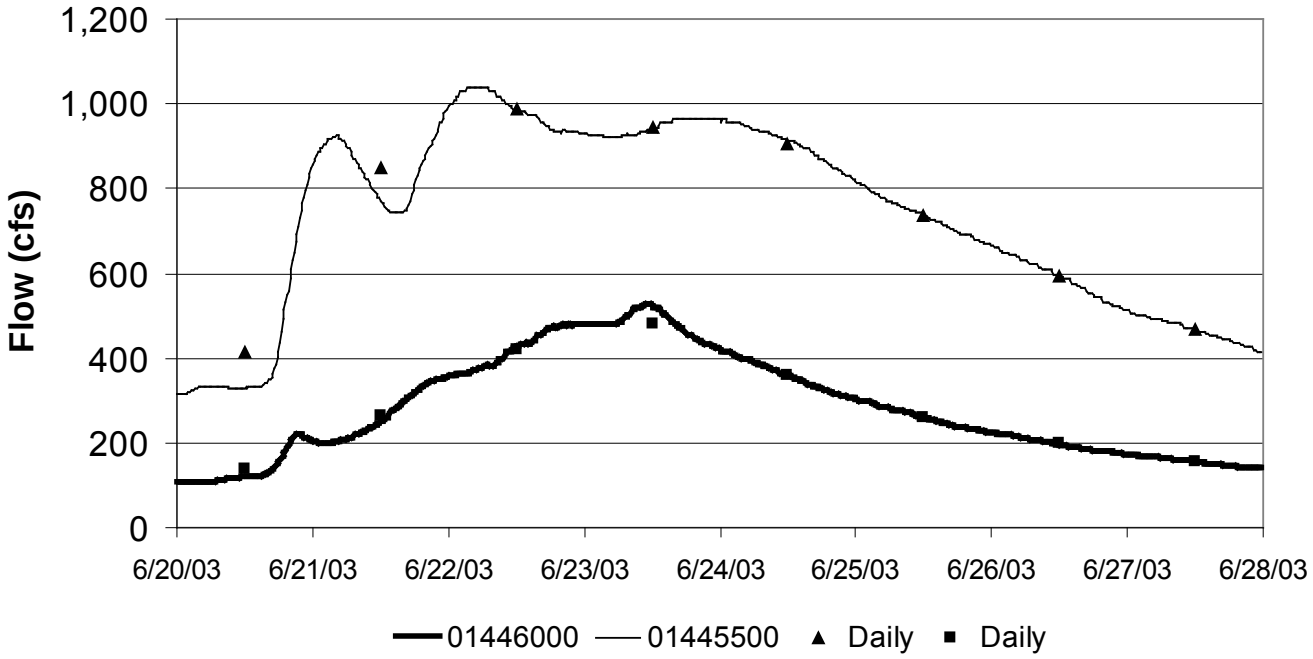

Figure 3.4a. Gage Pair 06 June 20-28, 2003

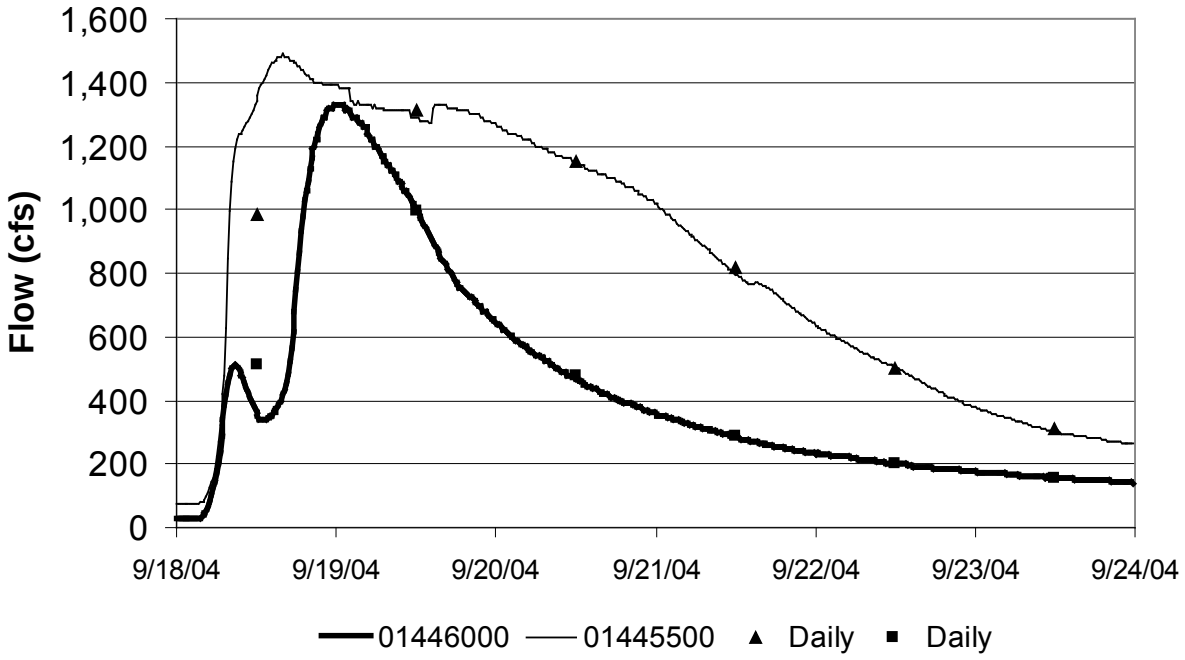

Figure 3.4b. Gage Pair 06 September 18-24, 2004

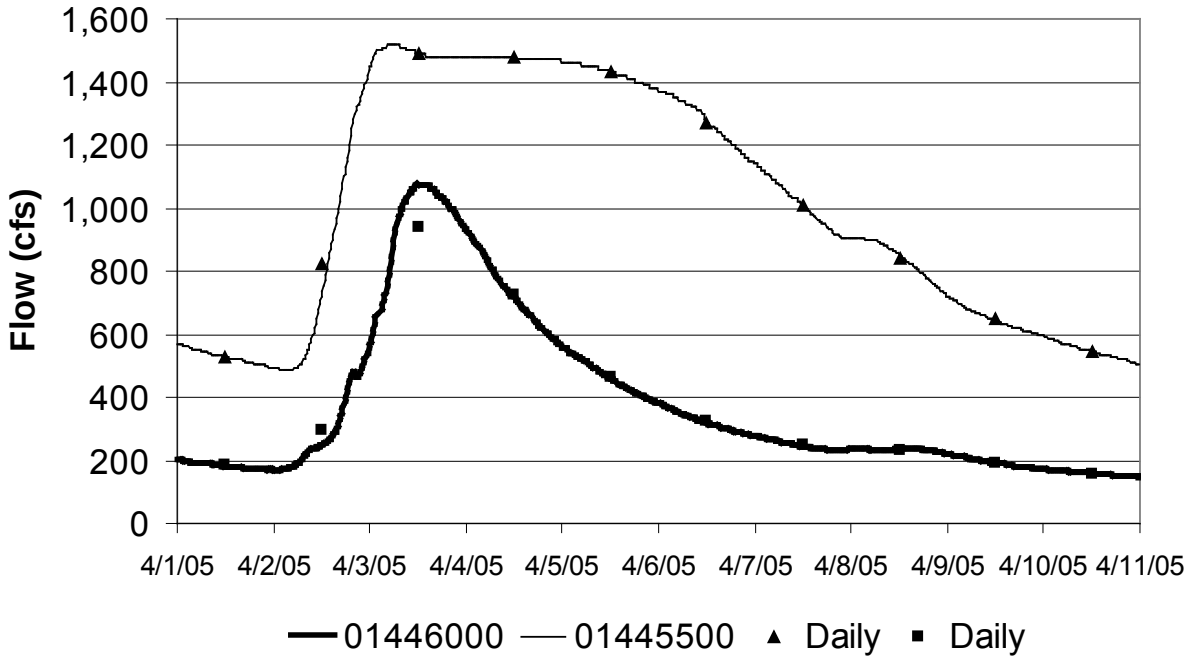

Figure 3.4c. Gage Pair 06 April 1-11, 2005

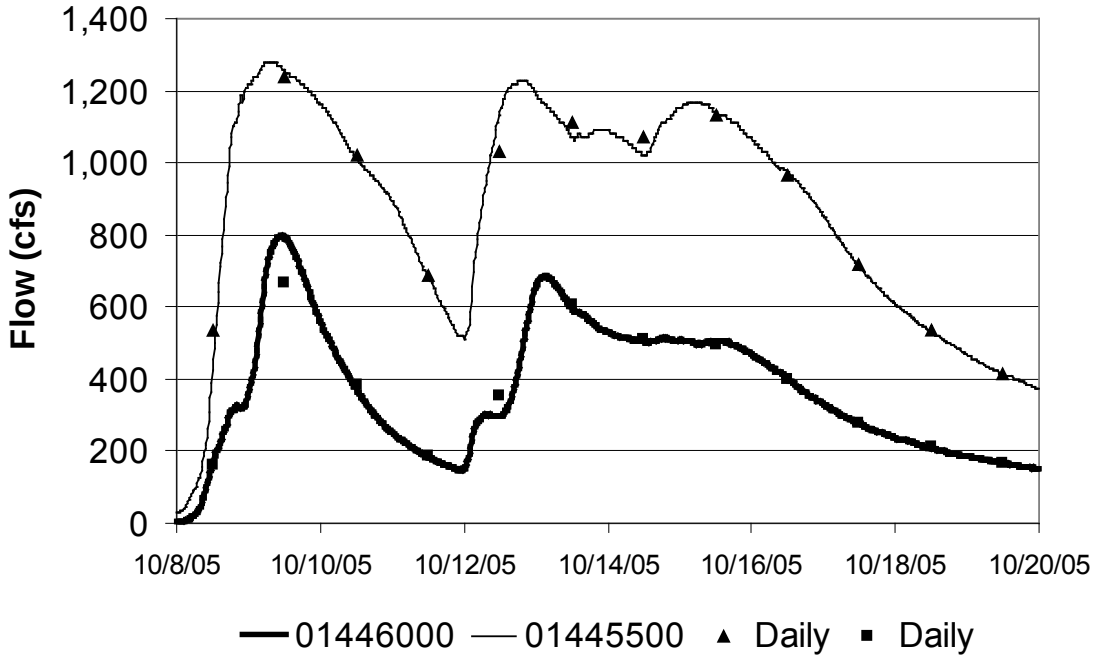

Figure 3.4d. Gage Pair 06 October 8-20, 2005

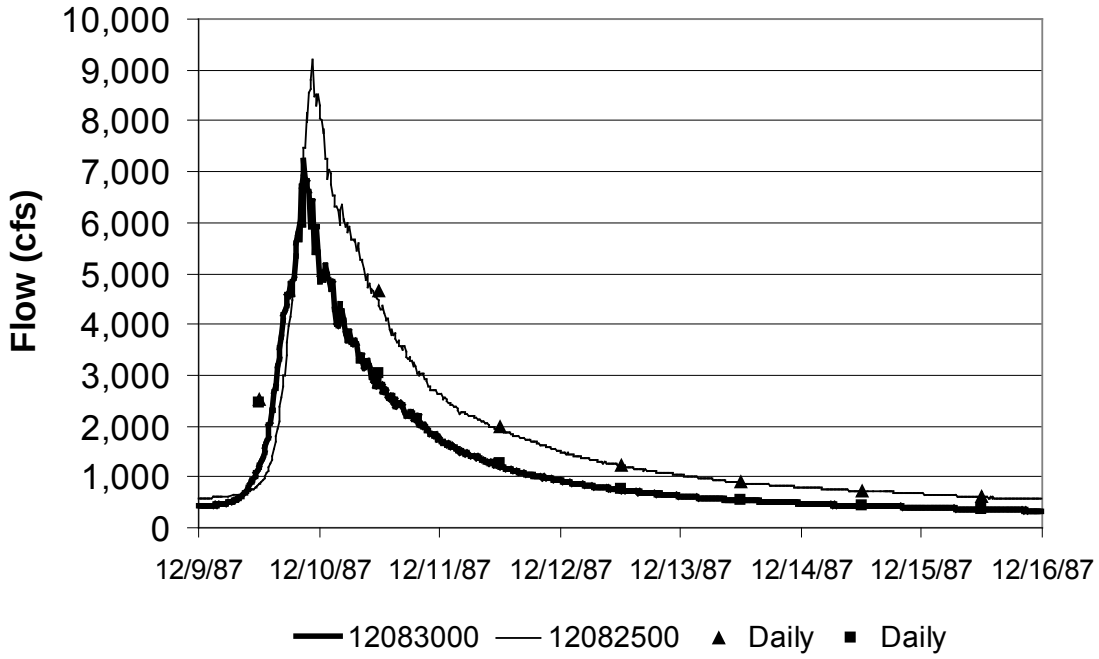

Figure 3.5a. Gage Pair 43 December 9-16, 1987

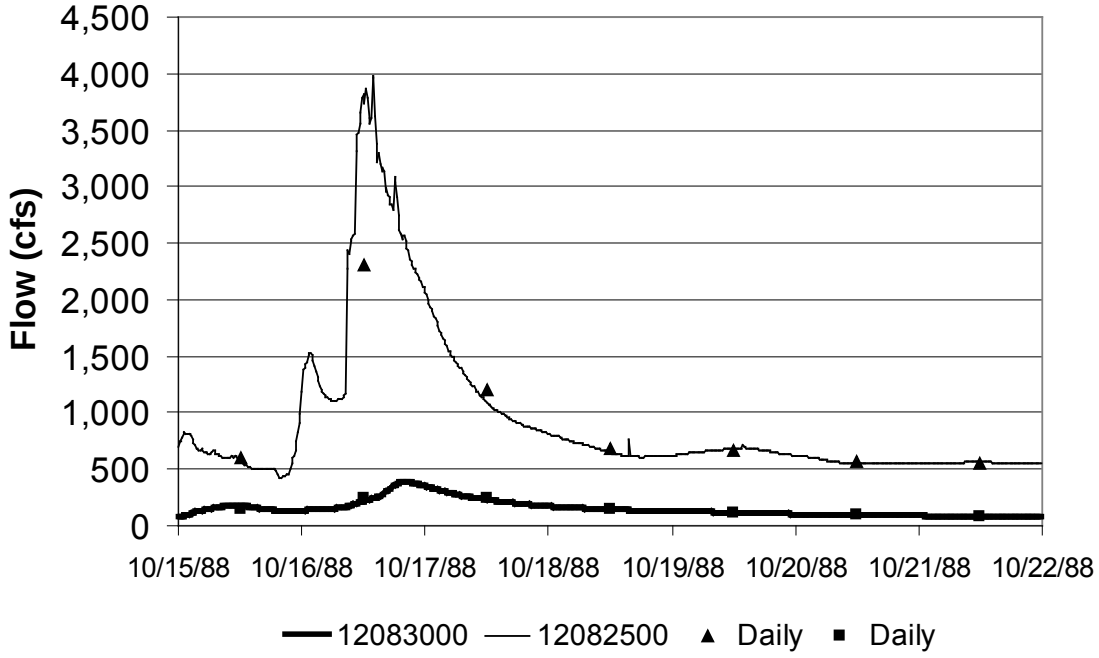

Figure 3.5b. Gage Pair 43 October 15-22, 1988

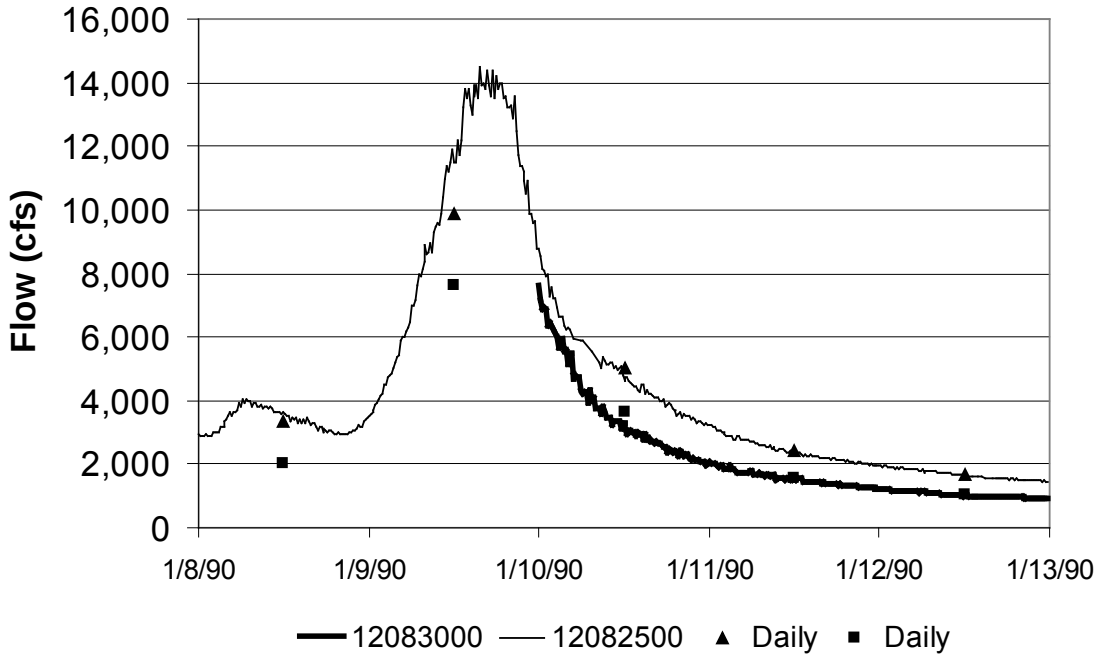

Figure 3.5c. Gage Pair 43 January 8-13, 1990

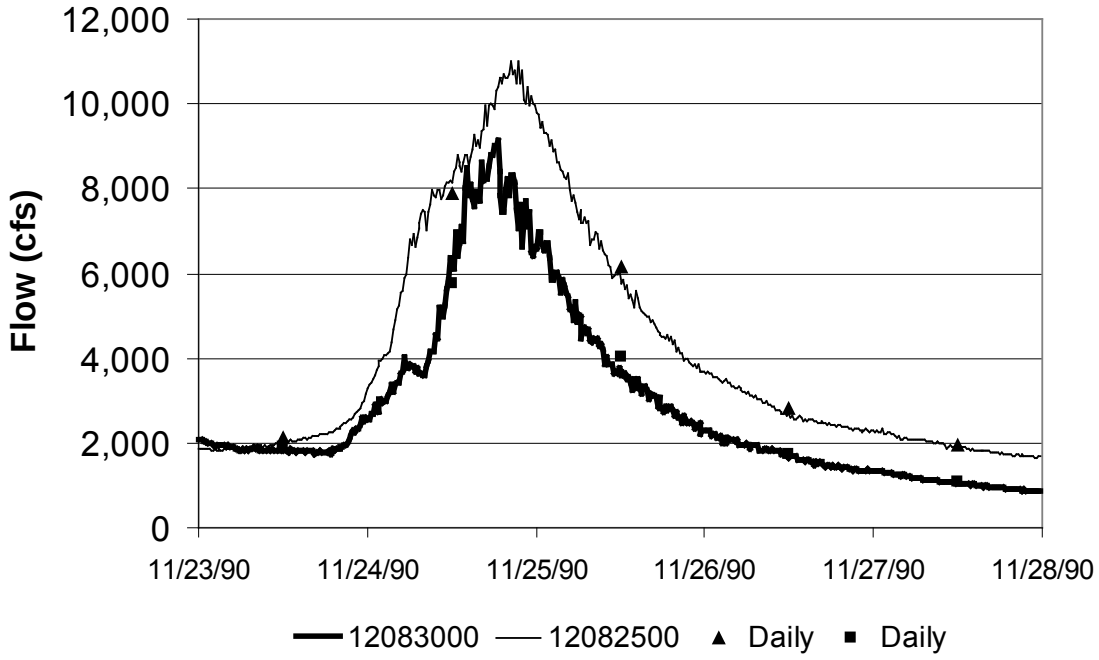

Figure 3.5d. Gage Pair 43 November 23-28, 1990

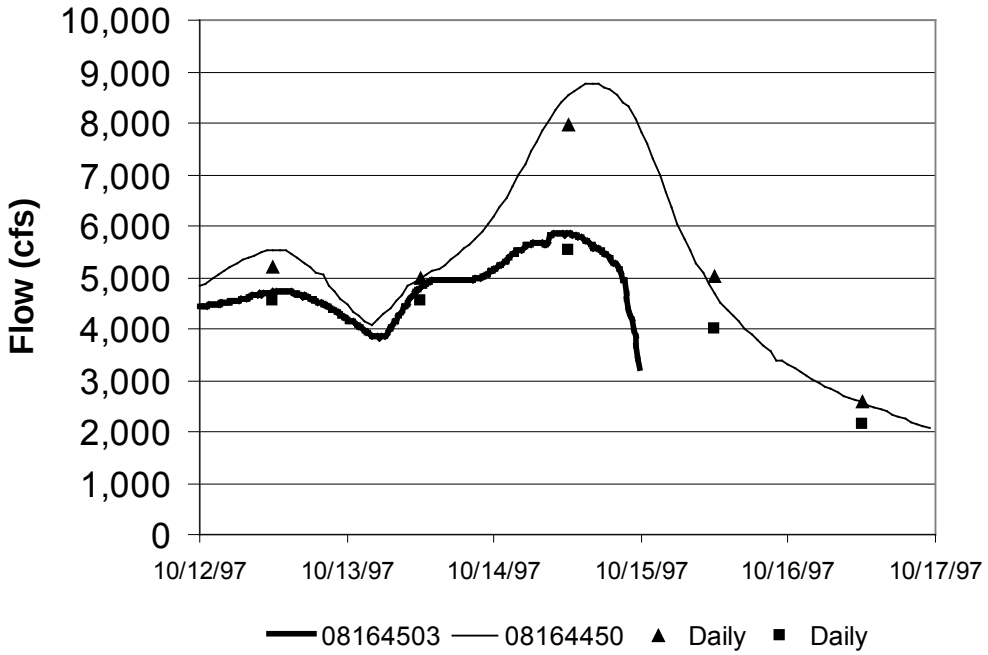

Figure 3.6a. Gage Pair 76 October 12-17, 1997

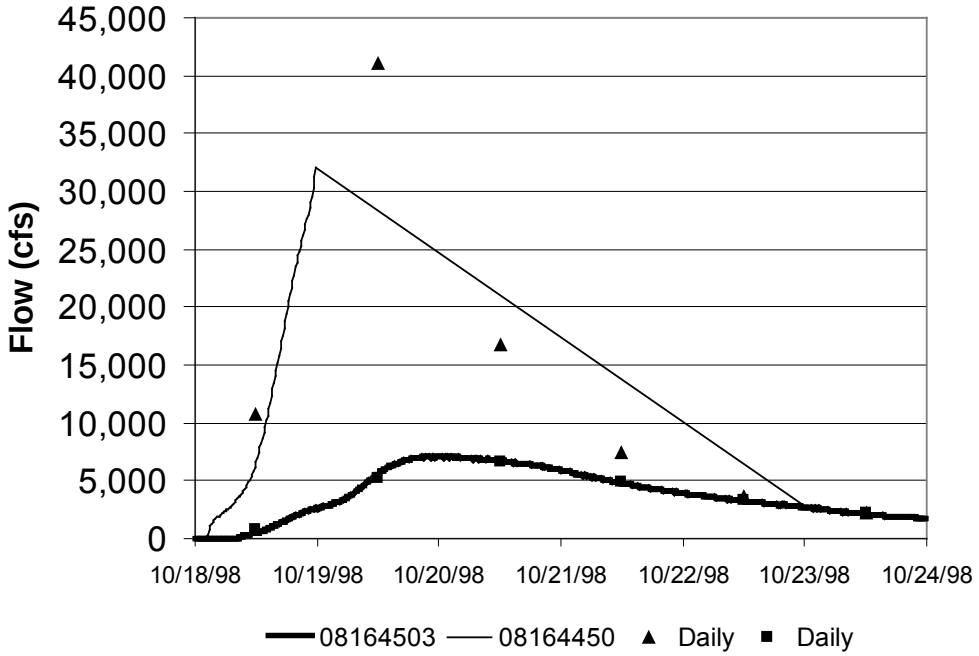

Figure 3.6b. Gage Pair 76 October 18-24, 1998

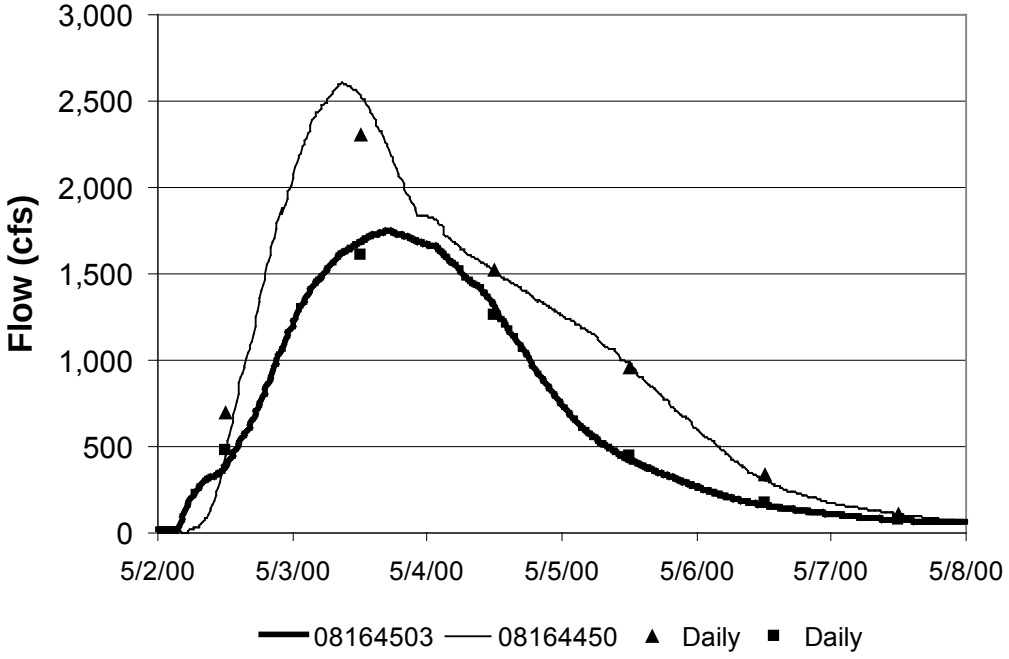

Figure 3.6c. Gage Pair 76 May 2-8, 2000

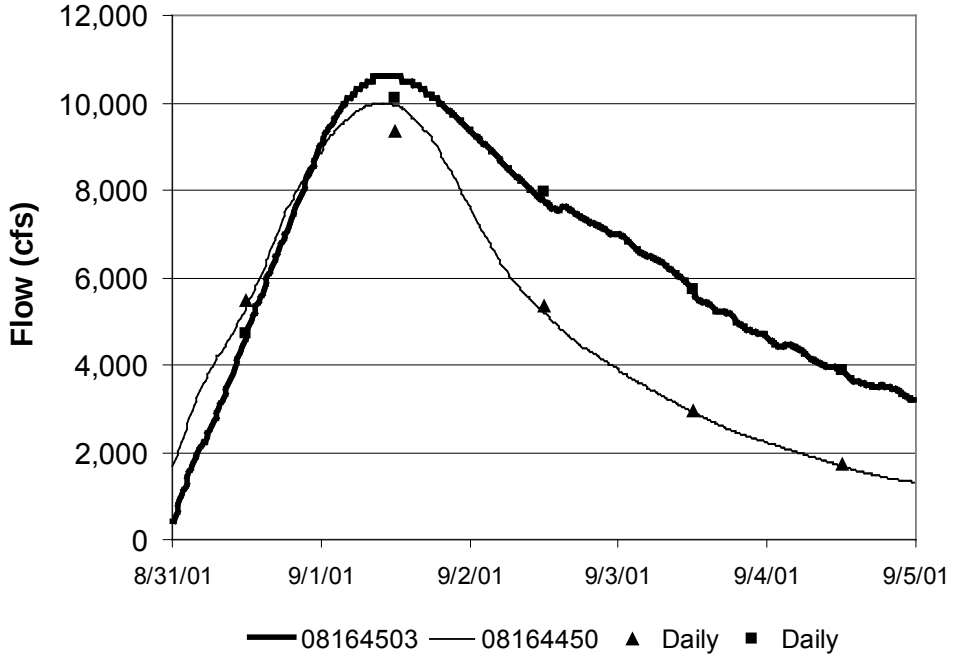

Figure 3.6d. Gage Pair 76 August 31-September 5, 2001

As is documented in the next section, mean daily data will be used in this research to evaluate the potential joint probability strategies and their application. Therefore, it is appropriate to define coincident flows as those mean daily flows from the main and tributary streams occurring on the same day. Consideration was given to a broader definition that considered coincident peaks to be within plus or minus one day of each other. Application of this definition did not substantially change the correlative relations.

If sufficient instantaneous data had been available to use for this study, defining coincident as flows occurring at exactly the same instant may not have adequately represented the correlative relation between the two gages. The figures in this section illustrate that some time window, possibly plus or minus two hours, would have been required to create representative coincident flow datasets.

3.3 Mean Daily Versus Instantaneous Data

Mean daily flow data were widely available; instantaneous data were not. Therefore, it was important to determine whether mean daily data could reasonably be used as an alternative to instantaneous data. If mean daily data behave similarly to instantaneous data, the daily data could be used for this research. Three questions were addressed:

1. Does the mean daily flow record correlate with instantaneous peaks?
2. Does an annual peak series derived from instantaneous records correlate with an annual peak series derived from daily records?
3. Will paired main/tributary datasets show similar correlations whether derived from daily versus instantaneous records?

The questions were addressed by detailed analysis of a selection of prototype watersheds intended to be representative of the paired database. Additional gage pairs with smaller watersheds were also considered because it was suspected that if the use of daily data would be problematic, such problems would most likely be revealed with smaller watersheds.

3.3.1 Mean Daily Flow/Instantaneous Peak Correlation

Comparison of the instantaneous hydrographs with mean daily data in Figures 3.4a through 3.6d revealed that in some cases, for example in Figure 3.4a, the mean daily and instantaneous peaks are relatively close, while in other cases, such as in Figure 3.5b they are not. As might be expected, the mean daily flow values on peak days are substantially lower than the instantaneous peaks when there is a rapid rise or decline in flow. However, the daily peaks do generally occur on the same day as the instantaneous peak and track much of the instantaneous hydrograph relatively well. In order to use mean daily data, these data must retain the same statistical properties, for example, correlation relations, as the instantaneous data even if they are not the same numerical values.

One measure is whether the set of instantaneous annual peaks is correlated with the mean daily data occurring on the same day. The result of correlation (Pearson's ρ) analyses for three prototype gages is shown in Table 3.4. Overall, there is strong correlation between the annual peaks and the corresponding mean daily flow with the lowest correlation being 0.874.

Table 3.4. Correlation Between Instantaneous Annual Peaks and Same-day Mean Daily Flow

Pair	Gage	Observations	Correlation
06	01445500	57	0.956
06	01446000	57	0.979
43	12082500	90	0.967
43	12083000	90	0.874
76	08164450	29	0.986
76	08164503	29	0.998

3.3.2 Annual Series Correlation

Another means of evaluating whether the mean daily data can be relied on to retain the statistical character of the flood hydrographs is to correlate the annual peak series data from instantaneous data with the annual peak series data from mean daily data. For this comparison, the instantaneous annual peak series data is the traditional series prepared by the USGS. The annual peak series from mean daily data were derived from the daily flow record as part of this study. The correlation is assessed through two quantitative measures, Pearson's ρ and Kendall's τ. (See Appendix C.5 for a description of these measures.)

Plots of annual peaks derived from instantaneous and mean daily data for the main and tributary gages of Gage Pair 05 are displayed in Figures 3.7 and 3.8, respectively. A high correlation for the main stream gage, which has a drainage area of 784 mi^2, is shown in Figure 3.7. Although correlation is indicated, a weaker relationship between instantaneous and daily data for the 2 mi^2 tributary watershed is shown in Figure 3.8. The same evaluation was completed for five additional gage pairs. The correlation plots for these pairs are found in Appendix F. The results are summarized in Table 3.5 and include the p-value (level of significance) for the correlation estimates. All of the p-values are low indicating the results are statistically significant.

The correlation results are plotted against the watershed drainage area in Figure 3.9. For Pearson's ρ, the correlation is 0.8 or greater for all of the gages except for the two smallest watersheds. Similarly, Kendall's τ is 0.7 or greater for all but the two smallest watersheds. Although climate and other factors will influence these observations, the data suggest that the annual series based on daily data is highly correlated with annual series based on instantaneous data for watersheds greater than about 6.5 square miles. Lower correlations, though not appreciably lower in some cases, are observed for smaller watersheds.

Seven gages from seven gage pairs in the study database are smaller than 6.5 mi^2. The smallest watershed is the tributary in Gage Pair 21 at 1.2 mi^2. A concern with the use of daily data is a loss of information, particularly for smaller watersheds. The results summarized in Table 3.5 and Figure 3.9 suggest that conclusions based on daily data for smaller watersheds should be cautiously interpreted, but that correlations are good for larger watersheds. Further analyses of small watersheds were conducted (see Appendix F.1.3.2). Based on these assessments, it was concluded that the use of daily data for smaller watersheds (1.0 to 6.5 mi^2) was appropriate.

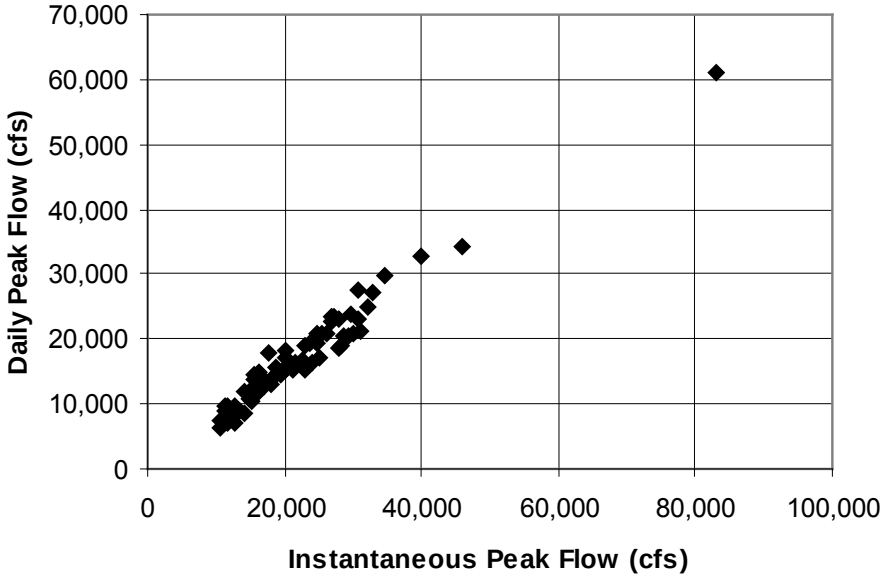

Figure 3.7. Daily and Instantaneous Peaks for 01403060 (Gage Pair 05)

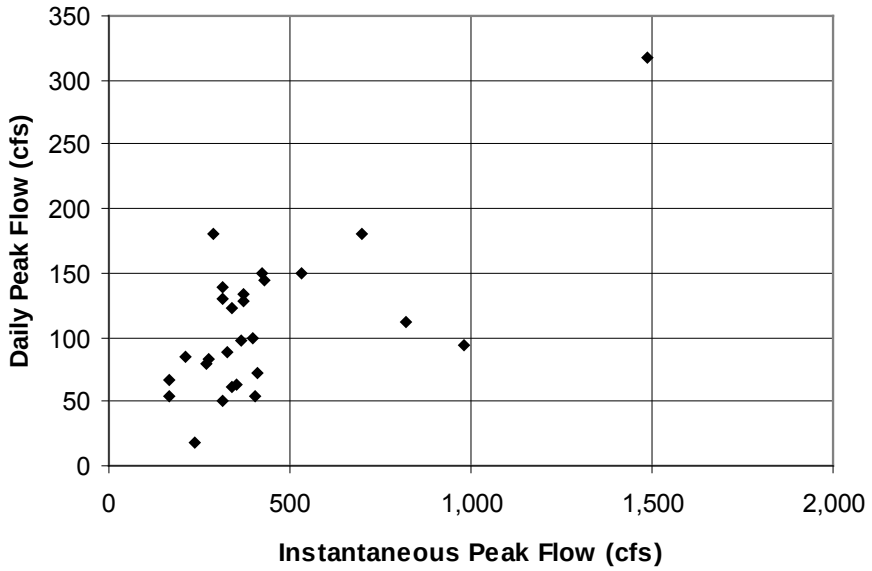

Figure 3.8. Daily and Instantaneous Peaks for 01403150 (Gage Pair 05)

Table 3.5. Daily Data Correlation with Instantaneous Data

		Drainage Area (mi^2)	Number of Obser-vations	Pearson's ρ	Kendall's τ	p-value (ρ)	p-value (τ)
Gage Pair 05	Main	784	67	0.978	0.846	0.00E+00	0.00E+00
	Tributary	2	27	0.690	0.382	6.90E-05	5.60E-03
Gage Pair 06	Main	106	85	0.970	0.868	0.00E+00	0.00E+00
	Tributary	36.7	42	0.802	0.705	1.80E-10	0.00E+00
Gage Pair 08	Main	8.5	42	0.909	0.699	0.00E+00	0.00E+00
	Tributary	6.9	25	0.920	0.858	7.50E-11	0.00E+00
Gage Pair 21	Main	117	32	0.987	0.907	0.00E+00	0.00E+00
	Tributary	1.2	27	0.677	0.644	1.10E-04	4.10E-06
Gage Pair 43	Main	133	64	0.968	0.834	0.00E+00	0.00E+00
	Tributary	70.3	64	0.874	0.718	0.00E+00	0.00E+00
Gage Pair 76	Tributary	289	29	0.986	0.927	0.00E+00	0.00E+00
	Tributary	178	29	0.998	0.969	0.00E+00	0.00E+00

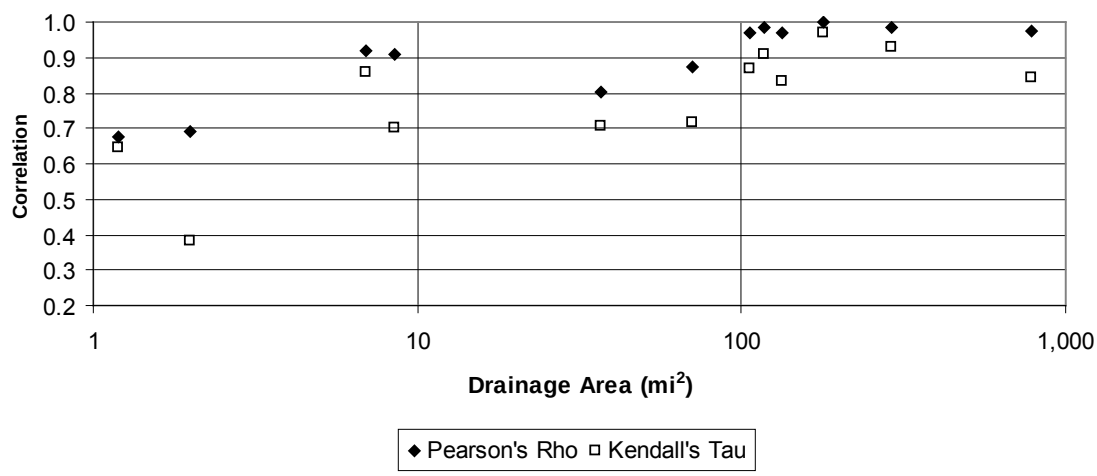

Figure 3.9. Daily–Instantaneous Correlation versus Drainage Area

3.3.3 Paired Dataset Correlation

Correlations between the Peak on Main (POM) and Peak on Tributary (POT) paired datasets generated from instantaneous data and those generated from daily data were compared. Does the use of daily data obscure a correlation between the flows on the main stream and the tributary, if one exists?

To compile the POM datasets using instantaneous data, the annual peak series for the main stream was identified from the annual series data. In most years, the specific time of the peak is not recorded in the annual peak series data. Therefore, the instantaneous record for the main stream was acquired and searched for the peak value. The time of peak was recorded. Next, the tributary instantaneous record was searched for the flow that occurred at the same time. These data are plotted in Figure 3.10 for Gage Pair 05. With 21 data pairs in the POM dataset, the Pearson's ρ is –0.062 and the Kendall's τ is 0.203. A correlation between the main and tributary streams is not apparent.

Similarly, the instantaneous annual peak series for the tributary was identified for the POT dataset. The instantaneous record for the tributary was acquired and searched for the peak value. The time of peak was recorded. Next, the main stream instantaneous record was searched for the flow that occurred at the same time. These data are summarized in Figure 3.11 for Gage Pair 05. With 27 data pairs in the POT dataset, the Pearson's ρ is 0.062 and the Kendall's τ is -0.076. Again, a correlation between the main and tributary streams is not apparent for this dataset.

Performing the analogous evaluation for Gage Pair 05 using daily data results in Figures 3.12 and 3.13 for the POM and POT datasets, respectively. As with the instantaneous data, neither figure suggests a correlation between the main and tributary gages.

The correlation statistics for the instantaneous and mean daily data are summarized in Table 3.6. Although, the daily data correlation estimates differ numerically from the instantaneous data correlation estimates, both sets of analyses show low correlation between the two gages. This observation is reinforced by the high p-values in the table that indicate the estimates of Pearson's ρ and Kendall's τ are not significantly different from zero.

Completing the same assessment on other gage pairs would likely be worthwhile, but there were insufficient instantaneous data available to systematically do so. In the analysis of Gage Pair 05 the daily and instantaneous data both indicate no correlation between the tributary and main stream flows. However, a broader conclusion that agreement can be expected in other pairs cannot be drawn from the analysis of one pair.

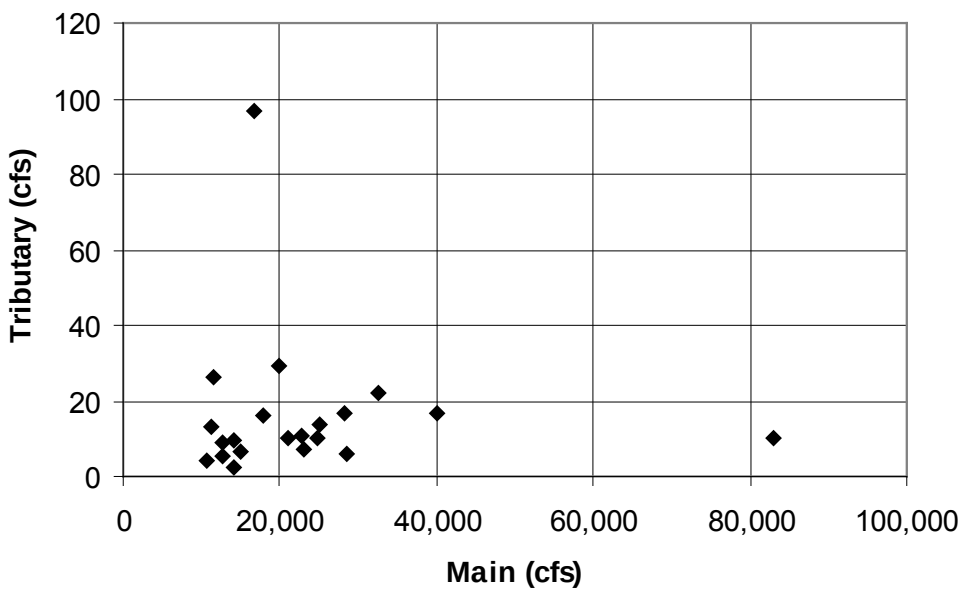

Figure 3.10. Peak on Main Dataset (Instantaneous Data, Pair 05)

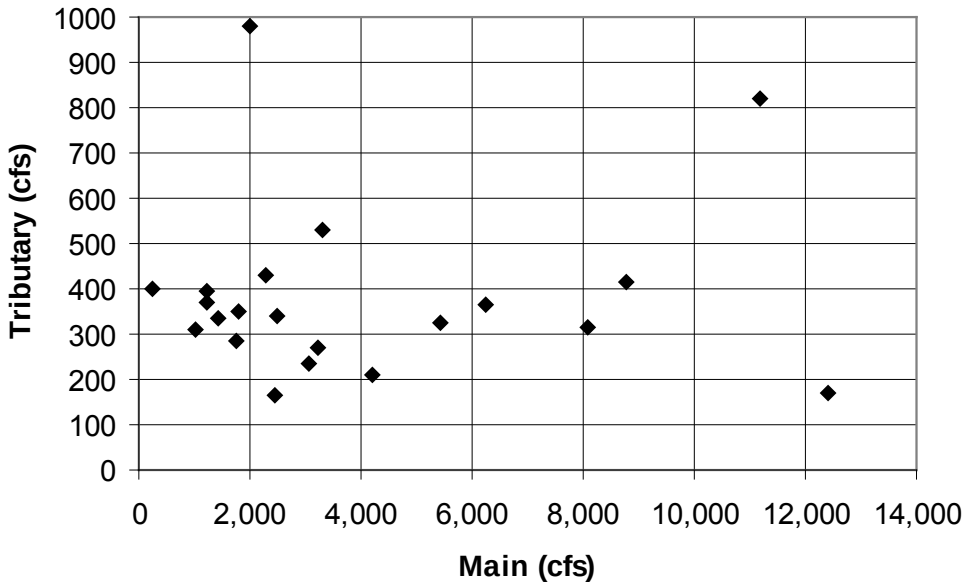

Figure 3.11. Peak on Tributary Dataset (Instantaneous Data, Pair 05)

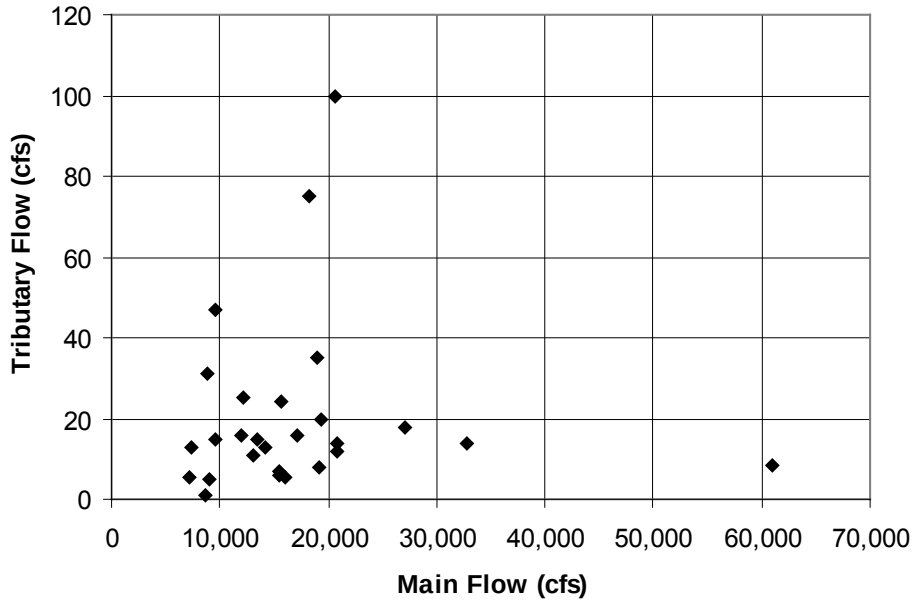

Figure 3.12. Peak on Main Dataset (Daily Data, Pair 05)

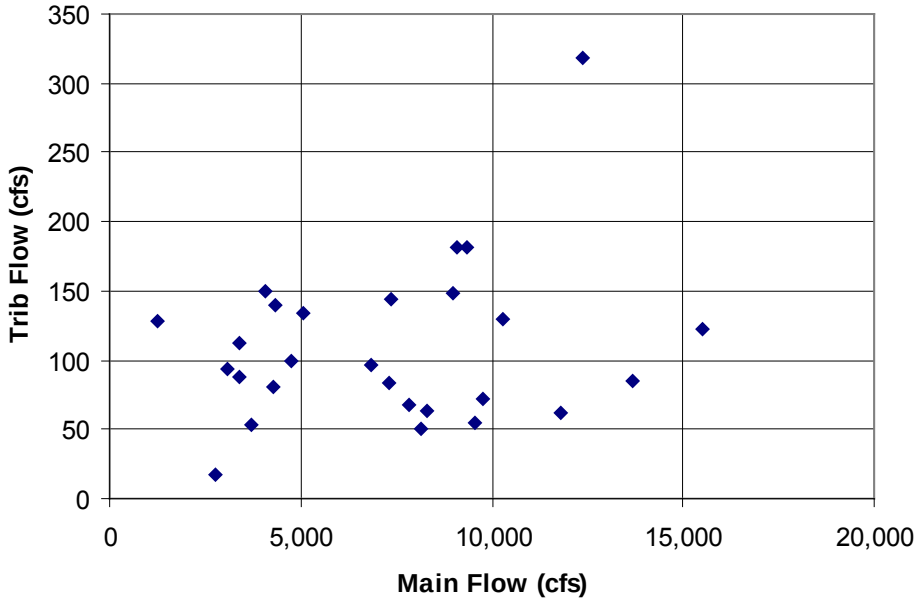

Figure 3.13. Peak on Tributary Dataset (Daily Data, Pair 05)

Table 3.6. Correlation Between Main and Tributary (Gage Pair 05)

	Instantaneous Data		Daily Data	
	POM	POT	POM	POT
n	21	21	27	27
Pearson's ρ	-0.062	0.062	0.004	0.244
Kendall's τ	0.203	-0.076	0.118	0.086
p-value (ρ)	7.9E-01	7.9E-01	9.8E-01	2.2E-01
p-value (τ)	2.1E-01	6.5E-01	4.0E-01	5.5E-01

4. Bivariate Distributions and Univariate Distributions with Copulas

Bivariate distributions are a subset of multivariate distributions involving two random variables. A bivariate distribution is characterized by a distribution function, which is a scalar-valued function of a vector-valued random variable (in this case two random variables). In the limit of each random variable, a univariate distribution function results, termed the marginal distribution. So, for a distribution function $H(x,y)$ there are two marginal distributions, $F(x)$ and $G(y)$. Most bivariate distributions have the same type of marginal distributions, that is, the same distribution function although with different parameter values. An extended discussion of the bivariate normal and Gumbel distributions is provided in Appendix C.2.

Copulas are a more general approach to bivariate (or multivariate) problems. The term copula refers to a function, called the dependence function, used to link two univariate distributions in such a way as to represent the bivariate (or multivariate) dependence between the two random variables. The potential of a copula is realized in that the copula is independent from the form of the univariate marginal distributions. Therefore, the marginal distributions can be chosen such that they provide a best-fit of the univariate random variables, with the copula used to model the dependence behavior. Many copulas are available for application to bivariate (and multivariate) random variables. Those of interest to hydrologists typically fall into the *Archimedean* family of copulas. For two random variables, X and Y with cumulative distribution functions of $F_X(x)$ and $F_Y(y)$, respectively, define $U = F_X(X)$ and $V = F_Y(Y)$. Then, U and V are uniformly distributed random variables and u will denote a specific value of U and v will denote a specific value of V. An extended discussion of copulas is provided in Appendix C.3.

The discussion of bivariate distributions and univariate distributions with copulas are combined in this chapter because of the similar nature of the analyses. The bivariate distributions evaluated are the log-transformed bivariate normal distribution and the Gumbel bivariate distribution. Along with these, the Gumbel-Hougaard (GH) and Frank copulas are examined. Development of the tools for application of bivariate distributions and univariate distributions with copulas is also described in the following sections.

4.1 Fitting the Distributions

Determination of which bivariate distributions or copulas might be most useful for addressing the joint probability problem depends, in part, on how well they fit observed data. Fits were evaluated using four of the prototype gage pairs: 05, 06, 43, and 76. Summary statistics for the annual peak series data for each gage were computed and listed in Table 4.1. Computed statistics for the concurrent daily mean datasets for both the peak on main (POM) and peak on tributary (POT) data are listed in Table 4.2. Significance values for various parameters are shown. Where these p-values are less than 0.01, the hypothesis that the computed parameter is not zero can be accepted at the 99 percent confidence level. The distribution fitting evaluation included: 1) a graphical evaluation/tabular application of the four methods and 2) a quantitative assessment of best-fit.

4.1.1 Graphical Assessment and Application

The fit of the data to the various distributions is assessed visually with a plot of the theoretical nonexceedance versus the empirical nonexceedance probability as computed by Equation C.39 (Appendix C). Figure 4.1 is an example of a plot using data from the POM dataset for Gage Pair 06. Additional examples are presented in Appendix C.6. A perfect fit is indicated when all data points fall on the 1:1 line.

Table 4.1. Annual Peak Summary Statistics

	Pair 05		Pair 06		Pair 43		Pair 76	
Statistic	01403060	01403150	01445500	01446000	12082500	12083000	08164450	08164503
Number of Observations (n)	67	27	85	42	64	64	29	29
Mean	17267	109	880	496	5175	3831	7561	5376
Standard Deviation	8189	58.6	351	246	2533	1647	8027	4192
Skew	2.51	1.72	0.782	1.26	1.69	0.923	3.00	1.74
Mean logs (natural)	9.67	4.56	6.7	6.09	8.45	8.16	8.57	8.33
Stand. Dev. logs (natural)	0.417	0.549	0.395	0.491	0.453	0.458	0.832	0.735
Skew logs (natural)	0.183	-0.732	0.0371	-0.064	0.0626	-0.657	0.301	0.019
Gumbel location, μ	13582	82.6	722	385	4035	3090	3949	3489
Gumbel scale, σ	6385	45.7	274	192	1975	1284	6259	3268

Table 4.2. Summary Statistics for Daily Data Concurrent Pair Datasets

	Pair 05		Pair 06		Pair 43		Pair 76	
	POM	POT	POM	POT	POM	POT	POM	POT
Concurrent Events (n)	27	27	43	42	64	64	29	29
Pearson's ρ	0.00173	0.244	0.639	0.778	0.724	0.722	0.553	0.832
p-value	0.993	0.219	4.20E-06	1.28E-09	1.33E-11	1.58E-11	1.86E-03	2.27E-08
Pearson's ρ (natural logs)	0.165	0.194	0.676	0.753	0.705	0.752	0.782	0.686
p-value	0.409	0.331	6.42E-07	8.82E-09	7.99E-11	7.60E-13	5.29E-07	4.06E-05
Kendall's τ	0.110	0.0856	0.518	0.547	0.504	0.534	0.660	0.484
p-value	0.427	0.532	1.07E-06	3.67E-07	4.36E-09	4.86E-10	4.51E-08	2.35E-04
Gumbel mixed model θ	0.003	0.388	0.962	n/a	n/a	n/a	0.843	n/a
Gumbel logistic model m	1.001	1.150	1.664	2.122	1.903	1.897	1.496	2.440
GH θ	1.124	1.094	2.075	2.208	2.016	2.146	2.941	1.938
Frank θ	1.000	0.775	6.070	6.660	5.810	6.390	9.790	5.450

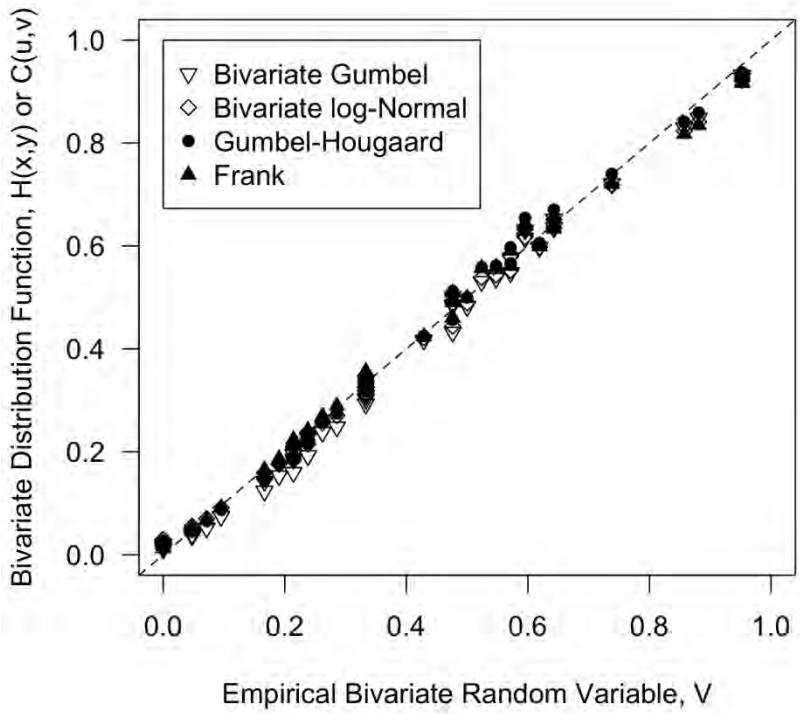

Figure 4.1. Fit for Gage Pair 06 (POM)

The ordinate in Figure 4.1 is computed from Equation C.21 for the Gumbel mixed model bivariate distribution and a numerical integration of Equation C.26 for the log-normal bivariate distribution. The ordinates for the copula methods are computed from Equation C.30 for the Gumbel-Hougaard copula and from Equation C.34 for the Frank copula.

An example of the joint distribution for Gage Pair 06 (POM) is represented in Figure 4.2. The abscissa and ordinate axes represent the cumulative nonexceedance probabilities for the main and tributary streams, respectively. The curves represent isolines of nonexceedance probabilities for each methodology. Taking the 0.7 isoline, for example, the probability of an event on the main stream and on the tributary being less than any given point on the line is 0.7. The individual data points are the estimated univariate nonexceedance values based on the Weibull plotting position formula.

The joint distribution plot implies a symmetrical distribution. However, the abscissa and ordinate axes only represent the univariate cumulative distribution function (CDF) for one of the variables. The other axis represents the cumulative nonexceedance probability for the complementary data in the dataset, not the annual peak series. For example, the abscissa represents the univariate CDF for the main stream because this dataset was created by identifying the annual peaks on the main stream and pairing those with the concurrent values on the tributary. The concurrent tributary values may or may not be annual peaks. Similarly, the ordinate axis for the POT dataset represents the univariate CDF while the abscissa does not. Without adjustment, treating both axes as univariate CDFs will result in an overestimate of the joint probability. Additional discussion of this issue is provided in Section 4.4.

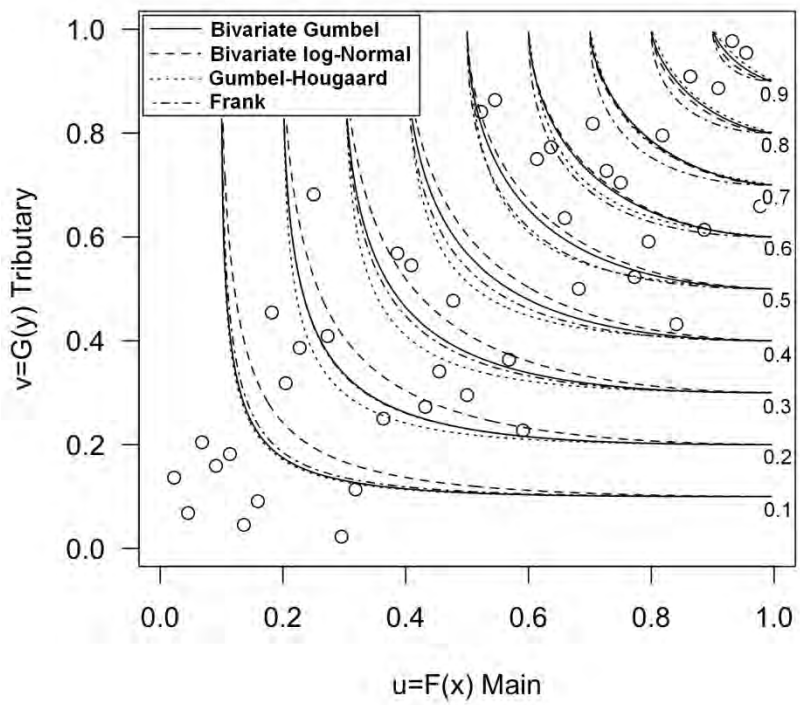

Figure 4.2. Joint Distribution for Gage Pair 06 (POM)

For design, the exceedance probability isolines are desired. Figure 4.3 is an example plot for Gage Pair 06 (POM) for the 10-percent exceedance probability. That is, for an event to be considered more extreme the condition on both the main and the tributary would be equaled or exceeded. (The abscissa and ordinate in Figure 4.3 represent nonexceedance probabilities as in Figure 4.2.)

Table 4.3 is an example of possible combinations of events from which the critical condition for a particular design location is obtained. The nonexceedance probabilities in the table are representative points calculated from the relation shown in Figure 4.3. For example, a nonexceedance probability of 0.5 on the main stream corresponds to a nonexceedance probability of 0.893 for the log-normal curve. (The numerical precision is achieved by computation, not reading the graph.) The designer evaluates the possible combinations from the preferred method to determine the combination of main and tributary flows that results in the most severe conditions. That combination is the design or “critical” combination.

To determine the appropriate discharges for each possible combination, the probability of exceedance must be transformed to a discharge. For the Gumbel mixed model bivariate distribution, the Gumbel univariate distribution is implicit and used for computing discharge. Similarly, for the normal bivariate distribution (log transformed), the univariate normal distribution is implicit and used for computing discharge.

Since copulas are independent of a marginal distribution, any appropriate marginal distribution is selected to convert exceedance probabilities to discharges. Log-Pearson Type III (LP3) would generally be a good choice because it is commonly taken to be applicable to flood flows and many of the USGS regression equations are based on quantiles determined using LP3.

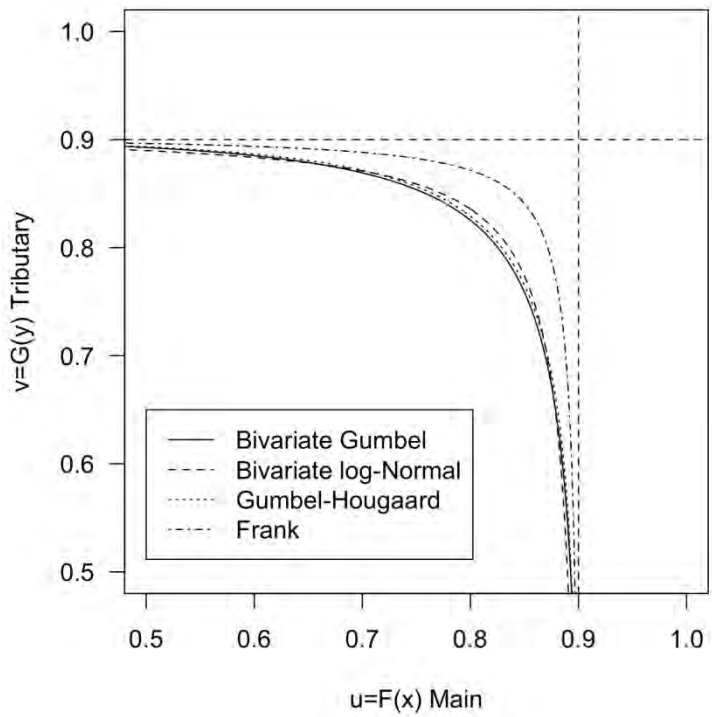

Figure 4.3. 10-percent Exceedance Probability for Gage Pair 06 (POM)

Table 4.3. 10-percent Exceedance Combinations for Gage Pair 06 (POM)

	Main stream		Tributary[1]	
Method (ID)	Nonexceedance Probability (u)	Return Period	Nonexceedance Probability (v)	Return Period
Log-normal (1)	0.500	2.0	0.893	9.3
Log-normal (2)	0.813	5.3	0.813	5.3
Log-normal (3)	0.890	9.1	0.540	2.2
Bivariate Gumbel (1)	0.500	2.0	0.890	9.1
Bivariate Gumbel (2)	0.825	5.7	0.825	5.7
Bivariate Gumbel (3)	0.880	8.3	0.636	2.7
GH copula (1)	0.500	2.0	0.896	9.6
GH copula (2)	0.846	6.5	0.846	6.5
GH copula (3)	0.896	9.6	0.520	2.1
Frank copula (1)	0.500	2.0	0.893	9.3
Frank copula (2)	0816	5.4	0.816	5.4
Frank copula (3)	0.890	9.1	0.557	2.3

[1]Not the annual series nonexceedance probabilities and return periods. These are conservative.

4.1.2 Best-Fit Distribution

The recommendation for a statistically based method should be based, in part, on which distribution best fits the observed data. The root mean squared error (RMSE) was adopted to

compare the four strategies. Because the focus of this study is on extreme events, only events with a probability of exceedance less than 0.5 were considered for this evaluation.

Theoretical and empirical values for the bivariate distribution function, as illustrated in Figure 4.1 for Gage Pair 06 (POM), were used for the comparison. The RMSE results are summarized in Table 4.4 for each of the four strategies for both the POM and POT datasets. Except for the bivariate Gumbel strategy, all 85 gage pairs are represented in the analysis. The bivariate Gumbel (mixed model) is limited to paired data where the Pearson's ρ is less the two-thirds, meaning this distribution was not defined for those gage pairs with a higher correlation.

Table 4.4. Summary of RMSE Computations

Parameter	Bivariate Gumbel (POM)	Bivariate Normal (POM)	Gumbel-Hougaard (POM)	Frank (POM)	Bivariate Gumbel (POT)	Bivariate Normal (POT)	Gumbel-Hougaard (POT)	Frank (POT)
Average	0.036	0.034	0.027	0.035	0.033	0.033	0.026	0.034
Median	0.038	0.032	0.023	0.030	0.031	0.030	0.023	0.031
Minimum	0.016	0.013	0.010	0.015	0.014	0.013	0.009	0.015
Maximum	0.064	0.076	0.079	0.080	0.069	0.109	0.071	0.086
n	31	85	85	85	23	85	85	85

For both the POM and POT datasets, the Gumbel-Hougaard copula results in the lowest average error of the four strategies, indicating the best fit. The same result is observed for the median and minimum errors. For the maximum error, the table reports that the bivariate Gumbel is the lowest for both the POM and POT datasets followed by the Gumbel-Hougaard. However, inspection of the data revealed that the RMSE error (Gage Pair 48) that is the largest for the Gumbel-Hougaard method is undefined for the bivariate Gumbel method because the Pearson's ρ is greater than two-thirds. Therefore, the RMSE analyses show that the Gumbel-Hougaard copula is the best-fit alternative.

4.2 Correlation Parameter Estimation

4.2.1 Objective

Application of bivariate distributions and copulas to the joint probability problem at ungaged watershed pairs requires an estimate of the correlation between the two watersheds. The objective of this part of the research effort was to develop relations to estimate Pearson's ρ and Kendall's τ from watershed and meteorological characteristics.

4.2.2 Data

Observed values of Pearson's ρ and Kendall's τ were computed from the POM and POT datasets for each gage pair. Recall that the POM dataset is the collection of annual peaks (daily data) on the main stream and the simultaneous (daily) flow on the tributary stream. The POT dataset is the collection of annual peaks on the tributary stream and the simultaneous (daily) flow on the main stream. Of interest is whether different conclusions regarding correlation result from the POM versus POT datasets.

4.2.2.1 Watershed, Meteorological, and Hydrological Data

Variables considered for testing in the analyses included watershed, meteorological, and hydrological parameters collected for this research. They are:

1. Watershed variables
   - 1.1 Hydrologic unit code
   - 1.2 Main and tributary watershed area (A_M and A_T)
   - 1.3 Main and tributary longest path length
   - 1.4 Main and tributary longest path elevation drop
   - 1.5 Main and tributary longest path slope
   - 1.6 Main and tributary length-slope parameter (T_{LS})
   - 1.7 Main/tributary area ratio (R_A)
   - 1.8 Length-slope parameter ratio (R_{LS})
   - 1.9 Total area (A_{TOT})
   - 1.10 Gage separation distance (d_g)
   - 1.11 Watershed centroid separation distance (d_c)
2. Meteorological variables
   - 2.1 Mean annual precipitation (P_M) (Average of value at main and tributary)
   - 2.2 24-h 2-yr storm ($P_{24\text{-}2}$) (Average of value at main and tributary)
3. Hydrological variables
   - 3.1 Main Q_2/Tributary Q_2 ratio
   - 3.2 Main Q_{100}/Tributary Q_{100} ratio

4.2.2.2 Estimated Correlation Parameters

The observed correlation parameters, Pearson's ρ and Kendall's τ, were calculated from the POM and POT datasets for each gage pair. The definition of these two parameters and the method to estimate each are provided in Appendix C.5. Tabular summaries of the statistics for the two datasets are presented in Appendix B.4.

Initial trials for developing relations for the correlation parameters revealed a broad scatter in the data. Obvious relations with the watershed, meteorological, and hydrological data were not apparent. This led to a close examination of selected gage pairs to consider if there would be any benefit to a modified definition of coincident event (same day). A search for outliers was also conducted.

The gage pairs selected for closer examination were identified by comparing the correlation parameters at each gage pair in the POM dataset with the values in the POT dataset. Gage pairs with substantial differences were examined manually.

A coincident event was defined as those flows occurring on the same day in both the tributary and main stream. Broadening that definition to include the highest value in either a two or three-day window had been considered, but was rejected. This decision was reconsidered to evaluate if any effect on the results might be expected. In addition, discharge outliers were identified and removed.

Changes made to the datasets because of these adjustments are detailed in Appendix C.7.1. The adjustments did not fundamentally change the analyses and the original definition of a coincident event was retained.

4.2.2.3 Correlation Matrix

Several of the independent variables considered for these analyses are cross-correlated, suggesting that caution is required when selecting a suite of variables for the relations. For example, area ratio and length-slope ratio are highly correlated and should not be used in the same relation. Similarly, mean annual precipitation and mean 24-h 2-yr storm are significantly correlated.

A correlation analysis between the dependent variables (ρ and τ) and independent variables (watershed, meteorological, and hydrological) was performed. Results are summarized in Tables 4.5a and 4.5b. Although none of the variables exhibit a correlation with Pearson's ρ and Kendall's τ greater than 0.5, the area ratio and length-slope parameter ratio exhibit the highest correlation with Pearson's ρ and Kendall's τ.

4.2.3 Correlation Parameter Regressions

Several analysis strategies were pursued to determine reasonable functional forms for estimating the correlation parameters Pearson's ρ and Kendall's τ:

- Filtering. Eliminating gage pairs based on characteristics that render them unsuitable for use in regression or considered to be outliers. However, none were identified.
- Functional adjustments. Transforming gage pair characteristics to improve the estimating accuracy of the regression equations.
- Grouping. Subdividing the dataset into groups that might exhibit characteristics fundamentally different from group to group.
- Alternative functional forms. Experimenting with equation forms other than log-space linear regression.
- Best-fit and envelope relations.

A detailed description of the analyses and their results is provided in Appendix C.7. Overall, regression analysis success was limited as measured by equation R^2 values. However, equations developed after subdividing the data based on the size of the tributary watershed showed promise. The two groups are based on the tributary (smaller) watershed area:

1. Drainage area of the tributary watershed is less than 400 mi^2. If the drainage area of the larger watershed is greater than 2000 mi^2, the larger watershed drainage area is taken as 2000 mi^2.
2. Drainage area of the tributary watershed is greater than 400 mi^2.

The recommended best-fit equations based on these two groups are summarized below. Separate equations were developed for the POM and POT datasets. See Appendix C.7 for more detail.

Table 4.5a. Correlation Matrix for the Peaks on Main (POM) Dataset

	Main Area (mi^2)	Main Long. Path Length (mi)	Main Long. Path Elev. Drop (ft)	Main Long. Path Slope (ft/ft)	Main LS Param-eter (h)	Trib. Area (mi^2)	Trib. Long. Path Length (mi)	Trib. Long. Path Elev. Drop (ft)	Trib. Long. Path Slope (ft/ft)	Trib. LS Param-eter (h)	Area Ratio (nd)	Total Area (mi^2)	Cen-troid Sepa-ration (mi)	LS Param-eter Ratio (nd)	Mean Annual Precip. (in)	Mean I24-2 (in)	Pear-son's ρ	Ken-dall's τ
Main Area (mi^2)	1.000																	
Main Longest Path Length (mi)	0.806	1.000																
Main Longest Path Elev. Drop (ft)	-0.020	-0.133	1.000															
Main Longest Path Slope (ft/ft)	-0.256	-0.431	0.610	1.000														
Main LS Parameter (h)	0.721	0.886	-0.366	-0.462	1.000													
Trib. Area (mi^2)	0.183	0.356	-0.033	-0.195	0.297	1.000												
Trib. Longest Path Length (mi)	0.163	0.474	-0.045	-0.227	0.421	0.841	1.000											
Trib. Longest Path Elev. Drop (ft)	-0.051	-0.157	0.786	0.463	-0.348	0.072	0.095	1.000										
Trib. Longest Path Slope (ft/ft)	-0.117	-0.308	0.640	0.568	-0.433	-0.279	-0.356	0.606	1.000									
Trib. LS Parameter (h)	0.176	0.517	-0.290	-0.332	0.576	0.717	0.891	-0.231	-0.453	1.000								
Area Ratio (nd)	0.504	0.348	-0.004	-0.150	0.269	-0.205	-0.276	-0.136	0.073	-0.207	1.000							
Total Area (mi^2)	0.979	0.833	-0.026	-0.281	0.740	0.379	0.327	-0.033	-0.168	0.313	0.432	1.000						
Centroid Separation (mi)	0.741	0.805	-0.112	-0.339	0.773	0.282	0.273	-0.127	-0.232	0.356	0.338	0.756	1.000					
LS Parameter Ratio (nd)	0.407	0.368	-0.094	-0.200	0.320	-0.222	-0.309	-0.180	0.062	-0.228	0.804	0.338	0.315	1.000				
Mean Annual Precip. (in)	-0.116	-0.163	0.202	0.153	-0.155	-0.118	-0.148	0.161	0.222	-0.172	-0.025	-0.133	-0.126	-0.010	1.000			
Mean I24-2 (in)	-0.023	0.066	-0.112	-0.198	0.119	-0.117	-0.028	-0.110	-0.143	0.027	0.010	-0.046	0.039	0.092	0.591	1.000		
Pearson's ρ	-0.278	-0.389	0.232	0.319	-0.355	-0.039	0.026	0.261	0.096	-0.057	-0.433	-0.270	-0.384	-0.531	0.114	-0.023	1.000	
Kendall's τ	-0.208	-0.331	0.153	0.257	-0.250	-0.018	0.048	0.153	0.012	0.023	-0.344	-0.199	-0.341	-0.375	0.154	0.093	0.852	1.000

Table 4.5b. Correlation Matrix for the Peaks on Tributary (POT) Dataset

	Main Area (mi^2)	Main Long. Path Length (mi)	Main Long. Path Elev. Drop (ft)	Main Long. Path Slope (ft/ft)	Main LS Param-eter (h)	Trib. Area (mi^2)	Trib. Long. Path Length (mi)	Trib. Long. Path Elev. Drop (ft)	Trib. Long. Path Slope (ft/ft)	Trib. LS Param-eter (h)	Area Ratio (nd)	Total Area (mi^2)	Cen-troid Sepa-ration (mi)	LS Param-eter Ratio (nd)	Mean Annual Precip. (in)	Mean I24-2 (in)	Pear-son's ρ	Ken-dall's τ
Main Area (mi^2)	1.000																	
Main Longest Path Length (mi)	0.806	1.000																
Main Longest Path Elev. Drop (ft)	-0.020	-0.133	1.000															
Main Longest Path Slope (ft/ft)	-0.256	-0.431	0.610	1.000														
Main LS Parameter (h)	0.721	0.886	-0.366	-0.462	1.000													
Trib. Area (mi^2)	0.183	0.356	-0.033	-0.195	0.297	1.000												
Trib. Longest Path Length (mi)	0.163	0.474	-0.045	-0.227	0.421	0.841	1.000											
Trib. Longest Path Elev. Drop (ft)	-0.051	-0.157	0.786	0.463	-0.348	0.072	0.095	1.000										
Trib. Longest Path Slope (ft/ft)	-0.117	-0.308	0.640	0.568	-0.433	-0.279	-0.356	0.606	1.000									
Trib. LS Parameter (h)	0.176	0.517	-0.290	-0.332	0.576	0.717	0.891	-0.231	-0.453	1.000								
Area Ratio (nd)	0.504	0.348	-0.004	-0.150	0.269	-0.205	-0.276	-0.136	0.073	-0.207	1.000							
Total Area (mi^2)	0.979	0.833	-0.026	-0.281	0.740	0.379	0.327	-0.033	-0.168	0.313	0.432	1.000						
Centroid Separation (mi)	0.741	0.805	-0.112	-0.339	0.773	0.282	0.273	-0.127	-0.232	0.356	0.338	0.756	1.000					
LS Parameter Ratio (nd)	0.407	0.368	-0.094	-0.200	0.320	-0.222	-0.309	-0.180	0.062	-0.228	0.804	0.338	0.315	1.000				
Mean Annual Precip. (in)	-0.116	-0.163	0.202	0.153	-0.155	-0.118	-0.148	0.161	0.222	-0.172	-0.025	-0.133	-0.126	-0.010	1.000			
Mean I24-2 (in)	-0.023	0.066	-0.112	-0.198	0.119	-0.117	-0.028	-0.110	-0.143	0.027	0.010	-0.046	0.039	0.092	0.591	1.000		
Pearson's ρ	-0.224	-0.293	0.166	0.268	-0.269	-0.031	0.079	0.233	0.082	-0.002	-0.335	-0.217	-0.289	-0.375	0.163	0.200	1.000	
Kendall's τ	-0.134	-0.170	0.145	0.254	-0.179	0.034	0.136	0.173	-0.041	0.055	-0.280	-0.119	-0.145	-0.273	0.136	0.176	0.828	1.000

For the 73 watershed pairs in Group One, drainage area ratio was recommended as the independent variable to estimate correlation at ungaged sites. The best-fit equation for estimating Pearson's ρ for the POM case was (R^2 = 0.443):

$$\rho = 0.96(R_A)^{-0.20} \tag{4.1}$$

where,

ρ = Pearson's ρ, dimensionless

R_A = drainage area ratio, dimensionless

The best-fit equation for estimating Pearson's ρ for the POT case was (R^2 = 0.310):

$$\rho = 0.86(R_A)^{-0.10} \tag{4.2}$$

For Kendall's τ, the POM analysis resulted in the best-fit equation of (R^2 = 0.340):

$$\tau = 0.69(R_A)^{-0.15} \tag{4.3}$$

where,

τ = Kendall's τ, dimensionless

The best-fit equation for estimating τ for the POT case was (R^2 = 0.255):

$$\tau = 0.66(R_A)^{-0.11} \tag{4.4}$$

All of the coefficients and exponents in Equations 4.1 through 4.4 have p-values less than 0.05 except for the 0.96 coefficient in Equation 4.1. Therefore, the coefficient for Equation 4.1 is not considered significantly different from one.

The main stream drainage area ranged from 8.5 to 8,753 square miles and the drainage area ratio ranged from 1 to 474. Capping the main stream drainage area for Group One was recommended because area ratios computed using these larger basins appeared to be misleading in predicting correlation between the main and tributary flows. This observation suggests that there is a practical limit for drainage area ratio beyond which there is no further decrease in correlation between confluent drainage areas. This practical limit might relate to the spatial extent of storms causing rare flood events. (The R^2 for the regression relation also improved.)

The adjusted dataset, with the capping implemented, included main stream drainage areas ranging from 8.5 to 2,000 square miles and drainage area ratios ranging from 1 to 400. Furthermore, if R_A is estimated to be greater than 400, then 400 should be used in the equations.

The study database did not include gage pairs where the smaller drainage area was less than one square mile. Therefore, the results of this research might not apply for watershed pairs with this characteristic. However, a professional engineer must exercise judgment on a site-specific basis to determine if Equations 4.1 through 4.4 might be appropriate or if alternative tools exist for such cases. It might be reasoned, for example, that treating a tributary watershed with a drainage area less than one square mile as one square mile when computing the drainage area

ratio would result in a lower area ratio and, therefore, a conservative estimate of the correlation parameter.

For the 12 watershed pairs in Group Two, a meteorological parameter was more useful than drainage area for estimating the correlation parameters. The hypothesis is that these watersheds are sufficiently large that the entire drainage area does not contribute to coincident events to the same degree as the drainage area of smaller watersheds. Mean annual precipitation, P_M, was selected as the independent variable. See Appendix C.7 for more detail.

The best-fit equation for estimating Pearson's ρ for the POM dataset was (R^2 = 0.651):

$$\rho = 0.046(P_M)^{0.75} \tag{4.5}$$

where,

P_M = Mean annual precipitation, in

For the POT dataset, the best-fit equation for Pearson's ρ was (R^2 = 0.593):

$$\rho = 0.16(P_M)^{0.42} \tag{4.6}$$

The best-fit equation for estimating Kendall's τ for the POM dataset was (R^2 = 0.531):

$$\tau = 0.035(P_M)^{0.76} \tag{4.7}$$

For the POT dataset, the best-fit equation for Kendall's τ was (R^2 = 0.503):

$$\tau = 0.12(P_M)^{0.46} \tag{4.8}$$

All of the coefficients and exponents in Equations 4.5 through 4.8 have p-values less than 0.05.

For Group Two watershed pairs, the mean annual precipitation, P_M, ranges from 15 to 52 in. If P_M is estimated to be greater than 52 inches, then 52 should be used in the equation. If P_M is estimated to be less than 15 inches, then 15 should be used. Although outside the range of mean annual precipitation in the study database, Equations 4.5 through 4.8 are still considered valid. For P_M less than 15, the estimated correlation is anticipated to be conservative and for P_M greater than 52, the estimated correlation approaches one and further increases in correlation will not change the specified combinations significantly.

In addition to the best-fit equations, envelope equations were also developed as described in Appendix C.7.6. Because the best-fit equations exhibit relatively low R^2 values, envelope equations were developed as a design option where the designer could be confident that the correlation would not be underestimated. The use of these equations for design is described in the applications guides in Appendix G and Appendix H. See Appendix F.2 for a discussion of the implications of watershed regulation.

The recommended envelope equations take the form of the best-fit equations with a constant to increase the estimated correlation estimate. The constant was selected to envelope 90 percent of the data.

Group One, Pearson's ρ, POM:

$$\rho = 0.96(R_A)^{-0.20} + 0.19 \tag{4.9}$$

Group One, Pearson's ρ, POT:

$$\rho = 0.86(R_A)^{-0.10} + 0.18 \tag{4.10}$$

Group One, Kendall's τ, POM:

$$\tau = 0.69(R_A)^{-0.15} + 0.23 \tag{4.11}$$

Group One, Kendall's τ, POT:

$$\tau = 0.66(R_A)^{-0.11} + 0.23 \tag{4.12}$$

Group Two, Pearson's ρ, POM:

$$\rho = 0.046(P_M)^{0.75} + 0.20 \tag{4.13}$$

Group Two, Pearson's ρ, POT:

$$\rho = 0.16(P_M)^{0.42} + 0.13 \tag{4.14}$$

Group Two, Kendall's τ, POM:

$$\tau = 0.035(P_M)^{0.76} + 0.21 \tag{4.15}$$

Group Two, Kendall's τ, POT:

$$\tau = 0.12(P_M)^{0.46} + 0.09 \tag{4.16}$$

Coincident flooding events are the result of the complex interaction between the storm and watershed characteristics. Therefore, low R^2 values for predictive equations for correlation between watershed pairs is not surprising. The availability of both best-fit and envelope equations allows a designer to assess the risks at a given site and choose between best-fit and more conservative envelope equations.

The following observations are apparent from the effort to develop regression relations:

1. "Observed" correlations computed from the gage pair data are sensitive to outliers, record length, and other limitations in the record. This sensitivity contributed to the difficulty in defining regression relations. However, the observed correlations did not seem to be sensitive to changing the window for defining a peak from 1 day to 3 days.
2. In many cases, filtering of the data improves the coefficient of determination (R^2), that is, the amount of variance described by the model, but does not substantially change the regression model parameters. This suggests the model parameters are robust.

4.2.4 Correlation Parameter Groupings

An alternative method of estimating the correlation parameter for a pair of confluent watersheds based on groupings of data was also evaluated. (See Appendix C.8 for detailed description of the analyses.) After several trials, the data were grouped based on whether the area ratio was greater or less than 7 and the total drainage area was greater or less than 350 mi^2.

The number of observations, the mean, standard deviation, and the 90th percentile (90% of the values in the group are less than) for each subgrouping are summarized in Tables 4.6 and 4.7 for Pearson's ρ and in Tables 4.8 and 4.9 for Kendall's τ. Correlation between the main and tributary, expressed by Pearson's ρ or Kendall's τ, decreases with increasing area ratio and increasing total area. Because of the small number of observations for $R_A > 7$ and $A_{TOT} <$ than

350 mi^2 comparisons between this grouping and others are not meaningful. The same pattern is observed in all four tables.

A t-test was performed to determine whether the mean values represent distinct populations. Except for those comparisons that included the grouping with only 2 observations ($R_A > 7$ and $A_{TOT} < 350$ mi^2), the differences were significant at the 5-percent level (Appendix C.8).

Table 4.6. Pearson's ρ (POM)

		$A_{TOT} < 350$ mi^2	$A_{TOT} > 350$ mi^2
$R_A < 7$	Number of pairs (n)	27	37
	Mean	0.85	0.74
	Standard Deviation	0.11	0.18
	90th Percentile	0.95	0.93
$R_A > 7$	Number of pairs (n)	2	19
	Mean	0.88	0.50
	Standard Deviation	-	0.24
	90th Percentile	-	0.75

Table 4.7. Pearson's ρ (POT)

		$A_{TOT} < 350$ mi^2	$A_{TOT} > 350$ mi^2
$R_A < 7$	Number of pairs (n)	27	37
	Mean	0.84	0.76
	Standard Deviation	0.12	0.14
	90th Percentile	0.95	0.91
$R_A > 7$	Number of pairs (n)	2	19
	Mean	0.67	0.63
	Standard Deviation	-	0.17
	90th Percentile	-	0.85

Table 4.8. Kendall's τ (POM)

		$A_{TOT} < 350$ mi^2	$A_{TOT} > 350$ mi^2
$R_A < 7$	Number of pairs (n)	27	37
	Mean	0.68	0.57
	Standard Deviation	0.12	0.17
	90th Percentile	0.81	0.77
$R_A > 7$	Number of pairs (n)	2	19
	Mean	0.56	0.39
	Standard Deviation	-	0.14
	90th Percentile	-	0.54

Table 4.9. Kendall's τ (POT)

		$A_{TOT} < 350$ mi^2	$A_{TOT} > 350$ mi^2
$R_A < 7$	Number of pairs (n)	27	37
	Mean	0.65	0.58
	Standard Deviation	0.13	0.15
	90th Percentile	0.80	0.75
$R_A > 7$	Number of pairs (n)	2	19
	Mean	0.44	0.45
	Standard Deviation	-	0.13
	90th Percentile	-	0.57

The recommended correlation parameters are summarized in Table 4.10. A designer would determine if a representative (mean) or envelope (90th percentile) estimate is appropriate given the risks involved in the design. If a representative correlation is most appropriate, the mean value from the applicable group is selected. If a conservative analysis is appropriate, the envelope value is recommended.

For Pearson's ρ, the designer must also consider whether the POM or POT value is most appropriate. The designer can choose to select the more conservative (largest) of the two. For Kendall's τ, there is little difference between the results of the POM and POT analyses, so the largest of the two is recommended and included in Table 4.10.

Table 4.10. Correlation Parameters Based on Grouping

Parameter	Method	Drainage Area Ratio	A_{TOT} < 350 mi^2	A_{TOT} > 350 mi^2
Pearson's ρ (POM)	Representative	$R_A < 7$	0.85	0.74
		$R_A > 7$	-	0.50
	Envelope	$R_A < 7$	0.95	0.93
		$R_A > 7$	-	0.75
Pearson's ρ (POT)	Representative	$R_A < 7$	0.84	0.76
		$R_A > 7$	-	0.63
	Envelope	$R_A < 7$	0.95	0.91
		$R_A > 7$	-	0.85
Kendall's τ (POM and POT)	Representative	$R_A < 7$	0.68	0.58
		$R_A > 7$	-	0.45
	Envelope	$R_A < 7$	0.81	0.77
		$R_A > 7$	-	0.57

There were insufficient data available to estimate correlation where drainage area ratio exceeds 7 and total drainage area is less than 350 mi^2. It might be reasonable to expect that estimates for watershed pairs in this group would be less than estimates for the group with the same total area, but an area ratio less than 7. It might also be reasonable to expect that estimates in this group would exceed estimates for the group with the same area ratio, but a greater total area. Rather than speculating, it is recommended that if a watershed pair has a drainage area ratio greater than 7 and a total area less than 350 mi^2, the correlation parameter be taken from the group where drainage area ratio is less than 7 and the total area is less than 350 mi^2. This is expected to be a conservative approach.

4.2.5 Recommended Method

Two methods for estimating correlation parameters for ungaged watershed pairs were developed; one is based on regression equations and one is based on data groupings. Both methods are considered valid and a designer may choose to use either based on discretionary judgment or unique characteristics of a site. However, if a designer lacks a preference the data groupings method is recommended.

As described in Appendix C.8.3, the regression method exhibited slightly smaller root mean square errors in some cases when estimating the correlation parameters for the study database. However, when the designer has no preference the data groupings method is recommended for three reasons:

1. It is slightly simpler to apply especially for Kendall's τ where the distinction between POM and POT conditions is eliminated.
2. No adjustments to the data are required before or after computing drainage area ratio.
3. It is consistent with the conditional probability matrix selection table used in the total probability method (Chapter 5). Both the selection table and the data groupings method are based on area ratio and total drainage area.

4.3 Event Combination Development

Using the Gumbel-Hougaard copula and noting that copulas are independent of the marginal distributions, potential combinations for the 10-, 25-, 50-, 100-, and 500-yr joint design frequency events can be directly computed. The equations needed are the dependence parameter and copula for Gumbel-Hougaard. The dependence parameter is calculated using Equation C.31 and the copula is provided in Equation C.30 (Appendix C.3.1). The exceedance probability equation, Equation C.8 from Appendix C, is used to determine combinations of events that represent the same exceedance probability.

For example, consider that Kendall's τ has been estimated as 0.5 and calculate the tributary return period given the 100-yr joint return period (0.01 exceedance probability) is of interest and the main stream exceedance probability is 0.02. Equation C.31 may be written to compute the generating function, θ:

$$\theta = \frac{1}{1-\tau} = \frac{1}{1-0.5} = 2.0$$

The tributary exceedance probability is calculated by simultaneously applying the general exceedance probability equation (Equation C.8) and the Gumbel-Hougaard copula (Equation C.30). Equation C.8 is repeated below with the following substitutions (see Appendix C for more detailed explanation: $C_\theta(u,v) = H(x,y)$, $u = F_X(x)$, and $v = F_Y(y)$.

$$P^{\cap}_{x,y} = 1 - u - v + C_\theta(u,v)$$

$P^{\cap}_{x,y}$ is the joint exceedance probability; u is the marginal probability of nonexceedance for the main stream; v is the marginal probaility of nonexceedance for the tributary stream, and $C_\theta(u,v)$ is the copula. Substituting the values for the example, $P^{\cap}_{x,y} = 0.01$ and u = 0.98, yields the following relation with two unknowns:

$$0.01 = 1 - 0.98 - v + C_\theta(u,v)$$

The Gumbel-Hougaard equation (Equation C.30) provides a second relation. For the example, the known values of u and θ are substituted as follows:

$$C_\theta(u,v) = \exp\left\{-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{1/\theta}\right\} = \exp\left\{-\left[(\ln(0.98))^{2.0} + (-\ln(v))^{2.0}\right]^{1/2.0}\right\}$$

Trial and error solution for the example results in v = 0.9853 (nonexceedance probability) for the tributary. Equivalently, this is an exceedance probability of 1-0.9853 = 0.0147 and a return period of 1/0.0147 = 68 years. Therefore, when Kendall's τ = 0.5, the combination of a 50-yr event on the main stream and a 68-yr event on the tributary stream represents an event with a joint exceedance probability of 0.01.

Results for the 10-, 25-, 50-, 100-, and 500-yr joint design frequency events for a range of Kendall's τ estimates are provided in Tables 4.11, 4.12, 4.13, 4.14, and 4.15, respectively. If Kendall's τ is estimated by one of the regression equations, the result is rounded to the nearest 0.1. Then, the appropriate table (4.11 through 4.15) is consulted to determine the potential combinations for hydraulic analysis. The column labeled "equal" represents the return period on both the main and tributary streams that is equivalent to the joint return period.

Potential combinations are also provided in Tables 4.16, 4.17, 4.18, 4.19, and 4.20 based on the Kendall's τ estimates generated from the groupings analysis. Within each table, the row labeled with the value of Kendall's τ for the project site contains combinations of events the designer might analyze with the hydraulic model to determine which is most critical for the design objective. The column labeled "equal" represents the return period on both the main and tributary streams that is equivalent to the joint return period.

Use of the combination tables is illustrated by continuing the previous example where a joint probability of exceedance of 0.01 (100-yr) is of interest and the Kendall's τ has been determined to be 0.5. The designer consults Table 4.14 for the 100-yr joint exceedance probability. Interpreting the top heading of numeric values as main stream return periods, the corresponding tributary stream return periods are taken from the row for τ = 0.5. The following combinations are apparent (main, tributary): (2,100); (5,98); (10,95); (25,87); (50,68), (59,59). Next, the roles of main stream and tributary stream are reversed resulting in the following additional combinations (main, tributary): (100,2); (98,5); (95,10); (87,25); and (68,50). The process of selecting and evaluating combinations is illustrated in detail in Appendix G.

Table 4.11. 10-yr Joint Return Periods

	Return Period on One Stream			
Kendall's τ	1.25	2	5	Equal
0.2	9	7	3	4
0.3	10	8	5	5
0.4	-	9	6	6
0.5	-	10	8	6
0.6	-	-	9	7
0.7	-	-	10	8
0.8	-	-	-	9
0.9	-	-	-	9

Table 4.12. 25-yr Joint Return Periods

	Return Period on One Stream				
Kendall's τ	1.25	2	5	10	Equal
0.2	23	20	13	7	8
0.3	24	22	17	11	11
0.4	25	24	21	16	13
0.5	-	25	23	20	15
0.6	-	-	24	22	17
0.7	-	-	25	24	19
0.8	-	-	-	25	21
0.9	-	-	-	-	23

Table 4.13. 50-yr Joint Return Periods

	Return Period on One Stream					
Kendall's τ	1.25	2	5	10	25	Equal
0.2	47	41	30	21	7	15
0.3	49	46	39	31	15	20
0.4	50	49	45	40	25	25
0.5	-	50	48	45	34	30
0.6	-	-	50	48	42	34
0.7	-	-	-	50	47	39
0.8	-	-	-	-	49	43
0.9	-	-	-	-	50	46

Table 4.14. 100-yr Joint Return Periods

	Return Period on One Stream						
Kendall's τ	1.25	2	5	10	25	50	Equal
0.2	95	86	67	53	32	12	28
0.3	98	94	84	74	54	28	39
0.4	100	98	94	88	73	48	49
0.5	-	100	98	95	87	68	59
0.6	-	-	100	99	95	83	68
0.7	-	-	-	100	99	93	77
0.8	-	-	-	-	100	99	85
0.9	-	-	-	-	-	100	93

Table 4.15. 500-yr Joint Return Periods

	Return Period on One Stream							
Kendall's τ	1.25	2	5	10	25	50	100	Equal
0.2	485	452	395	354	294	239	167	132
0.3	496	486	462	439	398	353	286	189
0.4	499	497	489	480	459	431	381	243
0.5	-	500	498	496	488	474	445	293
0.6	-	-	-	500	498	494	481	341
0.7	-	-	-	-	500	499	497	385
0.8	-	-	-	-	-	-	500	426
0.9	-	-	-	-	-	-	500	464

Table 4.16. 10-yr Joint Return Periods (Groupings)

				Return Period on One Stream			
			Kendall's τ	1.25	2	5	Equal
Best-fit	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.68	-	10	9	8
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.59	-	10	9	7
	$R_A > 7$	$A_{TOT} < 350$ mi^2					
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.45	10	9	7	6
Envelope	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.81	-	-	10	9
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.77	-	-	10	8
	$R_A > 7$	$A_{TOT} < 350$ mi^2					
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.57	-	10	8	7

Table 4.17. 25-yr Joint Return Periods (Groupings)

				Return Period on One Stream				
			Kendall's τ	1.25	2	5	10	Equal
Best-fit	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.68	-	-	25	24	19
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.59	-	25	24	22	17
	$R_A > 7$	$A_{TOT} < 350$ mi^2						
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.45	25	24	22	18	14
Envelope	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.81	-	-		25	22
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.77	-	-	-	25	21
	$R_A > 7$	$A_{TOT} < 350$ mi^2						
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.57		25	24	22	17

Table 4.18. 50-yr Joint Return Periods (Groupings)

				Return Period on One Stream					
			Kendall's τ	1.25	2	5	10	25	Equal
Best-fit	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.68	-	-	-	50	48	38
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.59	-	50	49	48	40	33
	$R_A > 7$	$A_{TOT} < 350$ mi^2							
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.45	50	49	47	43	30	27
Envelope	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.81	-	-	-		50	43
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.77	-	-	-	50	49	41
	$R_A > 7$	$A_{TOT} < 350$ mi^2							
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.57	-	50	49	48	40	33

Table 4.19. 100-yr Joint Return Periods (Groupings)

				Return Period on One Stream						
			Kendall's τ	1.25	2	5	10	25	50	Equal
Best-fit	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.68	-	-	-	100	98	92	75
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.59	-	-	100	98	94	80	67
	$R_A > 7$	$A_{TOT} < 350$ mi^2								
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.45	100	99	96	92	80	58	54
Envelope	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.81	-	-	-	-	100	99	86
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.77	-	-	-		100	98	83
	$R_A > 7$	$A_{TOT} < 350$ mi^2								
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.57	-	100	99	98	93	79	66

Table 4.20. 500-yr Joint Return Periods (Groupings)

				Return Period on One Stream							
			Kendall's τ	1.25	2	5	10	25	50	100	Equal
Best-fit	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.68	-	-	-	-	500	499	495	376
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.59	-	-	500	499	497	491	476	331
	$R_A > 7$	$A_{TOT} < 350$ mi^2									
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.45	500	499	495	490	477	456	417	269
Envelope	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.81	-	-	-	-	-	-	500	430
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.77	-	-	-	-	-	500	499	414
	$R_A > 7$	$A_{TOT} < 350$ mi^2									
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.57	-	-	500	499	496	490	473	327

4.4 Complementary Data in Paired Datasets

As noted in Table 4.3, the tributary nonexceedance probabilities and return periods are not based on the annual peak series, but are based on a series of observations that are coincident with the peak series on the main stream. In this POM example, the main stream gage is the primary gage and the tributary gage is the complementary gage. This distinction between the annual peak series of data and the complementary series of data might be important in some cases when the designer converts a nonexceedance probability (or return period) to a discharge for hydraulic analysis. The importance of this issue is illustrated using two gages from the research database.

Compilation of the POM datasets started with the annual peak series of the main stream (primary) gage and then paired each of those observations with the same day observations from the tributary stream (complementary) gage. Similarly, compilation of the POT datasets started with the annual peak series of the tributary (primary) gage and then paired each of those observations with the same day observations from the main stream (complementary) gage.

Three different annual series (all based on mean daily flow) for the main stream gage (01403060) from Gage Pair 05 are displayed in Figure 4.4. The series labeled “annual” is the complete record of annual maximum peaks (n=67) available for the gage. The series labeled “POM” are the values in the peak on main (POM) dataset. Since this gage is the primary gage in Gage Pair 05 when compiling the POM dataset, the POM series is also composed only of annual peaks. However, a smaller number of observations (n=27) are represented because the periods of record between the main and tributary gages must overlap. Even though the number

of observations is much less, the plotting positions for the Annual and POM series describe the same curve. The series label "POT" represents the complementary data from gage 01403060 that is coincident with the annual peaks from the tributary gage, which is the primary gage when the POT dataset is compiled. The plotting positions for this series follow a much different curve as shown in the figure.

When a joint probability strategy produces a probability of exceedance (or nonexceedance), it must be correctly interpreted to infer the appropriate discharge for design. Using gage 01403060 (Figure 4.4) as an example, a 0.1 exceedance probability on the POT series indicates a discharge of 12,500 ft^3/s while that same exceedance probability on the annual series indicates a discharge twice the magnitude at 25,000 ft^3/s.

The same analysis for the tributary gage from Gage Pair 06 (01446000) is summarized in Figure 4.5. In this case, it is expected that the annual (n=42) and the POT (n=42) describe the same relation. In fact, they are identical. However, the POM series (n=43), where the data are from the complementary gage, again shows a different curve, though not as distinct as previously observed.

Gage pair 05 has a much lower correlation than Gage Pair 06, which may explain the much greater difference between the curves for Gage Pair 06. The higher correlation in Gage Pair 06 makes it more likely that the same annual peaks will be included in all three series. An attempt was to try to relate the difference between the annual series and the complementary data series as a function of correlation between the paired watersheds. However, no relations were derived.

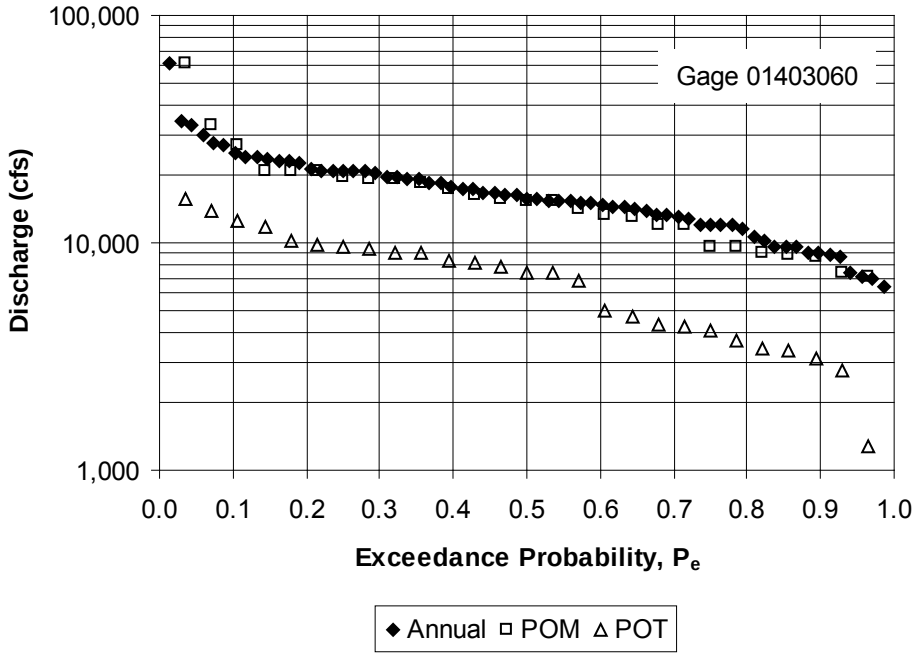

Figure 4.4. Flood Frequency Curves for Gage 01403060 (GP05)

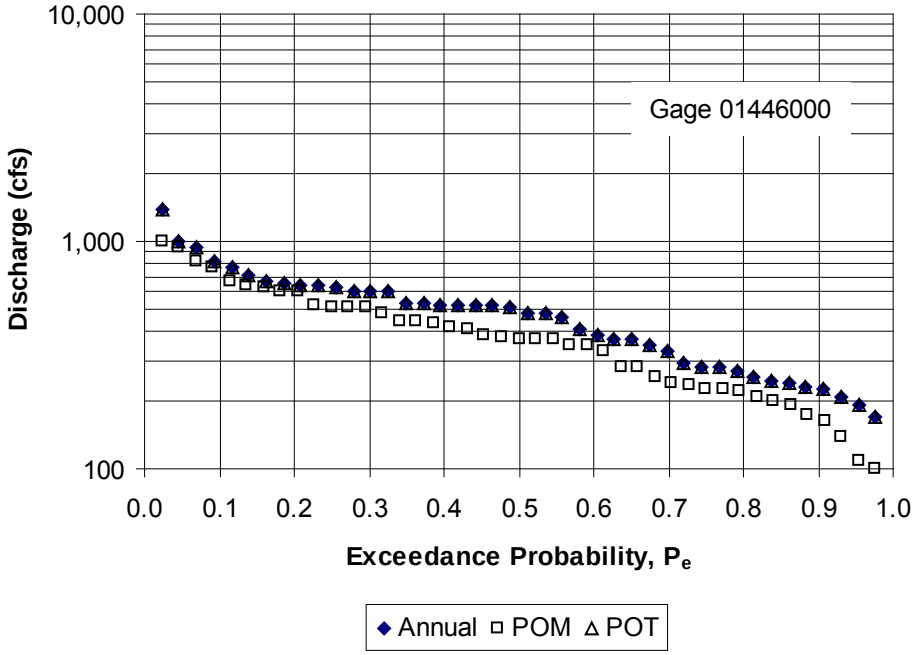

Figure 4.5. Flood Frequency Curves for Gage 01446000 (GP06)

5. Total Probability Method

The overall process of applying the total probability method is summarized in Appendix D.1. This chapter addresses the development of conditional probability matrices (CPMs) for use within the context of the total probability method. CPMs were developed from the expected correlation between the main and tributary streams derived from analyses of the gage pair database.

The data included annual maximum flow series records for 170 gages, as well as, peak on main (POM) and peak on tributary (POT) coincident flow datasets for 85 gage pairs. The total number of flow pairs included in the dataset was 3,382 for the POM database and 3,348 for the POT database. All data were derived from the mean daily flow records, as previously described.

5.1 Pair Count Matrices

Pair count matrices are a data presentation tool that facilitates assessment of coincident flow data. Annual exceedance probability (AEP) pair count matrices were developed for each gage pair. Each row and column in a matrix is based on a "bin" that holds a range of AEP values. Bin ranges were defined to provide a useful categorization of flows and are summarized in Table 5.1.

Table 5.1. AEP Bins

Bin number	Bin AEP	Bin AEP range
1	0.500	0.316 > AEP >= 1.000
2	0.200	0.141 > AEP >= 0.316
3	0.100	0.063 > AEP >= 0.141
4	0.040	0.028 > AEP >= 0.063
5	0.020	0.014 > AEP >= 0.028
6	0.010	0.004 > AEP >= 0.014
7	0.002	0.000 > AEP >= 0.004

The pair count matrices for Gage Pair 06 are summarized in Tables 5.2a and 5.2b for the POM and POT databases, respectively. Highly correlated flows result in most flow pairs near the diagonal of the matrix. Uncorrelated flows result in most flow pairs in the first column (POM) or first row (POT) of the matrix. Gage Pair 06 exhibits a high degree of correlation with many flow pairs located on the diagonal. Additional examples of gage pair count matrices are found in Appendix D.

Aggregate pair count matrices (one matrix for the POM case and one matrix for POT case) were also generated by summing the pair count matrices for multiple gage pairs. An overview of the behavior of multiple gage pairs might become apparent from viewing the aggregate matrices as a whole or by subdividing the aggregate pair count matrices by various filter criteria. The goals of applying filter criteria to generate aggregate pair count matrices were to:

- Identify watershed characteristics to which the correlation between main and tributary flows is sensitive.
- Identify sub ranges of watershed characteristics that result in unique aggregate pair count matrices that represent varying degrees of correlation between main and tributary flows.

Table 5.2a. Pair Counts for Gage Pair 06 (POM)

		Tributary AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Total
Main stream AEP	0.500	30	3	1					34
	0.200	3	2	1					6
	0.100			1	1				2
	0.040	1							1
	0.020								0
	0.010								0
	0.002								0
		34	5	3	1	0	0	0	43

Table 5.2b. Pair Counts for Gage Pair 06 (POT)

		Tributary AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Total
Main stream AEP	0.500	28	7	1					36
	0.200		2	1			1		4
	0.100			1	1				2
	0.040								0
	0.020								0
	0.010								0
	0.002								0
		28	9	3	1	0	1	0	42

Watershed characteristics used to filter gage pairs included:

- Drainage area.
- Drainage area ratio.
- Hydrologic unit code (HUC).
- Geographic region.
- Length/slope time parameter.
- Length/slope time parameter ratio.

The aggregate pair count matrices resulting from filtering were inspected visually for suitability as predictors of flow correlation. If varying the range of watershed characteristics did not result in a substantial change in the trend of correlation, then those characteristics were considered poor indicators of correlation. If varying the range of watershed characteristics resulted in a substantial change in correlation, then those characteristics were considered good indicators of correlation.

Simultaneous filtering based on the total drainage area and drainage area ratio resulted in aggregate pair count matrices that were the most revealing of the degree of correlation between main and tributary flows. Filtering based on both total drainage area and drainage area ratio (main stream to tributary) proved more effective than filtering on either of these characteristics alone. For pairs of small watersheds, correlation was observed in the form of AEP pairs near

the diagonal of the count matrix. For large watersheds, lower drainage area ratio appeared to be associated with higher correlation.

Other filtering strategies were examined. Filtering based strictly on HUC was attempted but revealed that there are simply not enough flow pairs in the dataset to adequately define the 18 hydrologic regions in the study. Hydrologic regions were grouped in an attempt to identify any systematic dependency of the structure of the count matrix on geographic region. The following general geographic regions (HUCs in parentheses) were defined: East coast (01-03), Midwest (04-12), arid Southwest (13-16), and West coast (17-18). Filtering based on geographic region alone did not reveal any obvious differences between regions. Further filtering, based on drainage area or other characteristics, was complicated by the lack of flow pairs for a single geographic region.

Length/slope time parameter and length/slope time parameter ratio were also evaluated as a filter. The rationale for using length/slope time parameter was that this characteristic may better account for the response time of the watershed than drainage area. However, using a length/slope time parameter filter resulted in less well-defined pair count matrices than using drainage area. It is likely that the suitability of length/slope time parameter for filtering of AEP pairs is limited to small watersheds in which spatial distribution of rainfall during a storm event is more nearly uniform, and therefore the timing of peak runoff is a stronger function of length/slope time parameter.

5.2 Conditional Probability Matrices

Conditional probability matrices (CPMs) were developed based on aggregated pair count matrices that offered the greatest utility in capturing gage pair correlation. The selected filter ranges for the pair count matrices are summarized in Table 5.3. Of the 85 gage pairs used in this assessment, 2, 4, 53, and 26 pairs fell in the 1 to 50, 50 to 100, 100 to 1000, and 1000 or greater categories, respectively. Because the number of gage pairs is limited, it was not possible to generate pair count matrices for each of the combinations of total area and area ratio shown in the table. Instead, four pair matrices were identified according to qualitative degrees of correlation: high, moderate, mixed, and poor.

Table 5.3. CPM Selection Table – Degree of Correlation

Area ratio	Total area (sq. mi.)			
	1 to 50	50 to 100	100 to 1000	1000 or greater
1 to 10	High	Moderate	Mixed	Mixed
10 to 100	High	Moderate	Mixed	Poor
100 or greater	High	Moderate	Poor	Poor

The database used for this study did not include data pairs where the smaller watershed drainage area was less than one square mile. Therefore, the results of this research may not apply for watershed pairs with this characteristic. However, a professional engineer must exercise judgment on a site-specific basis to determine if Table 5.3 is appropriate or if better alternatives exist. It might be reasoned, for example, that treating a tributary watershed with a drainage area less than one square mile as one square mile when computing the drainage area ratio would result in a lower area ratio and, therefore, selection of the same or a more conservative (higher correlation) CPM matrix.

The high, mixed, and poor pair count matrices were the result of aggregating the gage pairs that exhibit the combinations of total area and area ratio indicated in Table 5.3. The mixed conditional probability matrix was named because of evidence of a mixed population of moderately correlated and uncorrelated events. In the pair count matrices, this is seen as a

strong trend of pairs clustered near the diagonal of the matrix, along with a strong trend of pairs with AEP = 0.5.

There were insufficient gage pairs to define the moderate pair count matrix. To resolve this, the moderately correlated pair count pattern observed within the mixed pair count matrix was adopted as the pattern for the moderate matrix. In so doing, the moderate case provides a reasonable transition between the high and mixed cases.

The four aggregated pair count matrices were converted to CPMs by normalizing the pair counts in each matrix cell to the total number of counts for the row (POM) or column (POT). The resulting CPMs are summarized in Tables 5.4a through 5.7b. For the POM CPMs, the AEP for the main stream flow bin is given and the conditional probabilities for the tributary stream flow bins are given in each row. Similarly, for the POT CPMs, the AEP for the tributary stream flow bin is known and the conditional probabilities for the main stream flow bins are given in each column.

Minor adjustments to the resulting CPMs were then made to remove outliers and smooth the empirical distributions to provide well-behaved concurrent density functions. These adjustments were based on visual inspection of the matrices.

The *high* CPMs (Tables 5.4a and 5.4b) indicate that the design AEP condition is given by concurrent main and tributary flows of both AEP equal to the design AEP. This special case does not require total probability computations. Instead, for example, the AEP = 0.01 design condition is given by AEP =0.01 on the main stream and AEP = 0.01 on the tributary stream.

The *poor* CPMs (Tables 5.7a and 5.7b) were generated by populating the cells with the accumulated concurrent flow probability represented by the AEP bins. Smaller terms in these matrices could be aggregated with larger terms to reduce the number of non-zero cells (and required hydraulic model runs).

Application of the CPMs within the total probability method is illustrated with an example in Appendix G.

Table 5.4a. CPM for Highly Correlated Flows (POM)

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Sum
Main Stream AEP	0.500	1							1
	0.200		1						1
	0.100			1					1
	0.040				1				1
	0.020					1			1
	0.010						1		1
	0.002							1	1

Table 5.4b. CPM for Highly Correlated Flows (POT)

		Tributary Stream AEP						
		0.500	0.200	0.100	0.040	0.020	0.010	0.002
Main Stream AEP	0.500	1						
	0.200		1					
	0.100			1				
	0.040				1			
	0.020					1		
	0.010						1	
	0.002							1
	Sum	1	1	1	1	1	1	1

Table 5.5a. CPM for Moderately Correlated Flows (POM)

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Sum
Main Stream AEP	0.500	0.9	0.1						1
	0.200	0.4	0.5	0.1					1
	0.100	0.2	0.3	0.4	0.1				1
	0.040		0.2	0.3	0.4	0.1			1
	0.020			0.2	0.3	0.4	0.1		1
	0.010				0.2	0.3	0.4	0.1	1
	0.002					0.2	0.4	0.4	1

Table 5.5b. CPM for Moderately Correlated Flows (POT)

		Tributary Stream AEP						
		0.500	0.200	0.100	0.040	0.020	0.010	0.002
Main Stream AEP	0.500	0.9	0.4	0.2				
	0.200	0.1	0.5	0.3	0.2			
	0.100		0.1	0.4	0.3	0.2		
	0.040			0.1	0.4	0.3	0.2	
	0.020				0.1	0.4	0.3	0.2
	0.010					0.1	0.4	0.4
	0.002						0.1	0.4
	Sum	1	1	1	1	1	1	1

Table 5.6a. CPM for Mixed Correlated Flows (POM)

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Sum
Main Stream AEP	0.500	0.9	0.1						1
	0.200	0.5	0.4	0.1					1
	0.100	0.3	0.3	0.3	0.1				1
	0.040	0.3	0.1	0.2	0.3	0.1			1
	0.020	0.3		0.1	0.2	0.3	0.1		1
	0.010	0.3			0.1	0.2	0.3	0.1	1
	0.002	0.3				0.1	0.3	0.3	1

Table 5.6b. CPM for Mixed Correlated Flows (POT)

		Tributary Stream AEP						
		0.500	0.200	0.100	0.040	0.020	0.010	0.002
Main Stream AEP	0.500	0.9	0.5	0.3	0.3	0.3	0.3	0.3
	0.200	0.1	0.4	0.3	0.1			
	0.100		0.1	0.3	0.2	0.1		
	0.040			0.1	0.3	0.2	0.1	
	0.020				0.1	0.3	0.2	0.1
	0.010					0.1	0.3	0.3
	0.002						0.1	0.3
	Sum	1	1	1	1	1	1	1

Table 5.7a. CPM for Poorly Correlated Flows (POM)

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Sum
Main Stream AEP	0.500	0.684	0.175	0.078	0.035	0.014	0.010	0.004	1
	0.200	0.684	0.175	0.078	0.035	0.014	0.010	0.004	1
	0.100	0.684	0.175	0.078	0.035	0.014	0.010	0.004	1
	0.040	0.684	0.175	0.078	0.035	0.014	0.010	0.004	1
	0.020	0.684	0.175	0.078	0.035	0.014	0.010	0.004	1
	0.010	0.684	0.175	0.078	0.035	0.014	0.010	0.004	1
	0.002	0.684	0.175	0.078	0.035	0.014	0.010	0.004	1

Table 5.7b. CPM for Poorly Correlated Flows (POT)

		Tributary Stream AEP						
		0.500	0.200	0.100	0.040	0.020	0.010	0.002
Main Stream AEP	0.500	0.684	0.684	0.684	0.684	0.684	0.684	0.684
	0.200	0.175	0.175	0.175	0.175	0.175	0.175	0.175
	0.100	0.078	0.078	0.078	0.078	0.078	0.078	0.078
	0.040	0.035	0.035	0.035	0.035	0.035	0.035	0.035
	0.020	0.014	0.014	0.014	0.014	0.014	0.014	0.014
	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010
	0.002	0.004	0.004	0.004	0.004	0.004	0.004	0.004
	Sum	1	1	1	1	1	1	1

6. Conclusions and Recommendations

The research summarized in this report advances the state of the art for deriving appropriate discharges for coincident flooding on confluent streams. Previously, drainage area ratio had been the primary factor in describing coincident flooding and it was unclear how broadly applicable that guidance (see Section 2.2.7) was to be considered. Employing gaged data from across the coterminous United States, a better understanding of the factors influencing joint probability was achieved along with recommendations for design tools for practitioners and two applications guides (Appendix G and Appendix H).

Simplifying assumptions were made at various junctures of the research including those required because of the limits of data availability and study resources. Suggestions for further research are provided to address some of these assumptions.

6.1 Factors Influencing Joint Probability

Stream flow at a given point in a surface water network is a function of many hydrometeorological variables. Methods for estimating peak stream flow for design purposes strip the great many variables down to typically one (usually drainage area) to four or five depending on the methodology. Errors in predicting stream flow can be substantial even in the case of a single stream.

In the confluent stream case, the joint probability of flooding is based on all of those hydrometeorological variables for the separate watersheds plus the interrelation between the two. As previously noted, the drainage area ratio between the two watersheds has historically been the primary tool for providing that linkage.

This research examined a range of watershed and meteorological variables in an effort to gain insight into the problem. As part of the analyses for estimating Kendall' τ and Pearson's ρ for the bivariate and copula strategies, the factors determined to be most useful were drainage area ratio, total drainage area, and mean annual precipitation. Like the drainage area ratio, the length slope parameter ratio, provides a measure of the relative size of the paired watersheds. In fact, drainage area ratio and length slope parameter ratio are highly correlated and both demonstrated some value for estimating correlation. Although the length slope parameter ratio was slightly better, the improvement was not sufficient to recommend this parameter over area ratio. Area ratio is preferred because it is more easily estimated and has been previously linked with correlation relations.

While developing the conditional probability matrices for use in the total probability method, drainage area ratio and total area provided the greatest insight. While not identical results, there is overlap with the conclusions from the bivariate and copula analyses.

The joint probability problem was framed assuming the design point and influence reach are on the tributary (smaller) stream. However, the tools developed from this study may also be used when the design point and influence reach are on the main stream.

6.2 Recommended Tools

Two general areas of investigation yielded practical tools ready for application by practitioners. One is a direct method for determining a set of exceedance probability combinations associated with the desired joint probability for design based on copulas with univariate marginals. The second is a series of conditional probability matrices for use in the total probability method.

6.2.1 Copulas

As discussed in Section 4.1, the best-fit statistical method of the four distributions/copulas considered in detail was the Gumbel-Hougaard copula. In addition to data fit, other factors considered were ease of use and robustness of the methodology.

The Gumbel-Hougaard and Frank copulas require estimation of a single parameter for application to an ungaged watershed pair to characterize the dependence between the main and tributary streams. This dependence parameter is a function of Kendall's τ. This research resulted in methods to estimate Kendall's τ for ungaged watershed pairs based on watershed and meteorological characteristics.

The bivariate normal and Gumbel (mixed model) distributions require the estimation of five parameters: a location and scale parameter for each marginal distribution (main and tributary streams) and a dependence parameter. The dependence parameter is calculated as a function of Pearson's ρ. Because of the need for additional parameters, the bivariate normal and Gumbel (mixed model) distributions are not as easy to use as the copula methods.

The two copula methods offer advantages over the bivariate normal and mixed model Gumbel distributions. First, the mixed model bivariate Gumbel (mixed model) distribution may only be applied when the estimated Pearson's ρ correlation between the paired watersheds is less than two-thirds. Therefore, it cannot be applied to watersheds when the correlation coefficient exceeds two-thirds. Second, a fundamental advantage of the copula methods over both bivariate distributions is that underlying marginal distributions are not assumed. Therefore, a designer may choose any appropriate marginal distribution for both the main and tributary streams or two entirely different marginal distributions for the main and tributary streams. However, given its frequent application for annual peak flow, the Log Pearson III distribution for both the main and tributary streams would likely be a common selection. Both the normal and Gumbel bivariate distributions require that the marginal distributions are normal and Gumbel, respectively.

Considering the fit of the observed gage pair data to the bivariate distributions and copulas, the Gumbel-Hougaard method performed better than the normal and Gumbel bivariate distributions and the Frank copula. Considering the fit with observed data, simplicity (need to estimate a single dependence parameter), and flexibility (no assumption of the marginal distribution), the Gumbel-Hougaard copula was recommended for use. However, application of this method does not require a background in statistics because guidance developed as part of this research transformed the method to a series of tables. These tables provide the designer with a set of combinations of main and tributary stream conditions, each having a specified joint probability. Each combination is analyzed hydraulically to determine the critical design condition. Application guides for designers are provided in Appendix G and Appendix H.

6.2.2 Total Probability Method

The second method is the total probability method, which is based on the total probability theorem. The key element of this part of the research was the development of a series of conditional probability matrices for application within the context of the total probability method described in Section 4.2. Four sets of matrices for high, moderate, mixed, and poor correlation are presented. Full implementation of the total probability method involves evaluating several hydraulic scenarios to generate an exceedance probability curve for stage at the design location. This method may be preferred over the copula method if the objective is to evaluate several alternatives over a range of possible occurrences in the context of a cost-benefit analysis. An application guide for designers is provided in Appendix G.

6.3 Suggested Research

As with any research effort, this study was conducted with resource constraints established at the outset. Several decisions pertaining to the type of data used and how it was to be analyzed were made and documented. In addition, during the course of the research, unanticipated issues arose. Suggestions for additional research work are as follows:

1. Expanded use of instantaneous data.

   One of the primary reasons for focusing on daily rather than instantaneous data was the lack of instantaneous data. The USGS is gradually making more of these data available through the Instantaneous Data Archive (IDA). As the availability of these data expands, the comparison with daily data can be expanded with the potential for eventually using all instantaneous data. However, record lengths will likely limit the use of instantaneous data to some degree.

2. Expansion of results to watershed pairs where the smaller watershed has a drainage area less than 1 mi^2.

   The study database did not include watershed pairs with these smaller watersheds. Although it may be reasonable to extend the proposed tools to these cases, there is no evidence this is true.

3. Expansion of the watershed pairs to include larger drainage area ratios, especially for pairs with a combined watershed area less than 350 mi^2.

   The study database did not include watershed pairs with an area ratio greater than 500 and only two with an area ratio greater than 7 when the combined area is less than 350 mi^2. It appears likely that larger ratios will not show substantial reductions in correlations, but it would be useful to verify this hypothesis. For smaller combined watersheds, area ratios above 7 (and below 500) may have some effect on the correlation, but insufficient data were available to quantify the effect as part of this study.

4. Evaluation of alternative copulas.

   Because of their use in other hydrologic applications, two Archimedean copulas were featured in this study: Gumbel-Hougaard and Frank. Other copulas and other copula families from the non-Archimedean group may be of interest.

5. Development of better relations between watershed and meteorological characteristics and measures of correlation.

   The relations developed in this study were statistically weaker than would ultimately be desired. This was addressed, in part, by the development of best-ft and envelope equations and by estimates based on data groupings. Further testing may yield benefits.

6. Quantification of the relation between annual series data and complementary data series.

   In the POM datasets, the tributary series is complementary to the main stream peaks, but is not the tributary annual series. Conversely, in the POT datasets, the main stream series is complementary to the tributary peaks, but is not the main stream annual series. A predictive tool for estimating the difference between the complementary and annual series was not developed as part of this

study, but would help quantify the degree of conservatism embedded in the copula tables.

7. Consideration of how the problem formulation changes for design objectives other than stage, e.g. velocity.

   Exceedance probability for stage may not be equivalent to exceedance probability for velocity or other design considerations. How the formulation of the problem might change is a useful investigation. However, use of the total probability method, which considers all possible events, might be important for addressing other design objectives.

8. Further evaluation of the degree to which watershed regulation may affect application of the joint probability methods developed in this research.

7. References

AASHTO, 2000. *Model Drainage Manual*, American Association of State Highway and Transportation Officials, Chapter 13.

Asquith, W. H., 2002. *Effects of Regulation on L-Moments of Annual Peak Streamflow in Texas*, United States Geological Survey, WRIR 01-4243.

Benson, M. A., 1963. *Factors Influencing the Occurrence of Floods in a Humid Region of Diverse Terrain*, United Stated Geological Survey, Water-Supply Paper 1580-B.

Benson, M. A., 1964. *Factors Influencing the Occurrence of Floods in the Southwest*, United Stated Geological Survey, Water-Supply Paper 1580-D.

Bogardi, I., L. Duckstein, and F. Szidarovszky, 1975. "Hydrological System Reliability at the Confluence of Rivers," Application of Mathematical Models in Hydrology and Water Resources Systems, *Proceedings of the Bratislava Symposium*, Czechoslovakia, IASH-AISH Publication No. 115, International Association of Hydrological Sciences, pp. 36–44, September.

Bonnin, G., D. Martin, T. Parzybok, B. Lin, D. Riley, and M. Yekta, 2006a. *Precipitation Frequency Atlas of the United States*. NOAA Atlas 14 Vol. 1, Version 4.0, National Weather Service, Silver Spring, MD.

Bonnin, G., D. Martin, T. Parzybok, B. Lin, D. Riley, and M. Yekta, 2006b. *Precipitation Frequency Atlas of the United States*. NOAA Atlas 14 Vol. 2, Version 3.0, National Weather Service, Silver Spring, MD.

Bowker, A. H. and G. J. Lieberman, 1972. *Engineering Statistics, Second Edition,* Prentice-Hall, Inc., Englewood Cliffs, N.J.

Coles, S., 2001. *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag, Berlin, p. 48.

Cunnane, C., 1978. "Unbiased Plotting Positions-A Review," *Journal of Hydrology,* 37, 205–222.

De Michele, C., G. Salvadori, M. Canossi, A. Petaccia, and R. Rosso, 2005. "Bivariate Statistical Approach to Check Adequacy of Dam Spillway," *Journal of Hydrologic Engineering,* Vol. 10 No. 1, pp. 50–57, ASCE.

Dyhouse, G. R., 1985. "Stage Frequency Analysis at a Major River Junction," *Journal of Hydraulic Engineering*, Vol. 111, No. 4, ASCE, April.

Favre, A.C., S. El Adlouni, L. Perrault, N. Thiemonge, and B. Bobee, 2004. "Multivariate Hydrologic Frequency Analysis using Copulas," *Water Resources Research*, Vol. 40.

Genest, C. and L.P. Rivest, 1993. "Statistical Inference Procedures for Bivariate Archimedean Copulas," *Journal of the American Statistical Association*, Vol. 88, No. 423, pp. 1034–1043.

Genest, C. and A.C. Favre, 2007. "Everything You Always Wanted to Know About Copula Modeling But Were Afraid to Ask," *Journal of Hydrologic Engineering*, Vol. 12, No. 4, 347–368. ASCE

Genz, A., 2004. "Numerical Computation of Rectangular Bivariate and Trivariate Normal and t Probabilities," *Statistics and Computing*, Vol. 14, Issue 3, pp. 151–160.

EPA and USGS, 2007. "NHDPlus User Guide," http://www.horizon-systems.com/nhdplus/data/NHDPLUS_UserGuide.pdf

HEC, 1992. *Flood Flow Frequency Analysis, User's Manual (HEC-FFA)*, CPD-13, May.

Interagency Advisory Committee of Water Data, 1982. *Guidelines for Determining Flood Flow Frequency, Bulletin 17B*, U.S. Geological Survey, Reston, VA.

Loganathan, G. V., C. Y. Kuo, and J. Yannaccone, 1987. "Joint Probability Distribution of Stream Flows and Tides in Estuaries," *Nordic Hydrology*, Vol. 18, pp. 237–246.

Morris, C. D. and S. J. Calise, 1987. "Bivariate Analysis of Concurrent Flooding," International Symposium on Flood Frequency and Risk Analysis: Hydrologic Frequency Modeling, *Proceedings,* May 14–17, 1986, Baton Rouge, LA, ed. Singh V. P., Reidel: Dordrecht, The Netherlands, pp. 615–632.

Nelsen, R. B., 2006. *An Introduction to Copulas*, (Second edition) Springer Series in Statistics, New York: Springer-Verlag.

NOAA, 1961. *Rainfall Frequency Atlas of the United States for Durations from 30 minutes to 24 Hours and Return Periods from 1 to 100 Years,* Technical Paper 40.

NOAA, 1973. *Precipitation Frequency Atlas of the Western United States*, Atlas 2.

Pingel, N. and D. Ford, 2004. "Interior Floodplain Flood-Damage Reduction Study," *Journal of Water Resources Planning and Management*, Vol. 130, No. 2, pp. 123–130, ASCE.

Prior-Jones, R. L. and F. L. Beiboer, 1990. "Use of Joint Probability in Deriving Environmental Design Criteria," Environmental Forces on Offshore Structures and their Prediction, *Proceedings,* London, November 28–29.

Raynal, J. A. and J. D. Salas, 1987. "A Probabilistic Model for Flooding Downstream of the Junction of Two Rivers," International Symposium on Flood Frequency and Risk Analysis: Hydrologic Frequency Modeling, *Proceedings,* May 14–17, 1986, Baton Rouge, LA, ed. Singh V. P., Reidel: Dordrecht, The Netherlands, pp. 595–601.

Sackl, B. and H. Bergman, 1987. "A Bivariate Flood Model and its Application," International Symposium on Flood Frequency and Risk Analysis: Hydrologic Frequency Modeling, *Proceedings,* May 14–17, 1986, Baton Rouge, LA, ed. Singh V. P., Reidel: Dordrecht, The Netherlands, pp. 571–582.

Salvadori, G., C. De Michele, N. T. Kottegoda, and R. Rosso, 2007. *Extremes in Nature: An Approach Using Copulas*, Volume 56 of Water Science and Technology Library, Springer.

Shiau, J. T., H. Y. Wang, and C. T. Tsai, 2006. "Bivariate Frequency Analysis of Floods Using Copulas," *Journal of the American Water Resources Association,* Vol. 42, No. 6, pp. 1549–1564.

Stewart, D. W., A. Rea, and D. M. Wolock, 2006. http://water.usgs.gov/GIS/metadata/usgswrd/XML/streamgages.xml

Sumioka, S.S., D.L. Kresch, and K.D. Kasnick, 1998. *Magnitude and Frequency of Floods in Washington*, Water-Resources Investigations Report 97-4277, USGS.

USACE, 1993. *Hydrologic Frequency Analysis*, EM 1110-2-1415. (Primarily Chapter 11 "Frequency of Coincident Flow").

Yue, S., 2000. "The Bivariate Lognormal Distribution to Model a Multivariate Flood Episode," *Hydrological Processes*, Vol. 14, pp. 2575–2588.

Yue, S., T. Ouarda, B. Bobee, P. Legendre, and P. Bruneau, 1999. "The Gumbel Mixed Model for Flood Frequency Analysis," *Journal of Hydrology*, Vol. 226, pp. 88–100.

Yue, S., T. Ouarda, and B. Bobee, 2001. “A Review of Bivariate Gamma Distributions for Hydrological Application,” *Journal of Hydrology*, Vol. 246, pp. 1–18.

Yue, S. and P. Rasmussen, 2002. “Bivariate Frequency Analysis: Discussion of Some Useful Concepts in Hydrological Application,” *Hydrological Processes*, Vol. 16, pp. 2881–2898.

Zhang, L. and V.P. Singh, 2006. “Bivariate Flood Frequency Analysis Using the Copula Method,” *Journal of Hydrologic Engineering*, Vol. 11, No. 2, 150–164, ASCE.

Appendix A. Literature Review

Table of Contents

List of Figures ..... A-i

A.1 Bivariate Distributions.....A-1

A.1.1 Water Resources A-1

A.1.2 Other Fields A-4

A.2 Total Probability Method.....A-5

A.3 Other Joint Probability Methods A-5

A.4 Storm Cell Characterization and the Use of Radar A-6

A.4.1 Regional Variation in Storms A-6

A.4.2 Spatial and Temporal Characterization of Storm Cells A-9

A.4.3 Use of Radar for Storm Cell Characterization.....A-10

A.5 Bibliography A-14

List of Figures

Figure A.1. National WSR-88D Doppler Radar Network in the United States A-11

Figure A.2. Sample Verification for the Minneapolis October 6, 2005 Flood Event A-13

Figure A.3. Atmosphere-Truth Rainfall Field in the Minneapolis Simulations A-13

The research team conducted a comprehensive literature review of coincident flooding and potentially related topics. Literature was reviewed for:

1. Applications of bivariate probability distributions to water resource problems:
   - 1.1 Confluent streams
   - 1.2 Rain on snow events (precipitation and temperature)
   - 1.3 Interior flooding (high tailwater and local rainstorm)
   - 1.4 IFIM (fish habitat) depth and velocity
   - 1.5 Dam safety (flood and earthquake)
   - 1.6 Groundwater contamination (flow and contaminant release)
2. Applications of bivariate probability distributions to other disciplines:
   - 2.1 Traffic engineering
   - 2.2 Seismic events
   - 2.3 Signal processing
   - 2.4 Biometry
   - 2.5 Coastal storms and offshore structural design
3. Other joint probability characterizations.
4. Characterization of low-frequency (0.5 to 0.01 annual exceedance probability) precipitation spatial distributions, including radar-based estimates of precipitation fields.
5. Issues of hydrologic scale (runoff and rainfall).

A summary of the findings is provided in the following sections.

A.1 Bivariate Distributions

Numerous applications of bivariate distributions are found in the literature and are useful where subject variables are correlated. The most common distributions appear to be the Gumbel and normal bivariate distributions, but uses of the generalized extreme value (GEV), exponential, and Weibull distributions are found. Applications in water resources and other fields are discussed.

An alternative method of characterizing multivariate dependence is to use univariate distributions and combine them by use of a copula. Copulas are functions that connect multivariate probability distributions to their univariate marginal probability distributions. The advantage of copulas is that they provide greater flexibility in specifying the marginal distributions compared with true bivariate distributions.

A.1.1 Water Resources

Bogardi, et al. (1975) examined design of flood protection at the confluence of a tributary and a main water body (river, lake, or sea). Using the normal probability distribution, they tested their method on a levee system on a tributary/main river system in central Hungary.

Sackl and Bergmann (1987) described flooding events through application of a bivariate normal distribution to flood peaks and volumes. Flood peak and flood runoff volume were transformed so that the coefficient of skew of the marginal distributions approached zero. With an

unspecified dataset, they illustrated plotting of iso-probability density function lines as well as lines of probability of exceedance for different quadrants.

Shiau (2003) applied the bivariate extreme value distribution with Gumbel marginal distributions to model extreme flood events characterized by flood runoff volume and flood peak. Shiau applied the methodology to the Pachang River in Taiwan where a correlation coefficient of 0.403 between flood volume and peak was estimated. A set of curves of nonexceedance probability for the coincident events was developed.

Morris and Calise (1987) estimated flooding probabilities for a tributary river where the stage-flow relation on the tributary was influenced by backwater on the main river. A bivariate probability density function was used to describe the dependence between the main stream stage and the tributary stream discharge. Despite auto-correlation between daily values, Morris and Calise assumed there were a greater number of independent values in a 20-year daily record than using only annual maximums from 20 years (also assures concurrent events are captured). A three dimensional rating surface was used to summarize tributary flow and main river stage on the x- and y-axes with the resultant tributary stage represented on the z-axis. Multiple two-dimensional splines were used to interpolate between selected stage/flow combinations for which backwater calculations were completed. The total probability theorem was used to determine the probability of any tributary stage height. These daily probabilities were then converted to annual probabilities. The method was applied to the confluence of the Meramec River and the Mississippi River using daily flow and stage from a 56-year period of record. Tributary flows were log-transformed prior to fitting the bivariate normal distribution. HEC-2 was used to compute 50 tributary stage heights. This number was expanded to a 46 by 21 stage matrix grid using splines.

Raynal and Salas (1987) conducted a flood-frequency analysis downstream from the confluence of the Bear River and Dry Creek near Wheatland, California. They used a bivariate generalized extreme value (GEV) distribution to model the flow below the point of confluence. The bivariate cumulative distribution function was evaluated numerically. Raynal and Salas compared several methods for approximating the total flow below the point of confluence. These methods included: 1) adding flows that actually occurred, 2) assuming complete dependence, 3) assuming complete independence, and 4) using the joint distribution with a correlation coefficient of 0.86. Return periods for observed flows were similar to return periods estimated for complete dependence and the joint distribution. Results suggested that the correlation coefficient reasonably quantifies the degree of dependence between events. Raynal and Salas also suggested the bivariate normal distribution may be suitable for coincident events with greater than moderate dependence between events.

To develop boundary conditions for an estuarine numerical hydrodynamic model, Loganathan, et al. (1987) proposed to develop the joint probability distribution for stream flows and tides where stream flows and tides were not statistical independent. A joint distribution of stream flow and tidal height was developed using a Box-Cox transformation of the underlying variables and fitting the result to a bivariate normal distribution. The method was applied to 628 paired tidal stage/stream flow data points where the Rappahannock River flows into the Chesapeake Bay. An empirical joint probability distribution, an empirical probability distribution assuming independence between variables, and a theoretical joint probability distribution were computed. The joint distribution yielded a greater tidal stage level for the same stream flow than values obtained by assuming independence.

Beersma and Buishand (2004) addressed drought in The Netherlands where water can be supplied by direct precipitation or supplemented with flow from the Rhine River. Drought typically occurs when a precipitation deficit is combined low river discharges. Beersma and

Buishand modeled the joint distribution of precipitation deficit and low stream flow using a transformed bivariate normal distribution and a bivariate Gumbel distribution with a logistic dependence model. These models were related to copulas in order to derive the dependency structure. As an alternative method, nearest-neighbor resampling was used to construct a long synthetic record. This record was used to derive an empirical cumulative distribution function. For a 95-year record, discharge deficit on the Rhine and maximum annual precipitation deficit were calculated. Different distributions predicted different degrees of association between the two deficits. The bivariate normal predicted a week relation between large values, underestimating the joint exceedance probability of extreme values. The coincident occurrence of large values was better described by the limiting Gumbel distribution. The work of Beersma and Buishand highlighted the need to ensure the joint probability distribution captures dependence at low-frequency values as well as median values. In addition, a resampling method could possibly provide means to estimate joint probabilities in a framework similar to the use of a historical time series.

Ashkar, et al. (1998) addressed the bivariate analysis of low-flow events focusing on volume and duration. They considered two types of bivariate distributions whose marginals were exponential. They found the exponential marginals to be effective in characterizing the Lepreau River in Canada for use in hydroelectric planning. They further noted the exponential distribution to be practical because only one parameter must be estimated.

Prior-Jones and Beiboer (1990) explored bivariate probability analysis as applied to offshore structures in the North Sea. They examined wave and current probabilities using the bivariate Weibull distribution. Because their objective was an initial exploration of the subject, no definitive conclusions were offered.

Coles and Tawn (1994) investigated the probability of structure failure in response to interaction of multiple dependent variables. They outlined a procedure of: 1) identifying functional relations between design parameters and environmental variables, 2) describing dependency between variables, 3) establishing thresholds over which analysis is valid, and 4) determining the design parameter given the most likely parameter values for a given return period. They primarily focused on coastal flooding in response to storm surge, tides, and waves. The degree of flooding was dependent not only on the static sea height, but also on the quantity of water topping sea walls (surge and waves) during a given unit time.

Svensson and Jones (2004) investigated the possibility of elevated flood risk in response to dependence between sea surge and river flow in estuaries around Britain. Sea surge and river flows are both linked to mid-latitude cyclones. Daily mean flows for 72 rivers were matched with daily maximum storm surge levels as estimated from 19 observation points. A measure intended to assess the dependence between extremes was used to explore the relation between flow and surge at each site as well as spatial relations between each site for each single variable. Coles, et al. (1999) used the same dependence measure. A bootstrapping method was used to estimate confidence intervals for the dependence measure.

Several authors explored ways of linking marginal distributions using copulas. De Michele, et al. (2005) applied a copula from the Gumbel family for peak flow and volumes for dam spillway analysis. Favre, et al. (2004) used a similar method for a run-of-river hydroelectric facility examining peak flow and volume. Salvadori and De Michele (2004) investigated copulas in hydrologic applications.

Recently, Zhang and Singh (2006) used copulas for bivariate distributions of flood peak and volume, as well as flood volume and duration. They also compared the use of copulas with the Gumbel (mixed model) and the bivariate Box-Cox transformed normal distributions. At two

sites, one in Louisiana and one in Quebec, the copula-based distributions were found to be in better agreement with plotting position based frequency estimates.

Adamson, et al. (1999) applied a Monte Carlo method to bypass theoretical difficulties of bivariate distributions. The Gibb's sampler method was presented as a means to construct a bivariate distribution from only the conditional distributions. In essence, bivariate pairs were determined by generating an alternating sequence of each bivariate variable (for example, X_0, Y_0, X_1, Y_1…) where each sequential value depends on the former. The method was applied to the Mekong River in Laos where seasonal flow is dominated by a single annual flood hydrograph corresponding to the monsoon. Annual maximum peak discharge, flow volume, and flow duration were determined from a 79-year record. Adamson, et al. demonstrated the application of flood damage estimation as a function of primary damage attributed to large discharges and secondary damage attributed to duration.

A.1.2 Other Fields

Akkaya and Yucemen (2002) developed a risk model for earthquake damage. The model incorporates spatial and temporal correlations between strain energy in the earth's crust. An earthquake of a given magnitude is assumed to occur when the total strain energy over the potential rupture plane exceeds a given threshold. The model stochastically simulates strain energies along fault lines according to normal and lognormal models. The model was compared to data from the North Anatolian Fault Zone.

Coles, et al. (1999) investigates methods for quantifying dependencies in maximum temperatures arising from different processes driven by the same phenomenon or by the same process observed at different spatial locations. A method for measuring dependence of extremes is to calculate the conditional probability of observing an extreme value in one process given an extreme value of a concurrent process. This dependence function was related by Coles, et al. to standard models for bivariate extremes. While three different datasets were used to illustrate concepts, a time series of annual maxima temperatures at two sites in England was the only dataset that includes spatial dependencies. Application of the dependency function indicated a high degree of spatial dependency, encouraging use of a bivariate extreme value model.

Gazis and Edie (1968) provided a summary of models applied to transportation networks. The focus at that time was on simple analytic models because much of the theoretical work in operational research had not yet been applied to transportation engineering.

McNeil (1970) reviewed models used by traffic analysts and assumptions underlying those models, then developed appropriate distributions for modeling traffic queuing processes.

Cowan (1979) developed the distributional properties of merging traffic by studying the arrival processes of automobiles. He characterized the problem by considering the arrival of random-sized "bunches" of cars, separated by random time gaps.

Gordon (1995) used a Pareto process to represent the arrival rates of network traffic packets. Stochastic models were generated to represent network traffic and evaluate the dependability of particular pieces of networking equipment. Fractional Brownian motion, fractional ARIMA, and chaotic maps were studied.

Bunker and Troutbeck (2003) developed statistics using the models of Cowan (1975) to represent the merging characteristics of incoming freeway traffic.

Jagerman et al. (1997) provided a summary of stochastic and analytic modeling practice in the telecommunications field.

A.2 Total Probability Method

The total probability method is a well-established tool used in water resources analyses. It is described in various guidance documents for the U.S. Army Corps of Engineers (USACE 1987; USACE 1993).

Dyhouse (1985) conducted a joint probability analysis on the Illinois River to determine the feasibility of enhancing the levee system. Backwater from the Mississippi River influences Illinois River levels, complicating stage-frequency calculations. The total probability theorem was used to determine the probability of a certain tributary river stage elevation. Several measures of dependency were applied to coincident events (Mississippi River stage and Illinois River flow) that indicated the events were virtually independent. The method was applied to a 40-year data record and HEC-2 was used to develop water-surface profiles on the Illinois River for varying combinations of Mississippi River stage and Illinois River flow. To avoid calculating profiles for every possible combination of events, a set of curves was constructed from which values could be interpolated. To obtain consistent results, Dyhouse noted that one must select the "dominant" variable to establish the duration curve. Depending on the distance from the Mississippi River on the tributary, the Mississippi River stage level is not always the dominant variable on which the tributary stage was conditioned.

Pingel and Ford (2004) determined the expected value of annual damage for a location with interior flooding using the total probability theorem. Unlike a standard flooding analysis, the degree of interior flooding is dependent on runoff generated in the catchment as well as externally variable control on the catchment discharge rate, in this case, the stage of a river receiving the runoff. Pingel and Ford used a joint probability framework in which all possible degrees of exterior control were considered instead of only considering a worst case. Specifically, for a case in which interior runoff and the exterior control were independent, they calculate a discrete probability distribution for interior stage height using the total probability theorem. The method was applied to a 40 km^2 watershed in Sacramento, California in which a levee blocks the natural drainage into the Lower American River. Pingel and Ford illustrated multiple steps in calculation of the interior stage probability function using actual observations and information calculated from standard hydrology and hydraulics models. The method was directly relevant to cases in which multiple, but statistically independent, conditions dictated a process outcome.

A.3 Other Joint Probability Methods

Tabular relations for addressing joint probability were found. The most common relation is exemplified in AASHTO (2000). Many other state and local agency guidance documents display the same tabular relation thought to be developed for a site-specific condition by the Corps of Engineers.

Wilson (1980) studied the probability of concurrent flooding at the confluence of the Mississippi and Meramec Rivers near St. Louis, Missouri. He used an exhaustive search technique to map all combinations of observed annual maximum stage on both rivers. The Mississippi and Meramec River stages were simulated by operating HEC-2 using every combination of discharge from each river. These results were used to develop a refined method by which the critical combination of discharges could be identified.

Morris and Wilson (1987) estimated flooding probabilities for a tributary river where the stage-flow relation on the tributary is influenced by backwater. A "critical combination method" is used to determine the highest stage on the tributary in each year of record. To avoid computing computationally intensive backwater calculations for every daily combination of main river stage and tributary flooding, critical combinations were selected. That is, for each year in the record,

the peak tributary flow and corresponding main stream stage were identified. These combinations were used to develop a three-dimensional "rating" surface relating main stage and tributary flow on the x- and y-axes to tributary stage height on the z-axis. Daily stage and flow data were substituted into the rating surface to identify the peak tributary stage each year. The method was applied to the Meramec River where it enters the Mississippi River. A two-tailed Student's *t* test indicated dependence between Meramec River flow and Mississippi River stage. Once annual peak stages were identified, a log Pearson Type III distribution was fitted to the series to determine exceedance probabilities for tributary stage.

Koltun and Sherwood (1998) performed an analysis of Ohio streams to generate joint probability relations. This study has limited direct application because they assumed data were independent if the drainage area ratio was greater than 10:1 and they paired streams whether or not they were adjacent to one another.

Singh (2000) provided a description of entropy theory and how it can be applied in water resources. As described by Singh, entropy theory is a versatile, robust, and efficient technique to determine the least biased probability distribution of a random variable.

Fernandez and Salas (1999a and 1999b) presented a method for estimating the return period of hydrologic events and the associated risk of failure when the underlying hydrologic series are auto-correlated. When annual peak flows were analyzed, they were not considered to be auto-correlated. However, if partial duration series or other types of "events" were considered, the potential for auto-correlation should be considered.

Alila and Mtiraoui (2002) considered heterogeneous flood frequency distributions composed of a mixture of two or more subpopulations in which floods are generated by different processes. All flood event magnitudes were considered independent, but not necessarily mutually exclusive, on an annual basis. They examined three standard means of dealing with mixtures. Inherently, a mixture assumes independence between populations. As implemented, their method only involved peak events when no additive effects must be considered. For the Gila River Basin, Arizona, Alila and Mtiraoui fit observed annual floods with a heterogeneous distribution composed of two lognormal distributions. The method is only applicable to cases of independent events, but is useful in cases where flooding could be caused by more than one process in a given year. A common example would be snowmelt in spring and thunderstorms in summer.

The particle-tracking runoff model of Cleveland, et al. (2005) used simple hydraulics to model the movement of runoff "particles" across a watershed represented by a digital elevation model. Rainfall-runoff dynamics were simple. Adjacent watersheds could be modeled using this method to assist understanding of the correlation, or conditional probability structure of runoff between confluent watersheds.

A.4 Storm Cell Characterization and the Use of Radar

The literature search included investigation of the characterization of storm cells and how an understanding of storms may inform the issue of flooding at confluent watersheds. Three major topic areas were included: 1) regional variation in storm characteristics across the United States, 2) the spatial and temporal characterization of storm cells, and 3) the use of radar for describing storm cell characteristics and their regional variation.

A.4.1 Regional Variation in Storms

There are at least two precipitation patterns that produce exceptional precipitation events that lead to extreme flows in the United States. The first pattern is driven by moderate intensity rainfalls over extremely long periods (days), a generally stratiform precipitation pattern, which

can be observed in the western United States, for example, California, Oregon, and Washington. These events can easily encompass an area greater than 10,000 km^2 and cause coincident flood flows on downstream confluences from river basins being hundreds of kilometers from each other. This specifically describes the California flood events of 1995 and 1997.

The other precipitation pattern, which contributes flood flows over most of the United States, is the convective complex (mesoscale), which generally leads to very high precipitation over a shorter period (hours), and usually over a relatively small area. When these events occur near mountain ranges, the process of accelerated runoff collection because of the steep slope of the terrain can produce large flows in a short time frame, leading to the term “flash flood.”

Flood events occurring on the west coast of the United States are generally comprised of area wide precipitation events that are stationary. The exception to this is the southern west coast where flood events are more dominated by highly convective situations where surface heating and warmer saturated air becomes more of a factor. This holds true for the southern tier states that are generally removed from the polar jet stream. Precipitation extremes for these regions are more of a summer event that is driven by heat and convection created from the diurnal heating process.

The inner mountain areas of the west, including the Rocky Mountains, could be described as a mixture of both orographic and convective events. The floods of record, such as the Big Thompson Flood, are often caused by a summertime convective system characterized by short duration and extreme precipitation rates. As one moves east away from the influence of the Pacific Ocean, the main mechanism for flood producing events is largely convective. However, there are notable exceptions such as the Great Flood of 1993 in the central part of the nation that consisted of convective elements produced by the meeting of cooler air carried by a consistent jet stream with the warmer moist air being pulled up from the Gulf of Mexico.

Numerous publications describe unique characteristics of specific storms. A sampling of representative publications of this type illustrates variations in storms across the U.S., as well as the similarities within certain regions. Other publications highlight a national comparison of storm and flood characteristics.

O’Connor and Costa (2004) identified topographic and climate features that result in high unit flows. They identified the top ten percent of peak annual discharges uninfluenced by anthropogenic modifications from 18,735 USGS gaging stations with at least five years of record. Flows were converted to flow per unit area. Maps were generated delineating basins with one or more 90th percentile unit discharges. Areas of high topographic relief combined with areas subject to high rainfall produced the highest unit discharges. Their work provided an interesting overview of the spatial pattern of flooding, but was limited in value to the current research since it does not indicate the cross-correlation in flow among gages at the same time.

Caracena, et al. (1979) provided a meteorological investigation and description of the synoptic and meso-synoptic characteristics that contributed to the Big Thompson storm on 31 July 1976. The storm complex that contributed to the Big Thompson flood was convective and enhanced by orographic uplift as the moist air was forced up the eastern slope of the Front Range. The orographic uplift created destabilization. A continuous flow of moist air continued the process for an extended period. Microphysical processes such as latent heat exchange caused by the condensation of cloud water acted as an engine to fuel even greater development of the convective complex. Radar underestimated rainfall estimates. The radar beam was overshooting the maximum reflectivity region, which yielded the lower reflectivity.

Maddox et al. (1978) compared two noteworthy flash flood events and analyzed the similarities. The two events were the Big Thompson Canyon flood of 31 July 1976 and the Rapid City floods of 9 June 1972. Both events produced 10 to 15 inches of precipitation during the evening hours. Contributing mechanisms were associated with a somewhat stationary negatively tilted upper-level ridge. Weak southeasterly steering winds allowed the storms to remain over the same area for two to four hours. Orographic lift contributed, as moisture was released due to the cloud condensation process.

Elsner, et al. (1989) documented the weather patterns and characteristics of a convective storm in Milwaukee, Wisconsin on 6 August 1986. The study was designed to determine the meteorological precursors to this heavy rain-producing event. The event produced 6 inches of rain over a 2-hour period. The importance of this work was the furthering of evidence that heavy precipitation rates can occur within convective complexes that have relatively warm cloud tops. This is different from what is found in meso-convective complexes (MCCs) that are often thought to account for extreme rain-producing events. In MCCs, the cloud is usually very deep and the cloud structure exhibits very cold cloud tops. The mechanism for this event was not orographically induced, but rather stimulated by sheer atmospheric dynamics. An interesting side note was that radar observations underestimated the rate of precipitation.

Randerson (1976) described the meteorological characteristics of the Las Vegas flood of July 1975. The areal extent of this event was 550 km^2 and it comprised two heavy rain-producing systems within the geographical area. The area included several washes and creeks that combined to provide a confluent flow to flood Las Vegas. Three inches of rainfall were estimated over a two to four hour period. The greatest precipitation amount captured by a rain gage was one inch. Radar data were used to estimate the three-inch maximum (in two different areas). Exact reasons why this event occurred were not offered, though it was thought that a strong moisture gradient was making its way into the area. Surface heating was probably sufficient to initiate the instability and produce vertical velocities necessary to convert moisture vapor to condensate. This rainfall amounts were produced without any assistance from orography or mechanical lifting. Vertical motion was initiated from surface heating, which then created further instability attributed to the latent heat release as the available moisture was being processed into condensate.

Masutani and Leetmaa (1999) provided a meteorological explanation of the floods of 1995 in California, and further developed the relation between El Nino and California wintertime rainfall. The flood events of 1995 differed from the common conception on the linkage between El Nino and floods in California. Sea surface temperature anomalies (SSTAs) were observed in the central Pacific during the winter of 1994/1995. During the warmest SSTAs (two standard deviations above normal) in November and December, no flooding or significant weather events occurred. The SSTAs weakened considerably by the March 1995 flood event. A link is explained between the tropical inter-seasonal oscillation (TIO) and March flooding in California. The TIO indicated that the region of heating moved from the middle of the Indian Ocean to the central Pacific Ocean. General Circulation Model (GCM) experiments showed that heating over both the Indian and central Pacific Oceans affects rainfall over California, but in opposite ways. The event of 1995 caused widespread precipitation with many instances of floods on rivers over a large region (>50 miles). Though many of the rivers are controlled by reservoir operations, the general event produced many examples of flooding on confluent streams.

Maddox, et al. (1980) determined meteorological characteristics from 61 flash-flood events in the western United States. The study was intended to assist forecasters in identifying these events through storm typing and general pattern recognition. Maddox, et al. explained that the meteorological characteristics of flash floods and heavy precipitation events that occurred over the western United States were often considerably different than those associated with flash

floods in the eastern United States. Western events had substantially less rainfall amounts than the eastern events, and the western events tended to occur during the afternoon and early evening, rather than after dark as in the east.

A.4.2 Spatial and Temporal Characterization of Storm Cells

Storms are created by a variety of influences. An understanding of the contribution of mesoscale (50 kilometers) and synoptic (100's of kilometers) influences on precipitation characteristics such as the duration, spatial distribution, and intensity of precipitation is necessary to characterize storm cells. Several references have been made in the literature to storm cell characterization.

Casson and Coles (2000) assessed joint distribution of wind speeds at pairs of sites, thus indicating the degree of spatial cross-correlation. Because of the severe damage caused by hurricane winds, accurate estimates of hurricane wind severity are essential. Although such extreme winds are difficult to measure, atmospheric pressure records are more complete and provide a basis for modeling the spatial-temporal patterns of wind. To examine spatial dependencies, they calculated a spatial dependency measure (following Coles, et al., 1999) between sites 50, 250, and 1000 nautical miles apart assuming the sites can be related by a joint distribution. They assumed asymptotic dependence; the likelihood of coincident extreme events increased as the event magnitude became larger. They examined 534 hurricane events occurring along the Gulf coast and eastern seaboard of the U. S. The assessment of spatial dependence revealed strong dependence at 50 miles for all regions and strong dependence at 250 miles in the northeast. The research provided a possible method to relate point measurements of rainfall to assess degree of spatial cross correlation between sites during a storm event. In addition, the paper provided insight into behavior (extent, typical trajectories, etc.) of large storms in the eastern United States.

De Lannoy, et al. (2005) attempted to parameterize a stochastic rainfall simulator at the mesoscale based on an analysis of the direction, velocity, and dimensions of frontal storm systems in the Netherlands. The authors obtained radar images with a 2.4 km spatial resolution spanning a 480 km square over the Netherlands at 15-minute intervals for 20 precipitation events. To track movement, a method was used that identified the optimal correlation between time lagged images. Dimensions of the irregularly shaped storms were quantified most successfully using a geometric method that established a coordinate system parallel and perpendicular to the storm and then found the point of intersection with the major axis of the coordinate system. The approach illustrated a possible method to describe the cross-correlation of rainfall inputs to adjacent watersheds. The main requirement would be to have a suitable record of storms producing significant rainfall.

Because of a lack of reliable discharge records, design floods in the Huai River basin were approximated from precipitation data. Svensson (1999) conducted statistical and correlation analyses to identify extreme rainfall characteristics in terms of intensity, duration, and range. For a 79,000 km^2 area in the Huai River Basin, daily data were obtained for 78 rain gages for 1957 through 1986. Auto-correlation was calculated for 14 sites for lags from 1 to 20 days. Cross-correlations were calculated between 78 sites for lag zero (to assess shape and orientation) and lag one (to asses movement). From the auto-correlation analysis, rainfall was seasonal with most rain arising from cold fronts between June and September. The largest storm events were a result of typhoons. For lag zero cross-correlation between stations, maps were developed to indicate the degree of correlation at a given distance in multiple directions. There was a 0.6 correlation over a 50 km range with stronger correlation on a northeast axis. The work provided an example of cross-correlation analysis between rainfall amounts. Given that most rainfall is a result of frontal systems and typhoons, there is a relatively strong spatial

correlation on an event basis. A similar analysis could be carried out in a region of interest to assess the likelihood of flooding in adjacent watersheds.

Fernandez, et al. (1999) investigated the temporal and spatial variability of precipitation in a 625 km^2 area in central west Argentina using a dense hydrometeorological network of 24 stations. Precipitation zones were defined and procedures for the regional analysis of convective storms and development of a standard project storm were presented.

Matsubayashi and Takagi (1987) attempted to establish the relation between the probabilistic characteristics of point and areal rainfalls. They demonstrated that the shift of the probability density function of point rainfall towards lower rainfall intensity might give a good approximation to that of area rainfall.

Wilson (1986) developed a spatial-temporal stochastic precipitation model for a small watershed in central Missouri. The size of the simulated area was approximately 40 square miles. This work required defining the cross-correlations between adjacent meteorologic stations and developing the joint probability distributions for sampling and generating precipitation sequences.

Hershfield (1973) presented the substance and limitations of methods used to estimate extreme rainfall-event probabilities. He also assembled a number of references on the subject of "extreme-value statistics." In the work, Hershfield presented the term "return period" and the relation between risk of exceedance and return period. The use of such terms as "once-in-a-century" has a rational basis if it is based on a long series of observations. However, the use of the term 100-yr storm is only justified if it refers to a probability estimate based on a series of data from a rain gage and represents only a relatively small area nearby the rain gage. The 100-yr term is not applicable for describing the magnitude of a storm over a substantial-size area because of a general lack of procedure for determining the exceedance probability of a major rainstorm over such an area. This observation is most important when studying the link between rainfall and runoff production and the frequency of areal rainfall patterns, which is of vital concern in flood events.

A.4.3 Use of Radar for Storm Cell Characterization

For determining joint probabilities for design coincident flows at confluences, a method of extracting precipitation data from radar might be useful. In the past, radar based precipitation estimation has led to an underestimation of precipitation. However, in many cases it has provided a suitable surrogate in the absence of a rain gage network in a particular area. Radar has also provided significant information on storm characteristics, particularly, the area engaged and the hot spots within the storm.

The use of National Weather Service (NWS) WSR-88D Doppler-derived rainfall to address the need for spatial, temporal and quantitative rainfall estimation in hydrologic modeling is growing. Brandes, et al. (1991) identified the basic opportunities to use the WSR-88D to a variety of applied research and operational needs. The national network of WSR-88D radar is shown in Figure A.1 and demonstrates that most of the United States is covered by the radar network. Many portions of the country are covered by overlapping radar that insures the observation of most significant extreme precipitation events.

Hunter (1996) described the capabilities, limitation and opportunities for improvement of WSR-88D rainfall estimations for hydrological and operational applications. Seo (1998) elaborated on these opportunities and generally supported the accuracy of the radar-derived temporal and spatial rainfall fields when used in combination with surface rain gage network data. Their work supports the use of the spatial and temporal derived rainfall fields, but identified that the early radar-rainfall derived fields lacked quantitative accuracy.

Vieux and Bendient (1998) studied the use of radar, specifically the WSR-88D radar for estimating rainfall. Their objective was to explain how radar observations could provide misleading precipitation amounts, especially underestimation. The authors investigated why the WSR-88D provided low estimates of precipitation. For example, in a case in Texas, the radar was only 23 km from the area in question and data could be used at the lowest levels. This would be a “best-case” scenario for radar estimation of rainfall. Yet, the radar still underestimated rainfall. The authors tried various Z-R relationships, and found a different Z-R relationship worked better than that prescribed at the time by the National Weather Service. This highlighted the uncertainty of using radar for precipitation estimation, especially in extreme events. The differences in Z-R were attributed to the higher rainfall rates. The paper concluded that rain gages should be used to calibrate the radar-rainfall estimates.

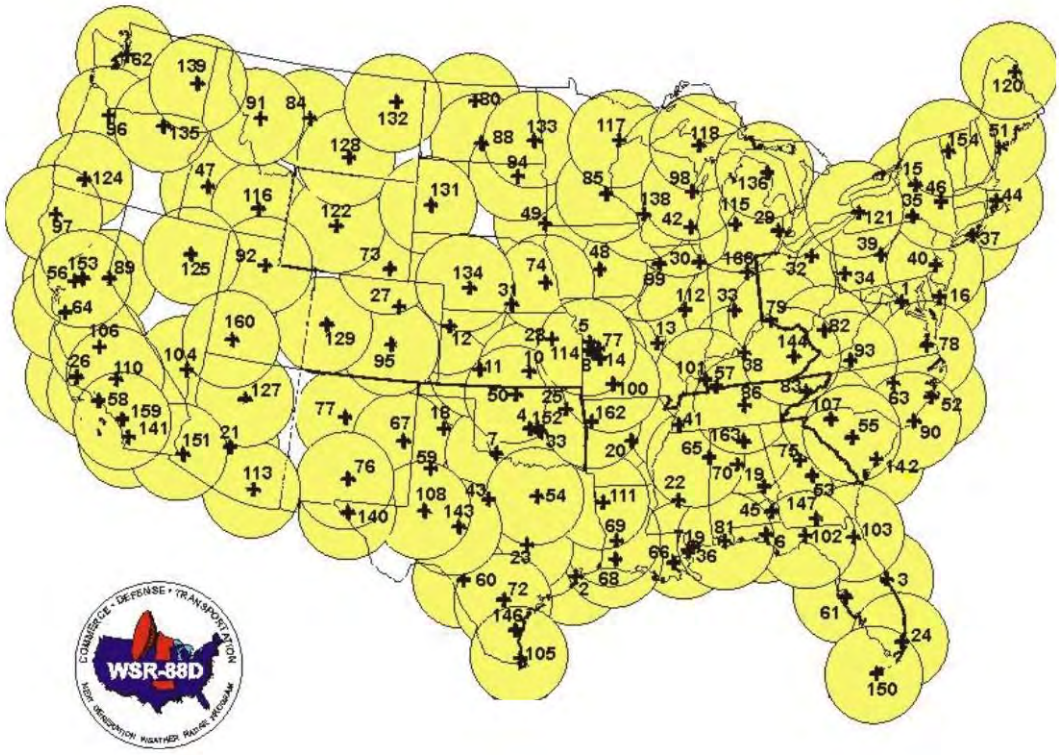

Figure A.1. National WSR-88D Doppler Radar Network in the United States

Henz (1996a, 1996b, 1996c) investigated the use of an atmosphere-truthed radar-rainfall estimation method to improve the quantitative aspects of the radar-rainfall estimation process. His method is based on the relationship of the peak observed rainfall rate of a thunderstorm to the portions of the storm that exhibit radar reflectivity of 50 dBZ or greater. A logarithmic down-stepped technique is used to assign rain rates to radar reflectivity of 30 dBZ to 50dBZ. The technique was used to provide estimates of hail size and damage paths, as well as quantitative rainfall estimation in operational flash flood prediction programs in Denver, Colorado and Phoenix, Arizona. The technique was also used in the reconstruction of historical flooding events. The atmosphere-truthed method provided excellent correlation to gage-observed

rainfall fields in dense ALERT networks maintained by flood control districts in Denver and Phoenix.

Seo, et al. (2000) updated his earlier findings and offered real-time adjustments of range-dependant bias in the WSR-88D rainfall fields that significantly reduced the error introduced by the distance of the storm from the radar. In a similar study, Wang, et al. (2000) provided a detailed comparison of the mean areal precipitation estimates for watersheds obtained from operational and historic rain gage networks and radar-derived rainfall fields. They investigated the impacts of each dataset on the simulated runoff modeled in each of eight watersheds of varying sizes. Their conclusions provided convincing evidence that radar-derived and gage-adjusted rainfall provided the best runoff simulations and was superior to either radar or rain gage datasets without the support of the other. This conclusion was independent of the size of the basins.

Hardegree, et al. (2003) produced a multi-watershed evaluation of WSR-88D radar-precipitation products at agricultural research station watershed research locations in Idaho, Arizona, Oklahoma, Georgia, and Mississippi. They evaluated the utility of these data for hydrologic and natural resources modeling applications. The WSR-88D precipitation estimates under-estimated gage readings in all locations except Arizona. However, they found the utility of the data for modeling extreme precipitation events, erosion, and runoff was significantly enhanced when gage-adjusted radar derived rainfall fields were used for individual storm events and specific watersheds.

Jorgerson and Julian (2005) tested the operational value of using WSR-88D radar rainfall estimates with gage-adjustment for the peak flow forecasting of two 1995 flood events on the Hassayampa River in Arizona. They used the rainfall fields as input to the CASC2D distributed watershed model to predict the magnitude and timing of the peak flow. The model successfully predicted the peak flow and added four to five hours notice to the timing of the peak flow event.

Christopherson and Henz (2006) performed a detailed meteorological and hydrologic evaluation of an October 6, 2005 extreme precipitation event in Minneapolis Minnesota. Henz used an atmosphere-truthed radar-rainfall relationship to perform a GIS-based radar estimation of the rainfall fields at a grid resolution of 1 km by 1 km per 5 to 6 minute time step. The rainfall field developed was compared to four independent rainfall observation sites from nearby first order National Weather Service observation gages. An example of the verification is shown in Figure A.2 for 15-minute time steps. The verification noted a 5 percent over-estimate of the observed rainfall by the radar-derived rainfall.

The gridded and summed rainfall field is shown in Figure A.3 for the basins and sub-basins used in the XP SWMM model. Alternative rainfall fields were entered into the XP SWMM model in 15-minute time steps: rain gage only, radar/gage-adjusted, and radar/atmosphere-truthed. Model simulations indicated that the atmosphere-truthed rainfall fields provided the best runoff simulation.

The consensus of these papers was that radar-estimated rainfall provides reliable spatial and temporal definition of rainfall events and gage-adjusted radar rainfall provides reliable quantitative rainfall estimates that improve model simulation of runoff, especially for extreme precipitation events.

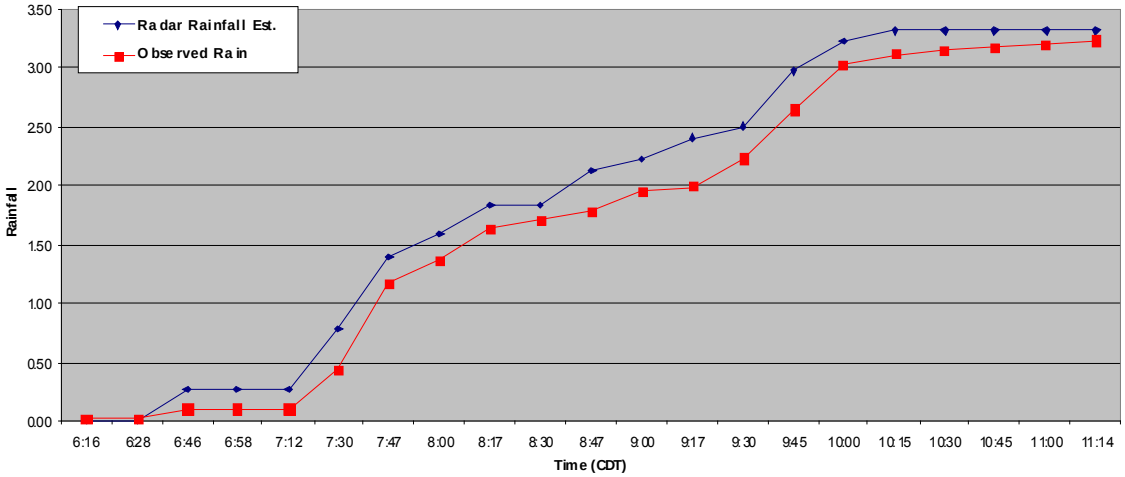

Figure A.2. Sample Verification for the Minneapolis October 6, 2005 Flood Event

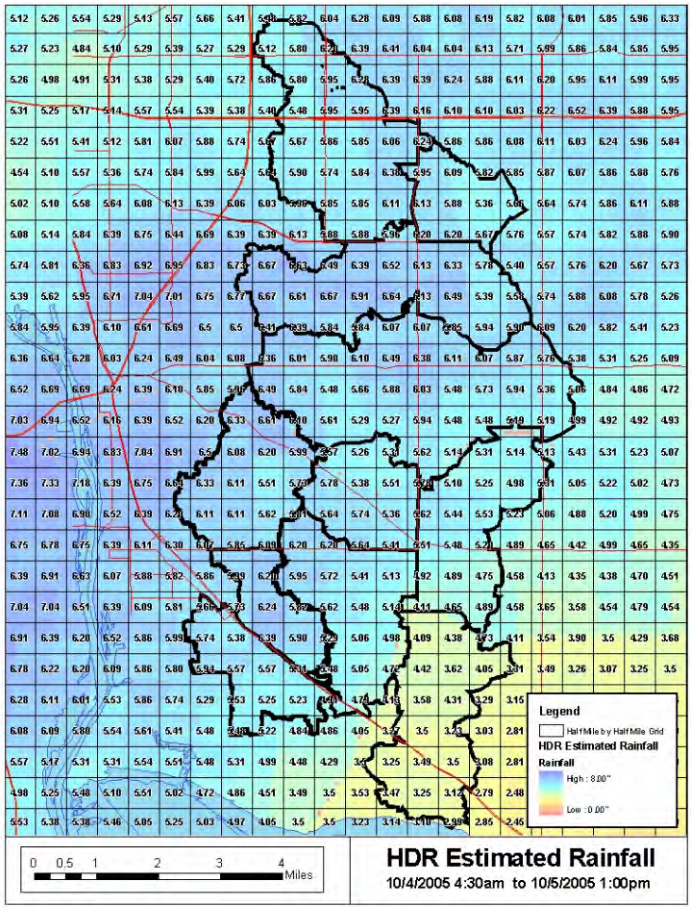

Figure A.3. Atmosphere-truth Rainfall Field in the Minneapolis Simulations

A.5 Bibliography

AASHTO, 2000. *Model Drainage Manual*, Chapter 13.

Adamson, P. T., A. V. Metcalfe, and B. Parmentier, 1999. "Bivariate extreme value distributions: An application of the Gibb's sampler to the analysis of floods," *Water Resources Research*, Vol. 35, No. 9.

Akkaya, A. D., and M. S. Yucemen, 2002. "Stochastic Modeling of Earthquake Occurrences and Estimation of Seismic Hazard: a Random Field Approach." *Probabilistic Engineering Mechanics*, Vol. 17, pp. 1–13.

Alila, Y., and A. Mtiraoui, 2002. "Implications of Heterogeneous Flood-frequency Distributions on Traditional Stream-discharge Prediction Techniques." *Hydrological Processes*, Vol. 16, No. 5, pp. 1065–1084.

Andrieu, H., J. D. Creutin, G. Delrieu, and D. Faure, 1997. "Use of a Weather Radar for the Hydrology of a Mountainous Area: 1. Radar Measurement Interpretation," *Journal of Hydrology*, 193, pp. 1–25.

Ang, A. H. S. and W. H. Tang, 1975. *Probability Concepts in Engineering Planning and Design, Volume I: Basic Principles*, John Wiley & Sons, New York.

Ashkar, F., N. El Jabi, and M. Issa, 1998. "A Bivariate Analysis of the Volume and Duration of Low-flow Events," *Stochastic Hydrology and Hydraulics*, Vol. 12, No. 2, pp. 97-116, July.

Balocki, J. B. and S. J. Burges, 1991. "Preliminary Examination of Relationships between Catchment Characteristics and Volumes of Infrequent Large Floods," SER/TR-130, Washington University, Seattle, WA, U.S.G.S. (Seattle), August.

Beersma, J. J. and T. A. Buishand, 2004. "Joint Probability of Precipitation and Discharge Deficits in the Netherlands," *Water Resources Research*, Vol. 40.

Bogardi, I., L. Duckstein, and F. Szidarovszky, 1975. "Hydrological System Reliability at the Confluence of Rivers," Application of Mathematical Models in Hydrology and Water Resources Systems, *Proceedings of the Bratislava Symposium*, Czechoslovakia, IASH-AISH Publication No. 115, International Association of Hydrological Sciences, pp. 36–44, September.

Brandes, E. A., D. S. Zrnic, G. E. Klazura, C. F. Suprenant and D. Sirmans, 1991. "The Next Generation Weather Radar (WSR-88D) as an Applied Research Tool," *Proceedings of the 25th International Conference on Radar Meteorology*, American Meteorological Society, Boston, pp. 47–50.

Bunker, J. and R. Troutbeck, 2003. "Prediction of Minor Stream Delays at a Limited Priority Freeway Merge," *Transportation Research Part B*, Vol. 37, pp. 719–735.

Casson, E, and S. Coles, 2000. "Simulation and Extremal Analysis of Hurricane Events," *Applied Statistics*, Vol. 48, No. 2, 227–245.

Caracena, F., R. A. Maddox, L. R. Hoxit and C. F. Chappell, 1979. "Mesoanalysis of the Big Thompson Storm," *Monthly Weather Review*, Vol. 107, No. 1, pp. 1–17.

Christopherson, J. and J. F. Henz, 2006. "Integrating Radar Rainfall Data, GIS, and XPSWMM to Model Storm Events," Minnesota Water 2006 and Annual Water Resources Joint Conference, October 24–25, 2006, Brooklyn Center, MN. Water Resources Center.

Cleveland, T. G., X. Fang, and D. B. Thompson, 2005. “Timing Parameter Estimation by a Particle Tracking Method. Technical Report TechMRT Report 0–4696–3, Texas Tech University, Lubbock, TX.

Coles, S., G., J. Heffernan, and J. Tawn, 1999. “Dependence Measures for Extreme Value Analyses,” *Extremes*, Vol. 2, No. 4, pp. 339–365.

Coles, S. G., and J. Tawn, 1994. “Statistical Methods for Multivariate Extremes: An Application to Structural Design,” *Applied Statistics*, Vol. 43, No. 1, pp. 1–48.

Cowan, R., 1979. “The Uncontrolled Traffic Merge,” *Journal of Applied Probability*, Vol. 16, No. 2, pp. 384–392.

Cowan, R. J., 1975. “Useful Headway Models,” *Transportation Research*, Vol. 9, No. 6, 371–375.

De Lannoy, G. J. M., N. E. C. Verhoest, and F. P. de Troch, 2005. “Characteristics of Rainstorms over a Temperate Region Derived from Multiple Time Series of Weather Radar Images,” *Journal of Hydrology*, Vol. 307, pp. 126–144.

De Michele, C., G. Salvadori, M. Canossi, A. Petaccia, and R. Rosso, 2005. “Bivariate Statistical Approach to Check Adequacy of Dam Spillway,” *Journal of Hydrologic Engineering,* Vol. 10, No. 1, pp. 50–57, ASCE.

Dolph, J. and D. Marks, 1992. “Characterizing the Distribution of Observed Precipitation and Runoff over the Continental United States,” EPA/600/J-92/458.

Dyhouse, G. R., 1985. “Stage Frequency Analysis at a Major River Junction,” *Journal of Hydraulic Engineering*, Vol. 111, No. 4, ASCE, April.

Elsner, J. B., W. H. Drag and J. K. Last, 1989. “Synoptic Weather Patterns Associated with the Milwaukee, Wisconsin Flash Flood of 6 August 1986**,”** *Weather and Forecasting*, Vol. 4, No. 4, pp. 537–554.

Farajalla, N. S. and B. E. Vieux. “Temporal and Spatial Aggregation of NEXRAD Rainfall Estimates on Distributed Storm Runoff Simulation,” http://www.ncgia.ucsb.edu.

Favre, A. C., S. El Adlouni, L. Perrault, N. Thiemonge, and B. Bobee, 2004. “Multivariate Hydrologic Frequency Analysis using Copulas,” *Water Resources Research*, Vol. 40.

Feld, G. and J. Wolfram, 1995. “Using Wind, Wave, and Current Time Histories to Estimate Realistic Extreme Joint Loading Conditions,” International Conference on Seakeeping and Weather, London, February 28–March 1.

Fernandez, B. and J. D. Salas, 1999a. “Return Period and Risk of Hydrologic Events. I: Mathematical Formulation,” *Journal of Hydrologic Engineering*, October, ASCE.

Fernandez, B. and J. D. Salas, 1999b. “Return Period and Risk of Hydrologic Events. II: Applications,” *Journal of Hydrologic Engineering*, October, ASCE.

Fernandez, P. C., S. Fattorelli, S. Rodriguez, and L. Fornero, 1999. “Regional Analysis of Convective Storms,” *Journal of Hydrologic Engineering*, October, ASCE.

Fricke, T. J., M. R. Kennedy, and N. B. Wellington, 1983. “The Use of Rainfall Correlation in Determining Design Storms of Waterways on a Long Railway Line,” Proceedings of Hydrology and Water Resources Symposium, Hobart, Australia, November 8–10.

Gazis, D. C. and L. C. Edie, 1968. “Traffic Flow Theory,” *Proceedings of the IEEE*, Vol. 56, No. 4, pp. 458–471.

Gordon, J., 1995. “Pareto Process as a Model of Self-similar Packet Traffic,” in Global Telecommunications Conference, pp. 2232–2236, IEEE.

Gupta, S. and C. S. Manohar, 2005. “Multivariate Extreme Value Distributions for Random Vibration Applications,” *Journal of Engineering Mechanics*, Vol. 131, No. 7, July, ASCE.

Haan, C. T., and B. N. Wilson, 1987. “Another Look at the Joint Probability of Rainfall and Runoff,” in Hydrologic Frequency Modeling: Proceedings of the International Symposium on Flood Frequency and Risk Analysis, May 14–17, 1986, Baton Rouge, LA, ed. Singh V. P., Reidel, Dordrecht, The Netherlands.

Hardegree, S., S. Van Vactor, K. Healy, C. Alonso, J. Bonta, D. Bosch, D. Fischer, D. Goodrich, D. Harmel, J. Steiner and M. Van Liew, 2003. “Multi-Watershed Evaluation of WSR-88D (NEXRAD) Radar-Precipitation Products, pp. 486–491.

Harold, T. I., A. Sharma, and S. Sheather, 2001. “Selection of a Kernel Bandwidth for Measuring Dependence in Hydrologic Time Series Using the Mutual Information Criterion,” *Stochastic Environmental Research and Risk Assessment*, Vol. 15, No. 4, pp. 310–324, August.

Hawkes, P. J., 2005. “Use of Joint Probability Methods in Flood Management: Guide to Best Practice,” R & D Technical Report FD2308/TR2, Defra/Environment Agency, Flood and Coastal Defence R & D Programme.

He, R. R. and B. Ran, 2000. “Calibration and Validation of a Dynamic Traffic Assignment Model,” *Transportation Research Record: Journal of the Transportation Research Board 1733*, TRB, National Research Council, Washington, D.C. pp. 56–62

Henz, J. F., 1996a. “A New Approach to Predicting the Size of Damaging Hail,” 18th Conference on Severe Local Storms, February 19–23, 1996, San Francisco, California. *American Meteorological Society*, pp. 86–87.

Henz, J. F., 1996b. “Innovative Re-Constitution of a Western Flash Flood Using Cloud-to-Ground Lightning, Radar Reflectivity and Surface Rainfall Relationships,” 18th Conference on Severe Local Storms, February 19–23, 1996, San Francisco, CA. American Meteorological Society, pp. 474–479.

Henz, J. F., 1996c. “Quantitative Convective Precipitation Prediction Applications to Recent 1994 and 1995 Flash Flooding Events,” 18th Conference on Severe Local Storms, February 19–23, 1996, San Francisco, CA. American Meteorological Society, pp. 667–672.

Hershfield, D. M., 1973. “On the Probability of Extreme Rainfall Events,” *Bulletin of the American Meteorological Society*, Vol. 54, No. 10, pp. 1013–1018.

Huff, F. A., 1968. “Spatial Distribution of Heavy Storm Rainfalls in Illinois,” *Water Resources Research*, February, pp. 47–54.

Hunter, S. M., 1996. “WSR-88D Radar Rainfall Estimation: Capabilities, Limitations, and Potential Improvements,” *National Weather Digest*, Vol. 20, pp. 26–38.

Hogg, R. V. and A. T. Craig, 1978. *Introduction to Mathematical Statistics*, (4th ed.) New York: MacMillan Publishing Company.

Hughes, W. C., 1977. “Peak Discharge Frequency from Rainfall Information,” *Journal of the Hydraulics Division*, Vol. 103, No. HY1, pp. 39–50, January, ASCE.

Jagerman, D. L., B. Melamed, and W. Willinger, 1997. “Stochastic Modeling of Traffic Processes,” Technical Report RRR 7–97, Rutgers Center for Operations Research, New Brunswick, NJ.

Jenkins, J. D. and H. M. Johnson, 1978. “Flood Profiles in Combined Tidal-freshwater Zones,” *Journal of the Hydraulics Division*, Vol. 104, No. 6, ASCE.

Johnson, R. A., J. W. Evans, and D. W. Green, 1999. “Some Bivariate Distributions for Modeling the Strength Properties of Lumber,” Forest Service Research Paper, FPL-RP-575, August.

Jorgeson, J. and P. Julien, 2005. “Peak Flow Forecasting with Radar Precipitation and the Distributed Model CASC2D,” *Water International*, Vol. 30, No. 1. International Water Resources Association, pp. 40–49.

Koltun, G. F. and J. M. Sherwood, 1998. “Factors Related to the Joint Probability of Flooding on Paired Streams,” USGS Water-Resources Investigations Report 98-4238.

Kubik, H. E., 1990. “Annual Extreme Lake Elevations by Total Probability Theorem,” Proceedings of the Great Lakes Water Level Forecasting and Statistics Symposium, May 17–18, Windsor, Ontario, Canada.

Lambert, M. and G. Kuczera, 1996. “Approximate Joint Probability Analysis of Extreme Water Levels in Coastal Catchments,” *Stochastic Hydraulics*, Ticide, Goulter, Xu, Wasimi & Bouchart (Editors), Balkema, Rotterdam, The Netherlands.

Liu, D. and C. Wang, 1996. “Extreme Wave Prediction in Markov Chain Condition,” *Shipbuilding of China*, No. 133, May.

Loganathan, G. V., C. Y. Kuo, and J. Yannaccone, 1987. “Joint Probability Distribution of Stream Flows and Tides in Estuaries,” *Nordic Hydrology*, Vol. 18, pp. 237–246.

Long, C. E., 1998. “Joint Wind Wave Height-Frequency-Direction Statistics at Two Disparate Sites,” WES/TR/CHL-98-10, May.

Maddox, R. A., L. R., Hoxit, C. F. Chappell and F. Caracena, 1978. “Comparison of Meteorological Aspects of the Big Thompson and Rapid City Flash Floods,” *Monthly Weather Review*, Vol. 106, No. 3, pp. 375–389.

Maddox, R. A., F. Canova, and L. R. Hoxit, 1980. “Meteorological Characteristics of Flash Flood Events over the Western United States,” *Monthly Weather Review*, Vol. 108, No. 11, pp. 1866–1877.

Maddox, R. A., J. Zhang, J. J. Gourley, and K. W. Howard, 2002. “Weather Radar Coverage Over the Contiguous United States,” *Weather and Forecasting*, Vol. 17, pp. 927–934.

Mark, D. J., 1996. “Southern Guam Typhoon Stage-Frequency Analysis,” MP-CERC-96-7, November.

Masutani, M. and A. Leetmaa, 1999. “Dynamical Mechanisms of the 1995 California Floods,” *Journal of Climate*, Vol. 12, No. 11, pp. 3220–3236.

Matsubayashi U. and F. Takagi, 1987. “On the Probabilistic Characteristics of Point and Areal Rainfall,” in Hydrologic Frequency Modeling: Proceedings of the International Symposium on Flood Frequency and Risk Analysis, May 14–17, 1986, Baton Rouge, LA, ed. Singh V. P., Reidel, Dordrecht, The Netherlands.

Matthews, W. T. and D. M. Neal, 1992. “Assessment of Helicopter Component Statistical Reliability Computations,” MTL-TR-92-71, September.

McNeil, D. R., 1970. “A Theoretical Analysis of Congestion in a Two-lane Freeway.” *Journal of Applied Probability,* Vol. 7, No. 2, pp. 304–315.

Meyer, P. L., 1970. *Introductory Probability and Statistical Applications*, (2nd ed.), Addison-Wesley Publishing Company, Reading, MA.

Morris, C. D. and S. J. Calise, 1987. "Bivariate Analysis of Concurrent Flooding," in Hydrologic Frequency Modeling: Proceedings of the International Symposium on Flood Frequency and Risk Analysis, May 14–17, 1986, Baton Rouge, LA, ed. Singh V. P., Reidel, Dordrecht, The Netherlands, pp. 615–632.

Morris, C. D. and L. C. Wilson, 1987. "Concurrent Flooding Probabilities," in Hydrologic Frequency Modeling: Proceedings of the International Symposium on Flood Frequency and Risk Analysis, May 14–17, 1986, Baton Rouge, LA, ed. Singh V. P., Reidel, Dordrecht, The Netherlands.

O'Connor, J. E. and J. E. Costa, 2004. "Spatial Distribution of the Largest Rainfall-Runoff Floods from Basins Between 2.6 and 26000 km^2 in the United States and Puerto Rico," *Water Resources Research*, Vol. 40.

Olivera, F. and T. Gill, 2004. "GIS Static Storm Model Development: Literature Review and Progress Report," FHWA/TX-05/0-4642-1, September.

Panofsky, H. A. and G. W. Brier, 1958, "Some Applications of Statistics to Meteorology," The Pennsylvania State University, State College, PA, p. 81, p. 94.

Pingel, N. and D. Ford, 2004. "Interior Floodplain Flood-Damage Reduction Study," *Journal of Water Resources Planning and Management*, Vol. 130, No. 2, pp. 123–130, ASCE.

Prior-Jones, R. L. and F. L. Beiboer, 1990. "Use of Joint Probability in Deriving Environmental Design Criteria," Environmental Forces on Offshore Structures and their Prediction, London, November 28–29.

Racicot, R. W. and F. Moses, 1972. "Filtered Poisson Process for Random Vibration Problems," *Journal of Soil Mechanics and Foundation Division*, Vol. 98, No. EM1, February, ASCE.

Randerson, D., 1976. "Meteorological Analysis for the Las Vegas, Nevada, Flood of 3 July 1975," *Monthly Weather Review*, Vol. 104, No. 6, pp. 719–727.

Raynal, J. A. and J. D. Salas, 1987. "A Probabilistic Model for Flooding Downstream of the Junction of Two Rivers," in Hydrologic Frequency Modeling: Proceedings of the International Symposium on Flood Frequency and Risk Analysis, May 14–17, 1986, Baton Rouge, LA, ed. Singh V. P., Reidel, Dordrecht, The Netherlands, pp. 595–601.

Ribeny, F. M., 1971. "On the Chance of Culvert Washouts on a Long Railway Line," Australia Institution of Engineers, Hydrology Papers, Sydney, Australia.

Sackl, B. and H. Bergman, 1987. "A Bivariate Flood Model and its Application," in Hydrologic Frequency Modeling: Proceedings of the International Symposium on Flood Frequency and Risk Analysis, May 14–17, 1986, Baton Rouge, LA, ed. Singh V. P., Reidel, Dordrecht, The Netherlands, pp. 571–582.

Salvadori, G., and C. De Michele, 2004. "Frequency Analysis via Copula: Theoretical Aspects and Applications to Hydrologic Events," *Water Resources Research*, Vol. 40.

Scheffner, N. W., L. E. Borgman, and D. J. Mark, 1996. "Empirical Simulation Technique Based Storm Surge Frequency Analyses," *Journal of Waterway, Port, Coastal, and Ocean Engineering*, Vol. 122, No. 2, March/April, ASCE.

Scheffner, N. W., D. J. Mark, L. Lin, W. A. Brandon, and M. C. Miller, 1999. "Development of Water-Surface Elevation Frequency-of-Occurrence Relationships for the Brunswick, North Carolina, Nuclear Power Plant Site," CHL-99-12, December.

Seo, D. J., 1996. “Nonlinear Estimation of Spatial Distribution of Rainfall – An Indicator Cokriging Approach,” *Stochastic Environmental Research and Risk Assessment*, Vol. 10, No. 2, pp. 127–150, May.

Seo, D. J., 1998. “Real-time Estimation of Rainfall Fields Using Radar Rainfall and Rain Gage Data,” *Journal of Hydrology*, Vol. 208, pp. 37–52.

Seo, D. J., J. Breidenbach, R. Fulton, D. Miller, and T. O’Bannon, 2000. “Real-time Adjustment of Range Dependent Biases in WSR88D Rainfall Estimates Due to Nonuniform Vertical Profile of Reflectivity,” *Journal of Hydrometeorology*, Vol. 1, pp. 222–240.

Shiau, J. T., 2003. “Return Period of Bivariate Distributed Extreme Hydrological Events,” *Stochastic Environmental Research and Risk Assessment*, Vol. 17, No. 1–2, pp. 42–57, May.

Skaugen, T. 1996. “Spatial Approach to the Modeling and Estimation of Areal Precipitation,” Doctoral Thesis, Oslo University, Norway.

Singh, V. P., 2000. “The Entropy Theory as a Tool for Modelling and Decision-Making in Environmental and Water Resources,” *Water SA*, Vol. 26, No. 1, pp. 1–12, January.

Sui, J., and G. Koehler, 2001. “Rain-on-Snow Induced Flood Events in Southern Germany,” *Journal of Hydrology*, Vol. 252, pp. 205–220.

Svensson, C., 1999. “Occurrence, Shape, and Movement of Daily Rainfall in the Huai Basin, China,” *Hydrological Processes*, Vol. 13, pp. 1197–1215.

Svensson, C., and D. A. Jones, 2004. “Dependence Between Sea Surge, River Flow, and Precipitation in South and West Britain,” *Hydrology and Earth System Sciences*, Vol. 8, No. 5, pp. 973–992.

Szidarovszky, F., I. Bogardi, and L. Duckstein, 1975. “Levee System Reliability Along a Confluence Reach,” *Journal of the Engineering Mechanics Division*, Vol. 101, No. 5, pp. 609–622, September/October, ASCE.

Tayfun, M. A., 1979. “Joint Occurrences in Coastal Flooding,” *Journal of the Waterway Port Coastal and Ocean Division*, Vol. 105, No. WW2, ASCE.

USACE, 1987. “Hydrologic Analysis of Interior Areas,” EM 1110-2-1413. (Primarily Section 4-8 “Coincident Frequency Methods”).

USACE, 1993. “Hydrologic Frequency Analysis,” EM 1110-2-1415. (Primarily Chapter 11 “Frequency of Coincident Flow”).

Van Mullem, J. A., 1998. “Coincident Frequency in Snowmelt Runoff Modeling,” The 1998 International Water Resources Engineering Conference, Part 2 (of 2), Memphis, TN, August, pp. 1303–1308.

Vieux, B. E. and P. B. Bendient, 1998. “Estimation of Rainfall for Flood Prediction from WSR-88D Reflectivity: A Case Study, 17–18 October 1994,” *Weather and Forecasting*: Vol. 13, No. 2, pp. 407–415.

Wang, D., M. B. Smith, Z. Zhang, S. Reed, V. I. Koren, 2000. “Statistical Comparison of Mean Areal Precipitation Estimates From WSR-88D, Operational and Historical Gage Networks,” 15th Conference on Hydrology, January 9–14, 2000, Long Beach CA. American Meteorological Society.

Wilks, D. S., 1995. “Statistical Methods in the Atmospheric Sciences,” Academic Press, San Diego, CA, pps. 11–12, 13, 14, 16–17, 211, 293.

Wilson, L. C., 1980. “The Probability of Concurrent Flooding on the Meramec and Mississippi Rivers,” Master’s Thesis, University of Missouri–Rolla.

Wilson, L. C., 1986. “A Continuous Spaciotemporal Stochastic Model for Short-Time-Increment Precipitation,” Ph.D. Thesis, University of Missouri–Rolla.

Wilson, L. L. and E. Foufoula-Georgiou, 1990. “Regional Rainfall Frequency Analysis Via Stochastic Storm Transposition,” *Journal of Hydraulic Engineering*, Vol. 116, No. 7, July, ASCE.

Wood, E. F., M. Sivapalan, and K. Beven, 1993. “Similarity and Scale in Catchment Storm Response,” PAPER-89RG01615, Princeton University.

Zhang, L. and V. P. Singh, 2006. “Bivariate Flood Frequency Analysis Using the Copula Method,” *Journal of Hydrologic Engineering*, Vol. 11, No. 2, pp. 150–164, March/April, ASCE.

Appendix B. Databases

Table of Contents

List of Tables ..... B-i

B.1 Gage Pairs B-1

B.1.1 Gage Pair Evolution B-1

B.1.2 Discharge Quantiles B-1

B.2 Instantaneous Data B-2

B.3 Watershed Parameters B-2

B.4 Statistical Summaries.....B-3

List of Tables

Table B.1. Summary of Gage Pairs.....B-4

Table B.2. IDA Gages. B-11

Table B.3. Annual Peak Series Statistics from Daily Data.....B-12

Table B.4. Summary Statistics for Concurrent Data: Peak on Main (POM).....B-16

Table B.5. Summary Statistics for Concurrent Data: Peak on Tributary (POT). B-18

B.1 Gage Pairs

The analyses in this research were based on a set of gage pairs from the coterminous United States. The following sections describe the final selection of gage pairs and the computation of discharge quantiles for each gage record.

B.1.1 Gage Pair Evolution

Table B.1 summarizes 93 gage pairs intended for this research. After working with the data, seven of these pairs were dropped as unsuitable leaving 85 gage pairs for the research. The gage pairs dropped and the corresponding reasons are as follows:

- Pair 07. One of the streams was regulated.
- Pair 10. One of the streams was regulated.
- Pair 30. Coincident sample size too small (n = 6).
- Pair 50. Coincident sample size too small (n = 3).
- Pair 63. Coincident sample size too small (n = 6).
- Pair 91. One of the streams was regulated.
- Pair 92. No overlap in gaged data.
- Pair 93. One of the streams was regulated.

B.1.2 Discharge Quantiles

Log Pearson Type III discharge quantiles following the guidelines of Bulletin 17B (IACWD, 1982) were computed using the Army Corps of Engineers software program HEC-FFA (HEC, 1992). For each stream gage, input data supplied to HEC-FFA were:

1. USGS annual peak data. These were downloaded from the NWIS Web server and were used for systematic events. All annual peak values were used in the analysis. Stream gages subject to upstream reservoir regulation were previously screened out of the gage set based on USGS data flags. Historic (as defined by HEC-FFA) events were not used in the analysis because a systematic method for supplementing annual peak data with historical event data was not identified.
2. Regional skew value. This was taken from Bulletin 17B Plate I "Generalized skew coefficients of logarithms of annual maximum stream flow." Plate I contours of constant skew were digitized using GIS tools. A skew surface (grid) was interpolated using the natural neighbor method. For each stream gage, the interpolated value of regional skew was read from the surface at the centroid of the gaged watershed.

HEC-FFA options specified for controlling computations were:

1. Computed quantiles for twelve values of percent chance exceedance: 0.2, 0.5, 1, 2, 5, 10, 20, 50, 80, 90, 95, and 99.
2. Adopted skew was weighted value of computed skew and regional skew. Weighting was based on mean-square errors of regional and station skew values. Mean-square error of regional skew value from Plate I was taken as 0.302 per Bulletin 17B.

3. High and low outliers were identified automatically by HEC-FFA, following the algorithm specified in Bulletin 17B. Outliers were removed by HEC-FFA from the data before final computations.

B.2 Instantaneous Data

Table B.2 provides a summary of the instantaneous data archive (IDA) data used in this research.

B.3 Watershed Parameters

NHDPlus (EPA and USGS, 2007) was used as the source for deriving watershed characteristics because it allowed for efficient delineation of watershed boundaries and computation of centroid locations, longest flow paths, and longest flow path average slopes. This was critical because identification of suitable stream gage pairs was a necessary and substantial effort and was made possible by taking advantage of USGS delineations of nearly 3 million small subbasins for the U.S., one for each NHDPlus stream segment, already included in NHDPlus.

Another advantage of using NHDPlus was that stream segments are networked. Through application of GIS tools, this allowed for identification of potential gage pairs in which one USGS stream gage is upstream of another. In those cases, the tributary watershed (smaller drainage area) is contained by the main watershed (larger drainage area). Identification of those cases allowed for further screening to determine whether inclusion or exclusion from the final set of gage pairs was appropriate.

Locations of gaged watershed centroids were needed, as these were, with few exceptions, generally not available from the USGS. Watershed centroid location (latitude/longitude) for each watershed was computed as follows:

1. Using GIS tools, identified NHDPlus stream segment nearest to gage location.
2. Using NHDPlus Value Added Attributes (VAA) tool, used the “navigate upstream with tributaries” tool feature to identify all stream segments upstream of segment nearest to gage.
3. Using GIS tools, selected all NHDPlus subbasins containing stream segments identified in previous step. Using GIS tools, merged selected subbasins into one watershed.
4. Using GIS tools, computed centroid of watershed upstream of gage.

The distance between watershed centroids was computed for each of the original 93 gage pairs as the Pythagorean distance based on projected coordinates of the two centroids. The Albers Equal Area Conic projection was used.

Maximum flow length of each watershed was computed as follows:

1. Using GIS tools, identified NHDPlus stream segment nearest to gage location.
2. Using NHDPlus Value Added Attributes (VAA) tool, used the “navigate upstream main stem” tool feature to identify all stream segments upstream of segment nearest to gage.
3. From the NHDPlus flowline attributes table, summed the lengths of all main stream segments to obtain watershed maximum flow length.

Watershed channel slope of each watershed was computed as follows:

1. From the NHDPlus flowline attributes table, recorded minimum and maximum elevation of all main stream segments.
2. Computed difference in elevation and divided by maximum flow length to obtain watershed channel slope. This is the average slope of the channel along the longest flow path.

A length slope parameter was computed to represent a quantity analogous to time of concentration. The method was essentially an application of the Kirpich equation, however, since many of the watersheds fall outside the appropriate range recommended for the Kirpich equation for time of concentration, it is simply called the length slope parameter. Whether this parameter is useful will be discussed elsewhere in this report. The equation is as follows:

$$T_{LS} = 0.00013 \frac{L^{0.77}}{S^{0.385}} \tag{B.1}$$

where,

T_{LS}	=	Length slope parameter, h
L	=	length of the basin along the watercourse from the point of interest to the farthest point on the basin divide, ft
S	=	average main channel slope for the basin, ft/ft

Values of *L* and *S* were taken as the watershed maximum flow length and watershed channel slope from NHDPlus.

B.4 Statistical Summaries

Statistical analyses of the annual peak series based on the daily data have been completed and are summarized for the 85 gage pairs used in this research in Table B.3. The table provides the number of years of data (n) along with the mean, standard deviation, and mean of the untransformed data and the mean, standard deviation, and skew of the log-transformed data. The final two columns provide the location and scale parameters for fitting the Gumbel distribution. These data are useful for defining annual exceedance probabilities for the marginal distributions.

Two concurrent datasets were assembled for each of the gage pairs based on daily data. One dataset was the annual peak on the main stream and the corresponding value on the tributary on the same day (POM). The second dataset was the annual peak on the tributary and the corresponding value on the main stream on the same day (POT). Tables B.4 and B.5 summarize the statistics for the two datasets. In some cases, the complementary daily value was zero. In order to compute some of the statistics in the tables, the zero values were adjusted to 0.1 ft^3/s.

For each pair the number of concurrent observations for each gage pair is reported. In a few cases, the number of years of concurrent data differs slightly between the POM and POT datasets when daily data are missing for the day on which a peak occurs on the other gage. The correlation measures of Pearson's ρ (untransformed and log (base 10) and Kendall's τ are summarized in the tables. The p-value to the right of each column is a measure of the significance of the correlation. A lower p-value indicates higher significance of the correlation estimate. In addition, Gumbel's θ, the Gumbel-Hougaard θ, and the Frank θ are provided. These data are used for fitting the concurrent datasets to the bivariate distributions and copulas.

Table B.1. Summary of Gage Pairs.

Gage Pair	Area Ratio	Gage Separation (mi)	Drainage Area (mi^2)	Station	State	Station Name	Longitude. (degrees)	Latitude (degrees)
1	1.8	3.0	2680.0	01010500	ME	St. John River at Dickey, Maine	-69.08814	47.11317
			1478.0	01011000	ME	Allagash River near Allagash, Maine	-69.07954	47.06961
2	82.3	6.8	385.0	01064500	NH	Saco River near Conway, NH	-71.09048	43.99084
			4.7	01064400	NH	LUCY BROOK NEAR NORTH CONWAY, NH	-71.17258	44.07016
3	54.3	4.0	94.0	01181000	MA	WEST BRANCH WESTFIELD RIVER AT HUNTINGTON, MA	-72.89594	42.23711
			1.7	01180000	MA	SYKES BROOK AT KNIGHTVILLE, MA	-72.87033	42.29094
4	2.8	0.5	19.9	01187300	CT	HUBBARD RIVER NR. WEST HARTLAND, CT.	-72.93959	42.03707
			7.0	01187400	CT	VALLEY BK NR WEST HARTLAND, CT.	-72.92991	42.03427
5	394.5	4.8	785.0	01403060	NJ	RARITAN RIVER BELOW CALCO DAM AT BOUND BROOK NJ	-74.54857	40.55162
			2.0	01403150	NJ	WEST BRANCH MIDDLE BROOK NEAR MARTINSVILLE NJ	-74.59065	40.61236
6	2.9	3.6	106.0	01445500	NJ	PEQUEST RIVER AT PEQUEST NJ	-74.97845	40.83054
			36.7	01446000	NJ	BEAVER BROOK NEAR BELVIDERE NJ	-75.04630	40.84327
7	634.4	2.2	9960.0	01536500	PA	Susquehanna River at Wilkes-Barre, PA	-75.880300	41.253875
			15.7	01537500	PA	Solomon Creek at Wilkes-Barre, PA	-75.905144	41.227497
8	1.2	1.2	8.5	01590000	MD	NORTH RIVER NEAR ANNAPOLIS, MD	-76.62220	38.98592
			6.9	01590500	MD	BACON RIDGE BRANCH AT CHESTERFIELD, MD	-76.61447	39.00202
9	3.5	0.5	57.3	01615000	VA	OPEQUON CREEK NEAR BERRYVILLE, VA	-78.07825	39.17448
			16.5	01616000	VA	ABRAMS CREEK NEAR WINCHESTER, VA	-78.08580	39.17785
10	107.7	2.1	9651.0	01638500	MD	POTOMAC RIVER AT POINT OF ROCKS, MD	-77.544277	39.272590
			89.6	01638480	VA	CATOCTIN CREEK AT TAYLORSTOWN, VA	-77.576454	39.254578
11	183.5	7.4	1081.0	01673000	VA	PAMUNKEY RIVER NEAR HANOVER, VA	-77.33221	37.76764
			5.9	01673500	VA	TOTOPOTOMOY CREEK NEAR ATLEE, VA	-77.38248	37.66932
12	206.0	8.2	6758.0	02037500	VA	JAMES RIVER NEAR RICHMOND, VA	-77.54664	37.56098
			32.8	02038000	VA	FALLING CREEK NEAR CHESTERFIELD, VA	-77.52247	37.44396
13	28.7	3.2	1131.0	03164000	VA	NEW RIVER NEAR GALAX, VA	-80.97765	36.64646
			39.4	03165000	VA	CHESTNUT CREEK AT GALAX, VA	-80.91934	36.64587

Gage Pair	Area Ratio	Gage Separation (mi)	Drainage Area (mi^2)	Station	State	Station Name	Longitude. (degrees)	Latitude (degrees)
14	2.4	3.7	306.0	03202400	WV	GUYANDOTTE RIVER NEAR BAILEYSVILLE, WV	-81.64417	37.60365
			126.0	03202750	WV	CLEAR FORK AT CLEAR FORK, WV	-81.70782	37.62329
15	2.0	3.1	129.0	03240000	OH	Little Miami River near Oldtown OH	-83.93144	39.74778
			63.2	03241500	OH	Massies Creek at Wilberforce OH	-83.88290	39.72255
16	1.8	2.9	84.7	03250100	KY	NORTH FORK TRIPLETT CREEK NEAR MOREHEAD, KY	-83.48106	38.19870
			47.5	03250000	KY	TRIPLETT CREEK AT MOREHEAD, KY	-83.42999	38.18415
17	3.3	3.9	202.0	03280600	KY	MIDDLE FORK KENTUCKY RIVER NEAR HYDEN, KY	-83.37091	37.13698
			61.3	03280700	KY	CUTSHIN CREEK AT WOOTON, KY	-83.30803	37.16497
18	2.8	1.7	1858.0	03455000	TN	FRENCH BROAD RIVER NEAR NEWPORT, TN	-83.16064	35.98203
			666.0	03461500	TN	PIGEON RIVER AT NEWPORT, TN	-83.17416	35.96043
19	4.2	2.5	82.5	03538225	TN	POPLAR CREEK NEAR OAK RIDGE, TN	-84.33939	35.99869
			19.5	03538250	TN	EAST FORK POPLAR CREEK NEAR OAK RIDGE, TN	-84.35823	35.96648
20	50.7	8.0	385.0	04117500	MI	THORNAPPLE RIVER NEAR HASTINGS, MI	-85.23625	42.61551
			7.6	04117000	MI	QUAKER BROOK NEAR NASHVILLE, MI	-85.09357	42.56587
21	101.7	0.9	117.0	04140500	MI	RIFLE RIVER AT STATE ROAD AT SELKIRK, MI	-84.06995	44.31342
			1.2	04141000	MI	SOUTH BRANCH SHEPARDS CREEK NEAR SELKIRK, MI	-84.08694	44.30777
22	1.6	3.9	63.6	05423500	WI	SOUTH BRANCH ROCK RIVER AT WAUPUN, WI	-88.72036	43.64146
			40.7	05423000	WI	WEST BRANCH ROCK RIVER NEAR WAUPUN, WI	-88.65244	43.66803
23	60.0	8.8	8753.0	05443500	IL	ROCK RIVER AT COMO, IL	-89.75040	41.78184
			146.0	05444000	IL	ELKHORN CREEK NEAR PENROSE, IL	-89.69672	41.90255
24	1.7	1.4	2256.0	05476750	IA	Des Moines River at Humboldt, IA	-94.22050	42.71934
			1308.0	05479000	IA	East Fork Des Moines River at Dakota City, IA	-94.19289	42.72369
25	67.5	8.3	1619.0	05482500	IA	North Raccoon River near Jefferson, IA	-94.37694	41.98796
			24.0	05483000	IA	East Fork Hardin Creek near Churdan, IA	-94.37025	42.10748
26	425.6	9.1	5150.0	05527500	IL	KANKAKEE RIVER NEAR WILMINGTON, IL	-88.18808	41.34607
			12.1	05526500	IL	TERRY CREEK NEAR CUSTER PARK, IL	-88.09843	41.23353

Gage Pair	Area Ratio	Gage Separation (mi)	Drainage Area (mi^2)	Station	State	Station Name	Longitude. (degrees)	Latitude (degrees)
27	5.8	3.3	212.0	05587900	IL	CAHOKIA CREEK AT EDWARDSVILLE, IL	-89.97485	38.82452
			36.7	05588000	IL	INDIAN CREEK AT WANDA, IL	-90.03296	38.84166
28	9.4	2.8	543.0	06193500	MT	Shields River at Clyde Park MT	-110.61800	45.88580
			57.9	06194000	MT	Brackett Creek near Clyde Park MT	-110.67022	45.86664
29	2.7	1.6	143.0	06211000	MT	Red Lodge Cr ab Cooney Re nr Boyd MT	-109.25334	45.43790
			53.3	06211500	MT	Willow Creek near Boyd MT	-109.23051	45.42218
30	3.3	3.7	803.0	06270000	WY	NOWOOD RIVER NEAR TENSLEEP, WY	-107.42836	44.01337
			247.0	06271000	WY	TENSLEEP CREEK NEAR TENSLEEP, WYO.	-107.38844	44.05793
31	1.1	0.6	2040.0	06776500	NE	DISMAL RIVER AT DUNNING, NE	-100.10047	41.82269
			1830.0	06775500	NE	MIDDLE LOUP RIVER AT DUNNING, NEBR.	-100.10067	41.83097
32	4.8	3.5	1400.0	06797500	NE	ELKHORN RIVER AT EWING, NEBR.	-98.33939	42.26823
			292.0	06798000	NE	SOUTH FORK ELKHORN RIVER AT EWING, NEBR.	-98.39785	42.24158
33	443.1	7.6	2840.0	06933500	MO	Gasconade River at Jerome, MO	-91.97669	37.92996
			6.4	06931500	MO	LITTLE BEAVER CR NR ROLLA, MO	-91.83634	37.93491
34	1.0	3.2	570.0	07058000	MO	Bryant Creek near Tecumseh, MO	-92.30639	36.62639
			561.0	07057500	MO	North Fork River near Tecumseh, MO	-92.24769	36.62294
35	8.4	1.3	474.0	08179000	TX	Medina Rv nr Pipe Creek, TX	-98.97607	29.67555
			56.3	08179100	TX	Red Bluff Ck nr Pipe Ck, TX	-98.95574	29.68122
36	1.9	3.2	117.0	08230500	CO	CARNERO CREEK NEAR LA GARITA, CO.	-106.31946	37.85975
			61.0	08231000	CO	LA GARITA CREEK NEAR LA GARITA, CO.	-106.31864	37.81340
37	1.5	2.1	167.0	08248000	NM	LOS PINOS RIVER NEAR ORTIZ, CO.	-106.07349	36.98194
			110.0	08247500	NM	SAN ANTONIO RIVER AT ORTIZ, CO.	-106.03836	36.99297
38	3.3	1.4	1390.0	08340500	NM	ARROYO CHICO NR GUADALUPE, NM	-107.18933	35.59222
			420.0	08334000	NM	RIO PUERCO ABV ARROYO CHICO NR GUADALUPE, NM	-107.16605	35.60087
39	1.0	0.4	295.0	08389500	NM	RIO BONITO AT HONDO, NM	-105.27509	33.38969
			290.0	08388000	NM	RIO RUIDOSO AT HONDO, NM	-105.27544	33.38351
40	160.9	6.2	5858.0	09508500	AZ	VERDE R BLW TANGLE CREEK, ABV HORSESHOE DAM, AZ.	-111.71605	34.07286
			36.4	09508300	AZ	WET BOTTOM CREEK NEAR CHILDS, AZ.	-111.69289	34.16076
41	474.2	7.3	3438.0	10170490	UT	COM FLW JORDAN RIVER & SURPLUS CANAL @ SLC, UT	-111.92684	40.72671
			7.3	10172200	UT	RED BUTTE CREEK AT FORT DOUGLAS, NEAR SLC, UT	-111.80535	40.78011

Gage Pair	Area Ratio	Gage Separation (mi)	Drainage Area (mi^2)	Station	State	Station Name	Longitude. (degrees)	Latitude (degrees)
42	1.9	0.6	134.0	10260500	CA	DEEP C NR HESPERIA CA	-117.22985	34.34220
			70.3	10261000	CA	WF MOJAVE R NR HESPERIA CA	-117.24080	34.34059
43	1.8	3.0	133.0	12082500	WA	NISQUALLY RIVER NEAR NATIONAL, WA	-122.08373	46.75222
			75.2	12083000	WA	MINERAL CREEK NEAR MINERAL, WA	-122.14542	46.74433
44	100.2	9.5	535.0	12134500	WA	SKYKOMISH RIVER NEAR GOLD BAR, WA	-121.66729	47.83696
			5.3	12147600	WA	SOUTH FORK TOLT RIVER NEAR INDEX, WA	-121.60017	47.70656
45	2.8	3.7	68.9	12500500	WA	NORTH FORK AHTANUM CREEK NEAR TAMPICO, WA	-120.91705	46.56391
			24.8	12501000	WA	SF AHTANUM CREEK AT CONRAD RANCH NR TAMPICO, WA	-120.91630	46.51074
46	1.9	3.9	323.0	13011900	WY	BUFFALO FORK AB LAVA CREEK NR MORAN WY	-110.44072	43.83772
			169.0	13011500	WY	PACIFIC CREEK AT MORAN WY	-110.51706	43.85086
47	3.5	3.6	59.6	14013000	WA	MILL CREEK NEAR WALLA WALLA, WA	-118.11882	46.00792
			17.0	14013500	WA	BLUE CREEK NEAR WALLA WALLA, WA	-118.14030	46.05756
48	1.3	2.7	176.0	14020300	OR	MEACHAM CREEK AT GIBBON, OR	-118.35667	45.68851
			131.0	14020000	OR	UMATILLA RIVER ABOVE MEACHAM CREEK, NR GIBBON, OR	-118.32352	45.71958
49	2.0	3.4	216.0	14178000	OR	NO SANTIAM R BLW BOULDER CRK, NR DETROIT, OR	-122.10085	44.70743
			108.0	14179000	OR	BREITENBUSH R ABV FRENCH CR NR DETROIT, OR.	-122.13190	44.75117
50	8.2	3.5	1162.0	14235000	WA	COWLITZ RIVER AT MOSSYROCK, WA	-122.49318	46.54985
			141.0	14236200	WA	TILTON RIVER AB BEAR CANYON CREEK NEAR CINEBAR, WA	-122.45945	46.59514
51	1.1	2.5	161.0	14301500	OR	WILSON RIVER NEAR TILLAMOOK, OR	-123.72442	45.47558
			145.0	14302500	OR	TRASK RIVER NEAR TILLAMOOK,OREG.	-123.71796	45.44012
52	3.4	6.2	152.0	01547950	PA	Beech Creek at Monument, PA	-77.70211	41.11151
			44.1	01547700	PA	Marsh Creek at Blanchard, PA	-77.60571	41.05964
53	1.8	7.3	231.0	02113850	NC	ARARAT RIVER AT ARARAT, NC	-80.56172	36.40458
			128.0	02113000	NC	FISHER RIVER NEAR COPELAND, NC	-80.67895	36.35696
54	12.5	8.1	1372.0	02126000	NC	ROCKY RIVER NEAR NORWOOD, NC	-80.17611	35.14850
			110.0	02127000	NC	BROWN CREEK NEAR POLKTON, NC	-80.14974	35.03385

Gage Pair	Area Ratio	Gage Separation (mi)	Drainage Area (mi^2)	Station	State	Station Name	Longitude. (degrees)	Latitude (degrees)
55	1.4	5.0	262.0	02220900	GA	LITTLE RIVER NEAR EATONTON, GA	-83.43700	33.31406
			190.0	02221525	GA	MURDER CREEK BELOW EATONTON, GA	-83.48132	33.25227
56	2.7	7.5	272.0	02344500	GA	FLINT RIVER NEAR GRIFFIN, GA	-84.42889	33.24389
			101.0	02344700	GA	LINE CREEK NEAR SENOIA, GA	-84.52211	33.31925
57	1.4	8.2	305.0	03173000	VA	WALKER CREEK AT BANE, VA	-80.70763	37.26924
			223.0	03175500	VA	WOLF CREEK NEAR NARROWS, VA	-80.84973	37.30614
58	4.9	7.0	1385.0	04114000	MI	GRAND RIVER AT PORTLAND, MI	-84.91248	42.85654
			281.0	04114500	MI	LOOKING GLASS RIVER AT HINMAN RD NEAR EAGLE, MI	-84.77853	42.82920
59	1.8	8.4	1153.0	04149000	MI	FLINT RIVER NEAR FOSTERS, MI	-83.95357	43.30858
			637.0	04145000	MI	SHIAWASSEE RIVER NEAR FERGUS, MI	-84.10451	43.25531
60	1.3	7.0	298.0	04196500	OH	Sandusky River near Upper Sandusky OH	-83.25630	40.85056
			229.0	04196800	OH	Tymochtee Creek at Crawford OH	-83.34905	40.92310
61	3.6	5.6	126.0	04234000	NY	FALL CREEK NEAR ITHACA NY	-76.47312	42.45340
			35.2	04233000	NY	CAYUGA INLET NEAR ITHACA NY	-76.54501	42.39315
62	2.6	9.7	1090.0	05367500	WI	RED CEDAR RIVER NEAR COLFAX, WI	-91.71190	45.05308
			418.0	05368000	WI	HAY RIVER AT WHEELER, WI	-91.91102	45.04785
63	1.3	4.5	101.0	05376000	MN	NORTH FORK WHITEWATER RIVER NEAR ELBA, MN	-92.06593	44.09235
			76.8	05376500	MN	SOUTH FORK WHITEWATER RIVER NEAR ALTURA, MN	-91.98066	44.06949
64	1.7	6.4	620.0	05500000	MO	South Fabius River near Taylor, MO	-91.57987	39.89678
			373.0	05501000	MO	North River at Palmyra, MO	-91.51752	39.81836
65	3.6	7.3	464.0	05594800	IL	SILVER CREEK NEAR FREEBURG, IL	-89.87296	38.39653
			129.0	05595200	IL	RICHLAND CREEK NEAR HECKER, IL	-89.97107	38.32375
66	5.3	5.7	310.0	06447500	SD	LITTLE WHITE R NEAR MARTIN SD	-101.63283	43.16689
			58.0	06448000	SD	LAKE CR ABOVE REFUGE NEAR TUTHILL SD	-101.60329	43.08712
67	1.7	6.9	424.0	05054500	ND	SHEYENNE RIVER ABOVE HARVEY, ND	-99.94865	47.70270
			253.0	06467600	ND	JAMES RIVER NR MANFRED, ND	-99.82880	47.64394
68	2.4	7.3	400.0	06719505	CO	CLEAR CREEK AT GOLDEN, CO.	-105.23526	39.75299
			164.0	06710500	CO	BEAR CREEK AT MORRISON, CO.	-105.19575	39.65281
69	11.0	6.5	594.0	06863500	KS	BIG C NR HAYS, KS	-99.31831	38.85247
			54.0	06863900	KS	NF BIG C NR VICTORIA, KS	-99.20597	38.88696

Gage Pair	Area Ratio	Gage Separation (mi)	Drainage Area (mi^2)	Station	State	Station Name	Longitude. (degrees)	Latitude (degrees)
70	2.5	8.0	849.0	06871000	KS	NF SOLOMON R AT GLADE, KS	-99.30922	39.67351
			341.0	06871500	KS	BOW C NR STOCKTON, KS	-99.28551	39.55974
71	2.3	8.5	1238.0	07050000	AR	WHITE RIVER AT BEAVER, ARK.	-93.76630	36.47298
			527.0	07050500	AR	Kings River near Berryville, AR	-93.62353	36.42628
72	2.0	8.6	302.0	07075000	AR	Middle Fork of Little Red River at Shirley, AR	-92.32029	35.65231
			148.0	07075300	AR	South Fork of Little Red River At Clinton, AR	-92.45497	35.59301
73	2.5	7.2	1037.0	07077500	AR	Cache River at Patterson, AR	-91.23763	35.26953
			421.0	07077700	AR	Bayou DeView near Morton, AR	-91.11076	35.25199
74	5.0	8.3	410.0	07261500	AR	Fourche LaFave River Near Gravelly, AR	-93.65629	34.87247
			81.4	07260000	AR	Dutch Creek at Waltreak, AR	-93.61283	34.98716
75	1.8	4.6	675	07346070	TX	Little Cypress Ck nr Jefferson, TX	-94.34606	32.71297
			365	07346045	TX	Black Cypress Bayou at Jefferson, TX	-94.35778	32.77901
76	1.6	7.7	289.0	08164450	TX	Sandy Ck nr Ganado, TX	-96.54628	29.16013
			178.0	08164503	TX	W Mustang Ck nr Ganado, TX	-96.46712	29.07177
77	8.9	9.3	817.0	08164000	TX	Lavaca Rv nr Edna, TX	-96.68662	28.95998
			91.7	08164600	TX	Garcitas Ck nr Inez, TX	-96.81918	28.89126
78	3.1	4.8	389.0	08195000	TX	Frio Rv at Concan, TX	-99.70478	29.48852
			126.0	08196000	TX	Dry Frio Rv nr Reagan Wells, TX	-99.78147	29.50468
79	2.4	8.5	289.0	09112500	CO	EAST RIVER AT ALMONT CO.	-106.84819	38.66435
			121.0	09113500	CO	OHIO CREEK NEAR BALDWIN, CO.	-106.99865	38.70197
80	10.2	7.2	457.0	09484600	AZ	PANTANO WASH NEAR VAIL, AZ.	-110.67744	32.03562
			44.8	09485000	AZ	RINCON CREEK NEAR TUCSON, AZ.	-110.62585	32.12963
81	4.0	5.5	249.0	10371500	OR	DEEP CREEK ABOVE ADEL,OREG.	-120.00192	42.18894
			63.0	10370000	OR	CAMAS CREEK NEAR LAKEVIEW,OREG.	-120.10245	42.21615
82	1.6	4.9	208.0	11383500	CA	DEER C NR VINA CA	-121.94814	40.01439
			131.0	11381500	CA	MILL C NR LOS MOLINOS CA	-122.02415	40.05468
83	4.0	5.4	739.0	11401500	CA	INDIAN C NR CRESCENT MILLS CA	-120.92824	40.07793
			184.0	11402000	CA	SPANISH C AB BLACKHAWK C AT KEDDIE CA	-120.95423	40.00282
84	5.0	5.4	262.0	12167000	WA	NF STILLAGUAMISH RIVER NEAR ARLINGTON, WA	-122.04754	48.26078
			52.0	12168500	WA	PILCHUCK CREEK NEAR BRYANT, WA	-122.16413	48.26605
85	2.7	9.9	714	12189500	WA	SAUK RIVER NEAR SAUK, WA	-121.56782	48.42478
			172	12182500	WA	CASCADE RIVER AT MARBLEMOUNT, WA	-121.41507	48.52658

Gage Pair	Area Ratio	Gage Separation (mi)	Drainage Area (mi^2)	Station	State	Station Name	Longitude. (degrees)	Latitude (degrees)
86	2.8	9.3	689	12424000	WA	HANGMAN CREEK AT SPOKANE, WA	-117.45000	47.65249
			665	12431000	WA	LITTLE SPOKANE RIVER AT DARTFORD, WA	-117.40450	47.78484
87	1.3	5.6	180.0	13309000	ID	BEAR VALLEY CREEK NR CAPE HORN ID	-115.28888	44.42929
			138.0	13308500	ID	MF SALMON RIVER NR CAPEHORN ID	-115.17896	44.40849
88	1.6	5.7	1910.0	13336500	ID	SELWAY RIVER NR LOWELL ID	-115.51360	46.08598
			1180.0	13337000	ID	LOCHSA RIVER NR LOWELL ID	-115.58669	46.15059
89	2.1	6.5	246.0	14147500	OR	N FK OF M FK WILLAMETTE R NR OAKRIDGE,OREG.	-122.50437	43.75701
			117.0	14146500	OR	SALMON CREEK NEAR OAKRIDGE,OREG.	-122.37309	43.76240
90	2.1	6.9	211	14154500	OR	ROW RIVER ABOVE PITCHER CREEK NEAR, DORENA, OREG	-122.87366	43.73518
			95.3	14156500	OR	MOSBY CR AT MOUTH, NR COTTAGE GROVE,OREG.	-122.99957	43.77632
91	1.7	0.6	6.5	07320500	OK	Sandstone Creek SWS 6 near Elk City, OK	-99.50335	35.48559
			3.9	07321000	OK	Sandstone Creek SWS 5 near Elk City, OK	-99.49236	35.48846
92	12.8	6.1	30.6	07176800	OK	Candy Creek near Wolco, OK	-96.04921	36.53494
			2.4	07175000	OK	Double Creek Sws 5 near Ramona, OK	-95.94144	36.51408
93	2.6	9.0	86.6	07333800	OK	McGee Creek near Stringtown, OK	-95.86965	34.44261
			32.7	07333500	OK	Chickasaw Creek near Stringtown, OK	-96.02680	34.46158

Table B.2. IDA Gages.

Pair_ID	USGS Station ID	Type	State	USGS station name	Area (mi.^{2})	Start Date	End Date	Time Step (min)
6	01445500	Main	NJ	PEQUEST RIVER AT PEQUEST NJ	106.0	10/1/81	9/30/06	15
	01446000	Trib	NJ	BEAVER BROOK NEAR BELVIDERE NJ	36.7	10/1/03	9/30/06	15
15	03240000	Main	OH	Little Miami River near Oldtown OH	129.0	10/1/90	9/30/06	15
	03241500	Trib	OH	Massies Creek at Wilberforce OH	63.2	10/1/91	9/30/06	30
20	04117500	Main	MI	THORNAPPLE RIVER NEAR HASTINGS, MI	385.0	10/1/89	9/30/06	60
	04117000	Trib	MI	QUAKER BROOK NEAR NASHVILLE, MI	7.6	10/1/94	9/30/06	30
27	05587900	Main	IL	CAHOKIA CREEK AT EDWARDSVILLE, IL	212.0	10/1/86	9/30/06	15
	05588000	Trib	IL	INDIAN CREEK AT WANDA, IL	36.7	9/24/86	9/30/06	15
43	12082500	Main	WA	NISQUALLY RIVER NEAR NATIONAL, WA	133.0	10/1/87	9/30/06	15
	12083000	Trib	WA	MINERAL CREEK NEAR MINERAL, WA	75.2	10/1/87	9/30/06	15
48	14020300	Main	OR	MEACHAM CREEK AT GIBBON, OR	176.0	10/1/88	9/30/06	30
	14020000	Trib	OR	UMATILLA RIVER ABOVE MEACHAM CREEK, NR GIBBON, OR	131.0	10/1/89	9/30/06	30
53	02113850	Main	NC	ARARAT RIVER AT ARARAT, NC	231.0	10/1/86	9/30/05	60
	02113000	Trib	NC	FISHER RIVER NEAR COPELAND, NC	128.0	10/1/86	9/30/05	30
55	02220900	Main	GA	LITTLE RIVER NEAR EATONTON, GA	262.0	10/1/89	9/30/06	60
	02221525	Trib	GA	MURDER CREEK BELOW EATONTON, GA	190.0	10/1/89	9/30/06	15
74	07261500	Main	AR	Fourche LaFave River Near Gravelly, AR	410.0	10/1/86	9/30/05	60
	07260000	Trib	AR	Dutch Creek at Waltreak, AR	81.4	10/1/99	9/30/05	60
76	08164450	Main	TX	Sandy Ck nr Ganado, TX	289.0	10/1/97	9/30/05	60
	08164503	Trib	TX	W Mustang Ck nr Ganado, TX	178.0	10/1/91	9/30/05	60
80	09484600	Main	AZ	PANTANO WASH NEAR VAIL, AZ.	457.0	10/1/89	9/30/06	15
	09485000	Trib	AZ	RINCON CREEK NEAR TUCSON, AZ.	44.8	10/1/87	9/30/06	15

Table B.3. Annual Peak Series Statistics from Daily Data.

Gage Pair	Gage Number	n	Mean	Standard Deviation	Skew	Mean of Logs[1]	Std dev. of Logs[1]	Skew of Logs[1]	Gumbel Location	Gumbel Scale
1	01010500	60	49262	15757	0.205	4.669	0.149	-0.482	42170	12286
1	01011000	75	15338	5430	0.687	4.159	0.157	-0.247	12894	4234
2	01064500	83	12463	5988	1.576	4.051	0.201	-0.128	9768	4669
2	01064400	27	249	143	0.758	2.316	0.286	-0.660	184	112
3	01181000	71	3258	2038	1.731	3.447	0.232	0.543	2341	1589
3	01180000	28	40	33	2.381	1.498	0.285	0.540	25	26
4	01187300	66	728	625	5.259	2.786	0.242	0.380	447	487
4	01187400	32	292	377	3.741	2.307	0.329	1.026	122	294
5	01403060	67	17267	8189	2.507	4.199	0.181	0.183	13581	6385
5	01403150	27	109	59	1.721	1.982	0.238	-0.732	83	46
6	01445500	85	880	351	0.782	2.911	0.172	0.037	722	274
6	01446000	42	496	246	1.256	2.645	0.213	-0.064	385	192
8	01590000	42	107	102	4.178	1.939	0.242	1.495	61	79
8	01590500	26	121	90	1.838	1.984	0.302	0.056	81	70
9	01615000	58	1357	890	1.504	3.043	0.296	-0.475	957	694
9	01616000	26	241	146	0.817	2.298	0.287	-0.248	176	114
11	01673000	65	10092	5931	2.201	3.938	0.250	-0.599	7423	4624
11	01673500	29	91	79	3.410	1.863	0.276	0.556	55	62
12	02037500	72	80976	47530	1.821	4.843	0.242	-0.127	59586	37059
12	02038000	39	664	635	3.183	2.696	0.322	0.381	378	495
13	03164000	77	18157	12259	3.113	4.193	0.230	0.444	12640	9559
13	03165000	62	923	464	0.741	2.908	0.233	-0.289	714	361
14	03202400	38	6266	3922	1.081	3.713	0.283	-0.199	4501	3058
14	03202750	32	3073	1389	0.785	3.444	0.205	-0.464	2448	1083
15	03240000	54	2303	1457	1.012	3.278	0.279	-0.104	1647	1136
15	03241500	54	1296	766	1.091	3.034	0.282	-0.701	951	598
16	03250100	27	3170	1443	2.169	3.466	0.173	0.387	2521	1125
16	03250000	43	2114	1091	1.729	3.273	0.222	-0.461	1623	851
17	03280600	32	6940	3569	1.108	3.788	0.221	-0.024	5334	2783
17	03280700	49	2236	1154	0.794	3.293	0.225	0.026	1717	900
18	03455000	83	22202	10783	0.969	4.298	0.209	-0.038	17349	8408
18	03461500	66	11683	6548	1.187	4.006	0.232	0.130	8736	5106
19	03538225	29	3574	1631	0.698	3.506	0.215	-0.631	2839	1272
19	03538250	27	821	443	1.021	2.858	0.224	0.209	622	346
20	04117500	62	2448	1195	1.178	3.338	0.218	-0.443	1910	932
20	04117000	33	81	42	1.086	1.848	0.237	-0.374	62	33
21	04140500	32	872	379	1.434	2.903	0.187	-0.349	701	296
21	04141000	27	20	9	0.237	1.242	0.295	-2.116	16	7
22	05423500	40	429	310	1.085	2.508	0.368	-0.880	289	242
22	05423000	24	340	250	0.362	2.356	0.463	-0.796	228	195
23	05443500	82	24479	10682	0.200	4.340	0.220	-0.744	19671	8329
23	05444000	67	2096	1282	0.652	3.225	0.315	-0.597	1518	1000
24	05476750	42	5624	3934	1.549	3.647	0.323	-0.652	3854	3067
24	05479000	66	5196	4237	1.744	3.583	0.361	-0.490	3289	3304

Gage Pair	Gage Number	n	Mean	Standard Deviation	Skew	Mean of Logs[1]	Std dev. of Logs[1]	Skew of Logs[1]	Gumbel Location	Gumbel Scale
25	05482500	66	8005	4935	0.654	3.787	0.375	-1.335	5785	3847
25	05483000	39	187	98	0.736	2.204	0.264	-0.709	143	76
26	05527500	89	25419	11275	0.692	4.361	0.201	-0.331	20344	8791
26	05526500	26	159	199	4.080	2.056	0.329	0.386	69	155
27	05587900	37	4333	1933	-0.180	3.578	0.253	-1.058	3462	1507
27	05588000	66	1206	805	1.120	2.972	0.335	-0.540	844	627
28	06193500	38	1171	666	1.674	3.009	0.230	0.001	871	519
28	06194000	25	224	139	1.699	2.283	0.241	0.249	161	108
29	06211000	68	565	488	2.157	2.614	0.355	-0.052	345	381
29	06211500	68	264	259	2.274	2.272	0.349	0.449	147	202
31	06776500	50	473	65	1.243	2.671	0.057	0.900	443	51
31	06775500	61	631	61	0.083	2.798	0.042	-0.211	604	47
32	06797500	59	2044	2030	1.411	3.084	0.479	-0.201	1131	1582
32	06798000	32	625	786	3.269	2.569	0.450	0.123	272	613
33	06933500	85	33525	21170	1.559	4.438	0.295	-0.660	23998	16506
33	06931500	27	189	123	1.835	2.207	0.245	0.467	134	96
34	07058000	52	9438	8190	2.882	3.837	0.368	-0.420	5752	6386
34	07057500	62	9732	8372	2.087	3.859	0.337	0.073	5964	6528
35	08179000	41	5677	7963	2.869	3.433	0.552	-0.005	2093	6209
35	08179100	23	919	1026	1.810	2.675	0.606	-1.097	457	800
36	08230500	39	81	73	1.358	1.739	0.401	0.015	48	57
36	08231000	46	83	71	2.094	1.783	0.354	0.121	51	56
37	08248000	87	1026	486	0.516	2.955	0.238	-0.807	807	379
37	08247500	66	342	217	0.757	2.427	0.342	-0.909	244	170
38	08340500	44	1326	906	1.359	3.025	0.308	-0.384	918	706
38	08334000	55	481	357	1.960	2.583	0.298	-0.169	320	279
39	08389500	25	495	741	3.340	2.430	0.456	0.639	162	578
39	08388000	25	412	837	3.742	2.201	0.554	0.736	35	652
40	09508500	60	15547	21183	2.431	3.835	0.604	-0.159	6014	16517
40	09508300	39	824	994	1.492	2.510	0.766	-1.306	377	775
41	10170490	63	1384	884	1.562	3.068	0.249	0.281	986	689
41	10172200	43	21	19	2.203	1.159	0.380	-0.158	12	15
42	10260500	91	2824	3440	1.708	3.050	0.691	-0.448	1276	2682
42	10261000	40	1537	2287	2.559	2.766	0.652	-0.008	508	1783
43	12082500	64	5175	2533	1.691	3.669	0.197	0.063	4035	1975
43	12083000	64	3831	1647	0.923	3.542	0.199	-0.657	3090	1284
44	12134500	78	32429	14983	1.031	4.466	0.202	-0.153	25686	11682
44	12147600	41	707	271	0.651	2.817	0.174	-0.401	586	211
45	12500500	52	378	187	1.928	2.528	0.220	-0.917	294	146
45	12501000	47	128	146	4.635	1.977	0.315	0.301	62	114
46	13011900	41	3756	1040	0.375	3.558	0.124	-0.297	3288	811
46	13011500	59	2341	739	0.601	3.348	0.139	-0.171	2008	576
47	14013000	68	821	523	2.362	2.850	0.228	0.556	586	408
47	14013500	32	237	151	1.759	2.306	0.243	0.315	169	118
48	14020300	34	2016	919	1.218	3.263	0.193	-0.050	1603	716
48	14020000	76	1885	921	1.516	3.232	0.191	0.402	1471	718

Gage Pair	Gage Number	n	Mean	Standard Deviation	Skew	Mean of Logs[1]	Std dev. of Logs[1]	Skew of Logs[1]	Gumbel Location	Gumbel Scale
49	14178000	80	6630	3307	1.545	3.773	0.207	-0.086	5142	2579
49	14179000	63	5115	2192	0.851	3.668	0.198	-0.572	4129	1709
51	14301500	75	13626	4834	0.873	4.107	0.159	-0.577	11451	3769
51	14302500	34	10155	2985	1.187	3.990	0.118	0.515	8812	2327
52	01547950	38	2461	1340	1.884	3.340	0.207	0.343	1858	1045
52	01547700	51	982	787	2.937	2.910	0.245	0.997	628	613
53	02113850	42	4115	2609	1.364	3.527	0.291	-0.348	2941	2034
53	02113000	75	3090	2005	1.790	3.408	0.272	-0.124	2188	1563
54	02126000	77	28967	13543	1.106	4.414	0.210	-0.326	22872	10559
54	02127000	34	2186	1954	4.178	3.249	0.272	0.188	1307	1523
55	02220900	29	4401	2601	1.002	3.566	0.276	-0.336	3231	2028
55	02221525	29	2923	1742	0.913	3.389	0.268	-0.100	2139	1358
56	02344500	69	5119	3833	3.022	3.616	0.290	-0.198	3395	2988
56	02344700	42	2790	2493	3.167	3.331	0.312	0.200	1668	1944
57	03173000	68	4849	2299	1.553	3.641	0.200	-0.236	3815	1792
57	03175500	75	4178	1841	0.479	3.575	0.211	-0.560	3350	1436
58	04114000	47	5576	2100	0.813	3.714	0.176	-0.828	4631	1637
58	04114500	52	1173	502	0.148	3.019	0.231	-1.194	947	391
59	04149000	55	6109	2942	1.324	3.736	0.219	-0.580	4784	2294
59	04145000	55	3355	1407	0.492	3.484	0.202	-0.577	2722	1097
60	04196500	62	4500	1695	0.252	3.620	0.178	-0.487	3737	1321
60	04196800	42	3458	1243	0.080	3.507	0.179	-0.816	2899	969
61	04234000	81	2542	1214	2.165	3.367	0.178	0.452	1995	947
61	04233000	69	631	344	1.191	2.740	0.230	0.052	476	268
62	05367500	48	5331	2595	1.032	3.673	0.229	-0.659	4163	2024
62	05368000	56	3336	2128	2.290	3.453	0.248	0.022	2379	1659
64	05500000	71	7728	3817	0.651	3.830	0.236	-0.468	6010	2976
64	05501000	71	9152	5947	1.097	3.859	0.323	-0.516	6475	4637
65	05594800	36	5707	3217	0.656	3.669	0.310	-0.903	4260	2509
65	05595200	37	3985	2679	2.321	3.526	0.257	-0.147	2779	2088
66	06447500	46	241	235	2.090	2.225	0.365	0.248	135	183
66	06448000	29	75	23	0.275	1.856	0.137	-0.111	65	18
67	05054500	51	173	178	2.047	2.002	0.507	-0.487	93	139
67	06467600	32	250	389	3.376	1.882	0.951	-1.597	75	303
68	06719505	32	1068	432	0.873	2.993	0.186	-0.681	873	336
68	06710500	88	311	272	1.844	2.357	0.339	0.290	188	212
69	06863500	60	1342	1863	3.199	2.788	0.616	-0.507	504	1452
69	06863900	24	458	942	3.562	1.913	0.975	-0.285	34	734
70	06871000	53	1241	1902	3.349	2.691	0.675	-0.483	384	1483
70	06871500	55	627	884	2.385	2.413	0.653	-0.441	229	689
71	07050000	38	31307	19412	1.417	4.423	0.252	0.149	22570	15135
71	07050500	50	13863	8645	1.011	4.058	0.282	-0.287	9973	6741
72	07075000	46	15129	12134	2.950	4.077	0.304	-0.103	9668	9460
72	07075300	38	7285	6857	3.622	3.748	0.307	0.331	4199	5346
73	07077500	59	6394	2678	0.589	3.768	0.183	0.014	5189	2088
73	07077700	46	3023	1112	1.121	3.453	0.155	-0.006	2522	867

Gage Pair	Gage Number	n	Mean	Standard Deviation	Skew	Mean of Logs[1]	Std dev. of Logs[1]	Skew of Logs[1]	Gumbel Location	Gumbel Scale
74	07261500	62	19268	12380	1.376	4.196	0.292	-0.378	13696	9653
74	07260000	37	3863	2133	0.918	3.519	0.256	-0.322	2903	1663
75	07346070	60	5799	5722	2.572	3.601	0.389	-0.281	3224	4461
75	07346045	38	3755	2656	1.011	3.457	0.348	-0.526	2559	2071
76	08164450	29	7561	8027	3.002	3.724	0.361	0.301	3948	6259
76	08164503	29	5376	4192	1.737	3.619	0.319	0.019	3489	3268
77	08164000	68	15130	17608	3.900	3.986	0.438	-0.597	7206	13729
77	08164600	36	4202	3082	0.899	3.464	0.440	-1.028	2815	2403
78	08195000	81	5156	8291	3.330	3.246	0.715	-0.273	1425	6464
78	08196000	54	1753	2298	2.544	2.855	0.679	-0.458	719	1791
79	09112500	84	2139	818	0.819	3.298	0.171	-0.420	1771	638
79	09113500	25	579	239	0.160	2.721	0.205	-0.695	471	187
80	09484600	32	413	408	3.064	2.469	0.364	-0.031	229	318
80	09485000	41	334	368	1.655	2.199	0.625	-0.503	168	287
81	10371500	60	1377	1104	1.822	3.019	0.329	-0.035	880	861
81	10370000	25	480	383	2.981	2.599	0.255	0.660	308	299
82	11383500	90	4452	3403	1.891	3.532	0.338	-0.590	2921	2653
82	11381500	78	3594	2800	1.817	3.440	0.328	-0.306	2334	2183
83	11401500	72	6454	6386	1.819	3.595	0.474	-0.443	3581	4979
83	11402000	73	4446	3537	1.334	3.488	0.420	-0.747	2854	2758
84	12167000	78	17610	6043	0.276	4.218	0.161	-0.496	14890	4711
84	12168500	26	2700	703	-0.242	3.415	0.126	-0.905	2384	548
85	12189500	78	26562	13421	1.315	4.377	0.201	0.306	20522	10464
85	12182500	51	5329	2106	0.899	3.696	0.164	0.251	4381	1642
86	12424000	57	5219	3388	1.440	3.617	0.333	-1.145	3695	2641
86	12431000	62	1289	635	1.082	3.055	0.231	-0.562	1003	495
87	13309000	32	1971	666	0.325	3.269	0.155	-0.364	1672	519
87	13308500	43	1556	497	0.302	3.169	0.146	-0.375	1333	387
88	13336500	77	24604	7365	0.595	4.372	0.128	0.106	21289	5742
88	13337000	79	18161	5503	0.647	4.240	0.130	0.025	15684	4291
89	14147500	60	6460	3383	1.342	3.756	0.221	-0.111	4938	2638
89	14146500	64	2952	1665	1.348	3.408	0.234	0.044	2203	1298
90	14154500	71	8410	3978	0.919	3.874	0.222	-0.544	6620	3101
90	14156500	34	3841	1807	0.930	3.535	0.222	-0.612	3028	1409

[1]Logarithms are base 10.

Table B.4. Summary Statistics for Concurrent Data: Peak on Main (POM).

Gage Pair	n	Pearson's ρ	p-value	Pearson's ρ of Logs[1]	p-value	Kendall's τ	p-value	Gumbel θ	GH θ	Frank θ
1	60	0.8385	0.0000	0.8509	0.0000	0.7073	0.0000	NA	3.4163	11.7526
2	28	0.4197	0.0262	0.6780	0.0001	0.4011	0.0028	0.6516	1.6696	4.1757
3	29	0.9512	0.0000	0.9097	0.0000	0.6734	0.0000	NA	3.0616	10.2888
4	31	0.9818	0.0000	0.8603	0.0000	0.6272	0.0000	NA	2.6821	8.7025
5	27	0.0017	0.9931	0.1632	0.4161	0.1097	0.4274	0.0029	1.1232	0.9968
6	43	0.6388	0.0000	0.6764	0.0000	0.5181	0.0000	0.9620	2.0750	6.0746
8	10	0.9057	0.0003	0.8870	0.0006	0.8222	0.0004	NA	5.6250	20.7132
9	26	0.8433	0.0000	0.8501	0.0000	0.7211	0.0000	NA	3.5856	12.4472
11	29	0.1268	0.5120	0.1791	0.3527	0.2172	0.1019	0.2050	1.2775	2.0337
12	39	0.2037	0.2135	0.3637	0.0229	0.2260	0.0433	0.3259	1.2920	2.1227
13	62	0.4004	0.0013	0.5025	0.0000	0.3530	0.0001	0.6233	1.5455	3.5458
14	32	0.8003	0.0000	0.8386	0.0000	0.6572	0.0000	NA	2.9174	9.6893
15	54	0.9135	0.0000	0.7376	0.0000	0.7407	0.0000	NA	3.8568	13.5550
16	17	0.3104	0.2253	0.4045	0.1073	0.3529	0.0518	0.4892	1.5455	3.5454
17	32	0.9408	0.0000	0.9209	0.0000	0.8052	0.0000	NA	5.1347	18.7356
18	53	0.7976	0.0000	0.7500	0.0000	0.5180	0.0000	NA	2.0748	6.0741
19	28	0.7821	0.0000	0.7743	0.0000	0.5748	0.0000	NA	2.3520	7.2936
20	34	0.4489	0.0077	0.6456	0.0000	0.5131	0.0000	0.6941	2.0538	5.9794
21	27	0.8103	0.0000	0.7647	0.0000	0.4499	0.0011	NA	1.8177	4.8920
22	21	0.9018	0.0000	0.9556	0.0000	0.8162	0.0000	NA	5.4416	19.9740
23	57	0.5318	0.0000	0.6955	0.0000	0.5217	0.0000	0.8129	2.0907	6.1450
24	42	0.6561	0.0000	0.8766	0.0000	0.6655	0.0000	0.9856	2.9896	9.9898
25	39	0.4338	0.0058	0.6216	0.0000	0.4487	0.0001	0.6721	1.8138	4.8735
26	26	0.7059	0.0001	0.8130	0.0000	0.6739	0.0000	NA	3.0664	10.3087
27	37	0.6997	0.0000	0.7719	0.0000	0.5455	0.0000	NA	2.2000	6.6301
28	25	0.5818	0.0023	0.6314	0.0007	0.4908	0.0006	0.8831	1.9639	5.5719
29	67	0.9332	0.0000	0.8829	0.0000	0.7230	0.0000	NA	3.6105	12.5489
31	50	0.3218	0.0227	0.3278	0.0201	0.2061	0.0357	0.5064	1.2596	1.9215
32	32	0.9112	0.0000	0.9183	0.0000	0.8093	0.0000	NA	5.2434	19.1743
33	28	0.0029	0.9882	0.4379	0.0198	0.2819	0.0375	0.0048	1.3926	2.7153
34	52	0.6679	0.0000	0.8102	0.0000	0.6918	0.0000	NA	3.2451	11.0480
35	27	0.2628	0.1854	0.1642	0.4130	0.2271	0.1018	0.4170	1.2939	2.1345
36	39	0.6614	0.0000	0.6670	0.0000	0.5231	0.0000	0.9928	2.0969	6.1727
37	83	0.8383	0.0000	0.8581	0.0000	0.6646	0.0000	NA	2.9816	9.9565
38	36	0.3262	0.0522	0.4150	0.0118	0.3203	0.0062	0.5130	1.4713	3.1515
39	25	0.9220	0.0000	0.6830	0.0002	0.3880	0.0067	NA	1.6339	3.9975
40	38	0.8349	0.0000	0.7768	0.0000	0.5712	0.0000	NA	2.3322	7.2078
41	43	0.6797	0.0000	0.5975	0.0000	0.3933	0.0002	NA	1.6482	4.0692
42	42	0.9240	0.0000	0.8721	0.0000	0.7802	0.0000	NA	4.5503	16.3724
43	64	0.7243	0.0000	0.7049	0.0000	0.5036	0.0000	NA	2.0145	5.8022
44	43	0.5514	0.0001	0.6850	0.0000	0.4126	0.0001	0.8405	1.7025	4.3379
45	48	0.9247	0.0000	0.9360	0.0000	0.7123	0.0000	NA	3.4761	11.9981
46	38	0.7311	0.0000	0.7151	0.0000	0.5096	0.0000	NA	2.0392	5.9137
47	32	0.8752	0.0000	0.8105	0.0000	0.5248	0.0000	NA	2.1045	6.2067
48	34	0.9285	0.0000	0.9230	0.0000	0.7375	0.0000	NA	3.8095	13.3669

Gage Pair	n	Pearson's ρ	p-value	Pearson's ρ of Logs[1]	p-value	Kendall's τ	p-value	Gumbel θ	GH θ	Frank θ
49	63	0.9531	0.0000	0.9490	0.0000	0.7872	0.0000	NA	4.6988	16.9737
51	35	0.9187	0.0000	0.8962	0.0000	0.7055	0.0000	NA	3.3954	11.6669
52	38	0.8215	0.0000	0.7784	0.0000	0.5969	0.0000	NA	2.4806	7.8466
53	42	0.9379	0.0000	0.9119	0.0000	0.7477	0.0000	NA	3.9631	13.9884
54	34	0.8685	0.0000	0.7362	0.0000	0.5323	0.0000	NA	2.1379	6.3557
55	29	0.8662	0.0000	0.9103	0.0000	0.7537	0.0000	NA	4.0600	14.3827
56	42	0.8661	0.0000	0.8683	0.0000	0.6833	0.0000	NA	3.1578	10.6874
57	67	0.7613	0.0000	0.7556	0.0000	0.5177	0.0000	NA	2.0732	6.0669
58	37	0.6351	0.0000	0.7896	0.0000	0.5403	0.0000	0.9568	2.1751	6.5204
59	54	0.7884	0.0000	0.8400	0.0000	0.6187	0.0000	NA	2.6225	8.4505
60	22	0.5779	0.0048	0.3581	0.1017	0.4946	0.0013	0.8777	1.9785	5.6386
61	69	0.7709	0.0000	0.7657	0.0000	0.6080	0.0000	NA	2.5512	8.1479
62	12	0.8219	0.0010	0.8353	0.0007	0.5152	0.0210	NA	2.0625	6.0187
64	71	0.3518	0.0026	0.4080	0.0004	0.2741	0.0007	0.5513	1.3777	2.6303
65	36	0.4717	0.0037	0.5474	0.0006	0.3556	0.0020	0.7270	1.5517	3.5780
66	29	0.6198	0.0003	0.4951	0.0063	0.2707	0.0422	0.9358	1.3712	2.5931
67	37	0.8235	0.0000	0.8102	0.0000	0.6100	0.0000	NA	2.5642	8.2034
68	32	0.8085	0.0000	0.7228	0.0000	0.5561	0.0000	NA	2.2528	6.8620
69	26	0.7229	0.0000	0.5210	0.0064	0.3399	0.0190	NA	1.5150	3.3854
70	54	0.6190	0.0000	0.6662	0.0000	0.4533	0.0000	0.9347	1.8292	4.9461
71	20	0.6418	0.0023	0.6525	0.0018	0.4908	0.0025	0.9661	1.9637	5.5710
72	24	0.9717	0.0000	0.9430	0.0000	0.7877	0.0000	NA	4.7094	17.0168
73	46	0.6213	0.0000	0.5117	0.0003	0.5165	0.0000	0.9379	2.0681	6.0440
74	37	0.8175	0.0000	0.8332	0.0000	0.6315	0.0000	NA	2.7137	8.8359
75	38	0.8158	0.0000	0.8734	0.0000	0.7137	0.0000	NA	3.4925	12.0657
76	29	0.5530	0.0019	0.7825	0.0000	0.6601	0.0000	0.8428	2.9420	9.7919
77	36	0.6474	0.0000	0.4624	0.0045	0.3129	0.0073	0.9738	1.4555	3.0656
78	54	0.8364	0.0000	0.6936	0.0000	0.5219	0.0000	NA	2.0914	6.1484
79	26	0.9020	0.0000	0.8997	0.0000	0.7562	0.0000	NA	4.1013	14.5507
80	31	0.1606	0.3880	0.2922	0.1106	0.2244	0.0879	0.2584	1.2894	2.1070
81	23	0.8966	0.0000	0.8888	0.0000	0.7312	0.0000	NA	3.7206	12.9992
82	78	0.9654	0.0000	0.9744	0.0000	0.8560	0.0000	NA	6.9468	26.0311
83	60	0.9164	0.0000	0.8936	0.0000	0.7078	0.0000	NA	3.4221	11.7764
84	30	0.6065	0.0004	0.6625	0.0001	0.4606	0.0004	0.9174	1.8541	5.0633
85	51	0.7885	0.0000	0.7888	0.0000	0.5632	0.0000	NA	2.2896	7.0224
86	57	0.3339	0.0111	0.3803	0.0035	0.2459	0.0070	0.5245	1.3261	2.3291
87	32	0.9321	0.0000	0.9309	0.0000	0.7842	0.0000	NA	4.6338	16.7107
88	77	0.9552	0.0000	0.9494	0.0000	0.7976	0.0000	NA	4.9406	17.9517
89	59	0.9547	0.0000	0.9612	0.0000	0.8338	0.0000	NA	6.0159	22.2875
90	34	0.9485	0.0000	0.9593	0.0000	0.7982	0.0000	NA	4.9558	18.0128

[1]Logarithm base 10.

Table B.5. Summary Statistics for Concurrent Data: Peak on Tributary (POT).

Gage Pair	n	Pearson's ρ	p-value	Pearson's ρ of Logs[1]	p-value	Kendall's τ	p-value	Gumbel θ	GH θ	Frank θ
1	60	0.8187	0.0000	0.8719	0.0000	0.6833	0.0000	NA	3.1573	10.6853
2	27	0.6368	0.0004	0.5980	0.0010	0.4006	0.0035	0.9592	1.6683	4.1690
3	28	0.9550	0.0000	0.8579	0.0000	0.6453	0.0000	NA	2.8192	9.2790
4	31	0.9786	0.0000	0.8777	0.0000	0.6127	0.0000	NA	2.5822	8.2796
5	27	0.2443	0.2194	0.1944	0.3311	0.0856	0.5316	0.3886	1.0936	0.7749
6	42	0.7786	0.0000	0.7529	0.0000	0.5473	0.0000	NA	2.2088	6.6687
8	10	0.8690	0.0011	0.7923	0.0063	0.5556	0.0286	NA	2.2500	6.8496
9	26	0.8202	0.0000	0.7789	0.0000	0.6410	0.0000	NA	2.7854	9.1372
11	29	0.6329	0.0002	0.6761	0.0001	0.5347	0.0001	0.9539	2.1490	6.4046
12	39	0.3893	0.0143	0.5663	0.0002	0.4165	0.0002	0.6069	1.7138	4.3929
13	62	0.5765	0.0000	0.6483	0.0000	0.4669	0.0000	0.8757	1.8757	5.1646
14	32	0.8460	0.0000	0.8798	0.0000	0.7313	0.0000	NA	3.7218	13.0042
15	54	0.9101	0.0000	0.8989	0.0000	0.7243	0.0000	NA	3.6277	12.6194
16	17	0.6993	0.0018	0.7537	0.0005	0.5441	0.0018	NA	2.1935	6.6017
17	32	0.9417	0.0000	0.8894	0.0000	0.7947	0.0000	NA	4.8719	17.6740
18	51	0.8325	0.0000	0.7946	0.0000	0.6187	0.0000	NA	2.6227	8.4514
19	27	0.8624	0.0000	0.8611	0.0000	0.6486	0.0000	NA	2.8455	9.3891
20	33	0.3892	0.0252	0.4951	0.0034	0.2765	0.0245	0.6068	1.3821	2.6556
21	27	0.3751	0.0539	0.4411	0.0213	0.2446	0.0785	0.5859	1.3238	2.3152
22	20	0.8657	0.0000	0.9151	0.0000	0.7474	0.0000	NA	3.9583	13.9689
23	60	0.5883	0.0000	0.5735	0.0000	0.4191	0.0000	0.8922	1.7213	4.4296
24	42	0.7811	0.0000	0.8131	0.0000	0.6081	0.0000	NA	2.5519	8.1511
25	39	0.7517	0.0000	0.7348	0.0000	0.5643	0.0000	NA	2.2950	7.0461
26	26	0.6958	0.0001	0.7736	0.0000	0.5323	0.0001	NA	2.1382	6.3566
27	37	0.8282	0.0000	0.7766	0.0000	0.6827	0.0000	NA	3.1517	10.6620
28	25	0.8441	0.0000	0.8049	0.0000	0.5109	0.0004	NA	2.0444	5.9371
29	68	0.9181	0.0000	0.8880	0.0000	0.7245	0.0000	NA	3.6294	12.6263
31	50	0.3875	0.0054	0.4123	0.0029	0.2619	0.0076	0.6042	1.3548	2.4983
32	32	0.7361	0.0000	0.9072	0.0000	0.7742	0.0000	NA	4.4286	15.8793
33	27	0.4992	0.0080	0.7038	0.0000	0.5726	0.0000	0.7665	2.3400	7.2415
34	53	0.7177	0.0000	0.8850	0.0000	0.7408	0.0000	NA	3.8583	13.5614
35	25	0.7788	0.0000	0.7847	0.0000	0.4975	0.0005	NA	1.9900	5.6910
36	38	0.8316	0.0000	0.8896	0.0000	0.7121	0.0000	NA	3.4738	11.9888
37	66	0.8423	0.0000	0.8511	0.0000	0.6808	0.0000	NA	3.1329	10.5846
38	36	0.6197	0.0001	0.7008	0.0000	0.4960	0.0000	0.9356	1.9842	5.6645
39	25	0.9124	0.0000	0.7389	0.0000	0.5109	0.0004	NA	2.0444	5.9371
40	39	0.8582	0.0000	0.7533	0.0000	0.6725	0.0000	NA	3.0536	10.2558
41	43	0.7118	0.0000	0.6587	0.0000	0.4606	0.0000	NA	1.8539	5.0626
42	41	0.8747	0.0000	0.8470	0.0000	0.6634	0.0000	NA	2.9710	9.9126
43	64	0.7226	0.0000	0.7524	0.0000	0.5341	0.0000	NA	2.1462	6.3924
44	41	0.5854	0.0001	0.6272	0.0000	0.4279	0.0001	0.8881	1.7479	4.5582
45	47	0.7600	0.0000	0.8235	0.0000	0.5383	0.0000	NA	2.1658	6.4793
46	38	0.5216	0.0008	0.5819	0.0001	0.3695	0.0011	0.7984	1.5860	3.7546
47	32	0.8328	0.0000	0.8012	0.0000	0.5466	0.0000	NA	2.2054	6.6537
48	34	0.8797	0.0000	0.8537	0.0000	0.6434	0.0000	NA	2.8045	9.2232

Gage Pair	n	Pearson's ρ	p-value	Pearson's ρ of Logs[1]	p-value	Kendall's τ	p-value	Gumbel θ	GH θ	Frank θ
49	63	0.8997	0.0000	0.9008	0.0000	0.7290	0.0000	NA	3.6900	12.8741
51	34	0.8900	0.0000	0.8499	0.0000	0.5970	0.0000	NA	2.4812	7.8491
52	38	0.8222	0.0000	0.7124	0.0000	0.4590	0.0001	NA	1.8485	5.0370
53	43	0.8962	0.0000	0.8768	0.0000	0.6885	0.0000	NA	3.2100	10.9030
54	34	0.8622	0.0000	0.7034	0.0000	0.5040	0.0000	NA	2.0162	5.8099
55	29	0.9330	0.0000	0.8984	0.0000	0.7586	0.0000	NA	4.1429	14.7196
56	42	0.7612	0.0000	0.7470	0.0000	0.5492	0.0000	NA	2.2181	6.7095
57	67	0.7856	0.0000	0.8237	0.0000	0.5898	0.0000	NA	2.4379	7.6636
58	38	0.6242	0.0000	0.7599	0.0000	0.4608	0.0000	0.9418	1.8545	5.0652
59	53	0.8247	0.0000	0.8035	0.0000	0.6254	0.0000	NA	2.6696	8.6497
60	23	0.5977	0.0026	0.6812	0.0003	0.4941	0.0007	0.9052	1.9766	5.6296
61	69	0.6859	0.0000	0.6361	0.0000	0.4921	0.0000	NA	1.9689	5.5945
62	12	0.6791	0.0152	0.5711	0.0524	0.6462	0.0038	NA	2.8267	9.3103
64	71	0.7496	0.0000	0.6842	0.0000	0.5876	0.0000	NA	2.4248	7.6074
65	36	0.7397	0.0000	0.7501	0.0000	0.4610	0.0001	NA	1.8555	5.0697
66	29	0.3607	0.0545	0.1633	0.3974	0.0842	0.5233	0.5646	1.0919	0.7618
67	34	0.3609	0.0360	0.7686	0.0000	0.5632	0.0000	0.5648	2.2895	7.0222
68	32	0.5434	0.0013	0.5537	0.0010	0.3215	0.0099	0.8292	1.4739	3.1658
69	24	0.3461	0.0976	0.6035	0.0018	0.4319	0.0032	0.5429	1.7604	4.6185
70	54	0.7361	0.0000	0.6830	0.0000	0.4906	0.0000	NA	1.9629	5.5672
71	19	0.9083	0.0000	0.8074	0.0000	0.6804	0.0000	NA	3.1285	10.5661
72	24	0.9763	0.0000	0.8817	0.0000	0.7731	0.0000	NA	4.4080	15.7959
73	46	0.7887	0.0000	0.7500	0.0000	0.5804	0.0000	NA	2.3834	7.4290
74	37	0.7354	0.0000	0.6673	0.0000	0.5723	0.0000	NA	2.3380	7.2330
75	38	0.7501	0.0000	0.7268	0.0000	0.5647	0.0000	NA	2.2974	7.0564
76	29	0.8317	0.0000	0.6856	0.0000	0.4840	0.0002	NA	1.9378	5.4520
77	36	0.5950	0.0001	0.6341	0.0000	0.5226	0.0000	0.9015	2.0948	6.1636
78	54	0.8569	0.0000	0.7796	0.0000	0.6254	0.0000	NA	2.6695	8.6492
79	25	0.7752	0.0000	0.6754	0.0002	0.6198	0.0000	NA	2.6300	8.4824
80	32	0.4725	0.0063	0.4055	0.0213	0.3192	0.0104	0.7281	1.4688	3.1383
81	23	0.9282	0.0000	0.9421	0.0000	0.8554	0.0000	NA	6.9179	25.9151
82	78	0.9586	0.0000	0.9606	0.0000	0.8479	0.0000	NA	6.5728	24.5282
83	60	0.8414	0.0000	0.8919	0.0000	0.7311	0.0000	NA	3.7192	12.9937
84	26	0.3778	0.0571	0.3923	0.0475	0.3117	0.0259	0.5899	1.4529	3.0516
85	51	0.8301	0.0000	0.8318	0.0000	0.6779	0.0000	NA	3.1049	10.4684
86	59	0.5383	0.0000	0.6358	0.0000	0.4295	0.0000	0.8220	1.7528	4.5822
87	32	0.9427	0.0000	0.9424	0.0000	0.7830	0.0000	NA	4.6075	16.6042
88	78	0.9090	0.0000	0.9030	0.0000	0.7455	0.0000	NA	3.9297	13.8520
89	58	0.9577	0.0000	0.9551	0.0000	0.8326	0.0000	NA	5.9747	22.1216
90	34	0.9534	0.0000	0.9610	0.0000	0.7957	0.0000	NA	4.8952	17.7681

[1]Logarithm base 10.

Appendix C. Bivariate Distributions and Copulas

Table of Contents

List of Tables ..... **C-ii**

List of Figures ..... **C-iii**

C.1 Univariate Distributions C-2

C.1.1 Normal Distribution C-2

C.1.2 Generalized Extreme Value Distributions C-3

C.1.3 Exponential Distribution C-4

C.2 Bivariate Distributions C-4

C.2.1 Bivariate Gumbel Distributions C-4

C.2.2 Bivariate Normal Distribution C-5

C.3 Copulas C-6

C.3.1 Gumbel-Hougaard Copula C-6

C.3.2 Frank Copula C-7

C.4 Empirical Distributions and Plotting Position C-7

C.4.1 Univariate Plotting Position C-7

C.4.2 Bivariate Empirical Distribution Functions C-8

C.5 Correlation Measures C-9

C.5.1 Pearson's ρ C-9

C.5.2 Kendall's τ C-10

C.6 Prototype Analyses C-12

C.6.1 Fitting Distributions C-12

C.6.2 Joint Distributions C-12

C.6.3 Exceedance Probability C-12

C.7 Correlation Parameter Regression C-29

C.7.1 Data Adjustments C-30

C.7.2 Regression on Area Ratio C-31

C.7.2.1 All Data C-31

C.7.2.2 Filters C-34

C.7.2.3 Functional Adjustments C-37

C.7.2.4 Groups C-38

C.7.3 Regression on Other/Multiple Variables C-39

C.7.4 Alternative Equation Forms C-42
C.7.5 Selection of Best-Fit Equations C-43
C.7.6 Selection of Envelope Equations C-45
C.8 Correlation Parameter Groupings C-46
C.8.1 Site-Specific Dependence C-46
C.8.2 Potential Combinations C-55
C.9 Comparison of Regression and Grouping C-58

List of Tables

Table C.1. Example Peak Flow Data C-10
Table C.2. Computation for Summation Terms C-10
Table C.3. Determination of Concordant and Discordant Pairs C-11
Table C.4a. 10-percent Exceedance Combinations for Gage Pair 05 (POM) C-23
Table C.4b. 10-percent Exceedance Combinations for Gage Pair 05 (POT) C-23
Table C.5a. 10-percent Exceedance Combinations for Gage Pair 06 (POM) C-25
Table C.5b. 10-percent Exceedance Combinations for Gage Pair 06 (POT) C-25
Table C.6a. 10-percent Exceedance Combinations for Gage Pair 43 (POM) C-27
Table C.6b. 10-percent Exceedance Combinations for Gage Pair 43 (POT) C-27
Table C.7a. 10-percent Exceedance Combinations for Gage Pair 76 (POM) C-29
Table C.7b. 10-percent Exceedance Combinations for Gage Pair 76 (POT) C-29
Table C.8. Correlation Adjustments: Peak on Main C-30
Table C.9. Correlation Adjustments: Peak on Tributary C-31
Table C.10. Summary of Selected Regressions on Area Ratio C-37
Table C.11a. Regression Scenarios for Pearson's ρ (POM) C-40
Table C.11b. Regression Scenarios for Kendall's τ (POM) C-40
Table C.11c. Regression Scenarios for Pearson's ρ (POT) C-41
Table C.11d. Regression Scenarios for Kendall's τ (POT) C-42
Table C.12. Equations for Pearson's ρ (Scenarios 8 and 11) C-43
Table C.13. Equations for Kendall's τ (Scenarios 8 and 11) C-43
Table C.14. Equations for Pearson's ρ (Scenario 12) C-44
Table C.15. Equations for Kendall's τ (Scenario 12) C-44
Table C.16. Equations for Pearson's ρ (Scenario 14) C-44
Table C.17. Equations for Kendall's τ (Scenario 14) C-44

Table C.18. Envelope Constants C-46
Table C.19. Pearson's ρ (POM) C-54
Table C.20. Pearson's ρ (POT) C-54
Table C.21. Kendall's τ (POM) C-54
Table C.22. Kendall's τ (POT) C-54
Table C.23. t-statistic for Pearson's ρ (POM) C-55
Table C.24. t-statistic for Pearson's ρ (POT) C-55
Table C.25. t-statistic for Kendall's τ (POM) C-55
Table C.26. t-statistic for Kendall's τ (POT) C-55
Table C.27. 10-yr Joint Return Periods (Groupings) C-56
Table C.28. 25-yr Joint Return Periods (Groupings) C-56
Table C.29. 50-yr Joint Return Periods (Groupings) C-56
Table C.30. 100-yr Joint Return Periods (Groupings) C-57
Table C.31. 500-yr Joint Return Periods (Groupings) C-57
Table C.32. Prediction Error Comparison C-58

List of Figures

Figure C.1a. Fit for Gage Pair 05 (POM) C-13
Figure C.1b. Fit for Gage Pair 05 (POT) C-13
Figure C.2a. Fit for Gage Pair 06 (POM) C-14
Figure C.2b. Fit for Gage Pair 06 (POT) C-14
Figure C.3a. Fit for Gage Pair 43 (POM) C-15
Figure C.3b. Fit for Gage Pair 43 (POT) C-15
Figure C.4a. Fit for Gage Pair 76 (POM) C-16
Figure C.4b. Fit for Gage Pair 76 (POT) C-16
Figure C.5a. Joint Distribution for Gage Pair 05 (POM) C-17
Figure C.5b. Joint Distribution for Gage Pair 05 (POT) C-17
Figure C.6a. Joint Distribution for Gage Pair 06 (POM) C-18
Figure C.6b. Joint Distribution for Gage Pair 06 (POT) C-18
Figure C.7a. Joint Distribution for Gage Pair 43 (POM) C-19
Figure C.7b. Joint Distribution for Gage Pair 43 (POT) C-19
Figure C.8a. Joint Distribution for Gage Pair 76 (POM) C-20
Figure C.8b. Joint Distribution for Gage Pair 76 (POT) C-20

Figure C.9a. 10-percent Exceedance Probability for Gage Pair 05 (POM) C-22
Figure C.9b. 10-percent Exceedance Probability for Gage Pair 05 (POT) C-22
Figure C.10a. 10-percent Exceedance Probability for Gage Pair 06 (POM) C-24
Figure C.10b. 10-percent Exceedance Probability for Gage Pair 06 (POT) C-24
Figure C.11a. 10-percent Exceedance Probability for Gage Pair 43 (POM) C-26
Figure C.11b. 10-percent Exceedance Probability for Gage Pair 43 (POT) C-26
Figure C.12a. 10-percent Exceedance Probability for Gage Pair 76 (POM) C-28
Figure C.12b. 10-percent Exceedance Probability for Gage Pair 76 (POT) C-28
Figure C.13a. Pearson's ρ versus Area Ratio (POM)..... C-32
Figure C.13b. Kendall's τ versus Area Ratio (POM)..... C-32
Figure C.13c. Pearson's ρ versus Area Ratio (POT)..... C-33
Figure C.13d. Kendall's τ versus Area Ratio (POT) C-33
Figure C.14a. Pearson's ρ versus Total Area (POM) C-34
Figure C.14b. Kendall's τ versus Total Area (POM)..... C-35
Figure C.14c. Pearson's ρ versus Total Area (POT) C-35
Figure C.14d. Kendall's τ versus Total Area (POT)..... C-36
Figure C.15a. Distribution of Pearson's ρ by Area Ratio (POM)..... C-48
Figure C.15b. Distribution of Pearson's ρ by Area Ratio (POT) C-48
Figure C.16a. Distribution of Kendall's τ by Area Ratio (POM)..... C-49
Figure C.16b. Distribution of Kendall's τ by Area Ratio (POT) C-49
Figure C.17a. Distribution of Pearson's ρ by Total Area (POM)..... C-50
Figure C.17b. Distribution of Pearson's ρ by Total Area (POT)..... C-50
Figure C.18a. Distribution of Kendall's τ by Total Area (POM)..... C-51
Figure C.18b. Distribution of Kendall's τ by Total Area (POT)..... C-51
Figure C.19a. Distribution of Pearson's ρ by Mean Annual Precipitation (POM)..... C-52
Figure C.19b. Distribution of Pearson's ρ by Mean Annual Precipitation (POT)..... C-52
Figure C.20a. Distribution of Kendall's τ by Mean Annual Precipitation (POM)..... C-53
Figure C.20b. Distribution of Kendall's τ by Mean Annual Precipitation (POT)..... C-53

Most statistical analyses of hydrologic variables apply univariate (single-valued) random variables, for example, analyses of the frequency distribution of annual peak discharges from a watershed or the distribution of rainfall depth over a fixed duration at a rain gage. However, a number of hydrologic problems involve two or more random variables. In those cases, either a simplying assumption is required to reduce the complexity to something treatable with univariate statistics, or more complex analyses involving multivariate statistics are required.

Salvadori, et al. (2007) discussed the multiple cases that can arise in hydrologic applications of multivariate statistics. First, consider the expression for an extreme event on one of the confluent streams (the marginal event) where "extreme" is defined as exceeding a threshold value.

$$E^{>}_{X,x} = \{X > x\} \tag{C.1}$$

where,

$E^{>}_{X,x}$	=	event where the occurrence, X, is greater than a threshold value x
X	=	magnitude of the event
x	=	threshold value for defining an event

Similarly, on the confluent stream:

$$E^{>}_{Y,y} = \{Y > y\} \tag{C.2}$$

where,

$E^{>}_{Y,y}$	=	event where the occurrence, Y, is greater than a threshold value y
Y	=	magnitude of the event
y	=	threshold value for defining an event

These expressions represent an event where exceedance of a threshold value occurs on each stream individually. Depending on the objective of the analysis, one may be interested in an extreme event occuring on X and Y simultaneously or, alternatively, an extreme event occuring on X or Y. The coincident extreme events for the "and" and "or" formulation, respectively, are:

$$E^{\cap}_{x,y} = \{X > x\} \cap \{Y > y\} \tag{C.3}$$

$$E^{\cup}_{x,y} = \{X > x\} \cup \{Y > y\} \tag{C.4}$$

The probability of the coincident extreme events defined in Equations C.3 and C.4 are defined as follows:

$$P^{\cap}_{x,y} = P(E^{\cap}_{x,y}) = P(X > x \cap Y > y) \tag{C.5}$$

$$P^{\cup}_{x,y} = P(E^{\cup}_{x,y}) = P(X > x \cup Y > y) \tag{C.6}$$

Let $H(x,y)$ denote a joint bivariate distribution of two random variables, X and Y. The joint probability distribution function is

$$H(x,y) = P[X < x, Y < y] \tag{C.7}$$

If the marginal distributions of the random variables X and Y are given by $F_X(x)$ and $F_Y(y)$, then

$$P_{x,y}^{\cap} = 1 - F_X(x) - F_Y(y) + H(x,y) \tag{C.8}$$

$$P_{x,y}^{\cup} = 1 - H(x,y) \tag{C.9}$$

Equation C.8 expresses exceedance probability with the "and" formulation while Equation C.9 expresses exceedance probability with the "or" formulation. As is apparent from these equations, the two formulations are related by the joint probability distribution H(x,y). Salvadori, et al. (2007) and Shiau, et al. (2006) discuss use of these probability statements in substantial detail. Shiau, et al. (2006) prefers to express Equation C.9 as $1 - P_{x,y}^{\cup}$ in order to express nonexceedance rather than exceedance probability. The "and" formulation, as expressed in Equation C.8 is applicable to the objectives of this research.

C.1 Univariate Distributions

A number of univariate distributions are used in hydrologic statistics. A few of them are useful as marginal distributions for bivariate hydrologic random variables and are presented here.

C.1.1 Normal Distribution

The normal distribution has a probability density function (PDF) given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left[-\frac{1}{2}\left(\frac{x - \mu_x}{\sigma_x} \right)^2 \right] \tag{C.10}$$

where,

x	=	the variate
μ_x	=	mean
σ_x	=	standard deviation

The distribution function, sometimes called the cumulative distribution function (CDF), is the integral of the density function,

$$F_{\mu_x,\sigma_x^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left[-\frac{(u - \mu_x)^2}{2\sigma_x^2} \right] du \tag{C.11}$$

There is no closed-form solution to the integral in Equation C.11. Some references express the distribution function in terms of the error function, but nonexceedance probabilities are derived from the normal distribution using numerical methods. Therefore, determination of nonexceedance probabilities is accomplished by numerical integration of the density function.

A concept used frequently in statistics is the *standard normal variate*. The standard normal variate is computed by scaling the observations by the mean and standard deviation to result in a normal distribution with mean zero and standard deviation unity. Mathematically,

$$Z = \frac{x - \mu_x}{\sigma_x}, \tag{C.12}$$

where,

Z	=	the standard normal deviate
x	=	the observation
μ_x	=	the mean of x (location parameter)
σ_x	=	the standard deviation of x (scale parameter)

This concept is useful because an analogue, the reduced variate, is used for extreme value distributions, which are discussed in the following section.

Because hydrologic data are often skewed, the normal distribution cannot be applied generally to untransformed values of discharge (or other hydrologic variables). However, the transform $y = \ln x$ (or a logarithm of another base) often removes enough skew in the dataset that the normal distribution is appropriate for fitting hydrologic variates. A normal distribution fit to the log-transformed values of a hydrologic random variable is referred to as the *log-normal* distribution.

C.1.2 Generalized Extreme Value Distributions

A family of distributions is based on the Generalized Extreme Value (GEV) distribution,

$$F(x) = \exp\left[-\left(1 - k\frac{x-\beta}{\alpha}\right)^{1/k}\right] \tag{C.13}$$

where,

$F(x)$	=	the distribution function of a random variable x ($-\infty \le x \le \infty$)
β	=	location parameter
α	=	scale parameter
k	=	shape parameter

The GEV is sub-categorized depending on the value of k. The Extreme Value Type I (EVI) distribution (the Gumbel distribution) arises when k approaches zero (Coles, 2001).

$$F(x) = \exp\left[-\exp\left(-\frac{x-\beta}{\alpha}\right)\right] \tag{C.14}$$

α and β are distribution parameters given by

$$\alpha = \frac{\sqrt{6}\sigma}{\pi} \tag{C.15}$$

and

$$\beta = \mu - 0.5772\alpha \tag{C.16}$$

where μ is the mean of the variate and σ is the standard deviation. The parameters β and α are the location and scale parameters, respectively.

The reduced variate (analogous to the standard normal deviate) is

$$y = \frac{x - \beta}{\alpha} \tag{C.17}$$

Quantiles can be extracted from the Gumbel distribution using the quantile function,

$$y = -\ln\left[\ln\left(\frac{1}{F(x)}\right)\right] \tag{C.18}$$

where F(x) is a vector of nonexceedance probabilities of interest. Given the quantiles, Equations C.17 and C.18 can be used to estimate discharges of interest.

C.1.3 Exponential Distribution

The exponential distribution arises as the marginal distribution of one of the Gumbel bivariate distributions. The density function of the exponential distribution is

$$f(x) = \lambda \exp(-\lambda x) \tag{C.19}$$

where x is the random variable that must be greater than or equal to zero and λ is the distribution parameter. The distribution function for the exponential distribution is

$$F(x) = 1 - \exp(-\lambda x) \tag{C.20}$$

C.2 Bivariate Distributions

Bivariate distributions are a subset of multivariate distributions involving two random variables. A bivariate distribution is characterized by a distribution function, which is a scalar-valued function of a vector-valued random variable (in this case two random variables). In the limit of each random variable, a univariate distribution function results, termed the marginal distribution. So, for a distribution function $H(x,y)$ there are two marginal distributions, $F(x)$ and $G(y)$. Most bivariate distributions have the same type of marginal distributions, that is, the same distribution function although with different parameter values. The bivariate normal and Gumbel distributions are discussed in the following sections.

C.2.1 Bivariate Gumbel Distributions

The univariate Gumbel distribution has multivariate extensions. Yue et al. (1999) presents results from application of the bivariate mixed-model Gumbel distribution to hydrologic problems. The bivariate mixed-model Gumbel distribution is given by

$$H(x,y) = F(x)F(y)\exp\left\{-\theta\left[\frac{1}{\ln F(x)} + \frac{1}{\ln F(y)}\right]^{-1}\right\} \tag{C.21}$$

where $F(x)$ and $F(y)$ are marginal Gumbel distributions and θ is a dependence parameter. The marginal distribution, F(x) and F(y), were given previously in Equation C.14. The dependence parameter is a function of Pearson's ρ,

$$\theta = 2\left[1 - \cos\left(\pi\sqrt{\frac{\rho}{6}}\right)\right] \tag{C.22}$$

The dependence parameter, θ, reaches unity when ρ is 2/3, indicating complete dependence in the Gumbel model. For values of ρ exceeding 2/3, the bivariate mixed-model Gumbel distribution does not apply because the dependence parameter is undefined.

Because the maximum value for ρ is 2/3 in the context of the bivariate Gumbel mixed model, an alternative Gumbel distribution was used by Yue et al. (2001) called the *bivariate logistic* distribution. The bivariate Gumbel logistic distribution is

$$H(x,y) = \exp\left\{-\left[\left(-\ln F(x)\right)^m + \left(-\ln F(y)\right)^m\right]^{1/m}\right\} \tag{C.23}$$

Similar to the bivariate Gumbel mixed model, the marginal distributions for the bivariate Gumbel logistic distribution, $F(x)$ and $F(y)$, are the Gumbel univariate distributions. The dependence parameter, m, is given by

$$m = \frac{1}{\sqrt{1-\rho}} \tag{C.24}$$

where $m \geq 1$ and ρ is Pearson's correlation coefficient. The model fails when ρ is unity.

A third bivariate Gumbel distribution is presented in Nelsen (2006). This distribution is called the Gumbel bivariate exponential distribution, given by

$$H(x,y) = 1 - e^{-x} - e^{-y} + e^{-(x+y+\theta xy)} \qquad (x,y) \geq 0 \tag{C.25}$$

where θ is the dependence parameter. Gumbel's bivariate exponential distribution has exponential marginals.

C.2.2 Bivariate Normal Distribution

Yue (2000) applied the bivariate normal distribution to the logarithms of hydrologic data to determine the joint probability structure of a bivariate dataset. The bivariate normal density function is

$$h(x,y) = \frac{1}{2\pi xy\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left[\frac{1}{1-\rho^2}\left(x^2 - 2\rho xy + y^2\right)\right] \tag{C.26}$$

where ρ is the correlation coefficient, $|\rho| \leq 1$, x and y are the variates, and σ_x and σ_y are the scale distribution parameters of their respective variates. The bivariate normal distribution location parameter (mean) is located at (0,0). Therefore, for univariate distributions with non-zero means of μ_x and μ_y, the data are adjusted by subtracting the means prior to fitting the bivariate distribution.

For the bivariate lognormal distribution, the logarithms (natural or another base) of the variates, $x = \ln(q)$ and $y = \ln(r)$ where q and r represent the observed values, are used to transform the variates and reduce the skew in the dataset.

As for the univariate normal distribution function, the bivariate normal distribution function has no closed-form analytic solution to the integral of the density function. Therefore, numerical integration of the density function is required to determine the joint nonexceedance probability. Genz (2004) developed numerical integration methods in a variety of languages to deal with this problem.

C.3 Copulas

Copulas are a more general approach to bivariate (or multivariate) problems. The term copula refers to a function, called the dependence function, used to link two univariate distributions in such a way as to represent the bivariate (or multivariate) dependence between the two random variables. The potential of a copula is realized in that the copula is independent from the form of the univariate marginal distributions. That is, the marginal distributions of a copula are uniformly distributed on the interval [0,1]. Therefore, the marginal distributions can be chosen such that they provide a best-fit of the univariate random variables, with the copula used to model the dependence behavior. Genest and Favre (2007) provide a useful overview on the use of copulas in hydrology.

Many copulas are available for application to bivariate (and multivariate) random variables. Those of interest to hydrologists typically fall into the *Archimedean* family of copulas. For two random variables, X and Y with cumulative distribution functions of $F_X(x)$ and $F_Y(y)$, respectively, define $U = F_X(X)$ and $V = F_Y(Y)$. Then, U and V are uniformly distributed random variables and u will denote a specific value of U and v will denote a specific value of V. As described by Zhang and Singh (2006) the one-parameter Archimedean copula can be described as follows:

$$C_\theta(u,v) = \phi^{-1}\{\phi(u) + \phi(v)\} \qquad 0 < u,v < 1 \tag{C.27}$$

where,

$C_\theta(u,v)$	=	the copula function with θ representing the dependence parameter
ϕ	=	the copula generating function
u,v	=	specific values of the uniformly distributed random variable U and V

The copula, $C_\theta(u,v)$, is analogous to the distribution function, H(x,y) discussed in the previous section. Genest and Favre (2007) use the concept of a bivariate probability integrated transform (BPIT) to evaluate suitability of a given copula. It is defined as:

$$K(t) = t - \frac{\phi(t)}{\phi'(t)} \tag{C.28}$$

where ϕ' is the derivative of ϕ.

C.3.1 Gumbel-Hougaard Copula

For the Gumbel-Hougaard family of copulas (Nelsen, 2006), the generating function is given by

$$\phi(t) = -\ln(t)^\theta \tag{C.29}$$

and the copula is

$$C_\theta(u,v) = \exp\left\{-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{1/\theta}\right\} \qquad \theta \in [1,\infty] \tag{C.30}$$

where θ is the dependence parameter. The relation between Kendall's τ and θ is given by

$$\tau = 1 - \theta^{-1} \tag{C.31}$$

For development of the BPIT, the derivative of ϕ is needed,

$$\phi'(t) = -\frac{\theta}{t}(-\ln t)^{\theta-1} \tag{C.32}$$

C.3.2 Frank Copula

The Frank family of copulas (Genest and Rivest, 1993; Nelsen, 2006) derives from the generating function

$$\phi(t) = -\ln\left(\frac{e^{-\theta t} - 1}{e^{-\theta} - 1}\right) \tag{C.33}$$

and the copula is

$$C_\theta(u,v) = -\frac{1}{\theta}\ln\left[1 + \frac{\left[e^{-\theta u} - 1\right]\left[e^{-\theta v} - 1\right]}{e^{\theta} - 1}\right] \tag{C.34}$$

where θ is the dependence parameter. The relation between Kendall's τ and θ is given by

$$\tau = 1 - \frac{4}{\theta}\left[1 - D_1(\theta)\right] \tag{C.35}$$

where $D_1()$ is the first order Debye function $D_k()$. For a positive argument, D_k is defined as:

$$D_k(\theta) = \frac{k}{\theta^k}\int_0^\theta \frac{t^k}{\exp(t) - 1}dt \quad \theta > 0 \tag{C.36a}$$

For a negative argument, D_k is defined as:

$$D_k(-\theta) = D_k(\theta) + \frac{k\theta}{k+1} \tag{C.36b}$$

For development of the BPIT, the derivative of ϕ is needed,

$$\phi'(t) = \frac{\theta e^{-\theta t}}{e^{-\theta t} - 1} \tag{C.37}$$

C.4 Empirical Distributions and Plotting Position

The plotting position of a sample from a random variable is used to locate the observation in probability space. *Probability space* is the set of coordinates wherein one coordinate (generally the ordinate) represents the value of the observation and the other coordinate (generally the abscissa) represents an estimate of the frequency of the event. The plotting position might be considered to represent the empirical distribution function of the random variable.

The frequency can be the exceedance or nonexceedance probability, as estimated by a plotting position formula. The probability coordinate can also be represented by a standardized value, such as the standard normal deviate or the reduced variate.

C.4.1 Univariate Plotting Position

Cunnane (1978) conducted an extensive review of plotting position formulae in common use at the time and suggested that a general form for plotting position formulae be:

$$P(X_i \le x) = \frac{i - \alpha}{n + 1 - 2\alpha} \tag{C.38}$$

where X_i is the event, i indicates the ith variate (sorted in ascending order), n is the number of observations in the dataset, and α is the plotting position factor.

For the Blom formula, $\alpha = 3/8$. For the Gringorten formula, $\alpha = 0.44$. For the Hazen formula, $\alpha = 0.5$. For the Beard formula, $\alpha = 0.31$. Finally, for the Weibull formula, $\alpha = 0$. Cunnane (1978) reported that the Blom formula provided good results for normal distributions and that the Gringorten formula provided good results for the univariate Gumbel distribution. The Weibull formula was best only for the uniform distribution.

Although Cunnane recommends against using the plotting position as an estimate of the nonexceedance frequency of observations in a sample, his advice is often ignored in subsequent literature. Yue, et al. (1999) and Yue and Rasmussen (2002) use the Gringorten formula for estimating nonexceedance probabilities. Zhang and Singh (2006) use the relative frequency for estimates of nonexceedance probability. Genest and Favre (2007) also use the relative frequency, less one degree of freedom, to estimate the nonexceedance probability.

Therefore, the Weibull plotting position formula is a reasonable method for this research. The use of the Weibull plotting position formula is not likely to affect the outcome of the work and its implementation is straightforward.

C.4.2 Bivariate Empirical Distribution Functions

Although most statisticians argue against using the results from a plotting position formula as an estimate of the distribution function, in application this is what is done. A distribution (or distributions) is fit to the observations and then a plotting position formula is used to determine how to plot the observations on the fitted distribution. The fraction of observations less than or equal to a particular observation is used to estimate the plotting position for the data point. This is an estimate of the distribution function.

A similar method can be used for multivariate random variables. If a set of observations is collected, then a random variable defined as the number of observations less than or equal to the observation can be computed. This method is defined in the context of copulas in the following paragraphs, but there appears to be no reason why it could not be applied for comparing results from other multivariate distributions as well.

Genest and Rivest (1993) develop the notion of a non-parametric estimate of H(x,y) based on the empirical distribution of the coincident events. Define a random variable, V, such that

$$V_i = \frac{\#\left\{(X_j, Y_j) \in (X_j < X_i, Y_j < Y_i)\right\}}{n-1} \tag{C.39}$$

where #{} indicates the count of elements in the set. The empirical bivariate probability integrated transform (BPIT) has the form

$$K_n(v) = \sum_{i=1}^{n} \frac{\delta(v - V_i)}{n} \tag{C.40}$$

where $\delta(x)$ is a step function such that

$$\delta(x) = \begin{Bmatrix} 0, x < 0 \\ 1, x \geq 0 \end{Bmatrix} \tag{C.41}$$

The BPIT represents the fraction of observed values less than or equal to v, where V_i is a function of the ranks of the (x,y) pairs of observations.

Genest and Favre (2007) present an alternative development. Let *I* be a variable defined such that

$$I_{i,j} = \begin{cases} 1, & X_j < X_i, Y_j < Y_i \\ 0, & \text{otherwise} \end{cases} \tag{C.42}$$

for arbitrary $i \neq j$ and let $I_{ii}=1$ for all $i \in (1,\ldots,n)$. Then define

$$V_i = \frac{1}{n}\sum_{j=1}^{N} I_{ij} \tag{C.43}$$

$$= \frac{1}{n}\#\{j : X_j < X_i, Y_j < Y_i\}.$$

V is a function of the ranks of the observations and is related to the distribution function of the copula.

Genest and Favre (2007) suggest that a particular copula can be fit to a dataset, and then plotted against the empirical copula as a visual test of fit. That is, the results of Equation C.28 are plotted against the empirical result of Equation C.40 (KK plot). If the empirical and fitted copulas agree, then the resulting plot (analogous to a qq-plot) should approximate a straight line. Any deviation from a straight line should be interpreted as a lack of fit.

However, the an analogous representation for Equation C.28 as applied to the bivariate Gumbel or normal distributions was not found such that comparisons of fit using a KK plot could not be accomplished. Therefore, fit was evaluated by comparing the empirical values of V (Equation C.39) with the computation of C (Equation C.27) for the Gumbel-Hougaard and Frank copulas and the same values of V with the computation of H for the bivariate Gumbel (Equation C.21) and H for the bivariate normal (numerical integration of Equation C.26).

C.5 Correlation Measures

Pearson's ρ and Kendall's τ are two measures of correlation used in previous sections. Each is described in more detail in this section, including a sample calculation.

C.5.1 Pearson's ρ

Pearson's ρ is a linear correlation measure between two variables. It ranges from –1 to 1. An absolute value of Pearson's ρ near unity is indicative of a substantial degree of linear association, while a low absolute value indicates a low degree of association. Pearson's ρ is estimated from the following equation (Bowker and Lieberman, 1972):

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{C.44}$$

where,

ρ	=	Pearson's ρ
x_i, y_i	=	individual observations from x and y, respectively
$\bar{x}, \bar{y}$	=	mean of the individual observations for x and y, respectively
n	=	number of observations

Computation of Pearson's ρ is illustrated using the table of peak flows given in Table C.1. Intermediate computations for the summation terms in Equation C.44 are shown in Table C.2.

Table C.1. Example Peak Flow Data.

Index	Main Flow (ft^3/s)	Tributary Flow (ft^3/s)
1	5,040	3,750
2	3,540	2,370
3	4,080	3,100
4	4,000	3,190
5	5,540	4,180
6	4,180	1,520
Mean	4,397	3,018

Table C.2. Computation for Summation Terms.

Index	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	643	413,878	732	535,336	470,706
2	-857	733,878	-648	420,336	555,406
3	-317	100,278	82	6,669	-25,861
4	-397	157,344	172	29,469	-68,094
5	1,143	1,307,211	1,162	1,349,469	1,328,172
6	-217	46,944	-1,498	2,245,003	324,639
Sum	0	2,759,533	0	4,586,283	2,584,967

Equation C.44 is applied:

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{2{,}584{,}967}{\sqrt{(2{,}759{,}533)(4{,}586{,}283)}} = 0.727$$

C.5.2 Kendall's τ

Kendall's τ is a rank correlation measure between two variables. It ranges from –1 to 1. As with Pearson's ρ, an absolute value of Kendall's τ near unity is indicative of a substantial degree of rank correlation while a low absolute value indicates a low degree of association. Kendall's τ places an emphasis on the rank order of data rather than their specific values. Kendall's τ is estimated from the following equation (Genest and Favre, 2007):

$$\tau = \frac{P_n - Q_n}{0.5(n)(n-1)} \tag{C.45}$$

where,

τ	=	Kendall's τ
P_n	=	number of concordant pairs
Q_n	=	number of discordant pairs
n	=	number of observations

Two pairs (x_i, y_i), (x_j, y_j) are said to be concordant when $(x_i - x_j)(y_i - y_j) > 0$ and discordant when $(x_i - x_j)(y_i - y_j) < 0$. It does not matter how much greater than or less than zero, only that the rank order is moving in the same direction. Ties (one or both values are equal and the resulting product equals 0) are not considered concordant or discordant. If ties occur, the denominator in Equation C.45 must be adjusted.

Determination of concordance or discordance for the sample dataset in Table C.1 is summarized in Table C.3. The number of pairs to test is determined by the denominator of Equation C.45. For the example, n = 6, therefore the number of pairs to test equals 0.5(6)(5) = 15.

Table C.3. Determination of Concordant and Discordant Pairs.

Comparison	i	j	(x_i-x_j)	(y_i-y_j)	$(x_i-x_j)(y_i-y_j)$	Concordant?	Discordant?
1	1	2	1,500	1,380	2,070,000	Y	
2	1	3	960	650	624,000	Y	
3	1	4	1,040	560	582,400	Y	
4	1	5	-500	-430	215,000	Y	
5	1	6	860	2,230	1,917,800	Y	
6	2	3	-540	-730	394,200	Y	
7	2	4	-460	-820	377,200	Y	
8	2	5	-2,000	-1,810	3,620,000	Y	
9	2	6	-640	850	-544,000		N
10	3	4	80	-90	-7,200		N
11	3	5	-1,460	-1,080	1,576,800	Y	
12	3	6	-100	1,580	-158,000		N
13	4	5	-1,540	-990	1,524,600	Y	
14	4	6	-180	1,670	-300,600		N
15	5	6	1,360	2,660	3,617,600	Y	

Comparison 1 is an evaluation of the first and second pair of values in Table C.1. In this case, the main and tributary flows are greater for the first pair compared with the second pair. This results in $(x_i-x_j)(y_i-y_j)$ = 2,070,000. Because this value is greater than zero, the pair is considered concordant as shown in Table C.3. Conversely, comparison 9 is an evaluation of the second and sixth pair of values in Table C.1. In this case, the main flow for the second pair is less than the sixth pair while the tributary flow for the second pair is greater than the sixth pair. This results in $(x_i-x_j)(y_i-y_j)$ = -544,000. Because this value is less than zero, the pair is considered discordant as shown in Table C.3. Overall, the number of concordant pairs, P_n, is 11 and the number of discordant pairs, Q_n, is 4. Equation C.45 can now be applied:

$$\tau = \frac{P_n - Q_n}{0.5(n)(n-1)} = \frac{11-4}{0.5(6)(6-1)} = 0.467$$

C.6 Prototype Analyses

The prototype gage pairs were analyzed to evaluate the suitability of candidate bivariate distributions and copulas to represent main and tributary flows. Steps in the evaluation process included: 1) a graphical evaluation of the theoretical fit of the four methods to the empirical data, 2) a review of the joint distributions, and 3) an examination of the nonexceedance probabilities.

C.6.1 Fitting Distributions

The "fit" of the data to various distributions was assessed visually with a plot of the theoretical nonexceedance versus the empirical nonexceedance probability. Figures C.1a through C.4b display these data for four prototype gages: GP05, GP06, GP43, and GP76.

The ordinate in the figures is computed from Equation C.21 for the Gumbel mixed model bivariate distribution and a numerical integration of Equation C.26 for the log-normal bivariate distribution. The ordinate for the copula methods are computed from Equation C.27. The generating function, ϕ, was computed from Equations C.29 and C.33 for the Gumbel-Hougaard and Frank copulas, respectively. The abscissa is computed from Equation C.39.

C.6.2 Joint Distributions

The joint distributions for the prototype gage pairs are shown in Figures C.5, C.6, C.7, and C.8. The abscissa and ordinate axes represent the cumulative nonexceedance probabilities for the main and tributary streams, respectively. The curves represent isolines of nonexceedance probabilities for each methodology. The individual data points are the estimated univariate nonexceedance values based on the Weibull plotting position formula and represented as circles.

The data for the POM dataset of Gage Pair 05 is shown in Figure C.5a. The individual data points are widely scattered and the isolines display a modest curvature. Both features indicate a low correlation between the main and tributary gages. Because all four distributions are relatively close together, a fifth curve (product) has been added that represents the probability isoline as a product of two independent probabilities. In Figure C.5b, the observed data for the POT dataset show a wider spread (lower correlation) than the POM dataset.

For Gage Pairs 06, 43, and 76 the higher correlation between the gages results in a clustering of the observed data points in a band from the lower left to upper right in the figure as shown in Figures C.6, C.7, and C.8, respectively. In addition, there is a greater curvature to the isolines compared with the isolines for Gage Pair 05.

C.6.3 Exceedance Probability

The 10-percent exceedance probability isolines for the prototype gage pairs 05, 06, 43, 76 are shown in Figure C.9, C.10, C.11, and C.12, respectively. For the POM plots, the abscissa represents the nonexceedance probability for the main stream and the ordinate represents the nonexceedance probability for the complementary data series for the tributary stream. Conversely, for the POT plots, the abscissa represents the nonexceedance probability for the complementary data series for the main stream and the ordinate represents the nonexceedance probability for the tributary stream. As discussed in Section 4.4, the nonexceedance probabilities for the complementary and annual data series are not the same unless the main and tributary discharges are completely dependent.

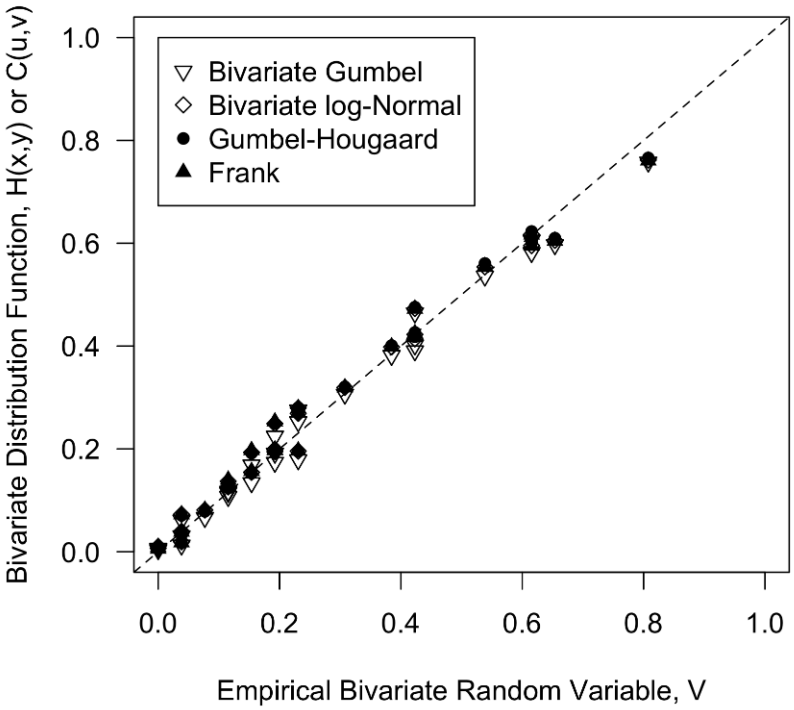

Figure C.1a. Fit for Gage Pair 05 (POM).

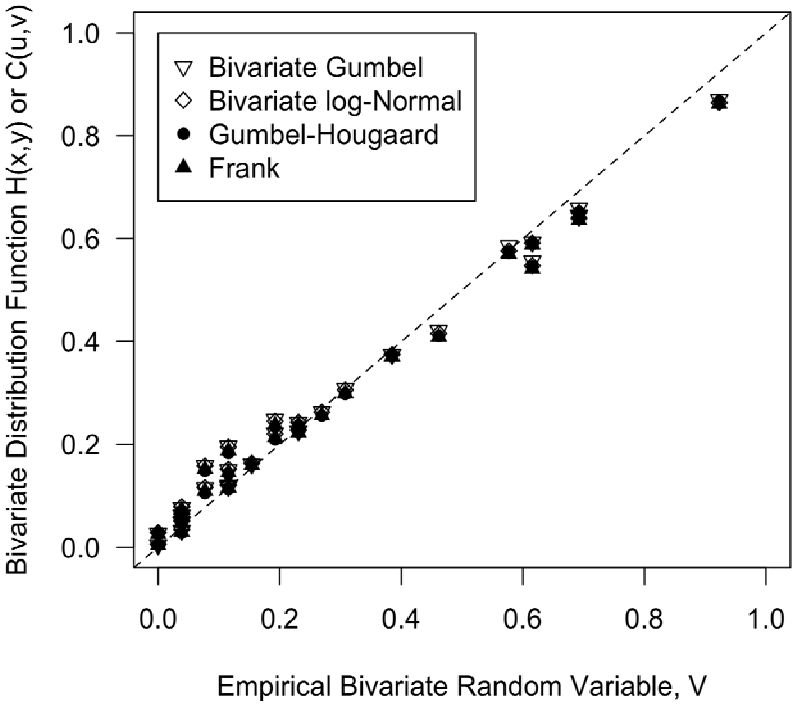

Figure C.1b. Fit for Gage Pair 05 (POT).

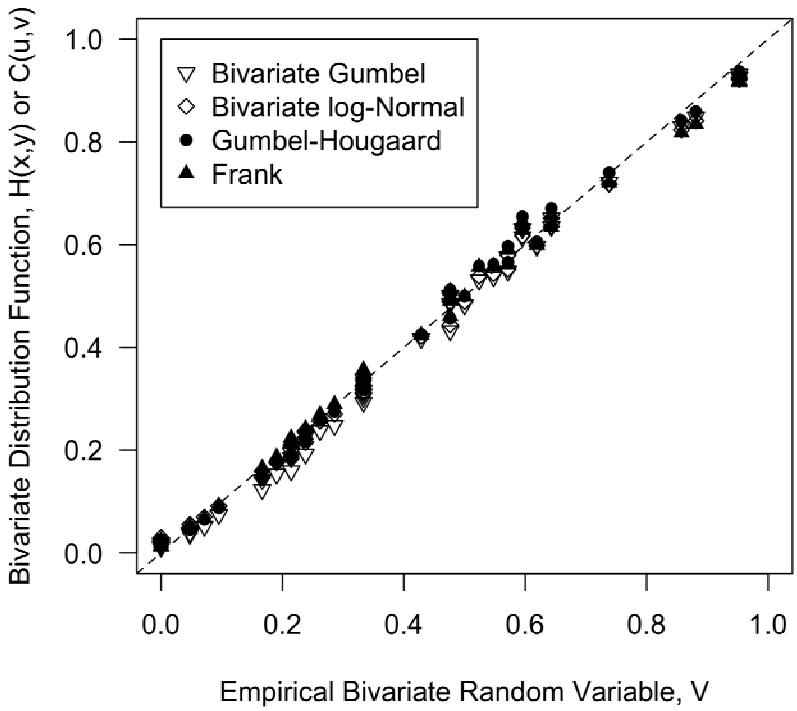

Figure C.2a. Fit for Gage Pair 06 (POM).

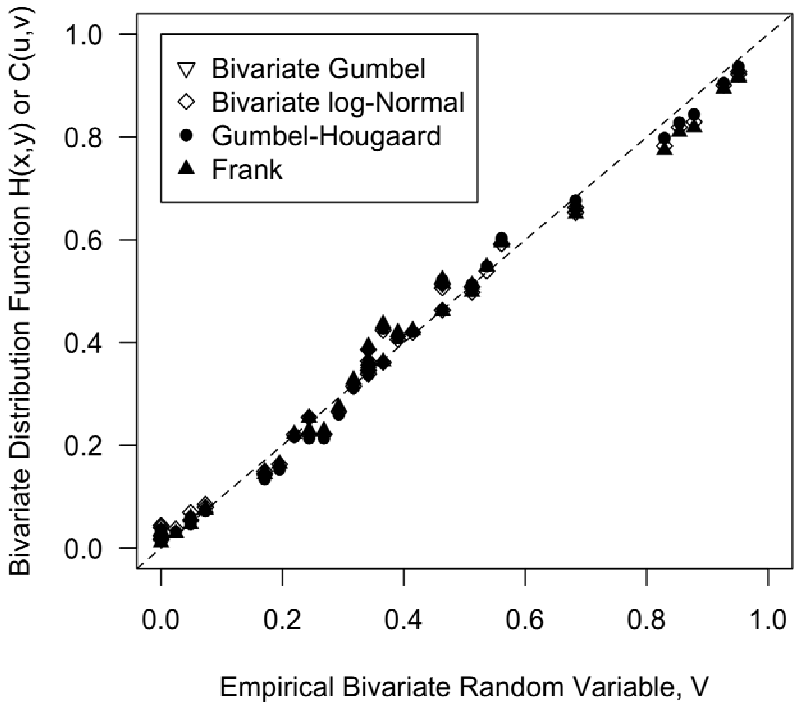

Figure C.2b. Fit for Gage Pair 06 (POT).

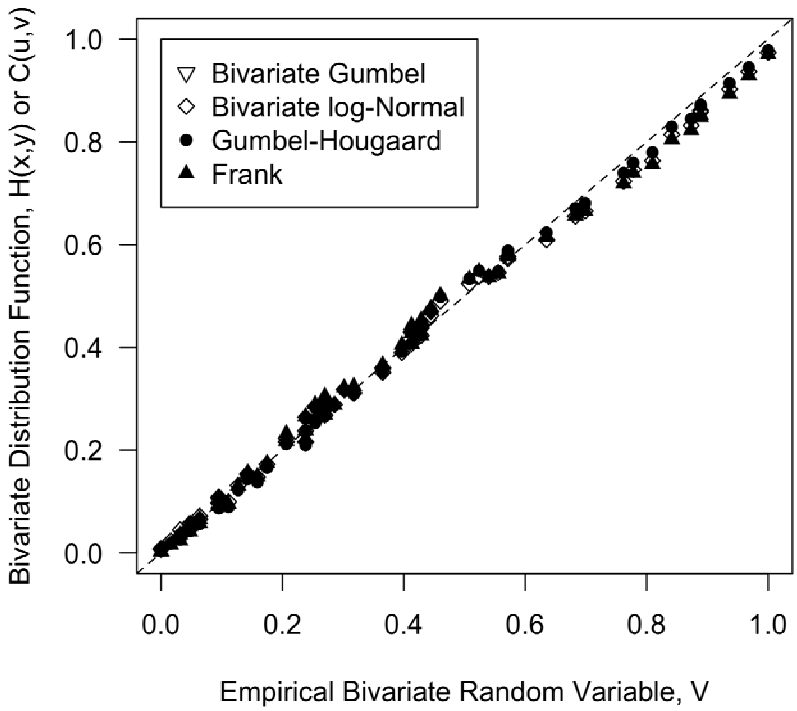

Figure C.3a. Fit for Gage Pair 43 (POM).

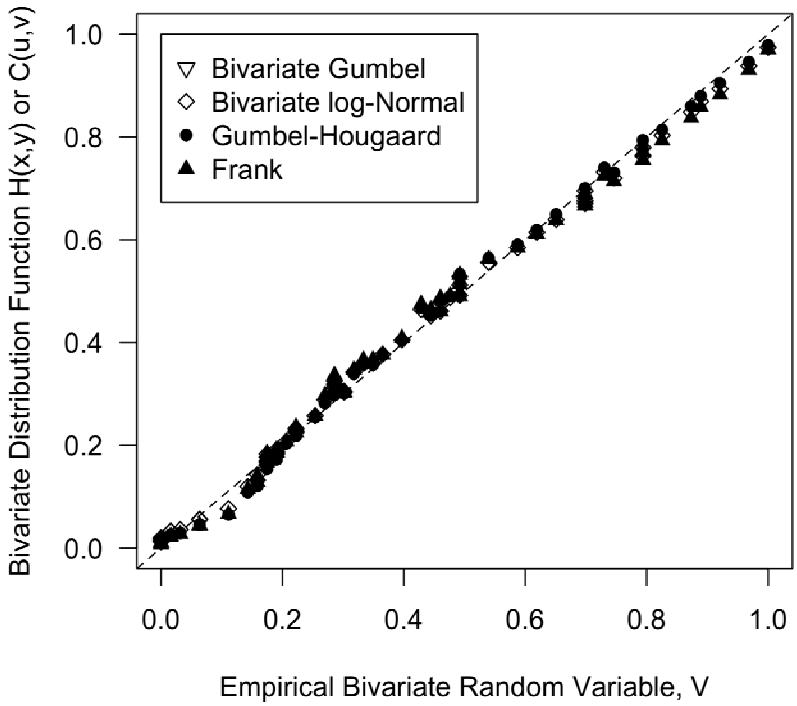

Figure C.3b. Fit for Gage Pair 43 (POT).

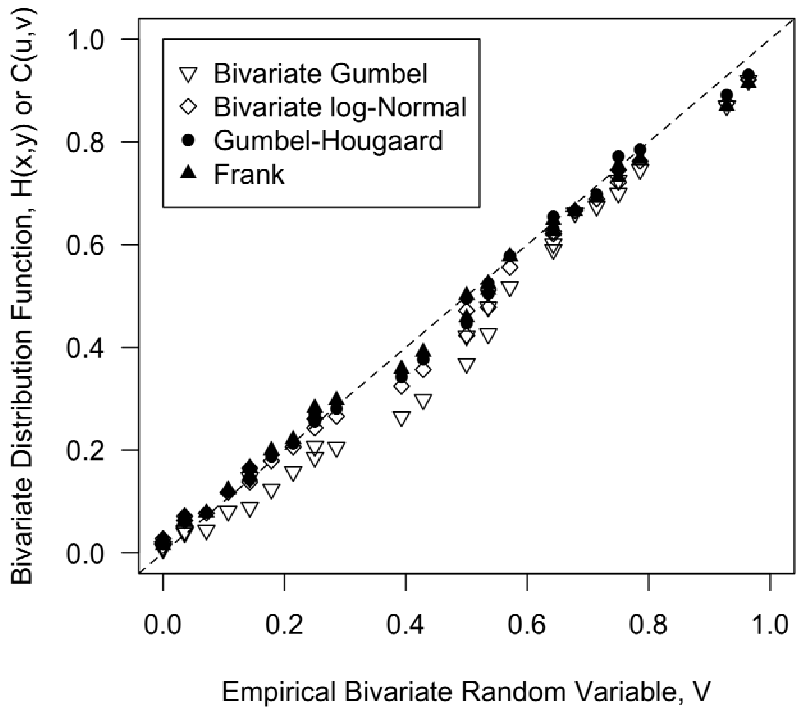

Figure C.4a. Fit for Gage Pair 76 (POM).

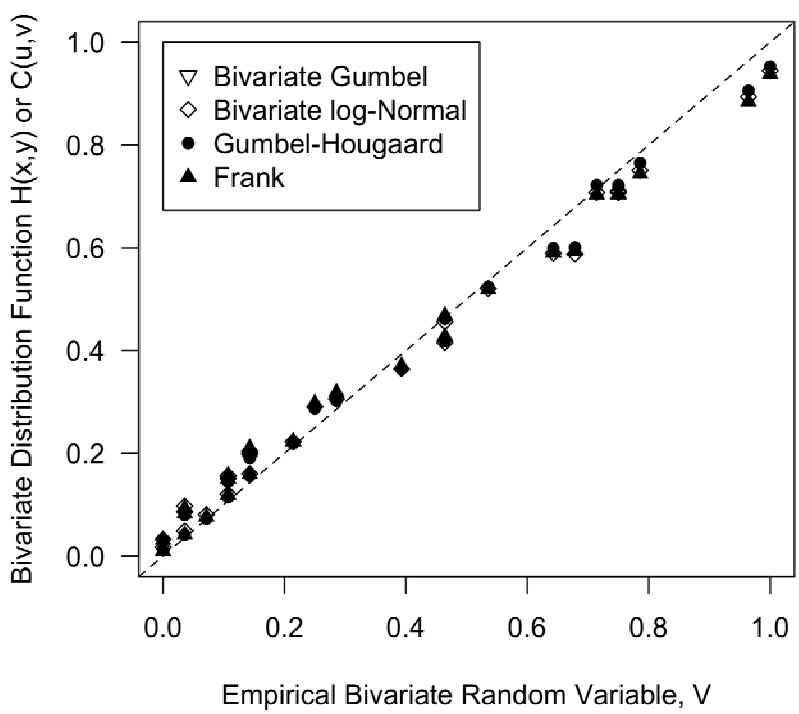

Figure C.4b. Fit for Gage Pair 76 (POT).

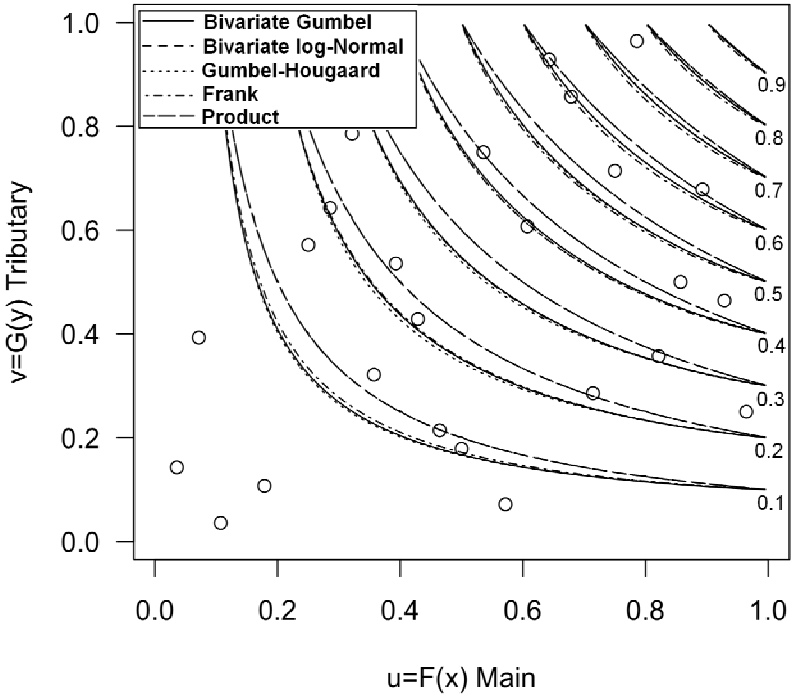

Figure C.5a. Joint Distribution for Gage Pair 05 (POM).

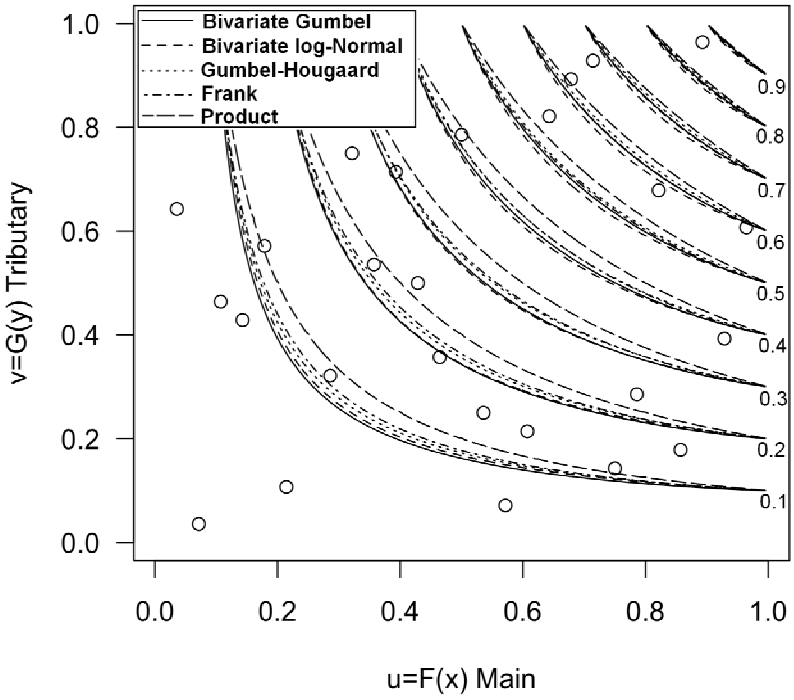

Figure C.5b. Joint Distribution for Gage Pair 05 (POT).

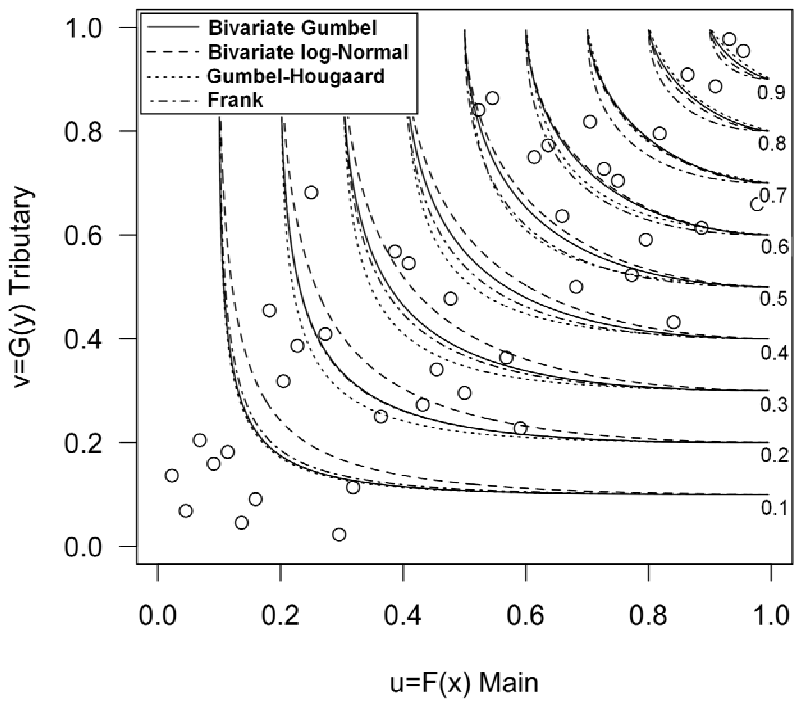

Figure C.6a. Joint Distribution for Gage Pair 06 (POM).

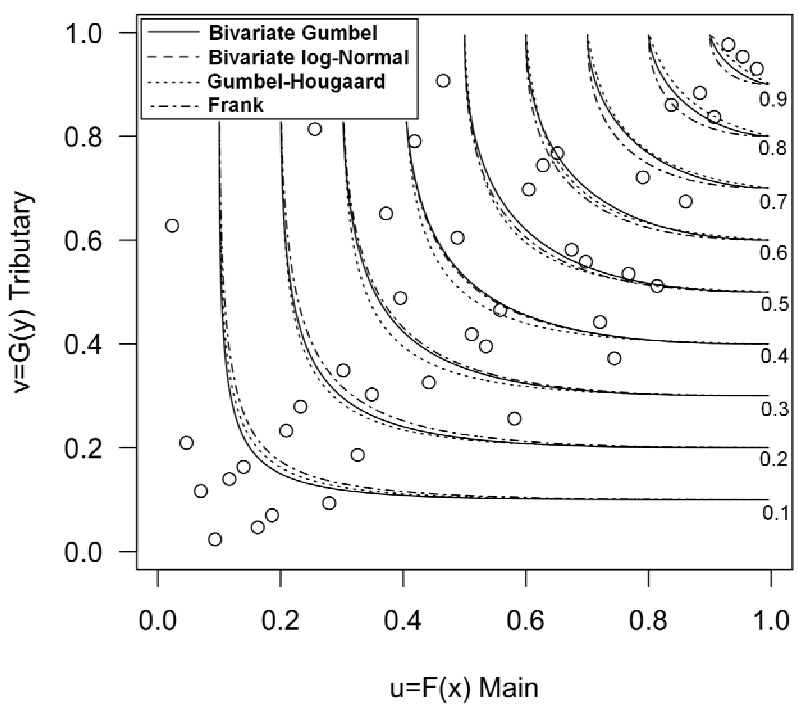

Figure C.6b. Joint Distribution for Gage Pair 06 (POT).

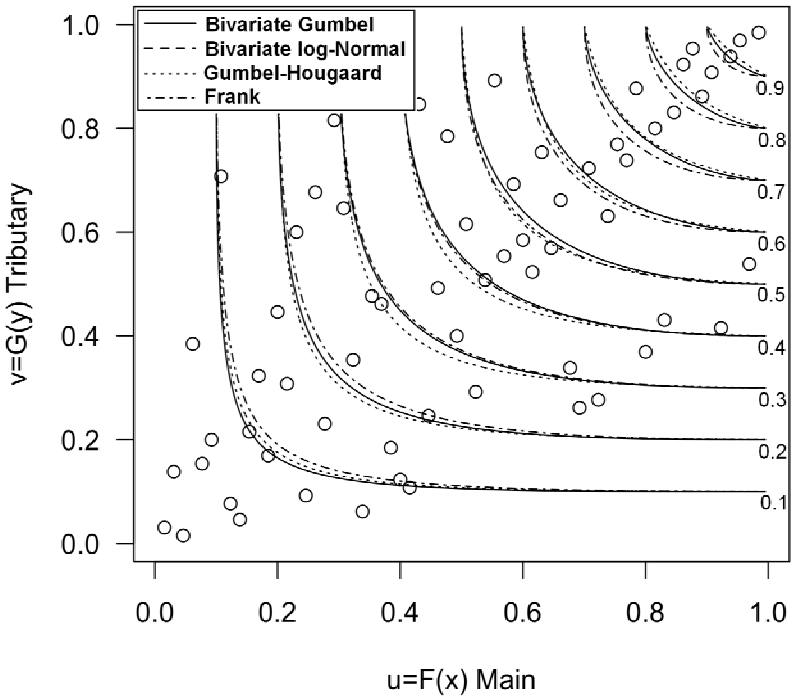

Figure C.7a. Joint Distribution for Gage Pair 43 (POM).

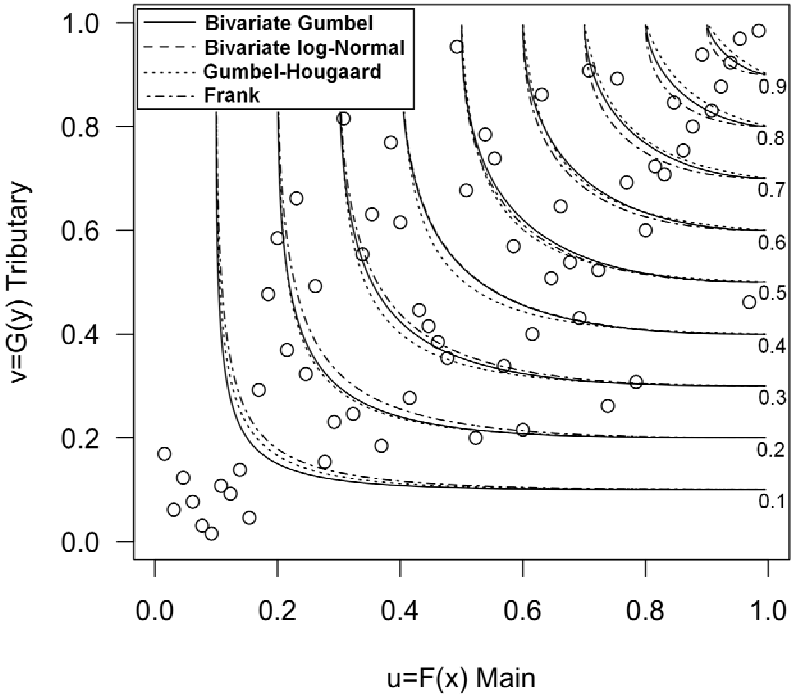

Figure C.7b. Joint Distribution for Gage Pair 43 (POT).

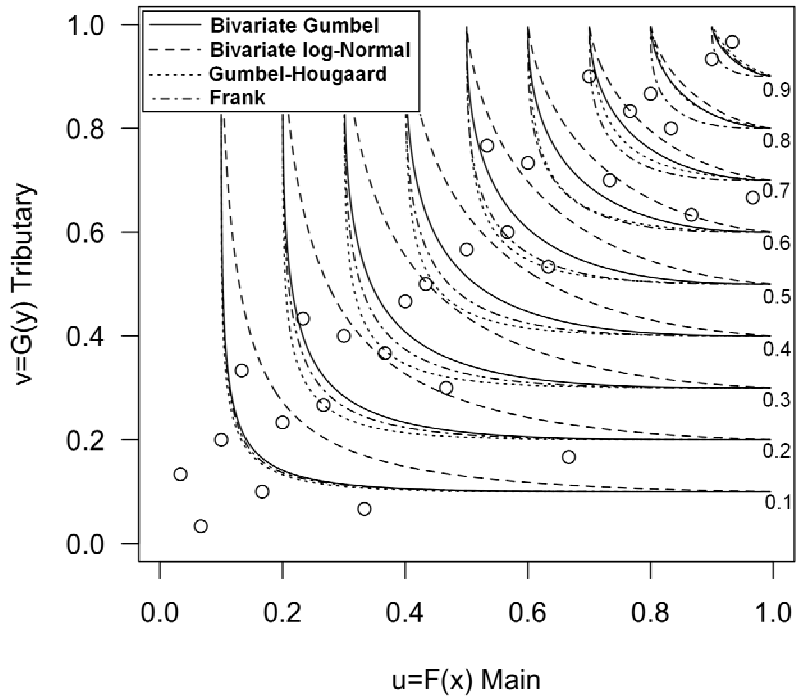

Figure C.8a. Joint Distribution for Gage Pair 76 (POM).

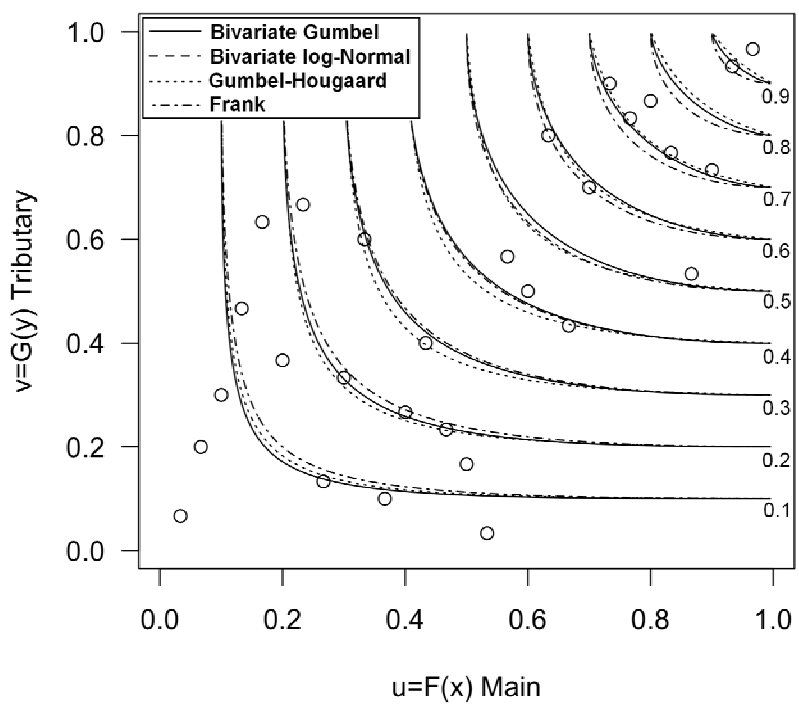

Figure C.8b. Joint Distribution for Gage Pair 76 (POT).

A single isoline for each method is shown in the figures to illustrate the practical application. All ordered pairs on a given curve represent a 10-percent exceedance probability as calculated by that method. Which combination may be critical for the design problem, depends on where the design location is on the tributary reach. Three combinations are taken from each curve, one at each end and one in the middle. For the Gumbel-Hougaard method, these are summarized in for the prototype gage pairs 05, 06, 43, and 76 in Tables C.4, C.5, C. 6, and C.7, respectively, as GH copula (1), GH copula (2), and GH copula (3).

The 10-percent exceedance probability isolines determined by alternative methods for Gage Pair 05 (POM) are displayed in Figure C.9a. All points on a given isoline represent the same joint exceedance probability. A summary of selected points from specific isolines is provided in Table C.4a for this gage pair. For example, three points for the bivariate Gumbel distribution isoline are given: 1) 0.5 nonexceedance probability on the main combined with 0.8 nonexceedance probability on the tributary, 2) 0.726 nonexceedance probability on the main combined with 0.726 nonexceedance probability on the tributary, and 3) 0.8 nonexceedance probability on the main combined with 0.5 nonexceedance probability on the tributary.

Return periods corresponding to each nonexceedance probability are also provided. These data are interpreted such that the first combination is a 2-yr event on the main combined with a 5-yr event on the tributary, the second combination is a 3.6-yr event on the main combined with a 3.6-yr event on the tributary, and the third combination is a 5-yr event on the main combined with a 2-yr event on the tributary.

Gage Pair 05 correlation measures (Pearson's ρ and Kendall's τ) are lower than the other prototype gage pairs. Two alternative frameworks for assessing joint probability when correlation is low were examined. First, treating the two gages as independent, the joint exceedance probability is the product of the independent exceedance probabilities. Using this assumption provides the results identified as "product" in Figure C.9a and Table C.4a. A second method for treating two independent gages is to ignore the confluence and assume that one stream experiences a design event while the other is not in a flood condition. Identified as the "univariate", results of applying this method are summarized in Table C.4.

Gage Pair 06 (POM) is more representative of the situation where the correlation between the main and tributary streams exhibits some dependency. Three 10-percent exceedance combinations for four methods are summarized in Table C.5a. Using the Gumbel-Hougaard copula, a 2-yr event on the main stream combined with a 9.6-yr event on the tributary stream represents a 10-yr joint return period event. A 6.5-yr event on both the main and tributary streams also represents a 10-yr event, as does a 9.6-yr event on the main stream combined with a 2.1-yr event on the tributary stream. The most severe of these three conditions would be considered the 10-yr event and used for design.

Comparing the four methods for Gage Pair 06 (POM), a 2-yr event on the main stream results in a return period event on the tributary ranging from 9.1 years for the Bivariate Gumbel to 9.6 years for the Gumbel-Hougaard copula. Looking at the event where the return period is equal on both the main and tributary streams, the return period ranges from 5.3 years for the Bivariate Gumbel to 6.5 years for the Gumbel-Hougaard copula. For this gage pair and the POM dataset the Gumbel-Hougaard copula results in more severe design conditions. However, the range is not significant for the 10-percent event. The range will be greater for the 1-percent event.

The same comparison may also be performed for Gage Pair 06 on the POT dataset based on the data in Table C.5b. Comparing the four methods yields the same observations noted for the POM dataset. The Gumbel-Hougaard copula results in the most severe design conditions and the Frank copula the least severe. In this case, the Bivariate Gumbel values could not be calculated because Pearson's ρ exceeded two-thirds.

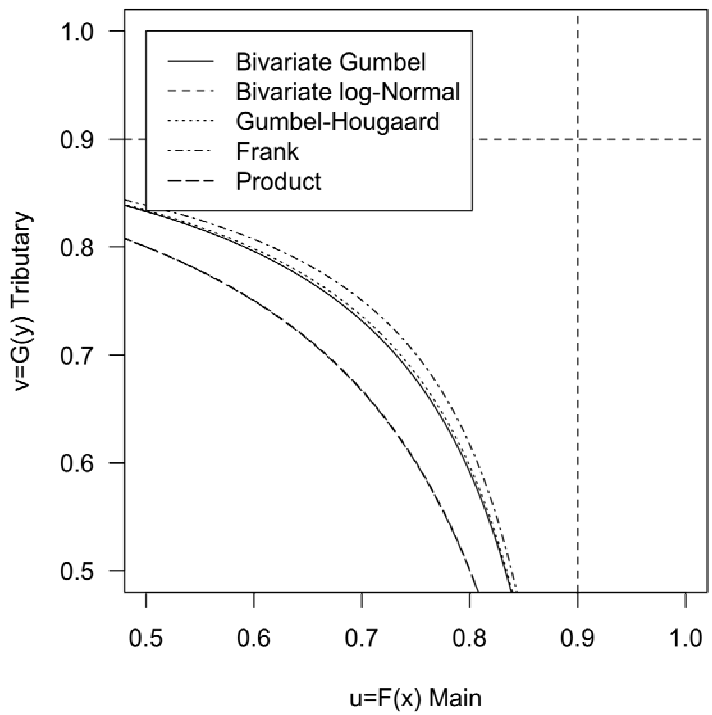

Figure C.9a. 10-percent Exceedance Probability for Gage Pair 05 (POM).

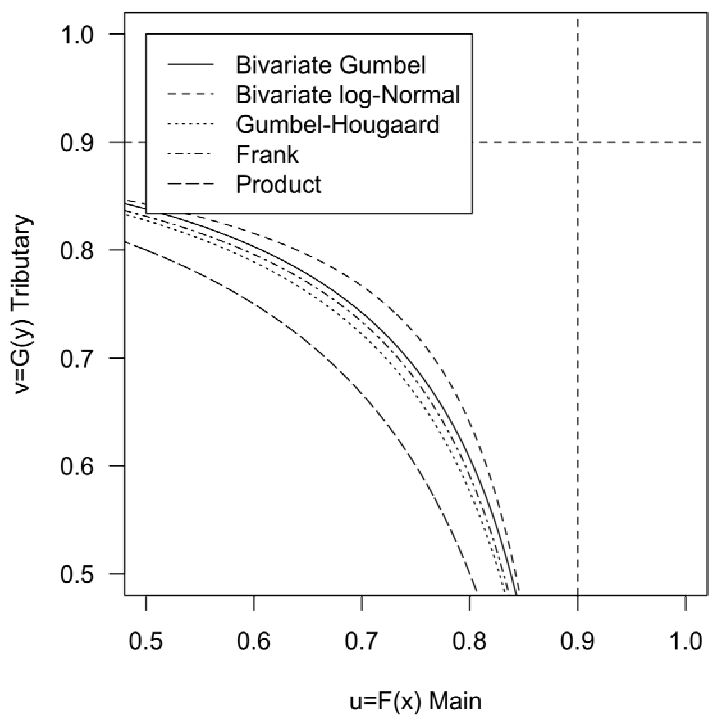

Figure C.9b. 10-percent Exceedance Probability for Gage Pair 05 (POT).

Table C.4a. 10-percent Exceedance Combinations for Gage Pair 05 (POM).

	Main		Tributary*	
Method (ID)	**Nonexceedance Probability (u)**	**Return Period**	**Nonexceedance Probability (v)**	**Return Period**
Product (1)	0.50	2.0	0.80	5.0
Product (2)	0.684	3.1	0.684	3.1
Product (3)	0.80	5.0	0.50	2.0
Bivariate Gumbel (1)	0.50	2.0	0.80	5.0
Bivariate Gumbel (2)	0.726	3.6	0.726	3.6
Bivariate Gumbel (3)	0.80	5.0	0.50	2.0
GH copula (1)	0.50	2.0	0.84	6.2
GH copula (2)	0.73	3.7	0.73	3.7
GH copula (3)	0.84	6.2	0.50	2.0
Univariate (1)	0.90	10	<0.50	<2.0
Univariate (2)	<0.50	<2.0	0.90	10

*Not the annual series nonexceedance probabilities and return periods. These are conservative.

Table C.4b. 10-percent Exceedance Combinations for Gage Pair 05 (POT).

	Main*		Tributary	
Method (ID)	**Nonexceedance Probability (u)**	**Return Period**	**Nonexceedance Probability (v)**	**Return Period**
Product (1)	0.50	2.0	0.80	5.0
Product (2)	0.684	3.1	0.684	3.1
Product (3)	0.80	5.0	0.50	2.0
Bivariate Gumbel (1)	0.50	2.0	0.84	6.2
Bivariate Gumbel (2)	0.744	3.9	0.744	3.9
Bivariate Gumbel (3)	0.88	6.2	0.50	2.0
GH copula (1)	0.50	2.0	0.83	5.9
GH copula (2)	0.726	3.6	0.726	3.6
GH copula (3)	0.83	5.9	0.50	2.0
Univariate (1)	0.90	10	<0.50	<2.0
Univariate (2)	<0.50	<2.0	0.90	10

*Not the annual series nonexceedance probabilities and return periods. These are conservative.

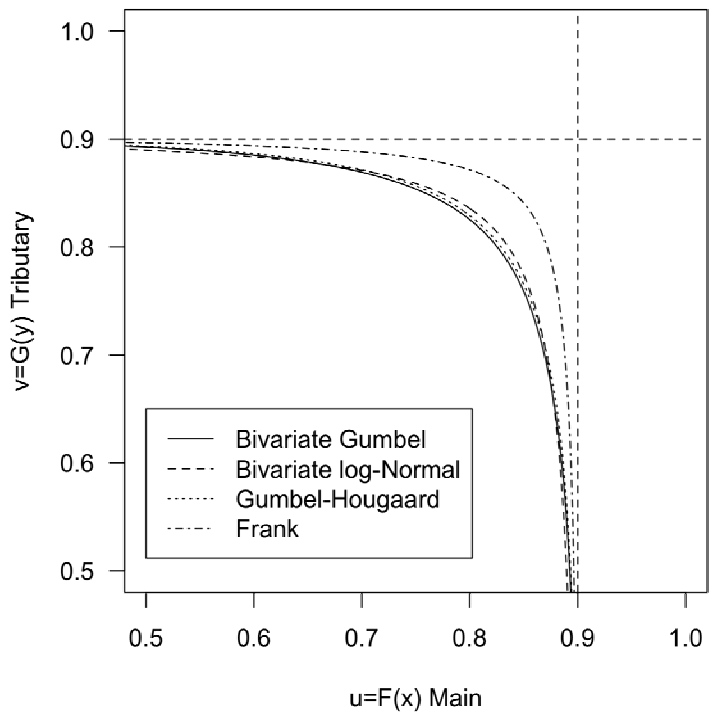

Figure C.10a. 10-percent Exceedance Probability for Gage Pair 06 (POM).

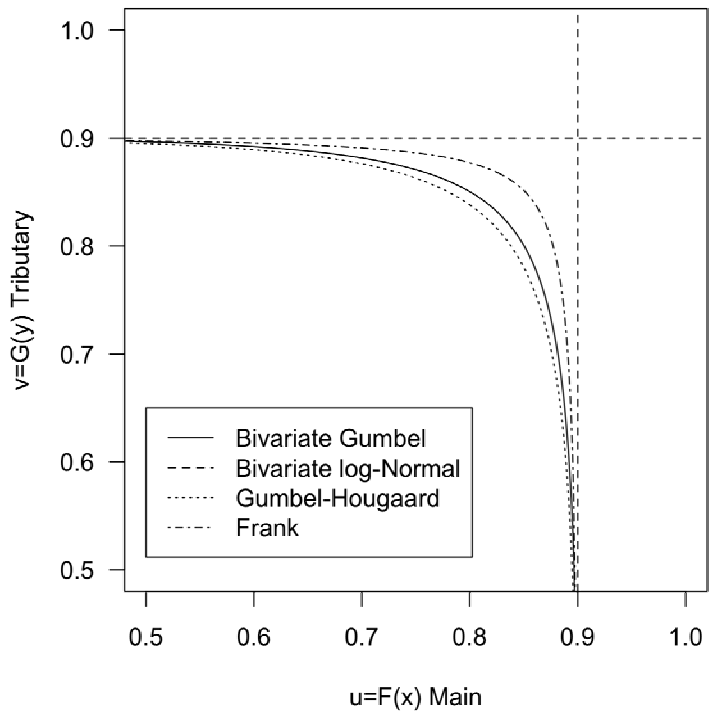

Figure C.10b. 10-percent Exceedance Probability for Gage Pair 06 (POT).

Table C.5a. 10-percent Exceedance Combinations for Gage Pair 06 (POM).

	Main		Tributary*	
Method (ID)	**Nonexceedance Probability (u)**	**Return Period**	**Nonexceedance Probability (v)**	**Return Period**
Log-normal (1)	0.50	2.0	0.893	9.3
Log-normal (2)	0.813	5.3	0.813	5.3
Log-normal (3)	0.890	9.1	0.540	2.2
Bivariate Gumbel (1)	0.50	2.0	0.890	9.1
Bivariate Gumbel (2)	0.825	5.7	0.825	5.7
Bivariate Gumbel (3)	0.880	8.3	0.636	2.7
GH copula (1)	0.50	2.0	0.896	9.6
GH copula (2)	0.846	6.5	0.846	6.5
GH copula (3)	0.896	9.6	0.520	2.1
Frank copula (1)	0.50	2.0	0.893	9.3
Frank copula (2)	0816	5.4	0.816	5.4
Frank copula (3)	0.890	9.1	0.557	2.3

*Not the annual series nonexceedance probabilities and return periods. These are conservative.

Table C.5b. 10-percent Exceedance Combinations for Gage Pair 06 (POT).

	Main*		Tributary	
Method (ID)	**Nonexceedance Probability (u)**	**Return Period**	**Nonexceedance Probability (v)**	**Return Period**
Log-normal (1)	0.50	2.0	0.897	9.7
Log-normal (2)	0.829	5.8	0.829	5.8
Log-normal (3)	0.890	9.1	0.628	2.7
Bivariate Gumbel (1)	na	na	na	na
Bivariate Gumbel (2)	na	na	na	na
Bivariate Gumbel (3)	na	na	na	na
GH copula (1)	0.50	2.0	0.897	9.7
GH copula (2)	0.850	6.7	0.850	6.7
GH copula (3)	0.896	9.6	0.574	2.3
Frank copula (1)	0.50	2.0	0.895	9.5
Frank copula (2)	0.821	5.6	0.821	5.6
Frank copula (3)	0.890	9.1	0.590	2.4

*Not the annual series nonexceedance probabilities and return periods. These are conservative.

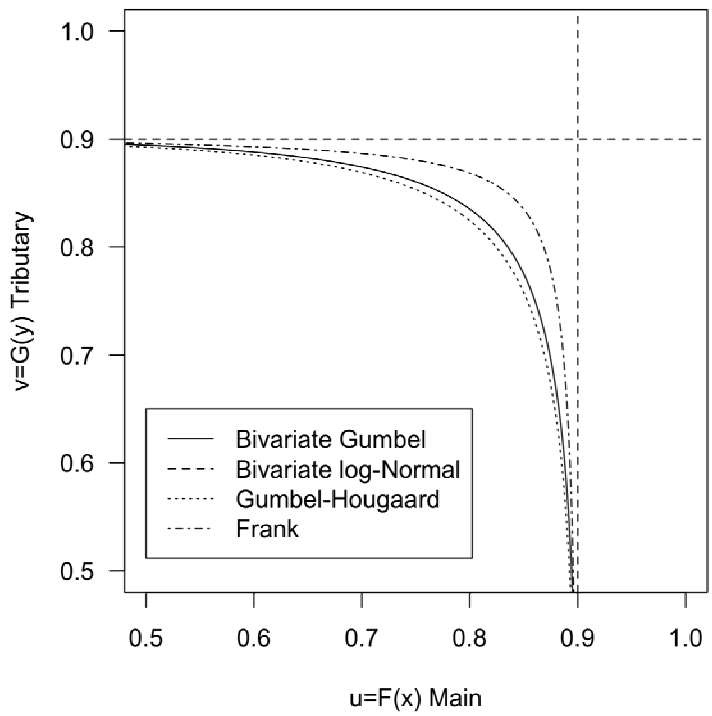

Figure C.11a. 10-percent Exceedance Probability for Gage Pair 43 (POM).

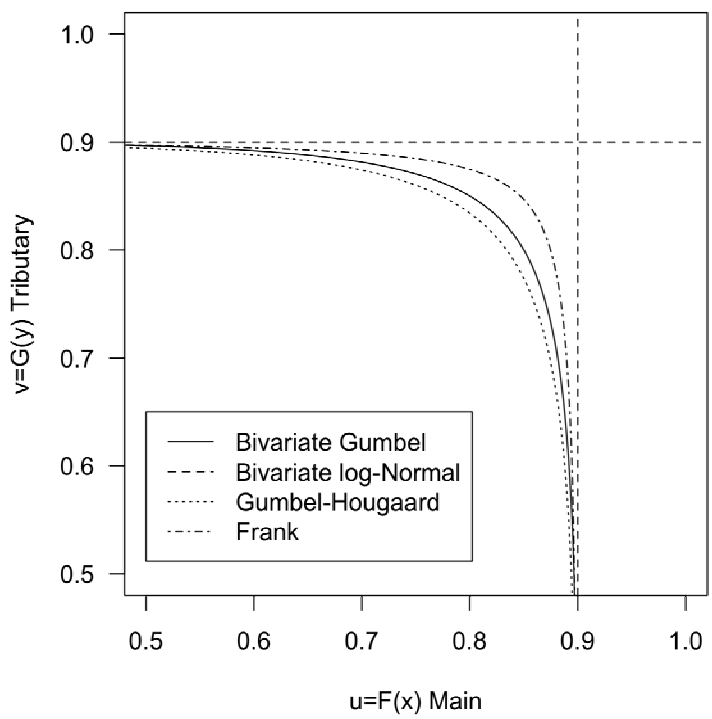

Figure C.11b. 10-percent Exceedance Probability for Gage Pair 43 (POT).

Table C.6a. 10-percent Exceedance Combinations for Gage Pair 43 (POM).

	Main		Tributary*	
Method (ID)	**Nonexceedance Probability (u)**	**Return Period**	**Nonexceedance Probability (v)**	**Return Period**
Log-normal (1)	0.50	2.0	0.894	9.4
Log-normal (2)	0.819	5.5	0.819	5.5
Log-normal (3)	0.890	9.1	0.575	2.4
Bivariate Gumbel (1)	na	na	na	na
Bivariate Gumbel (2)	na	na	na	na
Bivariate Gumbel (3)	na	na	na	na
GH copula (1)	0.50	2.0	0.896	9.6
GH copula (2)	0.843	6.4	0.843	6.4
GH copula (3)	0.890	9.1	0.656	2.9
Frank copula (1)	0.50	2.0	0.892	9.3
Frank copula (2)	0.813	5.3	0.813	5.3
Frank copula (3)	0.890	9.1	0.541	2.2

*Not the annual series nonexceedance probabilities and return periods. These are conservative.

Table C.6b. 10-percent Exceedance Combinations for Gage Pair 43 (POT).

	Main*		Tributary	
Method (ID)	**Nonexceedance Probability (u)**	**Return Period**	**Nonexceedance Probability (v)**	**Return Period**
Log-normal (1)	0.50	2.0	0.897	9.7
Log-normal (2)	0.828	5.8	0.828	5.8
Log-normal (3)	0.890	9.1	0.627	2.7
Bivariate Gumbel (1)	na	na	na	na
Bivariate Gumbel (2)	na	na	na	na
Bivariate Gumbel (3)	na	na	na	na
GH copula (1)	0.50	2.0	0.897	9.7
GH copula (2)	0.848	6.6	0.848	6.6
GH copula (3)	0.896	9.6	0.550	2.2
Frank copula (1)	0.50	2.0	0.894	9.4
Frank copula (2)	0.818	5.5	0.818	5.5
Frank copula (3)	0.890	9.1	0.575	2.4

*Not the annual series nonexceedance probabilities and return periods. These are conservative.

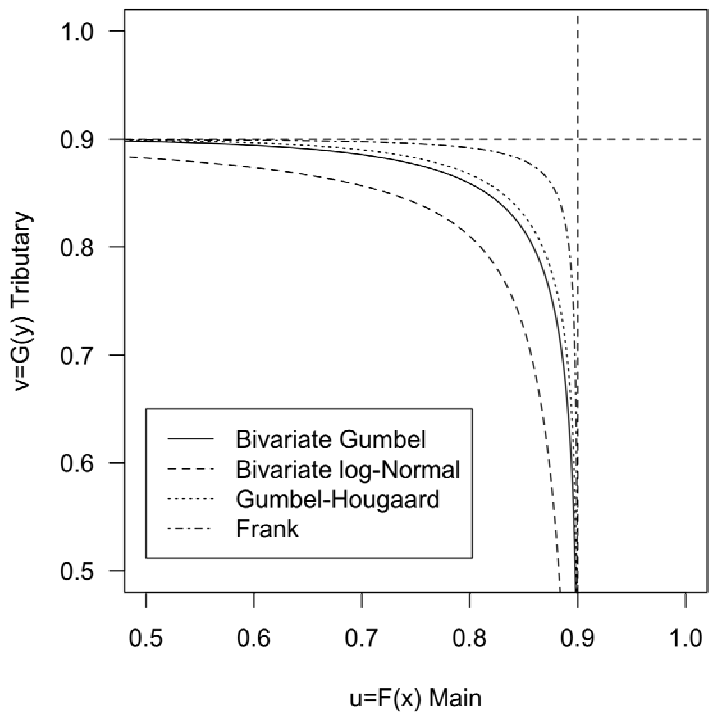

Figure C.12a. 10-percent Exceedance Probability for Gage Pair 76 (POM).

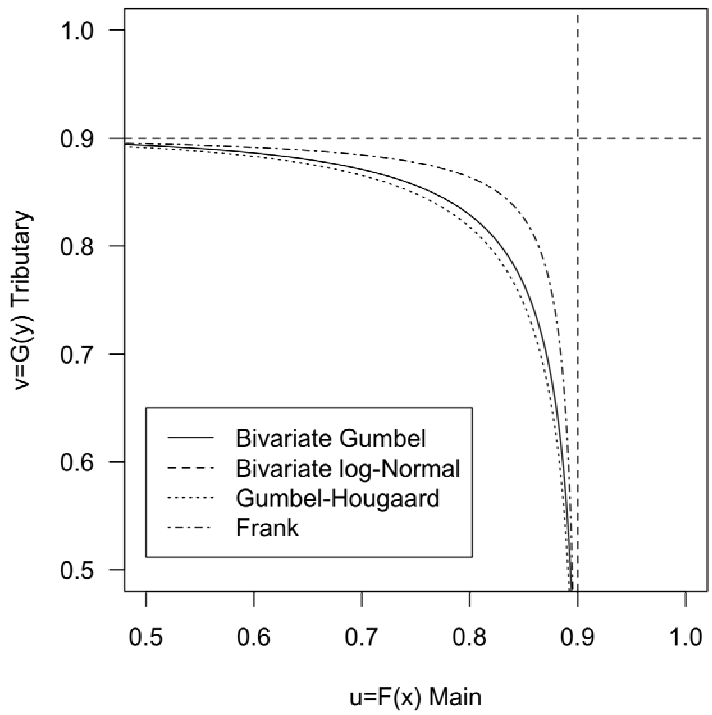

Figure C.12b. 10-percent Exceedance Probability for Gage Pair 76 (POT).

Table C.7a. 10-percent Exceedance Combinations for Gage Pair 76 (POM).

	Main		Tributary*	
Method (ID)	**Nonexceedance Probability (u)**	**Return Period**	**Nonexceedance Probability (v)**	**Return Period**
Log-normal (1)	0.50	2.0	0.898	9.8
Log-normal (2)	0.834	6.0	0.834	6.0
Log-normal (3)	0.890	9.1	0.660	2.9
Bivariate Gumbel (1)	0.50	2.0	0.882	8.5
Bivariate Gumbel (2)	0.825	5.7	0.825	5.7
Bivariate Gumbel (3)	0.880	8.3	0.533	2.1
GH copula (1)	0.50	2.0	0.900	10.0
GH copula (2)	0.868	7.6	0.868	7.6
GH copula (3)	0.890	9.1	0.814	5.4
Frank copula (1)	0.50	2.0	0.899	9.9
Frank copula (2)	0.840	6.3	0.840	6.3
Frank copula (3)	0.890	9.1	0.700	3.3

*Not the annual series nonexceedance probabilities and return periods. These are conservative.

Table C.7b. 10-percent Exceedance Combinations for Gage Pair 76 (POT).

	Main*		Tributary	
Method (ID)	**Nonexceedance Probability (u)**	**Return Period**	**Nonexceedance Probability (v)**	**Return Period**
Log-normal (1)	0.50	2.0	0.893	9.3
Log-normal (2)	0.815	5.4	0.815	5.4
Log-normal (3)	0.890	9.1	0.554	2.2
Bivariate Gumbel (1)	na	na	na	na
Bivariate Gumbel (2)	na	na	na	na
Bivariate Gumbel (3)	na	na	na	na
GH copula (1)	0.50	2.0	0.895	9.5
GH copula (2)	0.839	6.2	0.839	6.2
GH copula (3)	0.896	9.6	0.627	2.7
Frank copula (1)	0.50	2.0	0.891	9.2
Frank copula (2)	0.809	5.2	0.809	5.2
Frank copula (3)	0.890	9.1	0.518	2.1

*Not the annual series nonexceedance probabilities and return periods. These are conservative.

Comparing the results from the POM and POT datasets, it is observed that those for the POT dataset point to modestly more severe conditions than the POM dataset. For example, for the log-normal (3) case, the tributary return period from POM dataset is 9.3 years while for the POT dataset the comparable value is 9.7 years. The POT dataset has a higher Pearson's ρ, which results in more severe conditions for the log-normal method. Similarly, the POT dataset has a higher Kendall's τ, which results in more severe conditions for both copula methods.

C.7 Correlation Parameter Regression

Application of bivariate distributions or copulas for confluent watersheds without gages requires an estimate of Pearson's ρ for the former and Kendall's τ for the latter. This section describes the method and results of the regression analyses to develop relations between measurable

watershed and meteorological characteristics and the calculated values of Pearson's ρ and Kendall's τ from both the POM and POT datasets.

C.7.1 Data Adjustments

Initial evaluation of the computed Pearson's ρ and Kendall's τ from the POM and POT datasets revealed the presence of a small number of outliers. In addition, comparing correlations between the POM and POT datasets revealed that the definition of "coincident" as the daily flow occurring on the same day led to missing some coincident events. To address this, the coincident window was adjusted to the same day, or the day before or the day after, that is a three-day window. Overall, the adjustments were not substantial. The adjusted values for ρ and τ will be used in subsequent analyses.

The results on Pearson's ρ and Kendall's τ for the Peak on Main (POM) dataset are summarized in Table C.8. Pearson's ρ changed substantially for gage pairs 16, 35, and 80. For the latter two, the change primarily resulted from dropping the outlier rather than expanding the window for "coincident" to three days. For all gage pairs except for number 16, adjusting the coincident window did not have a pronounced effect on the Pearson's ρ. Neither changing the coincident window nor removal of outliers changed Kendall's τ substantially except for Gage Pair 16.

Table C.8. Correlation Adjustments: Peak on Main.

Gage Pair	n	# of Adjusted Events	# of Outliers Removed	ρ	Adjusted ρ	τ	Adjusted τ
5[1]	27	-	-	0.002	0.100	0.110	-
16	17	2	0	0.310	0.632	0.353	0.559
31	50	3	0	0.322	0.314	0.206	0.206
33[1]	28	-	-	0.003	0.100	0.289	-
35	27	3	1	0.263	0.637	0.227	0.277
38	36	1	0	0.326	0.334	0.320	0.327
39	25	1	0	0.922	0.922	0.388	0.375
64	71	14	0	0.352	0.414	0.274	0.309
65	36	10	0	0.472	0.637	0.356	0.305
66	29	4	0	0.620	0.628	0.271	0.232
67	37	8	0	0.824	0.832	0.610	0.602
69	26	2	0	0.723	0.723	0.340	0.360
80	31	3	1	0.161	0.419	0.224	0.246
84	30	5	0	0.606	0.652	0.461	0.472
86	57	8	0	0.334	0.404	0.246	0.272

1. ρ set to minimum of 0.100.

The results on Pearson's ρ and Kendall's τ for the Peak on Tributary (POT) datasets are summarized in Table C.9. Pearson's ρ changed substantially for gage pairs 67 and 69. The change primarily resulted from dropping the outlier rather than expanding the window for "coincident" to three days. For all gage pairs changing the window did not have a pronounced effect on the Pearson's ρ. Neither changing the coincident window nor removal of outliers changed Kendall's τ substantially.

Table C.9. Correlation Adjustments: Peak on Tributary.

Gage Pair	n	# of Adjusted Events	# of Outliers Removed	ρ	Adjusted ρ	τ	Adjusted τ
16	17	2	0	0.699	0.644	0.544	0.529
31	50	3	0	0.388	0.436	0.262	0.298
35	27	3	0	0.779	0.742	0.498	0.497
38	36	1	0	0.620	0.637	0.496	0.506
39	25	1	0	0.912	0.912	0.511	0.491
64	71	14	0	0.750	0.717	0.588	0.554
65	36	10	0	0.740	0.759	0.461	0.434
66	29	4	0	0.361	0.388	0.084	0.084
67	37	8	1	0.361	0.795	0.563	0.578
69	26	2	1	0.346	0.831	0.432	0.459
80	31	3	0	0.472	0.475	0.319	0.307
84	30	5	0	0.378	0.426	0.312	0.336
86	57	8	0	0.538	0.535	0.430	0.435

1. ρ set to minimum of 0.100.

C.7.2 Regression on Area Ratio

Regression analyses were initially attempted on a single variable that had historically been considered important in joint probability analysis and was relatively highly correlated. This established a baseline from which other regression relations would be evaluated.

C.7.2.1 All Data

For the peaks on main (POM) dataset a linear regression on area ratio, R_A, yielded the following result for Pearson's ρ (Equation C.46a) and Kendall's τ (Equation C.46b) with all 85 gage pairs.

$$\rho = 0.87(R_A)^{-0.16} \qquad (R^2=0.329) \tag{C.46a}$$

$$\tau = 0.64(R_A)^{-0.12} \qquad (R^2=0.225) \tag{C.46b}$$

where,

R_A = ratio of the main watershed area to the tributary watershed area

The R^2 values for these equations are weaker than would be ideal. A weak relation with area ratio is observed in Figures C.13a and C.13b, though the data are scattered.

For the POT dataset a linear regression on area ratio yields the following result for Pearson's ρ (Equation C.46c) and Kendall's τ (Equation C.46d) with all 85 gage pairs.

$$\rho = 0.83(R_A)^{-0.08} \qquad (R^2=0.236) \tag{C.46c}$$

$$\tau = 0.63(R_A)^{-0.09} \qquad (R^2=0.148) \tag{C.46d}$$

The R^2 values for these equations are also weaker than would be ideal. A weak relation with area ratio is observed in Figures C.13c and C.13d.

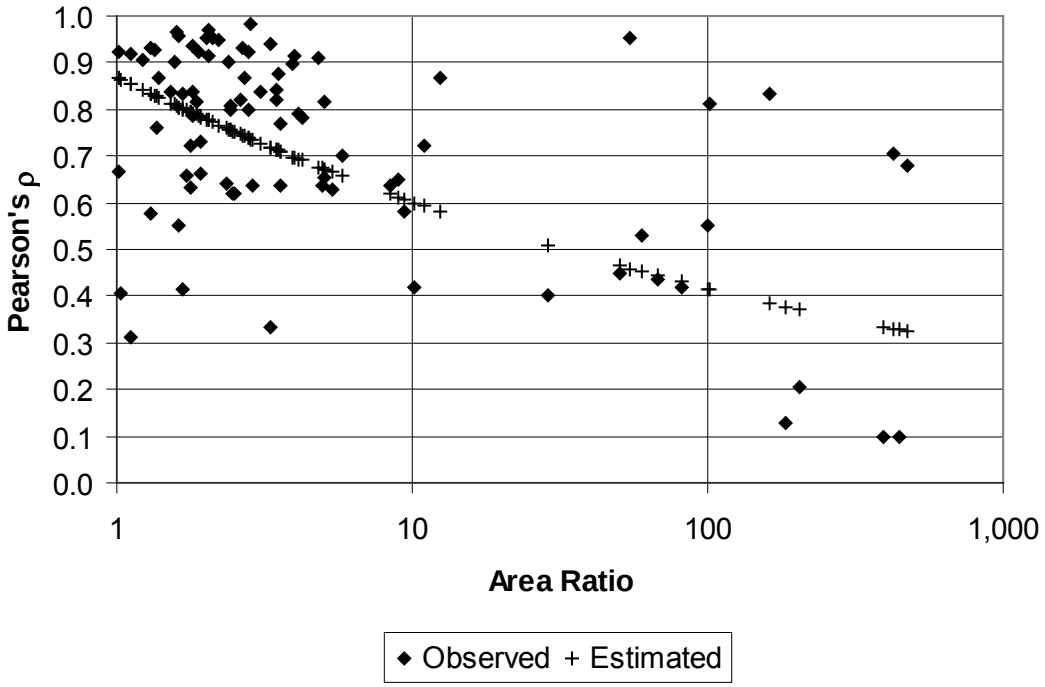

Figure C.13a. Pearson's ρ versus Area Ratio (POM).

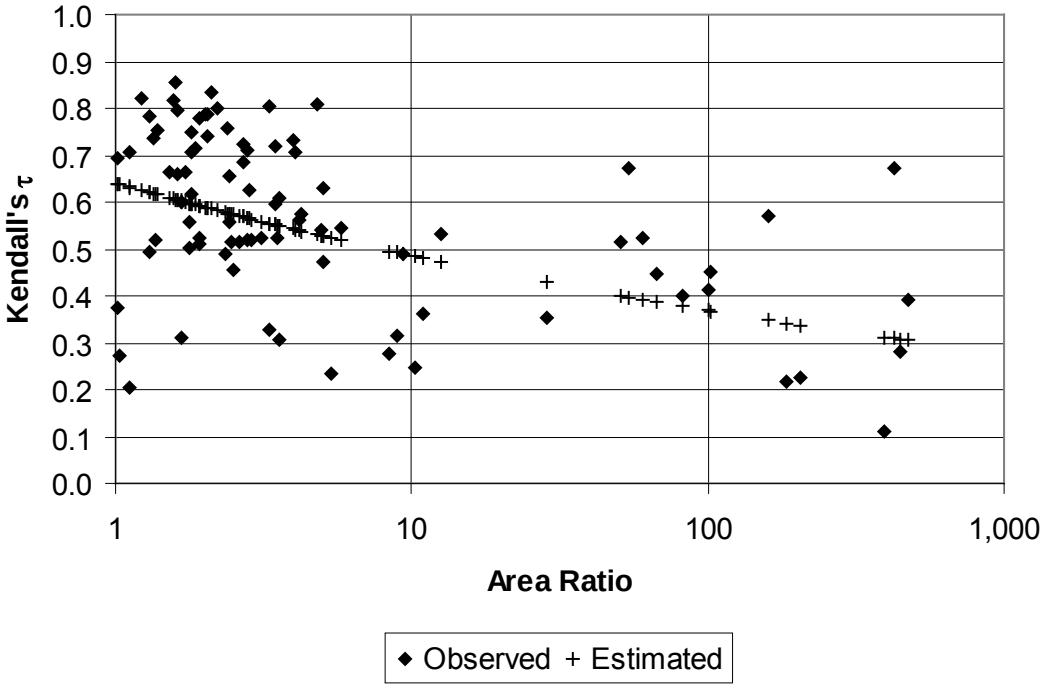

Figure C.13b. Kendall's τ versus Area Ratio (POM).

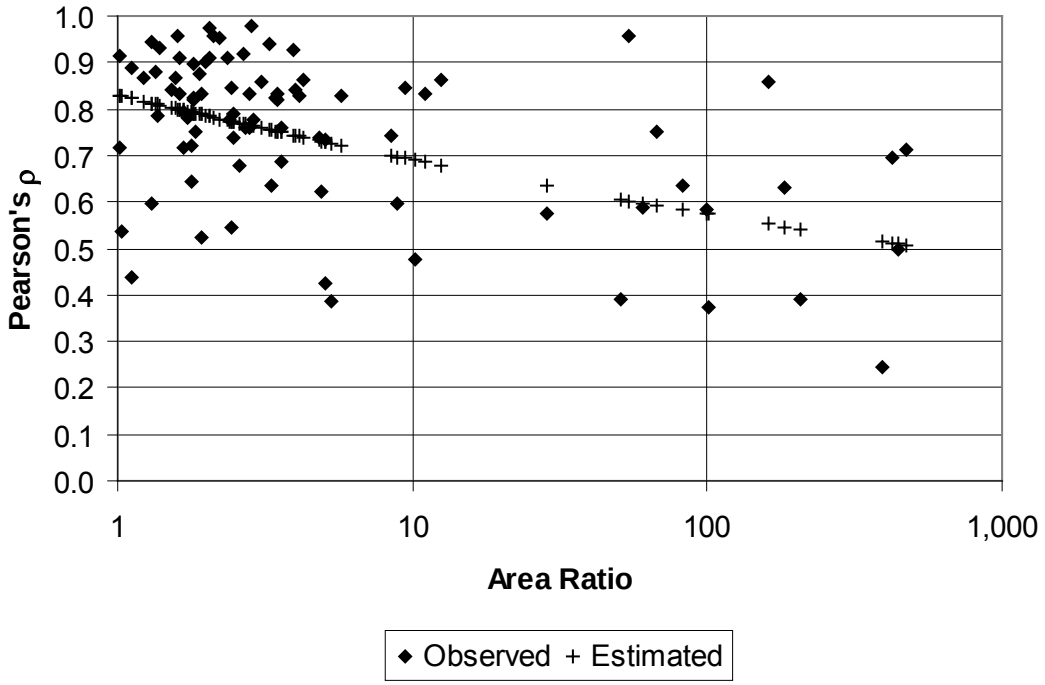

Figure C.13c. Pearson's ρ versus Area Ratio (POT).

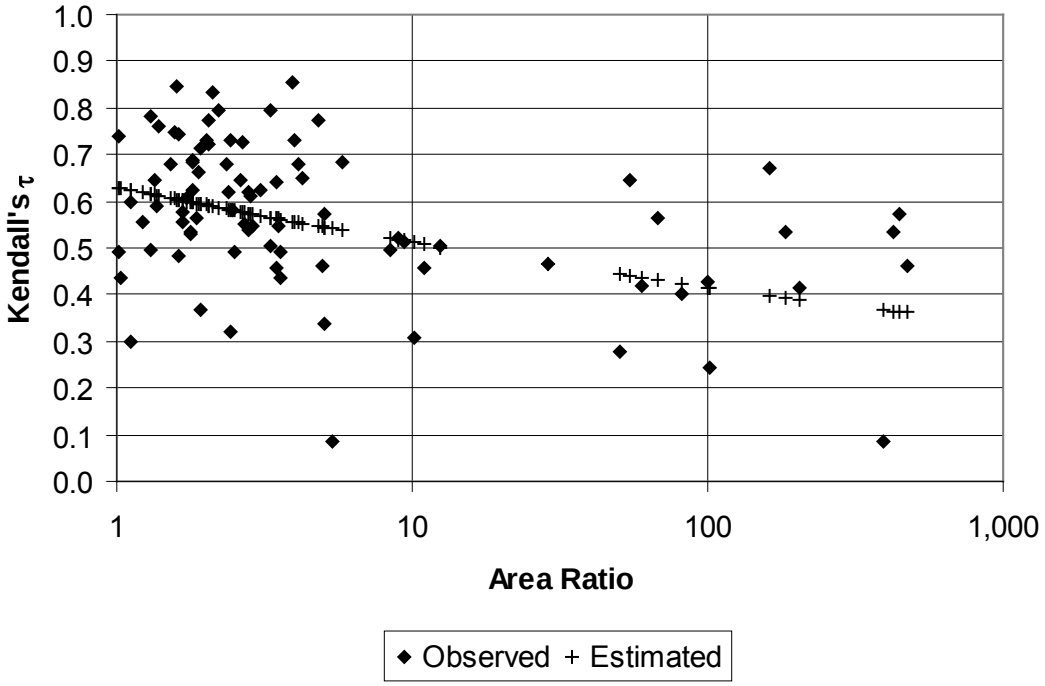

Figure C.13d. Kendall's τ versus Area Ratio (POT).

C.7.2.2 Filters

A series of filters were tested to examine if certain types of gage pairs obscured relations between the correlation parameters and area ratio. Several trials were made to screen out the following types of data:

1. A_M > 1000 mi^2; or A_M > 2000 mi^2. Hypothesis: The main watershed area (A_M) may be larger than the storm cells causing most flood events.
2. A_T > 400 mi^2. Hypothesis: The tributary watershed area (A_T) above a particular threshold may cause alternative driving factors for coincident flooding.
3. A_{TOT} > 1000 mi^2; or A_{TOT} > 2000 mi^2. Hypothesis: The total watershed area (A_{TOT}) may be larger than the storm cells causing most flood events.
4. Q_{2M}/Q_{2T} < 0.83. Hypothesis: If the estimated Q_2 of the main watershed was substantially lower than the Q_2 of the smaller tributary watershed, the gage pair is anomalous.
5. d_c > 20 mi; or d_c > 30 mi. Hypothesis: The distance between the main and tributary watershed centroids (d_c) may be too large compared with the storm cell size to be sensitive to coincident flooding.

The correlation parameters for the gage pairs versus the total watershed area, A_{TOT}, are shown in Figures C.14a, b, c, and d.

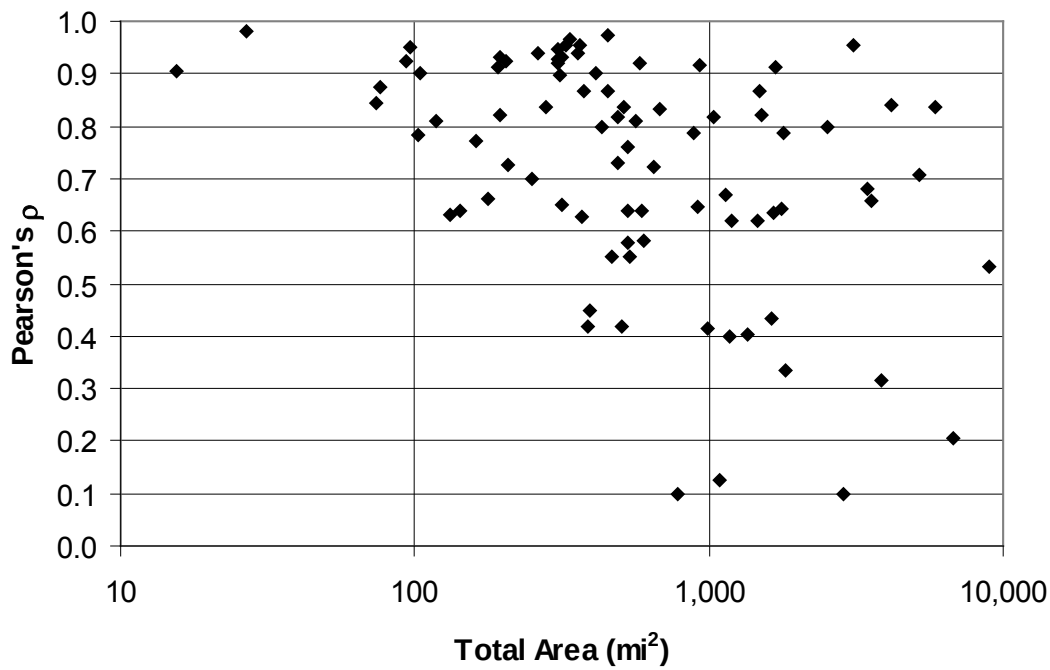

Figure C.14a. Pearson's ρ versus Total Area (POM).

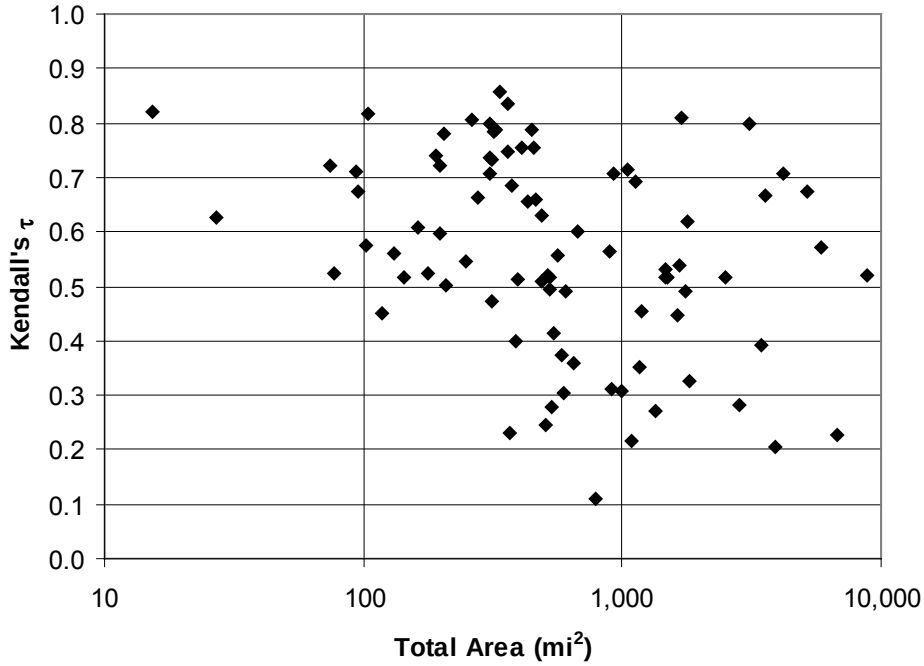

Figure C.14b. Kendall's τ versus Total Area (POM).

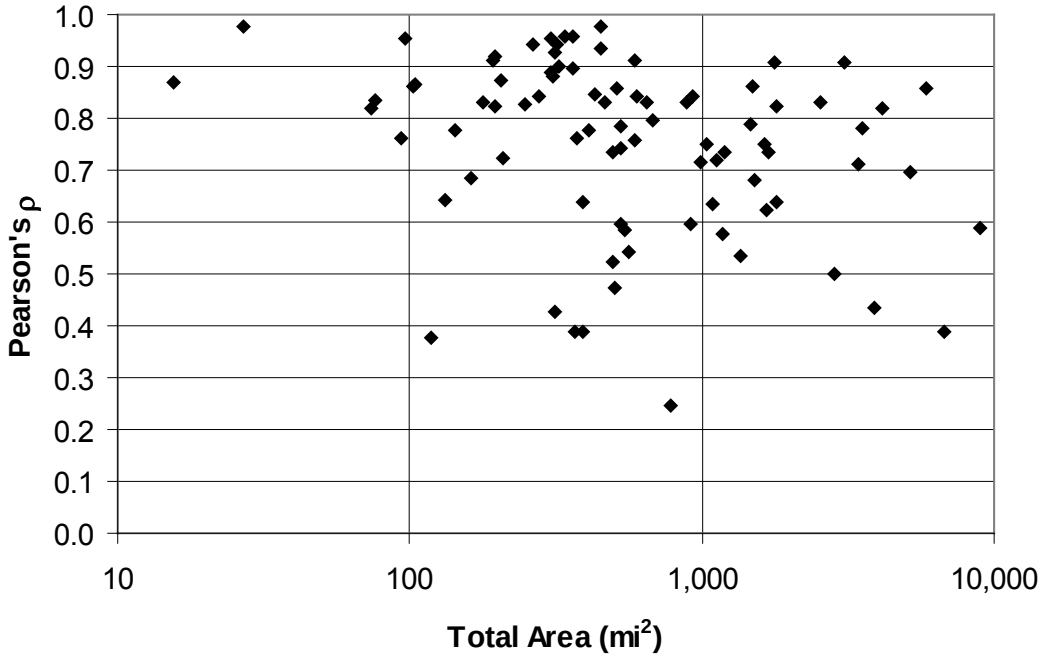

Figure C.14c. Pearson's ρ versus Total Area (POT).

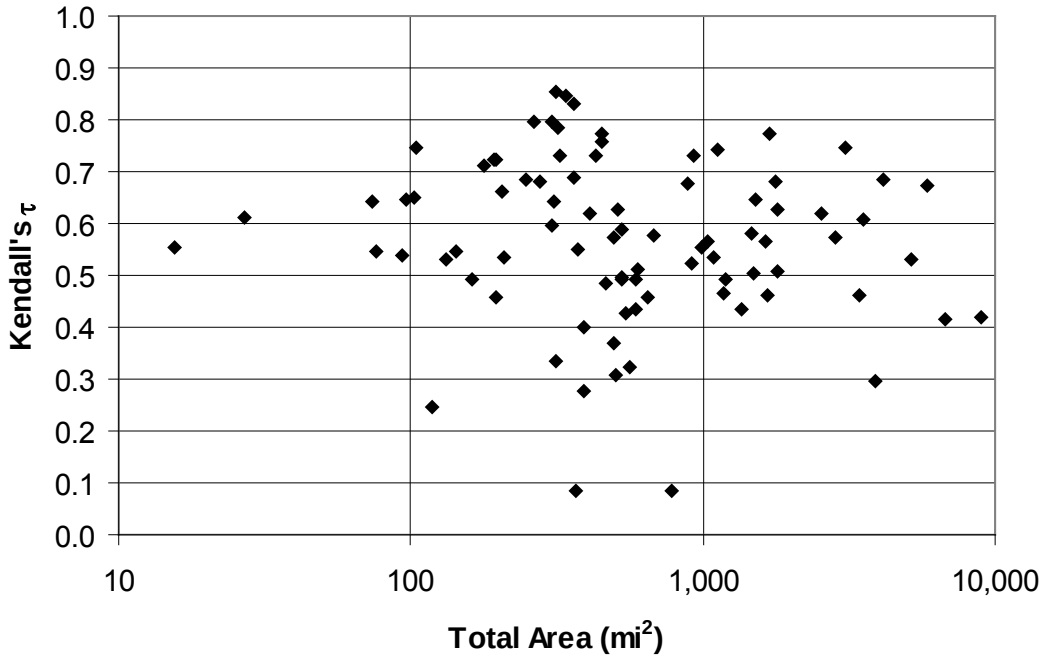

Figure C.14d. Kendall's τ versus Total Area (POT).

The most successful filter for the POM dataset, in terms of a high R^2, was to retain only the gage pairs with $A_M < 2000$ mi^2 and $A_T < 400$ mi^2. This filter eliminated 18 gage pairs leaving 67 gage pairs. The following equations were the result of this filter.

POM dataset:

$$\rho = 0.97(AR)^{-0.21} \qquad (R^2=0.475) \qquad \text{(C.47a)}$$

$$\tau = 0.71(AR)^{-0.18} \qquad (R^2=0.369) \qquad \text{(C.47b)}$$

POT dataset:

$$\rho = 0.88(AR)^{-0.12} \qquad (R^2=0.339) \qquad \text{(C.47c)}$$

$$\tau = 0.67(AR)^{-0.16} \qquad (R^2=0.256) \qquad \text{(C.47d)}$$

For the POT dataset, the most successful filter was to retain only the gage pairs with $d_c < 20$ miles. This filter eliminated 30 gage pairs leaving 55 gage pairs. The following equations were the result of this filter.

POM dataset:

$$\rho = 0.99(AR)^{-0.17} \qquad (R^2=0.388) \qquad \text{(C.48a)}$$

$$\tau = 0.71(AR)^{-0.16} \qquad (R^2=0.333) \qquad \text{(C.48b)}$$

POT dataset:

$$\rho = 0.91(AR)^{-0.15} \qquad (R^2=0.408) \qquad \text{(C.48c)}$$

$$\tau = 0.71(AR)^{-0.21} \qquad (R^2=0.351) \qquad \text{(C.48d)}$$

The information in Equations C.46, C.47, and C.48 is summarized in Table C.10. In all cases, filtering improved the R^2 values, quite substantially in some cases. Since filtering out "large" watersheds, defined by their drainage areas or centroid separations, always improved the equation fit, an argument may be made that above some size threshold, correlation relationships might be altered.

Table C.10. Summary of Selected Regressions on Area Ratio.

Dataset/correlation parameter	Equation	Coefficient	Exponent	R^2
Peaks on Main/ρ	C.46a	0.87	-0.16	0.329
	C.47a	0.97	-0.21	0.475
	C.48a	0.99	-0.17	0.388
Peaks on Main/τ	C.46b	0.64	-0.12	0.225
	C.47b	0.71	-0.18	0.369
	C.48b	0.71	-0.16	0.333
Peaks on Tributary/ρ	C.46c	0.83	-0.08	0.236
	C.47c	0.88	-0.12	0.339
	C.48c	0.91	-0.15	0.408
Peaks on Tributary/τ	C.46d	0.63	-0.09	0.148
	C.47d	0.67	-0.16	0.256
	C.48d	0.71	-0.21	0.351

C.7.2.3 Functional Adjustments

An alternative strategy to develop predictive relationships for the correlation parameters was with functional adjustments to the independent variables. Using similar reasoning to that for some of the filters, that is, watershed magnitude may influence correlation patterns, various functional adjustments were tested:

1. If $A_M > 2000$ mi^2 set $A_M = 2000$ mi^2.
2. If A_M (or A_T) > 1000 mi^2 set A_M (or A_T) = 0.26 (A_M) + 740.
3. If $A_{TOT} < 200$ mi^2 and $R_A > 5.0$ set $R_A = 5.0$.
4. If $A_{TOT} > 2000$ mi^2 and $R_A < 10.0$ set $R_A = 10.0$.

The first two adjustments attempted to mitigate the influence of very large watersheds by either capping the maximum size or by scaling the area down. The third adjustment was based on the observation that some of the less "well-behaved" gage had small areas with high area ratios that resulted in an under prediction in correlation. The fourth adjustment was based on the observation that some gage pairs with large areas and low drainage area ratios resulted in over predicting correlation.

All of these adjustments improved the R^2 of the resulting predictive relations. In most cases, the small improvements in R^2 did not justify the use of these functional adjustments. However, capping A_M at 2000 mi^2 was considered useful because the area ratios computed from the basins with drainage areas greater than 2000 mi^2 were not representative of the correlation in peak flows.

C.7.2.4 Groups

An additional strategy was to subdivide the gage pair dataset into groups according to a characteristic that might distinguish separate subpopulations. The grouping tested was based on geography. The gage pairs were divided into those in the eastern part of the U.S. versus those in the western part. The west was further subdivided based on whether the pair was in a "wet" or "dry" climate. Two definitions of wet and dry were evaluated based on: 1) the 24-h 2 yr rainfall intensity, and 2) the mean annual precipitation, P_M.

For estimating ρ from the Peaks on Main dataset and using the first definition, with 2.5 inches separating wet and dry, the following relations on area ratio resulted:

East (n = 44):

$$\rho = 0.93(R_A)^{-0.21} \qquad (R^2=0.495) \qquad \text{(C.49a)}$$

West, $I_{2\text{-}24} < 2.5$ (n = 24):

$$\rho = 0.71(R_A)^{-0.00} \qquad (R^2=0.000) \qquad \text{(C.49b)}$$

West, $I_{2\text{-}24} > 2.5$ (n = 17):

$$\rho = 0.90(R_A)^{-0.11} \qquad (R^2=0.393) \qquad \text{(C.49c)}$$

Comparing the results for the above equations with Equation C.46a, one might draw the conclusion that this grouping improves the predictions of Pearson's ρ for the East and West with $I_{2\text{-}24} > 2.5$ while revealing that for the West with $I_{2\text{-}24} < 2.5$ area ratio does not relate at all to Pearson's ρ.

Similarly, for estimating Pearson's ρ from the POM dataset and using the second definition, with 25 inches separating wet and dry, the following relations on area ratio resulted (there is no change in the equation for the East.):

East (n = 44)

$$\rho = 0.93(R_A)^{-0.21} \qquad (R^2=0.495) \qquad \text{(C.50a)}$$

West, $P_M < 25$ (n = 21)

$$\rho = 0.66(R_A)^{-0.01} \qquad (R^2=0.004) \qquad \text{(C.50b)}$$

West, $P_M > 25$ (n = 20)

$$\rho = 0.92(R_A)^{-0.12} \qquad (R^2=0.379) \qquad \text{(C.50c)}$$

Comparing the results for the above equations with Equation C.46a, one might again draw the conclusion that this grouping improves the predictions of Pearson's ρ for the East and West with $P_M > 25$ while revealing that for the West with $P_M < 25$ area ratio does not relate at all to Pearson's ρ. It is also apparent from comparing Equations C.49 b and c with Equations C.50 b and c that the alternative divisions into "wet" and "dry" do not influence the results. The equations are quite similar.

Does the regional grouping represent true differences in the population or does it reflect artifacts in the gage pair dataset? To investigate this, gage pairs were filtered to include only those pairs with $A_M < 2000$ mi^2 and $A_T < 400$ mi^2 and the regional analysis was repeated resulting in:

East (n = 32)

$$\rho = 0.97(R_A)^{-0.22} \qquad (R^2=0.484) \qquad \text{(C.51a)}$$

West, P_M < 25 (n = 16)

$$\rho = 0.93(R_A)^{-0.18} \qquad (R^2=0.370) \qquad \text{(C.51b)}$$

West, P_M > 25 (n = 19)

$$\rho = 0.91(R_A)^{-0.11} \qquad (R^2=0.365) \qquad \text{(C.51c)}$$

Equations C.50a, b, and c may be compared to Equations C.51a, b, and c. Although each group lost at least one gage pair to the filtering, application of the filter had a substantial effect on the West (P_M < 25) group. Area ratio is now a predictor variable and the coefficient, exponent, and R^2 are comparable to the other groupings. Furthermore, each of the group equations C.51a, b, and c are similar to Equation C.47a. Therefore, it is concluded that the apparent improvement by grouping is not indicative of separate populations, at least as it pertains to using area ratio as the predictive variable.

C.7.3 Regression on Other/Multiple Variables

Previous sections focused on area ratio as a predictor of the correlation parameters Pearson's ρ and Kendall's τ. In this section, regression with other single and multiple independent variables is explored.

Tables C.11a, b, c, and d summarize various regression scenarios for Pearson's ρ (POM dataset), Kendall's τ (POM dataset), Pearson's ρ (POT dataset), and Kendall's τ (POT dataset), respectively. The "intercept" column represents the linear regression intercept and the "Coeff." column is the intercept transformed to real space (Coeff = 10^intercept). The remaining columns represent the exponents for the independent variables included in the regression. If a p-value of greater than 0.05 was computed for an intercept or exponent, it is highlighted by shading. This indicates there is a greater than 5 percent chance that the value is zero.

Scenario 1 in each of the tables was given previously as Equations C.46a, b, c, and d, respectively, and is the point of reference for comparison with other scenarios. Scenarios 2 through 5 each add a second independent variable to Scenario 1. A scenario with R_{LS} is not included because it is highly correlated with R_A and would introduce significant cross-correlation. All improve the R^2, though several are not statistically significant as indicated in the tables by the shading when the p-values are greater than 0.05. For Scenarios 1 through 5 the R_A exponent is relatively consistent regardless of the addition of a second variable. For example, in Table C.11a, the exponent ranges only from 0.140 to 0.161 for Scenarios 1 through 5.

Scenario 6 was conceived to include any of the independent variables in Table C.11a if its p-value was less than 0.1. In this way, preconceptions regarding which parameters should be included were avoided. It was expected that R^2 for this scenario would be greater than any other scenarios using the entire database. Inspection of the four tables shows that R_{LS} is retained in all four regressions. However, the second, or in some cases third, independent variable differs from table to table.

R_{LS} was analyzed as a single independent variable as Scenario 7. For Pearson's ρ, the R^2 was close to that with R_A (Scenario 1), but for Kendall's τ, the R^2 improved indicating that R_{LS} may be a better predictor for Kendall's τ while R_A may be a better predictor for Pearson's ρ.

Table C.11a. Regression Scenarios for Pearson's ρ (POM).

ID	Description	n	R^2	Intercept	Coeff.	R_A	d_C	R_{LS}	P_M	$P_{24\text{-}2}$	A_{TOT}
1	All	85	0.329	-0.059	0.87	-0.161					
2	All	85	0.369	0.079	1.20	-0.145	-0.129				
3	All	85	0.334	-0.161	0.69	-0.161			0.069		
4	All	85	0.330	-0.040	0.91	-0.160				-0.045	
5	All	85	0.370	0.150	1.41	-0.141					-0.082
6	All	85	0.375	0.156	1.43			-0.249			-0.086
7	All	85	0.329	-0.065	0.86			-0.284			
8	A_T<400; A_M<2000	67	0.475	-0.014	0.97	-0.208					
9	A_T<400; A_M<2000	67	0.500	0.096	1.25	-0.202	-0.108				
10	A_T<400; A_M<2000	67	0.429	0.187	1.54			-0.285			-0.092
11	A_T<400; A_M<2000	67	0.389	-0.037	0.92			-0.314			
12	A_T>400	12	0.651	-1.341	0.046				0.747		
13	A_T<400	73	0.423	-0.026	0.94	-0.181					
14	A_T<400; A_M>2000 set to 2000	73	0.443	-0.016	0.96	-0.201					
	p-values greater than 0.05 when shaded.										

Table C.11b. Regression Scenarios for Kendall's τ (POM).

ID	Description	n	R^2	Intercept	Coeff.	R_A	d_C	R_{LS}	P_M	$P_{24\text{-}2}$	A_{TOT}
1	All	85	0.225	-0.197	0.64	-0.115					
2	All	85	0.276	-0.063	0.87	-0.100	-0.126				
3	All	85	0.268	-0.488	0.33	-0.116			0.187		
4	All	85	0.229	-0.227	0.59	-0.116				0.071	
5	All	85	0.258	-0.036	0.92	-0.099					-0.063
6	All	85	0.319	-0.344	0.45		-0.093	-0.193	0.158		
7	All	85	0.252	-0.197	0.64			-0.215			
8	A_T<400; A_M<2000	67	0.369	-0.151	0.71	-0.176					
9	A_T<400; A_M<2000	67	0.436	0.021	1.05	-0.166	-0.170				
10	A_T<400; A_M<2000	67	0.432	-0.208	0.62		-0.110	-0.281	0.101		
11	A_T<400; A_M<2000	67	0.386	-0.159	0.69			-0.301			
12	A_T>400	12	0.531	-1.454	0.035				0.757		
13	A_T<400	73	0.313	-0.172	0.67	-0.130					
14	A_T<400; A_M>2000 set to 2000	73	0.340	-0.164	0.69	-0.147					
	p-values greater than 0.05 when shaded.										

Table C.11c. Regression Scenarios for Pearson's ρ (POT).

ID	Description	n	R^2	Intercept	Coeff.	R_A	d_C	R_{LS}	P_M	$P_{24\text{-}2}$	A_{TOT}
1	All	85	0.236	-0.081	0.83	-0.079					
2	All	85	0.244	-0.045	0.90	-0.075	-0.034				
3	All	85	0.253	-0.202	0.63	-0.079			0.078		
4	All	85	0.256	-0.126	0.75	-0.081				0.105	
5	All	85	0.245	-0.026	0.94	-0.073					-0.022
6	All	85	0.258	-0.135	0.73			-0.143		0.120	
7	All	85	0.232	-0.085	0.82			-0.138			
8	A_T<400; A_M<2000	67	0.339	-0.054	0.88	-0.116					
9	A_T<400; A_M<2000	67	0.344	-0.020	0.96	-0.114	-0.034				
10	A_T<400; A_M<2000	67	0.335	-0.114	0.77			-0.191		0.118	
11	A_T<400; A_M<2000	67	0.310	-0.064	0.86			-0.186			
12	A_T>400	12	0.593	-0.783	0.16				0.425		
13	A_T<400	73	0.297	-0.068	0.86	-0.088					
14	A_T<400; A_M>2000 set to 2000	73	0.310	-0.064	0.86	-0.097					
	p-values greater than 0.05 when shaded.										

Scenario 8 was previously reported as Equations C.47a, b, c, and d and represented a filtered version of Scenario 1. A substantial improvement in R^2 was observed in all four cases.

Scenarios 9, 10, and 11 are filtered versions of Scenarios 2, 6, and 7. As expected, filtering improved R^2 in all cases. Comparison of the filtered scenarios revealed two general observations. First, either R_A or R_{LS} are the most substantial contributor to predicting correlation. R_A tends to be a better predictor for Pearson's ρ and R_{LS} tends to be a better predictor for Kendall's τ based on the R^2 values. Second, adding a second independent variable to either R_A or R_{LS} added little value.

Scenario 12 was designed to complement the filtered scenarios (8 through 11) by posing a relation applicable to gage pairs filtered out of those scenarios. A total of 12 gage pairs have a tributary area greater than 400 mi^2. For this group, relatively strong relations were found with mean annual precipitation, P_M, as the independent variable. A hypothesis for this success was that once the smaller watershed size exceeds a certain threshold, the relative size of the main and tributary watersheds is less of a predictor because of limits on the storm cell size.

Scenarios 13 and 14 were evaluated to explore an alternative grouping. The R^2 improves from Scenario 1 through Scenarios 13 and 14. Scenario 14 represents application of a functional adjustment to main areas greater than 2000 mi^2. The hypothesis was that drainage areas greater than 2000 mi^2 do not further affect the correlation. In all cases, the R^2 improved over Scenario 13. For the case of Kendall's τ for the POT dataset, the R^2 is notably lower than the other cases. Inspection of the data revealed that Gage Pair 66 might be an outlier. When removed, the R^2 improves as shown in Table C.11d, Scenario 14a.

Table C.11d. Regression Scenarios for Kendall's τ (POT).

ID	Description	n	R^2	Intercept	Coeff.	R_A	d_C	R_{LS}	P_M	$P_{24\text{-}2}$	A_{TOT}
1	All	85	0.148	-0.204	0.63	-0.091					
2	All	85	0.151	-0.233	0.59	-0.095	0.027				
3	All	85	0.183	-0.460	0.35	-0.092			0.164		
4	All	85	0.170	-0.273	0.53	-0.094				0.162	
5	All	85	0.152	-0.253	0.56	-0.096					0.019
6	All	85	0.224	-0.455	0.35			-0.184	0.164		
7	All	85	0.190	-0.200	0.63			-0.182			
8	A_T<400; A_M<2000	67	0.256	-0.171	0.67	-0.158					
9	A_T<400; A_M<2000	67	0.256	-0.157	0.70	-0.157	-0.014				
10	A_T<400; A_M<2000	67	0.335	-0.397	0.40			-0.293	0.145		
11	A_T<400; A_M<2000	67	0.311	-0.170	0.68			-0.292			
12	A_T>400	12	0.503	-0.932	0.12				0.460		
13	A_T<400	73	0.159	-0.200	0.63	-0.095					
14	A_T<400; A_M>2000 set to 2000	73	0.179	-0.192	0.64	-0.110					
14a	A_T<400; A_M>2000 set to 2000 (dropped GP66)	72	0.255	-0.181	0.66	-0.110					
	p-values greater than 0.05 when shaded.										

C.7.4 Alternative Equation Forms

The preceding analyses employed the conventional linear regression format in log-space:

$$Y = aX_1^{b_1}X_2^{b_2}...X_n^{b_n} \tag{C.52}$$

The use of functional adjustments provided modifiers to selected independent variables, but still employed this fundamental format. Other functional forms were evaluated:

Log-linear:

$$Y = a_0 + a_1\log(X_1) + a_2\log(X_2)...a_n\log(X_n) \tag{C.53}$$

Offset:

$$(Y - Y_0) = aX_1^{b_1}X_2^{b_2}...X_n^{b_n} \tag{C.54}$$

Many other forms are possible, but the limitation for any potential form is the ultimate relation between the independent and dependent variables. For the independent variables collected and developed for this study, the predictive power appears to be limited in any equation form.

C.7.5 Selection of Best-Fit Equations

Weighing the justification for the use of filters and groupings, significance of including multiple independent variables, and the anticipated ease of use for the designer, it was apparent that the most promising sets of equations derived from Scenarios 8, 11, 12, and 14.

Scenarios 8 and 11 applied to watershed pairs where the main stream area is less than 2000 mi^2 and the tributary stream area is less than 400 mi^2. The equations for Pearson's ρ and Kendall's τ are summarized in Table C.12 and Table C.13, respectively. The root mean square error (RMSE) was calculated for each equation and is included in the tables.

Table C.12. Equations for Pearson's ρ (Scenarios 8 and 11).

Dataset	Scenario	Equation	R^2	RMSE
POM	8	$\rho = 0.97(R_A)^{-0.21}$	0.475	0.191
POM	11	$\rho = 0.92(R_{LS})^{-0.31}$	0.389	0.190
POT	8	$\rho = 0.88(R_A)^{-0.12}$	0.339	0.156
POT	11	$\rho = 0.86(R_{LS})^{-0.19}$	0.310	0.154

Table C.13. Equations for Kendall's τ (Scenarios 8 and 11).

Dataset	Scenario	Equation	R^2	RMSE
POM	8	$\tau = 0.71(R_A)^{-0.18}$	0.369	0.147
POM	11	$\tau = 0.69(R_{LS})^{-0.30}$	0.386	0.146
POT	8	$\tau = 0.67(R_A)^{-0.16}$	0.256	0.147
POT	11	$\tau = 0.68(R_{LS})^{-0.29}$	0.311	0.141

Based on R^2, R_A was a better predictor of Pearson's ρ than R_{LS}. (See Table C.12.) Examination of the RMSE results does not support or contradict this conclusion. Similarly, it was previously concluded that R_{LS} was a better predictor of Kendall's τ than R_A. (See Table C.13.) RMSE results are essentially the same for all the dataset and scenario combinations and did not support or contradict this conclusion.

Scenario 12 applied to watershed pairs where the tributary area is greater than 400 mi^2. These equations for Pearson's ρ and Kendall's τ are summarized in Table C.14 and Table C.15, respectively.

Scenarios 8, 11, and 12 do not include a group of watershed pairs with a tributary watershed area of less than 400 mi^2 and a main stream watershed area greater than 2,000 mi^2. This group accounts for 6 of the 85 watershed pairs in the study database. A satisfactory predictive relation for this group was not found for either Pearson's ρ or Kendall's τ, therefore Scenario 14 was created to include these data points. The equations for scenario 14 are summarized in Table C.16 and Table C.17 for Pearson's ρ and Kendall's τ, respectively.

Table C.14. Equations for Pearson's ρ (Scenario 12).

Dataset	Scenario	Equation	R^2	RMSE
POM	12	$\rho = 0.046(P_M)^{0.75}$	0.651	0.143
POT	12	$\rho = 0.16(P_M)^{0.42}$	0.593	0.098

Table C.15. Equations for Kendall's τ (Scenario 12).

Dataset	Scenario	Equation	R^2	RMSE
POM	12	$\tau = 0.035(P_M)^{0.76}$	0.531	0.134
POT	12	$\tau = 0.12(P_M)^{0.46}$	0.503	0.087

Table C.16. Equations for Pearson's ρ (Scenario 14).

Dataset	Scenario	Equation	R^2	RMSE
POM	14	$\rho = 0.96(R_A)^{-0.20}$	0.443	0.206
POT	14	$\rho = 0.86(R_A)^{-0.10}$	0.310	0.150

Table C.17. Equations for Kendall's τ (Scenario 14).

Dataset	Scenario	Equation	R^2	RMSE
POM	14	$\tau = 0.69(R_A)^{-0.15}$	0.340	0.160
POT	14a	$\tau = 0.66(R_A)^{-0.11}$	0.255	0.137

Evaluation of the equations included R^2, RMSE, and practicality. In the final analysis it was recommended to consider the gage pair database to be divided in two groups emphasizing the equations from Scenarios 12 and 14. The two groups are based on the tributary (smaller) watershed area:

1. Drainage area of the tributary watershed is less than 400 mi^2. If the drainage area of the larger watershed is greater than 2000 mi^2, the larger watershed drainage area is taken as 2000 mi^2.
2. Drainage area of the tributary watershed is greater than 400 mi^2.

The recommended best-fit equations based on these two groups are summarized below. Separate equations were developed for the POM and POT datasets.

For the 73 watershed pairs in Group One, drainage area ratio is recommended as the independent variable to estimate correlation at ungaged sites based on Scenario 14. The best-fit equation for estimating Pearson's ρ for the POM case was (R^2 = 0.443):

$$\rho = 0.96(R_A)^{-0.20} \tag{C.55}$$

where,

ρ = Pearson's ρ, dimensionless

R_A = drainage area ratio, dimensionless

The best-fit equation for estimating Pearson's ρ for the POT case was (R^2 = 0.310):

$$\rho = 0.86(R_A)^{-0.10} \tag{C.56}$$

For Kendall's τ, the POM analysis resulted in the best-fit equation of (R^2 = 0.340):

$$\tau = 0.69(R_A)^{-0.15} \tag{C.57}$$

where,

τ = Kendall's τ, dimensionless

The best-fit equation for estimating τ for the POT case was (R^2 = 0.255):

$$\tau = 0.66(R_A)^{-0.11} \tag{C.58}$$

For the 12 watershed pairs in Group Two, a meteorological parameter was more useful than drainage area for estimating the correlation parameters. The recommended equations for this group were based on Scenario 12 and mean annual precipitation, P_M, was selected as the independent variable.

The best-fit equation for estimating Pearson's ρ for the POM dataset was (R^2 = 0.651):

$$\rho = 0.046(P_M)^{0.75} \tag{C.59}$$

where,

P_M = Mean annual precipitation, in

For the POT dataset, the best-fit equation for Pearson's ρ was (R^2 = 0.593):

$$\rho = 0.16(P_M)^{0.42} \tag{C.60}$$

The best-fit equation for estimating Kendall's τ for the POM dataset was (R^2 = 0.531):

$$\tau = 0.035(P_M)^{0.76} \tag{C.61}$$

For the POT dataset, the best-fit equation for Kendall's τ was (R^2 = 0.503):

$$\tau = 0.12(P_M)^{0.46} \tag{C.62}$$

C.7.6 Selection of Envelope Equations

The selected "best-fit" equations exhibit modest R^2 values and an estimate of the prediction error is given by the RMSE results provided in the previous section. Use of the best-fit equations will result in overestimates of the predicted correlation parameter in some cases and underestimates in others. For design purposes, it is desirable to develop a companion set of "envelope" equations that reduce the probability of underestimating the correlation.

Since higher correlation results in a greater likelihood of extreme events occurring simultaneously on both the main and tributary streams, underestimating the correlation creates a risk of under design while overestimating creates a risk of over design. To reduce the risk of

under design envelope equations that encompass 90 percent of the observed data were developed.

Three strategies for modifying the best-fit equations were considered:

1. Modifying the equation coefficient
2. Modifying the independent variable exponent
3. Adding a constant

Each strategy was compared in terms of a visual fit with the data and the resulting RMSE. The preferred method was adding a constant. The following equation defines the method:

$$\text{cor} = C(X)^e + B \tag{C.63}$$

where,

cor	=	represents either Pearson's ρ or Kendall's τ
X	=	independent variable
C	=	equation coefficient
e	=	independent variable exponent
B	=	envelope constant

A summary of envelope constants is found in Table C.18.

Table C.18. Envelope Constants.

Correlation Parameter	Group	Dataset	Envelope Constant
ρ	1	POM	0.19
ρ	1	POT	0.18
ρ	2	POM	0.20
ρ	2	POT	0.13
τ	1	POM	0.23
τ	1	POT	0.23
τ	2	POM	0.21
τ	2	POT	0.09

C.8 Correlation Parameter Groupings

An alternative to the regression relations for estimating correlation parameters was considered. The amount of variance described by the regression equations and measured using the R^2 statistic was such to provide motivation for an alternative method. This alternative groups the gage pairs by selected watershed characteristics to provide an estimate for correlation that is representative of the group.

C.8.1 Site-Specific Dependence

Regression relations for dependence (Pearson's ρ and Kendall's τ) were tested using several meteorological and watershed characteristics as described in Section C.7. In this analysis, three characteristics were explored for grouping the gage pairs: drainage area ratio, total area, and mean annual precipitation.

The distribution of Pearson's ρ versus drainage area ratio for the POM and POT datasets is summarized in Figures C.15a and C.15b, respectively. Similarly, the distribution of Kendall's τ versus drainage area ratio for the POM and POT datasets is summarized in Figures C.16a and C.16b. Although the data are scattered, it appear they can be separated into two populations: one for area ratios less than 7 and one for area ratios greater than 7. For Pearson's ρ, it is more likely to observe a higher dependence when area ratio is less than 7. For Kendall's τ, the tendency toward higher dependence when area ratio is less than 7 is more pronounced. Kendall's τ ranges between 0.2 and 0.9 when area ratio is less than 7 and between 0.2 and 0.7 when area ratio is greater than 7.

Distribution of the correlation parameters versus total area (sum of main and tributary drainage areas) is provided in Figures C.17 and C.18. It appears that gage pairs with smaller total area tend to exhibit higher dependence. Larger total areas appear to exhibit greater variability in their correlation parameter values. Considering Figures C.18a and C.18b, the values of Kendall's τ range between 0.4 and 0.9 for total drainage areas less than 350 mi^2 and between 0.2 and 0.9 for total drainage areas exceeding 350 mi^2.

Plots of correlation parameter versus mean annual precipitation are presented in Figures C.19 and C.20. For mean annual precipitation less than 60 inches there appears to be a wider range of values in the correlation. For Kendall's τ, the values range from 0.2 to 0.9 for mean annual precipitation less than 60, but for precipitation exceeding 60 inches, Kendall's τ ranges from only 0.4 to 0.8.

After analysis of several alternatives, the data were placed in subgroups based on whether the area ratio was greater or less than 7 and the total drainage area was greater or less than 350 mi^2. Subgroupings by mean annual precipitation (MAP) were discarded because only a few gage pairs exist with MAP greater than 60 inches and meaningful evaluation of whether the groups were distinct was not feasible.

The number of observations, mean, standard deviation, and 90th percentile (90% of the values in the group are lower) for the subgroupings are summarized in Tables C.19 and C.20 for Pearson's ρ and in Tables C.21 and C.22 for Kendall's τ. The results listed in the tables show that correlation between the main and tributary, expressed by Pearson's ρ or Kendall's τ, decreases with increasing area ratio and with increasing total area. For example, in Table C.21, the mean Kendall's τ for $R_A < 7$ and $A_{TOT} >$ than 350 mi^2 (0.57) is less than the mean τ for $R_A < 7$ and $A_{TOT} <$ than 350 mi^2 (0.68). Similarly, the mean τ for $R_A > 7$ and $A_{TOT} >$ than 350 mi^2 (0.39) is less than the mean τ for $R_A < 7$ and $A_{TOT} >$ than 350 mi^2 (0.57). Because of the small number of observations for $R_A > 7$ and $A_{TOT} <$ than 350 mi^2 comparisons between this grouping and the others are not meaningful. The same pattern is observed in all four tables.

The t-statistic was used to test whether the differences in the mean values are statistically significant, that is, do they represent separate populations. A summary of the computed t-statistic and the significance threshold (5 percent level) for the POM and POT datasets are provided in Tables C.23 and C.24 for Pearson's ρ and Tables C.25 and C.26 for Kendall's τ. If t is greater than t', the difference in the means is considered statistically significant. Without exception, the differences in the means are statistically significant.

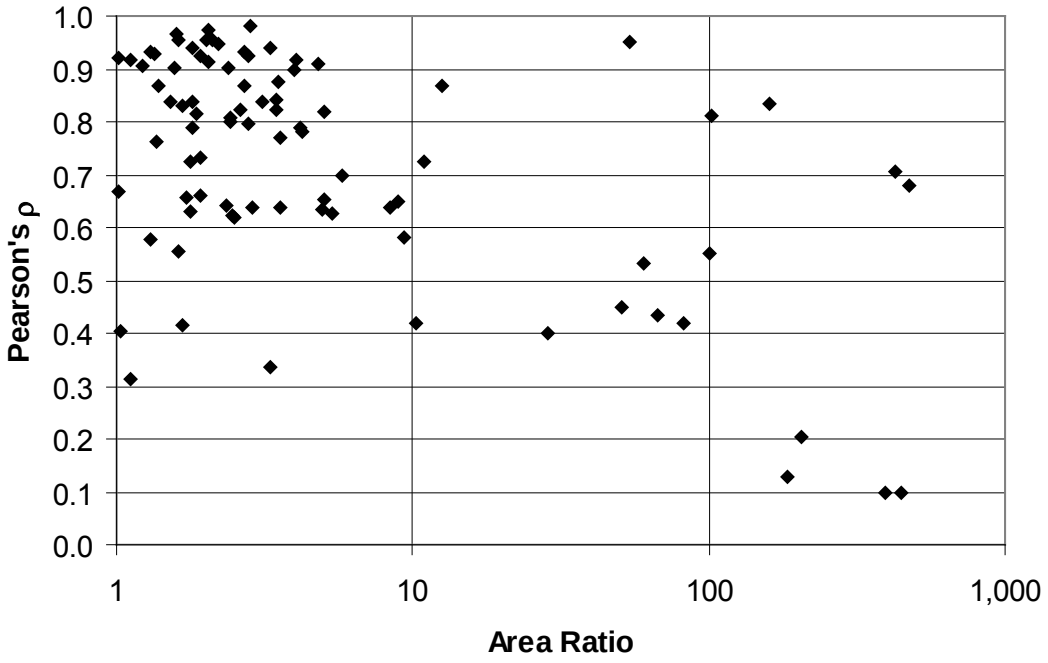

Figure C.15a. Distribution of Pearson's ρ by Area Ratio (POM).

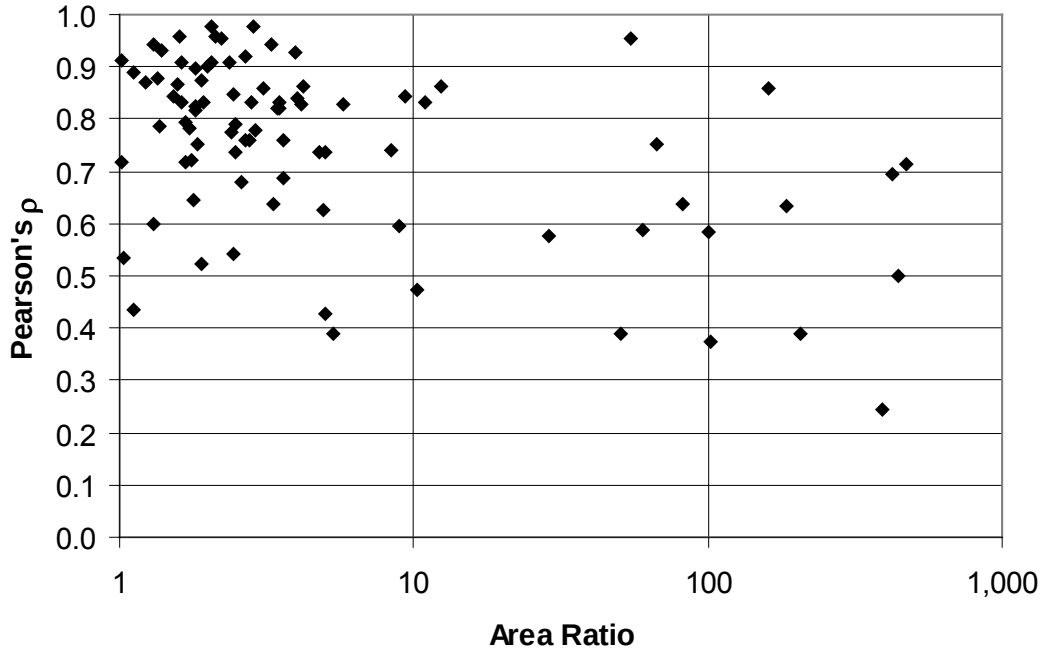

Figure C.15b. Distribution of Pearson's ρ by Area Ratio (POT).

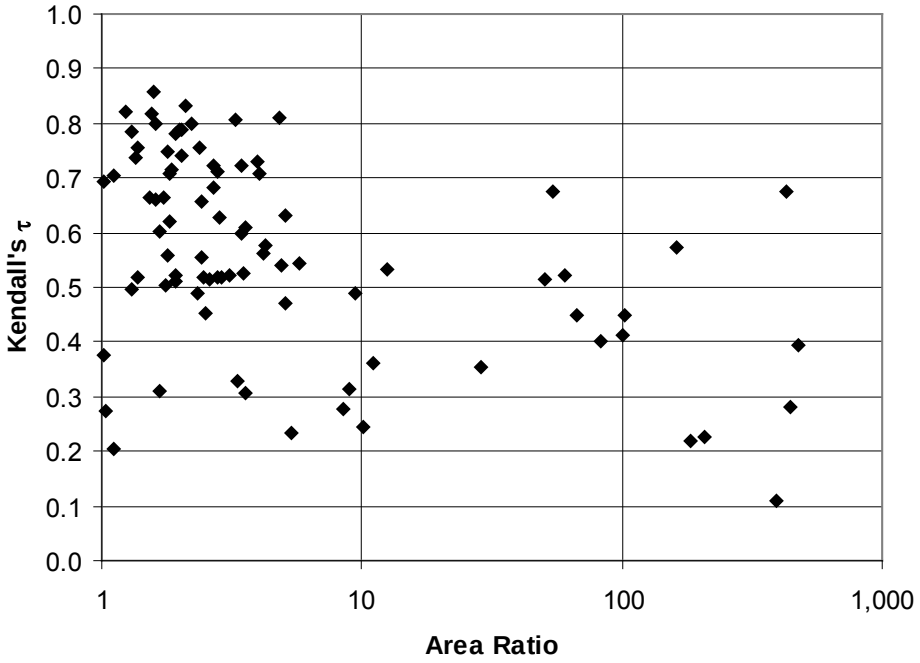

Figure C.16a. Distribution of Kendall's τ by Area Ratio (POM).

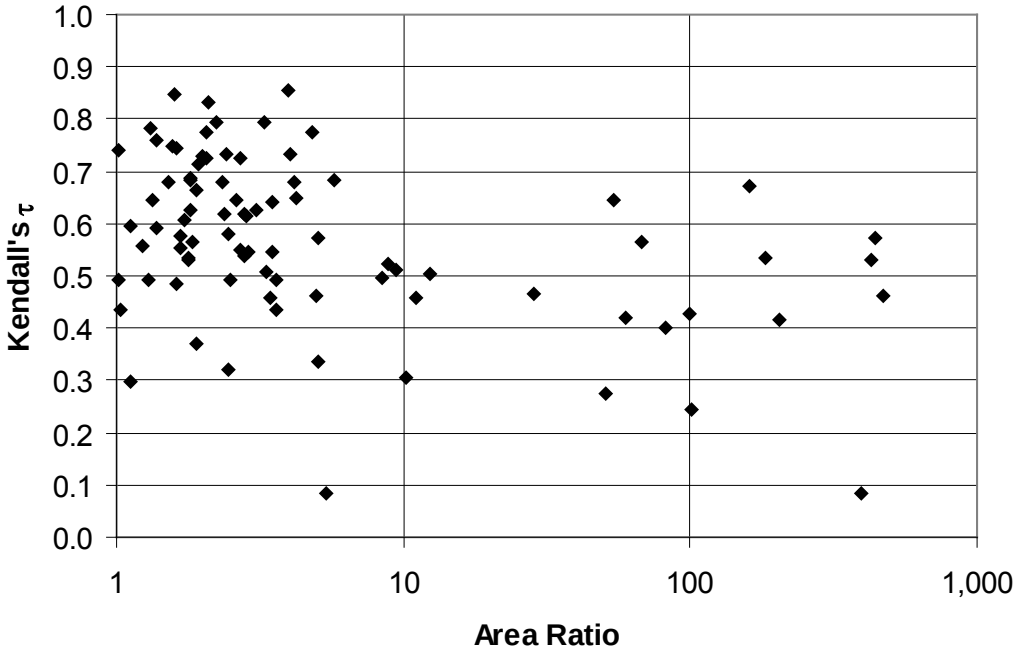

Figure C.16b. Distribution of Kendall's τ by Area Ratio (POT).

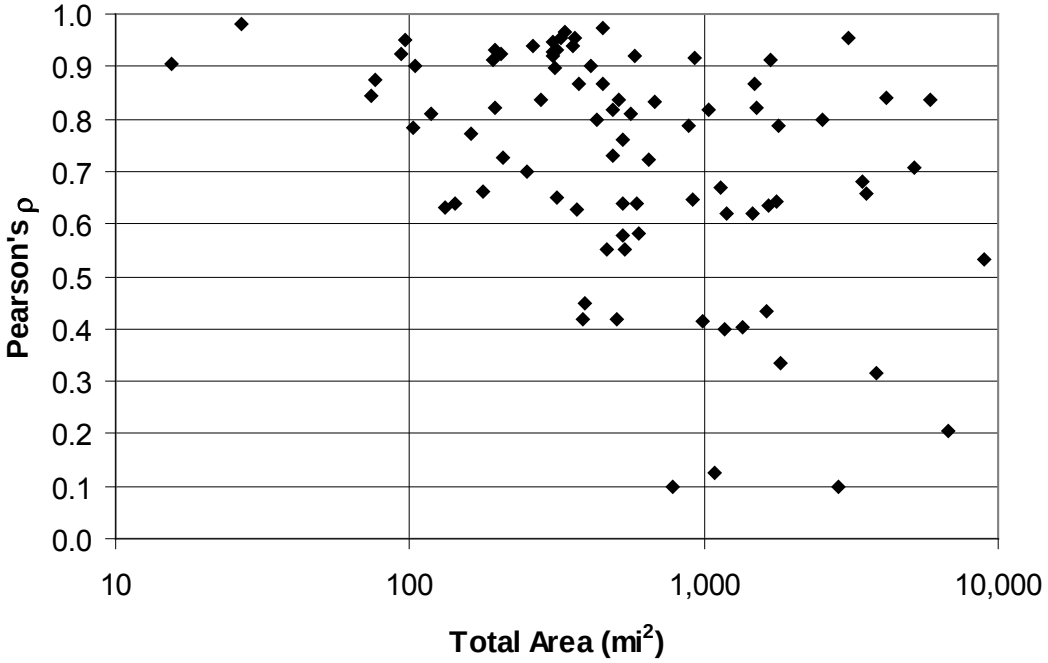

Figure C.17a. Distribution of Pearson's ρ by Total Area (POM).

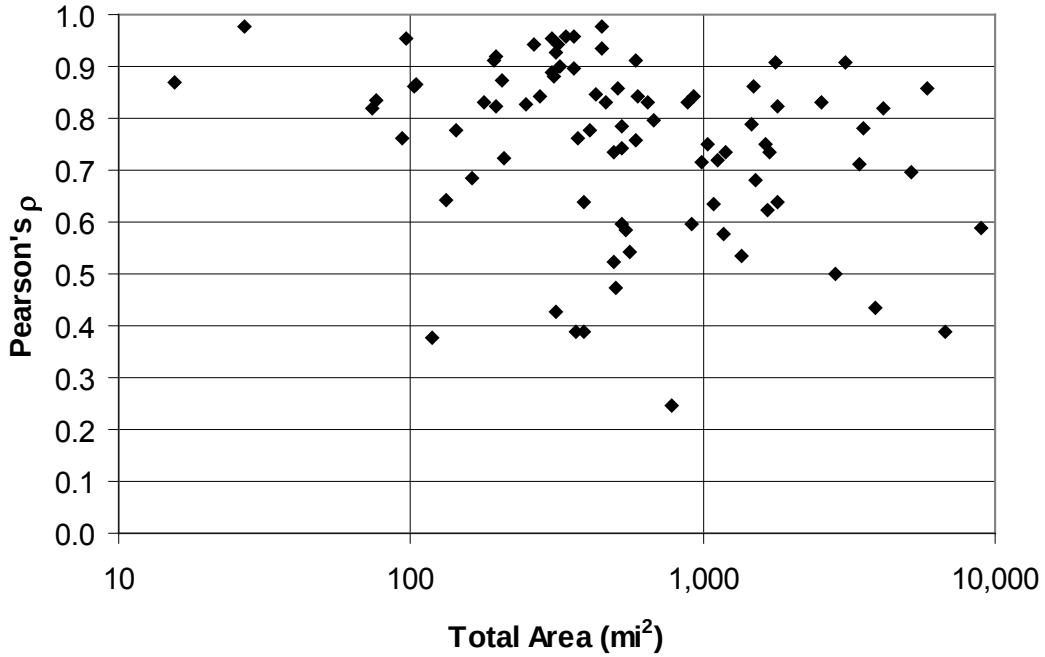

Figure C.17b. Distribution of Pearson's ρ by Total Area (POT).

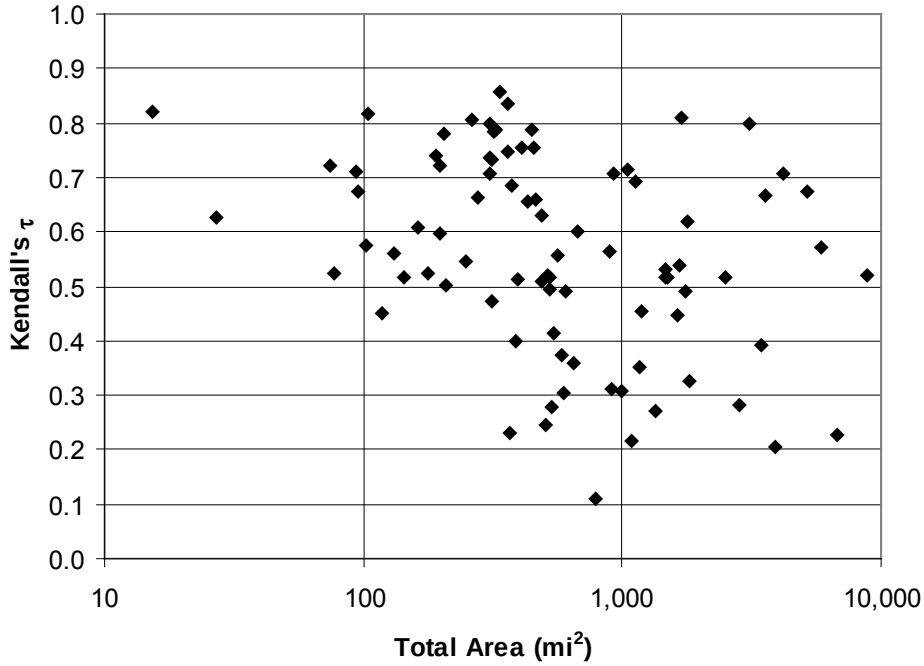

Figure C.18a. Distribution of Kendall's τ by Total Area (POM).

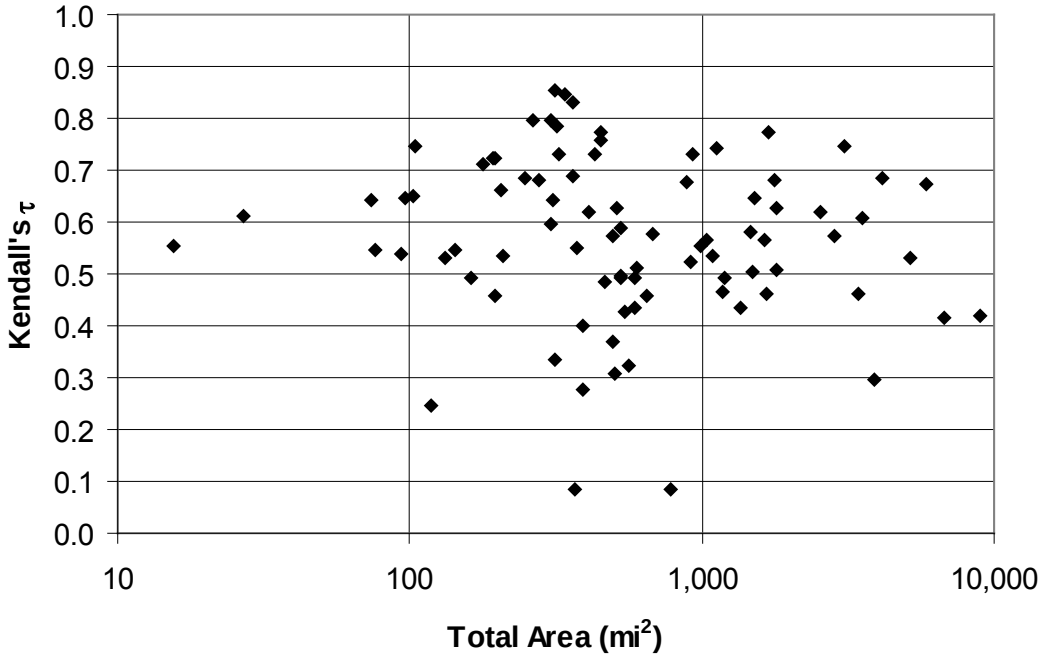

Figure C.18b. Distribution of Kendall's τ by Total Area (POT).

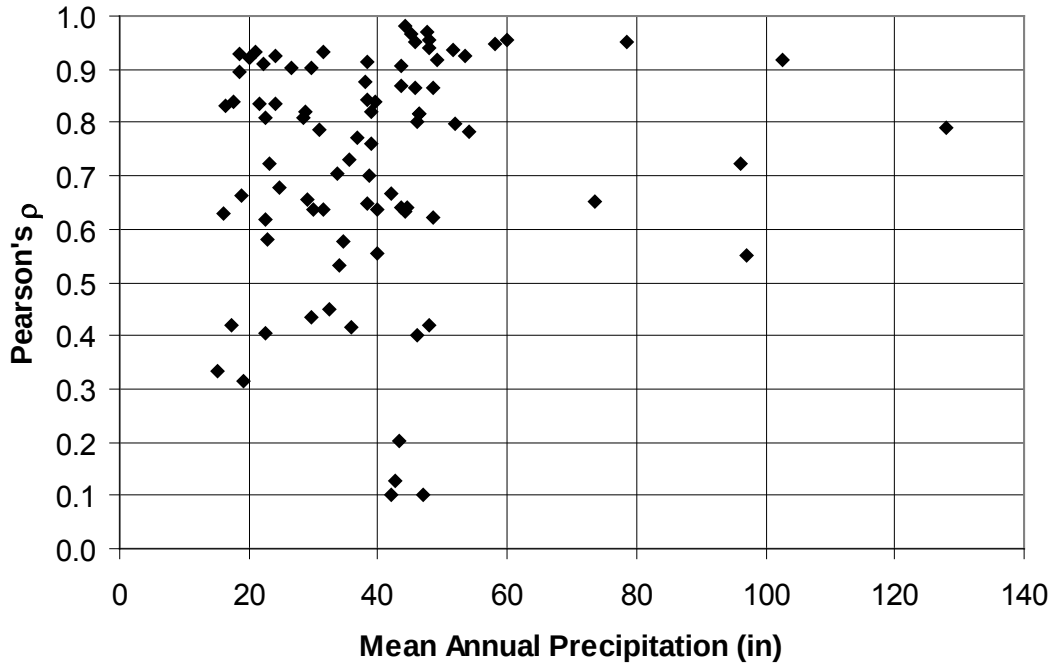

Figure C.19a. Distribution of Pearson's ρ by Mean Annual Precipitation (POM).

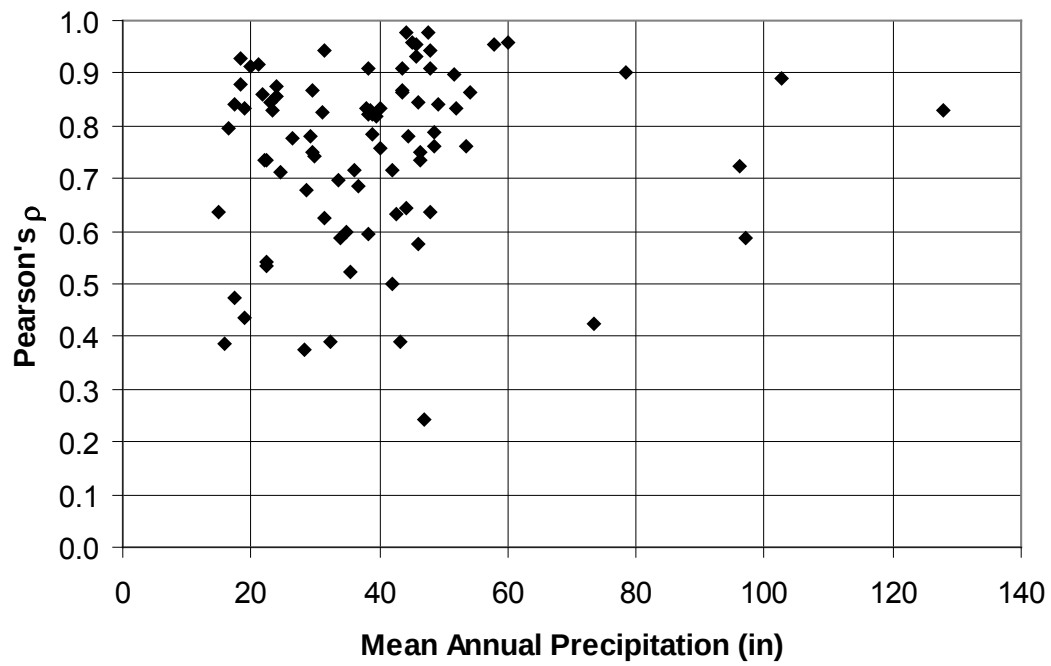

Figure C.19b. Distribution of Pearson's ρ by Mean Annual Precipitation (POT).

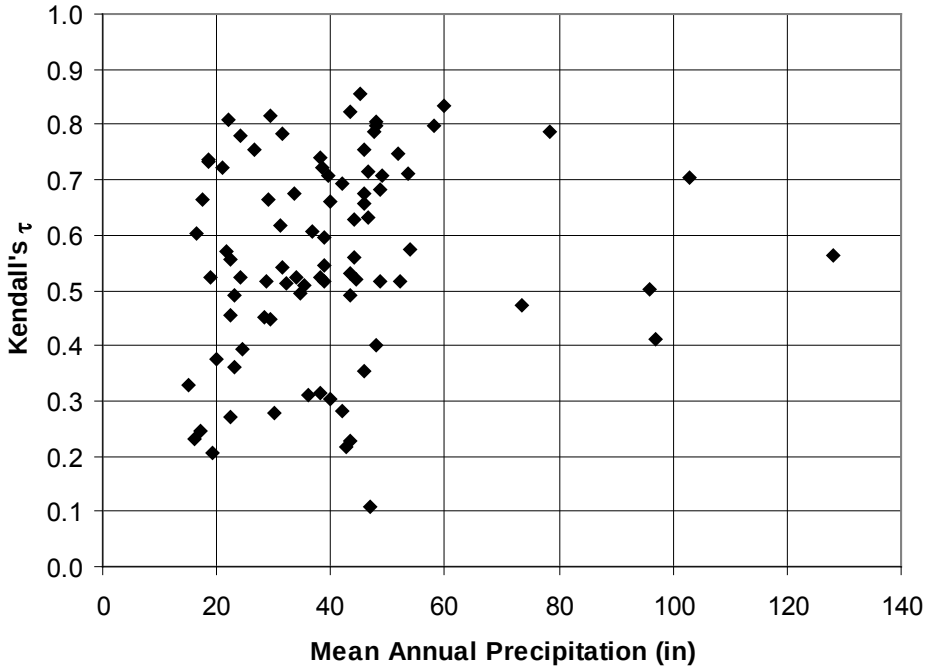

Figure C.20a. Distribution of Kendall's τ by Mean Annual Precipitation (POM).

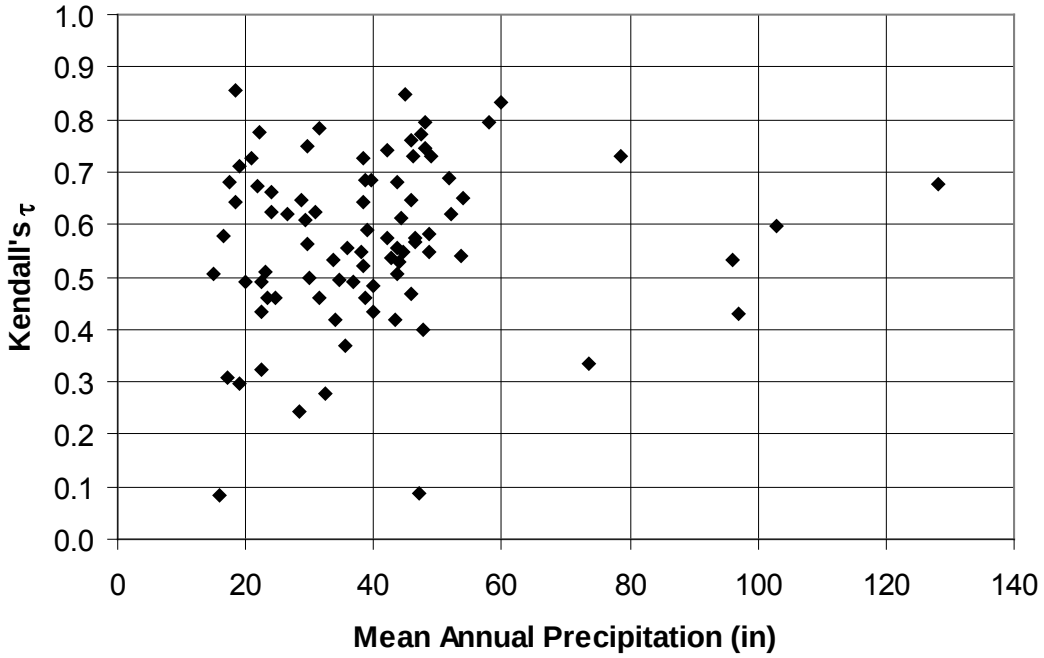

Figure C.20b. Distribution of Kendall's τ by Mean Annual Precipitation (POT).

Table C.19. Pearson's ρ (POM).

		A_{TOT} < 350 mi^2	A_{TOT} > 350 mi^2
R_A < 7	Number of pairs (n)	27	37
	Mean	0.85	0.74
	Standard Deviation	0.11	0.18
	90th Percentile	0.95	0.93
R_A > 7	Number of pairs (n)	2	19
	Mean	0.88	0.50
	Standard Deviation	-	0.24
	90th Percentile	-	0.75

Table C.20. Pearson's ρ (POT).

		A_{TOT} < 350 mi^2	A_{TOT} > 350 mi^2
R_A < 7	Number of pairs (n)	27	37
	Mean	0.84	0.76
	Standard Deviation	0.12	0.14
	90th Percentile	0.95	0.91
R_A > 7	Number of pairs (n)	2	19
	Mean	0.67	0.63
	Standard Deviation	-	0.17
	90th Percentile	-	0.85

Table C.21. Kendall's τ (POM).

		A_{TOT} < 350 mi^2	A_{TOT} > 350 mi^2
R_A < 7	Number of pairs (n)	27	37
	Mean	0.68	0.57
	Standard Deviation	0.12	0.17
	90th Percentile	0.81	0.77
R_A > 7	Number of pairs (n)	2	19
	Mean	0.56	0.39
	Standard Deviation	-	0.14
	90th Percentile	-	0.54

Table C.22. Kendall's τ (POT).

		A_{TOT} < 350 mi^2	A_{TOT} > 350 mi^2
R_A < 7	Number of pairs (n)	27	37
	Mean	0.65	0.58
	Standard Deviation	0.13	0.15
	90th Percentile	0.80	0.75
R_A > 7	Number of pairs (n)	2	19
	Mean	0.44	0.45
	Standard Deviation	-	0.13
	90th Percentile	-	0.57

Table C.23. t-statistic for Pearson's ρ (POM).

Comparison	t	t' (at 5% significance)	Significant (Yes or No)
$R_A < 7$ & $A_{TOT} < 350$ mi^2 vs. $R_A < 7$ & $A_{TOT} > 350$ mi^2	3.02	1.68	Yes
$R_A < 7$ & $A_{TOT} < 350$ mi^2 vs. $R_A > 7$ & $A_{TOT} > 350$ mi^2	6.06	1.71	Yes
$R_A < 7$ & $A_{TOT} > 350$ mi^2 vs. $R_A > 7$ & $A_{TOT} > 350$ mi^2	3.97	1.70	Yes

Table C.24. t-statistic for Pearson's ρ (POT).

Comparison	t	t' (at 5% significance)	Significant (Yes or No)
$R_A < 7$ & $A_{TOT} < 350$ mi^2 vs. $R_A < 7$ & $A_{TOT} > 350$ mi^2	2.48	1.68	Yes
$R_A < 7$ & $A_{TOT} < 350$ mi^2 vs. $R_A > 7$ & $A_{TOT} > 350$ mi^2	4.58	1.70	Yes
$R_A < 7$ & $A_{TOT} > 350$ mi^2 vs. $R_A > 7$ & $A_{TOT} > 350$ mi^2	2.82	1.70	Yes

Table C.25. t-statistic for Kendall's τ (POM).

Comparison	t	t' (at 5% significance)	Significant (Yes or No)
$R_A < 7$ & $A_{TOT} < 350$ mi^2 vs. $R_A < 7$ & $A_{TOT} > 350$ mi^2	3.07	1.68	Yes
$R_A < 7$ & $A_{TOT} < 350$ mi^2 vs. $R_A > 7$ & $A_{TOT} > 350$ mi^2	7.29	1.69	Yes
$R_A < 7$ & $A_{TOT} > 350$ mi^2 vs. $R_A > 7$ & $A_{TOT} > 350$ mi^2	4.13	1.69	Yes

Table C.26. t-statistic for Kendall's τ (POT).

Comparison	t	t' (at 5% significance)	Significant (Yes or No)
$R_A < 7$ & $A_{TOT} < 350$ mi^2 vs. $R_A < 7$ & $A_{TOT} > 350$ mi^2	2.00	1.68	Yes
$R_A < 7$ & $A_{TOT} < 350$ mi^2 vs. $R_A > 7$ & $A_{TOT} > 350$ mi^2	5.27	1.69	Yes
$R_A < 7$ & $A_{TOT} > 350$ mi^2 vs. $R_A > 7$ & $A_{TOT} > 350$ mi^2	3.35	1.69	Yes

C.8.2 Potential Combinations

With the statistical significance of the groupings established, potential combinations for the designer can be established for representative (mean) and envelope (90th percentile) selections. This analysis only addressed the use of Kendall's τ for in the copula method. As with the regression relations, the designer selects the representative or envelope approach based on the risks and costs associated with the project. With the groupings, the potential combinations are directly determined based on the drainage area ratio and total area of the site and the design joint return period.

Thus far, a distinction has been made between the POM and POT datasets. Comparison of the mean and 90th percentile values for Kendall's τ in Tables C.21 and C.22 reveals small differences between the results from the two datasets. Because these differences will not affect the resulting potential combinations, the highest mean and 90th percentile value from each group is used developing the potential combinations.

The return-period pairs for joint return periods ranging from the 10-yr to 500-yr are summarized in Tables C.27 through C.31. Values for Kendall's τ are provided in the tables for reference, but are not needed by the designer.

Table C.27. 10-yr Joint Return Periods (Groupings).

				Return Period on One Stream			
			Kendall's τ	1.25	2	5	Equal
Representative	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.68	-	10	9	8
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.59	-	10	9	7
	$R_A > 7$	$A_{TOT} < 350$ mi^2					
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.45	10	9	7	6
Envelope	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.81	-	-	10	9
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.77	-	-	10	8
	$R_A > 7$	$A_{TOT} < 350$ mi^2					
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.57	-	10	8	7

Table C.28. 25-yr Joint Return Periods (Groupings).

				Return Period on One Stream				
			Kendall's τ	1.25	2	5	10	Equal
Representative	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.68	-	-	25	24	19
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.59	-	25	24	22	17
	$R_A > 7$	$A_{TOT} < 350$ mi^2						
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.45	25	24	22	18	14
Envelope	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.81	-	-		25	22
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.77	-	-	-	25	21
	$R_A > 7$	$A_{TOT} < 350$ mi^2						
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.57		25	24	22	17

Table C.29. 50-yr Joint Return Periods (Groupings).

				Return Period on One Stream					
			Kendall's τ	1.25	2	5	10	25	Equal
Representative	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.68	-	-	-	50	48	38
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.59	-	50	49	48	40	33
	$R_A > 7$	$A_{TOT} < 350$ mi^2							
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.45	50	49	47	43	30	27
Envelope	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.81	-	-	-		50	43
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.77	-	-	-	50	49	41
	$R_A > 7$	$A_{TOT} < 350$ mi^2							
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.57	-	50	49	48	40	33

Table C.30. 100-yr Joint Return Periods (Groupings).

				Return Period on One Stream						
			Kendall's τ	1.25	2	5	10	25	50	Equal
Representative	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.68	-	-	-	100	98	92	75
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.59	-	-	100	98	94	80	67
	$R_A > 7$	$A_{TOT} < 350$ mi^2								
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.45	100	99	96	92	80	58	54
Envelope	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.81	-	-	-	-	100	99	86
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.77	-	-	-		100	98	83
	$R_A > 7$	$A_{TOT} < 350$ mi^2								
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.57	-	100	99	98	93	79	66

Table C.31. 500-yr Joint Return Periods (Groupings).

				Return Period on One Stream							
			Kendall's τ	1.25	2	5	10	25	50	100	Equal
Representative	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.68	-	-	-	-	500	499	495	376
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.59	-	-	500	499	497	491	476	331
	$R_A > 7$	$A_{TOT} < 350$ mi^2									
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.45	500	499	495	490	477	456	417	269
Envelope	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.81	-	-	-	-	-	-	500	430
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.77	-	-	-	-	-	500	499	414
	$R_A > 7$	$A_{TOT} < 350$ mi^2									
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.57	-	-	500	499	496	490	473	327

For example, if a designer determined that the representative value is appropriate and the joint return period required is 100-years, then the designer would consult Table C.30 for the potential combinations. Assuming the site has a drainage area ratio of 5 and a total area of 50 mi^2, the potential combinations are (main return period, tributary return period): (10,100), (25,98), (50,92), and (75,75). Three additional combinations are derived by transposing the return periods for the main and tributary streams: (100,10), (98,25), and (92,50). In total, this represents seven potential combinations. The designer might elect to observe that the 98-year return period event is sufficiently close to the 100-year event and collapse the first two pairs to (25,100), thereby reducing the potential combinations to five for hydraulic analysis.

The information presented in Tables C.27 through C.31 provides no guidance for appropriate combinations in cases where the area ratio exceeds 7 and the total area is less than 350 mi^2. Based on previous assessments, one would expect that the correlation for this case would be greater than when the area ratio is greater than 7 and the total area is greater than 350 mi^2, but less than when the area ratio is less than 7 and the total area is less than 350 mi^2.

In the absence of supporting data, a conservative strategy to apply the case where area ratio is less than 7 and the total area is less than 350 mi^2 is recommended. An alternative selection might be to use the case where the area ratio is less than 7 and the total area is greater than 350 mi^2.

C.9 Comparison of Regression and Grouping

Two methods for estimating the correlation parameters Pearson's ρ and Kendall's τ have been developed. Development of regression equations on watershed and meteorological characteristics was described in Section C.7. Development of the correlation parameters based on data groupings was presented in Section C.8. Both methods were applied to the database of 85 gage pairs to compare the performance of the two methods. The root mean square errors were determined and are summarized in Table C.32.

Table C.32. Prediction Error Comparison.

Correlation parameter	Method	POM Dataset RMSE	POT Dataset RMSE
Pearson's ρ	Regression	0.171	0.137
	Grouping	0.170	0.148
Kendall's τ	Regression	0.147	0.138
	Grouping	0.147	0.142

The root mean square errors (RMSE) for the two methods are nearly identical for the POM dataset. The RMSE for the grouping method is slightly lower than the regression method for Pearson's ρ, but identical for Kendall's τ. For the POT dataset, RMSEs are marginally lower for the regression method for both correlation parameters.

Appendix D. Total Probability Method

Table of Contents

List of Tables ..... **D-i**

List of Figures ..... **D-ii**

D.1 Theorem Application to River Stage D-1

D.2 Prototype Analyses D-3

D.2.1 Instantaneous Versus Mean Daily Data D-4

D.2.2 Daily Data Conditional Probability Matrices D-11

List of Tables

Table D.1. AEP Bins D-3

Table D.2. Prototype Gage Pair Watershed Characteristics D-4

Table D.3a. Instantaneous Data Pair Counts for Gage Pair 06 (POM) D-4

Table D.3b. Instantaneous Data Pair Counts for Gage Pair 06 (POT) D-5

Table D.4a. Instantaneous Data Pair Counts for Gage Pair 43 (POM) D-5

Table D.4b. Instantaneous Data Pair Counts for Gage Pair 43 (POT) D-5

Table D.5a. Instantaneous Data Pair Counts for Gage Pair 76 (POM) D-6

Table D.5b. Instantaneous Data Pair Counts for Gage Pair 76 (POT) D-6

Table D.6. Gage Statistics D-7

Table D.7a. Daily Data Pair Counts for Gage Pair 06 (POM) D-7

Table D.7b. Daily Data Pair Counts for Gage Pair 06 (POT) D-8

Table D.8a. Daily Data Pair Counts for Gage Pair 43 (POM) D-8

Table D.8b. Daily Data Pair Counts for Gage Pair 43 (POT) D-8

Table D.9a. Daily Data Pair Counts for Gage Pair 76 (POM) D-9

Table D.9b. Daily Data Pair Counts for Gage Pair 76 (POT) D-9

Table D.10a. Instantaneous Data Event Distribution for Gage Pair 43 (POM) D-9

Table D.10b. Instantaneous Data Event Distribution for Gage Pair 43 (POT) D-10

Table D.11a. Daily Data Event Distribution for Gage Pair 43 (POM) D-10

Table D.11b. Daily Data Event Distribution for Gage Pair 43 (POT) D-10

Table D.12a. Daily Data Pair Counts for Gage Pair 05 (POM) D-11

Table D.12b. Daily Data Pair Counts for Gage Pair 05 (POT) D-11

Table D.13a. CPM for Gage Pair 05 (POM) D-12

Table D.13b. CPM for Gage Pair 05 (POT) D-12
Table D.14a. CPM for Gage Pair 06 (POM) D-12
Table D.14b. CPM for Gage Pair 06 (POT) D-12
Table D.15a. CPM for Gage Pair 43 (POM) D-13
Table D.15b. CPM for Gage Pair 43 (POT) D-13
Table D.16a. CPM for Gage Pair 76 (POM) D-13
Table D.16b. CPM for Gage Pair 76 (POT) D-13

List of Figures

Figure D.1. Hypothetical Probability Density Function D-2
Figure D.2. Hypothetical Cumulative Distribution Function D-2

The objective of this portion of the research was to generate conditional probability matrices for the total probability method. The application of the total probability method to river stage and the prototype analyses leading to the development of the conditional probability matrices is provided in this appendix.

D.1 Theorem Application to River Stage

The total probability theorem has been applied at river confluences (e.g. Dyhouse, 1985; Pingel and Ford, 2004). For application of the total probability method for joint probability, the procedure described in EM 1110-2-1415 (USACE, 1993) is used. The objective of the method, as applied in this research, is to produce a cumulative distribution function for tributary stage.

The foundation of the total probability method is the total probability theorem:

$$P[B] = \sum_{i=1}^{n} P[B \mid A_i] P[A_i] \tag{D.1}$$

where,

$P[B]$	=	probability that B will occur
$P[B\|A_i]$	=	probability that B will occur given A_i has occurred
$P[A_i]$	=	probability that A_i will occur

For continuous variables, Equation D.1 may be written as follows:

$$P[B] = \int_A P[B \mid A] P[A] dA \tag{D.2}$$

Application of Equation D.1 or D.2 over the range of values for the variable B results in a probability density function (PDF) for B as shown in Figure D.1. The shape of the PDF in the figure is hypothetical; the actual shape will depend on the statistical nature of the variable.

The PDF is transformed into the cumulative distribution function (CDF) by estimating the probability of nonexceedance for each value of B. Conceptually, the probability of nonexceedance is the area under the PDF to the left of the value of B. Performing this analysis over the range of B results in a CDF as is shown in Figure D.2. Again, the shape in the figure is hypothetical and depends on the shape of the PDF.

It is often desirable to express the CDF in terms of probability of exceedance, P_e, rather than probability of nonexceedance P_{ne}. It is straightforward to achieve this transformation since $P_e + P_{ne} = 1$. The result of this relation is shown in Figure D.2 by the addition of the second vertical axis (P_e) with the numerical values reversed from the first vertical axis (P_{ne}).

As stated previously, the application of the total probability method of interest for this research is to develop the CDF for stage of the tributary at a site of interest. Therefore, the variable B in the previous discussion is tributary stage, y_T. Equation D.1 becomes:

$$P[y_T] = \sum_{i=1}^{n} P[y_T \mid Q_{M,i}] P[Q_{M,i}] \tag{D.3}$$

where,

$P[y_T]$	=	probability that stage y_T will occur
$P[y_T \mid Q_{M,i}]$	=	probability that stage y_T will occur given flow on the main stream, $Q_{M,i}$ has occurred
$P[Q_{M,i}]$	=	probability that main stream flow, $Q_{M,i}$ will occur

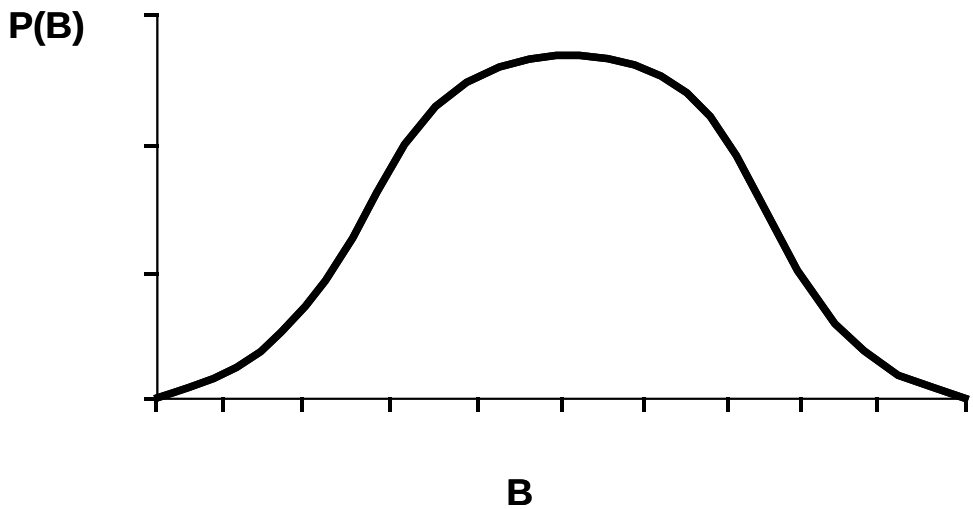

Figure D.1. Hypothetical Probability Density Function.

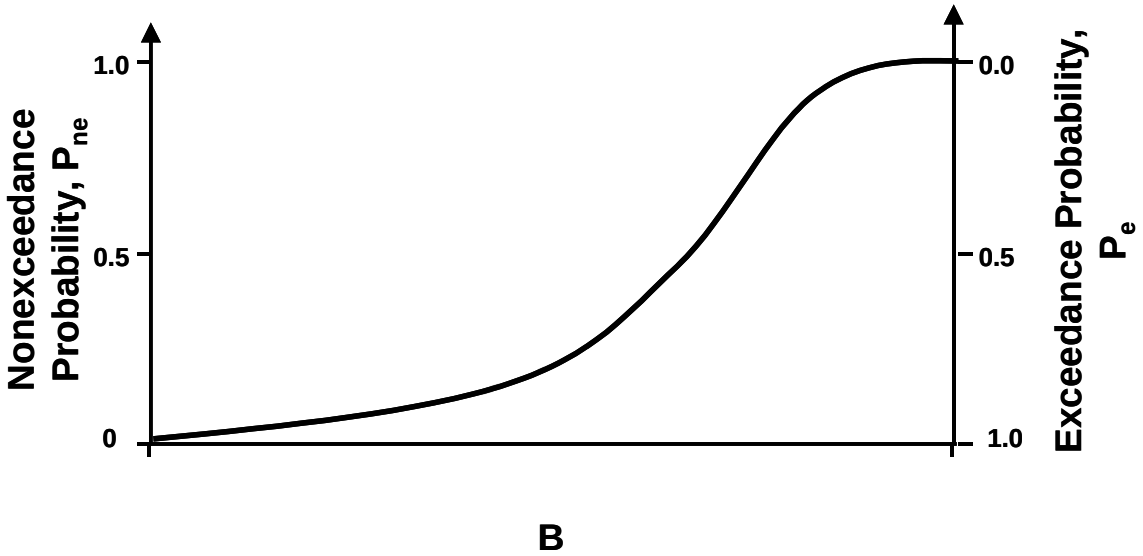

Figure D.2. Hypothetical Cumulative Distribution Function.

Because the tributary stage, y_T, is a function of both Q_M and Q_T, Equation D.3 might have been stated with Q_T rather than Q_M. The preferred representation depends on which of the two discharges is considered the "dominant" variable. Close to the confluence, Q_M may be the dominant variable while closer to the upstream end of the influence reach, Q_T may be the dominant variable.

The subscript, i, represents an index ranging from 1 to n where n is the number groups (bins) into which the variable Q_M is divided. $Q_{M,i}$ represents the index value for that bin. Bins for this analysis are defined based on annual exceedance probabilities (AEPs) of interest to designers analyzing coincident flooding. Seven bins defined by an index value and range are summarized in Table D.1. The "bin AEP Index Value" value is the average of the logs (base 10) of the upper and lower bounds of the bin AEP range. Bin number 3, for example, includes flows that range from an AEP of 0.063 to an AEP of 0.141 with the index value being AEP equal to 0.1. In terms of return period, bin number 3 includes flows that range from a 16-yr return period to a 7-yr return period with and index value of the 10-yr flow.

Table D.1. AEP Bins.

Bin number	Bin AEP (Index Value)	Bin AEP range	$P[Q_{M,i}]$
1	0.500	0.316 > AEP >= 1.000	0.684
2	0.200	0.141 > AEP >= 0.316	0.175
3	0.100	0.063 > AEP >= 0.141	0.078
4	0.040	0.028 > AEP >= 0.063	0.035
5	0.020	0.014 > AEP >= 0.028	0.014
6	0.010	0.004 > AEP >= 0.014	0.010
7	0.002	0.000 > AEP >= 0.004	0.004

To apply Equation D.3, $P[Q_{M,i}]$ and $P[y_T|Q_{M,i}]$ must be defined. $P[Q_{M,i}]$ is defined in Table D.1 and is calculated as the range of the AEP bin. For example, the probability of a flow falling in bin 3 is equal to the width of the range (0.141-0.063 = 0.078).

$P[y_T|Q_{M,i}]$ is the conditional probability of experiencing a stage of y_T given a flow on the main stream in bin i. However, with coincident flows y_T is a function of both Q_M and the flow on the tributary, Q_T. Referred to as a "linking relationship," there is some function used to calculate the tributary stage, y_T. Typically, this linking relationship is provided with a HEC-RAS model of the confluence with Q_T and Q_M as inputs. Functionally, the relation is expressed as follows:

$$y_T = f(Q_T, Q_M) \tag{D.4}$$

where,

y_T = Tributary stream stage

Q_T = Tributary stream flow

Q_M = Main stream flow

Although necessary, the linking relationship does not provide the conditional probability, $P[y_T|Q_{M,i}]$, needed to apply the total probability theorem. Development of the conditional probabilities (conditional probability matrices) is the focus of this part of the research.

D.2 Prototype Analyses

The prototype gage pairs were used to evaluate data issues and demonstrate concepts. The basic data for the prototype gage pairs are summarized in Table D.2.

Table D.2. Prototype Gage Pair Watershed Characteristics.

Characteristic	Pair 05	Pair 06	Pair 43	Pair 76
Main watershed				
Gage ID	01403060	01445500	12082500	08164450
Area (sq mi)	785	106	133	289
Tributary watershed				
Gage ID	01403150	01446000	12083000	08164503
Area (sq mi)	2	37	75	178
Joint Characteristics				
Total Area (sq mi)	787	143	208	467
Drainage Area Ratio	394.5	2.9	1.8	1.6
Centroid Separation (mi)	7.5	7.0	11.0	14.8

D.2.1 Instantaneous Versus Mean Daily Data

For three of the prototype gage pairs, peak flow dates were obtained from the annual and partial duration flow series. (Gage Pair 05 was not used for this assessment because no instantaneous data were available.) The times of peak were then obtained by reviewing the instantaneous data for those dates. For each gage pair, two sets of concurrent flow pairs were extracted. One was using the peak flows on the main with the complementary tributary flows (POM) and the other was using the peak flows on the tributary with the complementary main flows (POT). From these two sets of flow pairs two conditional probability matrices for each gage pair were developed.

To explore the use of daily versus instantaneous data, the process was conducted twice, once extracting flow pairs from instantaneous (15-minute) flow data and once extracting flow pairs from mean-daily flow data. The AEPs were computed separately for the instantaneous and daily data so that each value could be placed in its proper bin based on the appropriate AEP.

The resulting empirically derived AEP pair count tables for the instantaneous data are shown in Tables D.3 through D.5. A higher degree of correlation between the main and tributary watersheds is indicated when a higher frequency of observations fall on the diagonal indicated in the tables. As has been noted elsewhere, the instantaneous data are limited in their scope. Consequently, the instantaneous pair counts for gage pairs 06 and 76 are less than 10.

Table D.3a. Instantaneous Data Pair Counts for Gage Pair 06 (POM).

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Total
Main Stream AEP	0.500	1							1
	0.200		1						1
	0.100	1	1						2
	0.040								0
	0.020								0
	0.010								0
	0.002								0
	Total	2	2	0	0	0	0	0	4

Table D.3b. Instantaneous Data Pair Counts for Gage Pair 06 (POT).

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Total
Main Stream AEP	0.500	1							1
	0.200		1						1
	0.100								0
	0.040			1					1
	0.020			1					1
	0.010								0
	0.002								0
	Total	1	1	2	0	0	0	0	4

Table D.4a. Instantaneous Data Pair Counts for Gage Pair 43 (POM).

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Total
Main Stream AEP	0.500	27			1				28
	0.200	3	4						7
	0.100			1					1
	0.040			1					1
	0.020								0
	0.010								0
	0.002								0
	Total	30	4	2	1	0	0	0	37

Table D.4b. Instantaneous Data Pair Counts for Gage Pair 43 (POT).

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Total
Main Stream AEP	0.500	17	5		2				24
	0.200	1	2	1	1				5
	0.100								0
	0.040								0
	0.020								0
	0.010								0
	0.002								0
	Total	18	7	1	3	0	0	0	29

Table D.5a. Instantaneous Data Pair Counts for Gage Pair 76 (POM).

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Total
Main Stream AEP	0.500	3							3
	0.200	1	1	1					3
	0.100								0
	0.040			1					1
	0.020								0
	0.010	1							1
	0.002								0
	Total	5	1	2	0	0	0	0	8

Table D.5b. Instantaneous Data Pair Counts for Gage Pair 76 (POT).

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Total
Main Stream AEP	0.500	4	1						5
	0.200			1					1
	0.100								0
	0.040								0
	0.020		1						1
	0.010								0
	0.002								0
	Total	4	2	1	0	0	0	0	7

AEP estimates are required for the mean daily data series. As with the instantaneous data series, Bulletin 17B procedures were applied to the annual maxima mean daily flow values to compute mean daily flow quantiles for each of the gages.

The adopted skew coefficient was computed as the weighted value of the regional skew and station skew, rounded to the nearest tenth. The regional skew was taken as the value estimated from Plate 1 of Bulletin 17B. The estimates of regional skew provided on Plate 1 are based on (instantaneous) annual maxima flow values, not mean daily flow series. Therefore, it was implicitly assumed that the regional skew for the mean daily flow series is the same as that for instantaneous flows series. The station skew was computed from the mean daily flow series for the gage. Following Bulletin 17B, the regional and station skew coefficients were weighted by inverse proportion of their mean square errors. The mean square error for station skew was computed from the mean daily series, and the mean square error for the regional skew was taken to be 0.302 per Bulletin 17B. The resulting number of events (excluding outliers), computed mean, standard deviation, and skew values for each gage are listed in Table D.6.

For each gage, the set of mean daily POM and POT flow pairs was converted to AEP pairs by applying the previously computed gage statistical parameters (mean, standard deviation) to the log Pearson Type-III (LP3) distribution. This process presumed that the LP3 distribution, adopted by Bulletin 17B for instantaneous flows, is also appropriate to describe mean daily flows.

Table D.6. Gage Statistics.

Characteristic	Pair 05	Pair 06	Pair 43	Pair 76
Main Stream				
Gage ID	01403060	01445500	12082500	08164450
Events	67	85	64	29
Mean	4.1986	2.9110	3.6694	3.7235
Standard Deviation	0.1813	0.1716	0.1969	0.3614
Regional Skew	0.6890	0.5840	0.0000	-0.2910
Station Skew	0.1833	0.0371	0.0627	0.3009
Adopted Skew	0.3	0.1	0.0	0.1
Tributary Stream				
Gage ID	01403150	01446000	12083000	08164503
Events	27	42	64	29
Mean	2.0004	2.6452	3.5549	3.6189
Standard Deviation	0.1950	0.2131	0.1707	0.3191
Regional Skew	0.6740	0.5610	0.0000	-0.2980
Station Skew	0.3978	-0.0640	0.0437	0.0190
Adopted Skew	0.5	0.1	0.0	-0.1

The resulting AEP pair counts for the daily data are shown in Tables D.7 through D.9 for the POM and POT datasets for three prototype gage pairs. If the distribution of pair counts in both the instantaneous and daily data are similar, the use of daily data is validated. However, in two of the three prototype gage pairs, 06 and 76, there are insufficient instantaneous data observations for a meaningful comparison.

A comparison between the instantaneous and daily data pair counts for Gage Pair 43 is feasible. Because the instantaneous data include pair counts numbering 37 (POM) and 29 (POT) while the daily data include pair counts of 64 (POM and POT) the counts are normalized to probabilities to illustrate their distribution. The distributions for the POM and POT datasets are provided in Tables D.10a and D.10b for the instantaneous data and in Tables D.11a and D.11b for the daily data.

Table D.7a. Daily Data Pair Counts for Gage Pair 06 (POM).

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Total
Main Stream AEP	0.500	30	3	1					34
	0.200	3	2	1					6
	0.100			1	1				2
	0.040	1							1
	0.020								0
	0.010								0
	0.002								0
	Total	34	5	3	1	0	0	0	43

Table D.7b. Daily Data Pair Counts for Gage Pair 06 (POT).

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Total
Main Stream AEP	0.500	28	7	1					36
	0.200		2	1			1		4
	0.100			1	1				2
	0.040								0
	0.020								0
	0.010								0
	0.002								0
	Total	28	9	3	1	0	1	0	42

Table D.8a. Daily Data Pair Counts for Gage Pair 43 (POM).

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Total
Main Stream AEP	0.500	42	5						47
	0.200	4	3						7
	0.100	1	2	2	1				6
	0.040				2				2
	0.020	1							1
	0.010								0
	0.002						1		1
	Total	48	10	2	3	0	1	0	64

Table D.8b. Daily Data Pair Counts for Gage Pair 43 (POT).

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Total
Main Stream AEP	0.500	43	5	1	2				51
	0.200	2	3						5
	0.100		2	1	1				4
	0.040				2				2
	0.020	1							1
	0.010								0
	0.002						1		1
	Total	46	10	2	5	0	1	0	64

Table D.9a. Daily Data Pair Counts for Gage Pair 76 (POM).

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Total
Main Stream AEP	0.500	18	2						20
	0.200	2	2	2					6
	0.100			1					1
	0.040					1			1
	0.020								0
	0.010	1							1
	0.002								0
	Total	21	4	3	0	1	0	0	29

Table D.9b. Daily Data Pair Counts for Gage Pair 76 (POT).

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Total
Main Stream AEP	0.500	19	2						21
	0.200	1	2	2					5
	0.100		1		1				2
	0.040					1			1
	0.020								0
	0.010								0
	0.002								0
	Total	20	5	2	1	1	0	0	29

Table D.10a. Instantaneous Data Event Distribution for Gage Pair 43 (POM).

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Total
Main Stream AEP	0.500	0.73			0.03				0.76
	0.200	0.08	0.11	0.00					0.19
	0.100			0.03					0.03
	0.040			0.03					0.03
	0.020								0.00
	0.010								0.00
	0.002								0.00
	Total	0.81	0.11	0.05	0.03	0.00	0.00	0.00	1.00

Table D.10b. Instantaneous Data Event Distribution for Gage Pair 43 (POT).

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Total
Main Stream AEP	0.500	0.59	0.17		0.07				0.83
	0.200	0.03	0.07	0.03	0.03				0.17
	0.100								0.00
	0.040								0.00
	0.020								0.00
	0.010								0.00
	0.002								0.00
	Total	0.62	0.24	0.03	0.10	0.00	0.00	0.00	1.00

Table D.11a. Daily Data Event Distribution for Gage Pair 43 (POM).

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Total
Main Stream AEP	0.500	0.66	0.08						0.73
	0.200	0.06	0.05						0.11
	0.100	0.02	0.03	0.03	0.02				0.09
	0.040				0.03				0.03
	0.020	0.02							0.02
	0.010								0.00
	0.002						0.02		0.02
	Total	0.75	0.16	0.03	0.05	0.00	0.02	0.00	1.00

Table D.11b. Daily Data Event Distribution for Gage Pair 43 (POT).

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Total
Main Stream AEP	0.500	0.67	0.08	0.02	0.03				0.80
	0.200	0.03	0.05						0.08
	0.100		0.03	0.02	0.02				0.06
	0.040				0.03				0.03
	0.020	0.02							0.02
	0.010								0.00
	0.002						0.02		0.02
	Total	0.72	0.16	0.03	0.08	0.00	0.02	0.00	1.00

Strategies for comparing the instantaneous and daily data distributions for Gage Pair 43 include: 1) examine the marginal distributions in the respective “total” columns and rows and 2) compare the total portion of the events located on the diagonal, which would indicate comparable representations of correlation. For example, in Table D.10a (instantaneous), 87 percent of the observations are on the diagonal while in Table D.11a (daily) 77 percent are on the diagonal. Qualitatively, both comparative methods indicate that the two sets of tables are “close.”

Given that there are approximately twice the number of observations in the daily data for Gage Pair 43 and that instantaneous data are limited for other gage pairs it can neither be definitively concluded that the daily data and instantaneous data carry similar statistical properties nor that they do not. Since analysis is not feasible with instantaneous data, daily data were used.

D.2.2 Daily Data Conditional Probability Matrices

This section illustrates development of conditional probability matrices (CPMs) based on mean daily data from four prototype gage pairs (05, 06, 43, and 76). Separate CPMs for the peak on main (POM) and peak on tributary (POT) cases are developed to capture the asymmetry (above the diagonal in one case, below the diagonal in the other case) represented by these two cases.

The event counts were previously presented in Tables D.7 through D.9 for gage pairs 06, 43, and 76. A fourth gage pair, 05, was added to this analysis to assess the CPMs for a gage pair where the correlation between the main and tributary flows is not statistically significant. Tables D.12a and D.12b summarize the event counts for Gage Pair 05 for the POM and POT datasets, respectively. Of the four prototype gage pairs only Gage Pair 05 shows substantial asymmetry.

Table D.12a. Daily Data Pair Counts for Gage Pair 05 (POM).

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Total
Main Stream AEP	0.500	18							18
	0.200	6							6
	0.100	1							1
	0.040	1							1
	0.020								0
	0.010								0
	0.002	1							1
	Total	27	0	0	0	0	0	0	27

Table D.12b. Daily Data Pair Counts for Gage Pair 05 (POT).

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Total
Main Stream AEP	0.500	16	8	2			1		27
	0.200								0
	0.100								0
	0.040								0
	0.020								0
	0.010								0
	0.002								0
	Total	16	8	2	0	0	1	0	27

CPMs for each gage are shown in Tables D.13 through D.16. CPMs for the POM case were obtained by dividing the AEP pair count in each cell by the total pair count for the row. CPMs for the POT case were obtained by dividing the AEP pair count in each cell by the total pair count for the column. The approach described with the prototype gages was expanded to all gage pairs to develop recommended CPMs for ungaged watersheds. See Section 5.2 in the main report for a summary of this development.

Table D.13a. CPM for Gage Pair 05 (POM).

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Sum
Main Stream AEP	0.500	1.00							1.00
	0.200	1.00							1.00
	0.100	1.00							1.00
	0.040	1.00							1.00
	0.020								0.00
	0.010								0.00
	0.002	1.00							1.00

Table D.13b. CPM for Gage Pair 05 (POT).

		Tributary Stream AEP						
		0.500	0.200	0.100	0.040	0.020	0.010	0.002
Main Stream AEP	0.500	1.00	1.00	1.00			1.00	
	0.200							
	0.100							
	0.040							
	0.020							
	0.010							
	0.002							
	Sum	1.00	1.00	1.00	0.00	0.00	1.00	0.00

Table D.14a. CPM for Gage Pair 06 (POM).

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Sum
Main Stream AEP	0.500	0.88	0.09	0.03					1.00
	0.200	0.50	0.33	0.17					1.00
	0.100			0.50	0.50				1.00
	0.040	1.00							1.00
	0.020								0.00
	0.010								0.00
	0.002								0.00

Table D.14b. CPM for Gage Pair 06 (POT).

		Tributary Stream AEP						
		0.500	0.200	0.100	0.040	0.020	0.010	0.002
Main Stream AEP	0.500	1.00	0.78	0.33	0.00			
	0.200		0.22	0.33	0.00		1.00	
	0.100			0.33	1.00			
	0.040							
	0.020							
	0.010							
	0.002							
	Sum	1.00	1.00	1.00	1.00	0.00	1.00	0.00

Table D.15a. CPM for Gage Pair 43 (POM).

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Sum
Main Stream AEP	0.500	0.89	0.11						1.00
	0.200	0.57	0.43						1.00
	0.100	0.17	0.33	0.33	0.17				1.00
	0.040				1.00				1.00
	0.020	1.00							1.00
	0.010								0.00
	0.002	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00

Table D.15b. CPM for Gage Pair 43 (POT).

		Tributary Stream AEP						
		0.500	0.200	0.100	0.040	0.020	0.010	0.002
Main Stream AEP	0.500	0.93	0.50	0.50	0.40			
	0.200	0.04	0.30					
	0.100		0.20	0.50	0.20			
	0.040				0.40			
	0.020	0.02						
	0.010							
	0.002						1.00	
	Sum	1.00	1.00	1.00	1.00	0.00	1.00	0.00

Table D.16a. CPM for Gage Pair 76 (POM).

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Sum
Main Stream AEP	0.500	0.90	0.10						1.00
	0.200	0.33	0.33	0.33					1.00
	0.100			1.00					1.00
	0.040					1.00			1.00
	0.020								0.00
	0.010	1.00							1.00
	0.002								0.00

Table D.16b. CPM for Gage Pair 76 (POT).

		Tributary Stream AEP						
		0.500	0.200	0.100	0.040	0.020	0.010	0.002
Main Stream AEP	0.500	0.95	0.40					
	0.200	0.05	0.40	1.00				
	0.100		0.20		1.00			
	0.040					1.00		
	0.020							
	0.010							
	0.002							
	Sum	1.00	1.00	1.00	1.00	1.00	0.00	0.00

Appendix E. Other Strategies

Table of Contents

List of Tables ..... **E-i**

List of Figures ..... **E-i**

E.1 Marginal Analysis E-1

E.2 Tabular/Graphical E-5

List of Tables

Table E.1. Gage Pair 06 (New Jersey) Regression Data and Results E-2

Table E.2. Gage Pair 43 (Washington) Regression Data and Results E-2

Table E.3. Gage Pair 76 (Texas) Regression Data and Results E-3

List of Figures

Figure E.1. Iso-Probability Lines for Gage Pair 06 (New Jersey) E-4

Figure E.2. Iso-Probability Lines for Gage Pair 43 (Washington) E-4

Figure E.3. Iso-Probability Lines for Gage Pair 76 (Texas) E-5

Figure E.4. Iso-Probability Lines for Area Ratio 100:1 ..... E-6

Figure E.5. Iso-Probability Lines for Area Ratio 1:1 ..... E-6

E.1 Marginal Analysis

In Section 2.2.5, it was concluded that marginal analysis does not provide a useful framework for addressing the joint probability question that is the subject of this research. However, before drawing this conclusion, application of the method on several prototype watersheds was completed. Because the method may have utility in other contexts, this work is summarized in this appendix. The general procedure is as follows:

1. Estimate applicable peak flow for each watershed independently.
2. Estimate peak flow for the combined watershed.
3. Determine the design iso-probability line.
4. Determine the critical combination.

Step 1. Estimate applicable peak flow for each watershed independently.

For the selected prototype gage pair watersheds the USGS regression equations are applicable and applied. For Gage Pair 06, the statewide equation for New Jersey applies:

$$Q_T = aA^{b_1}S^{b_2}St^{b_3}I^{b_4} \tag{E.1}$$

where,

Q_T	=	design flow for return period, T
A	=	drainage area, mi^2
S	=	channel slope, ft/mi
St	=	storage area, percent
I	=	impervious cover, percent, $I = 0.117D^{0.792-0.039\log D}$
D	=	population density, $persons/mi^2$

Gage Pair 43 is located in USGS Washington region 2 for which the following regression equation applies:

$$Q_T = aA^{b_1}p^{b_2} \tag{E.2}$$

where,

Q_T	=	design flow for return period, T
A	=	drainage area, mi^2
p	=	mean annual precipitation, in

Gage pair 76 is located in USGS Texas region 9 for which the following regression equation applies:

$$Q_T = aA^{b_1}SL^{b_2}SH^{b_2} \tag{E.3}$$

where,

Q_T	=	design flow for return period, T
A	=	drainage area, mi^2
SL	=	stream slope, ft/mi
SH	=	basin shape factor, dimensionless

Tables E.1 through E.3 summarize the input variables and results for the main and tributary gages for the New Jersey, Washington, and Texas gage pairs, respectively.

Step 2. Estimate peak flow for the combined watershed.

Estimates of the peak flow quantiles for the combined watersheds use the same USGS regression equations applied for the main and tributary watersheds. The applicable data and results are summarized in the previous tables E.1 through E.3.

Table E.1. Gage Pair 06 (New Jersey) Regression Data and Results.

	Main 01445500	Tributary 01446000	Combined	Tributary Residual	Main Residual
Name	Pequest River	Beaver Brook			
A (mi^2)	106	36.7	142.7		
S (ft/mi)	7.5	22.8	7.5		
St (%)	0	0	0		
D (p/mi^2)	282	282	282		
I (%)	5.95	5.95	5.95		
Q_2 (ft^3/s)	2,728	1,401	3,554	826	2,153
Q_{10} (ft^3/s)	5,701	3,027	7,406	1,705	4,379
Q_{25} (ft^3/s)	7,512	4,074	9,701	2,189	5,627
Q_{50} (ft^3/s)	9,334	5,059	12,018	2,684	6,959
Q_{100} (ft^3/s)	11,637	6,374	14,938	3,301	8,563

Table E.2. Gage Pair 43 (Washington) Regression Data and Results.

	Main 12082500	Tributary 12083000	Combined	Tributary Residual	Main Residual
Name	Nisqually River	Mineral Creek			
A (mi^2)	133	70.3	203.3		
MAP (in)	94	98	95.4		
Q_2 (ft^3/s)	6,256	3,809	9,279	3,023	5,470
Q_{10} (ft^3/s)	11,269	6,918	16,666	5,396	9,748
Q_{25} (ft^3/s)	13,885	8,552	20,505	6,620	11,952
Q_{50} (ft^3/s)	16,380	10,111	24,176	7,796	14,066
Q_{100} (ft^3/s)	18,435	11,391	27,202	8,767	15,811

Table E.3. Gage Pair 76 (Texas) Regression Data and Results.

	Main 08164450	Tributary 08164503	Combined	Tributary Residual	Main Residual
Name	Sandy Creek	West Mustang Creek			
A (mi^2)	289	178	467		
SL (ft/mi)	5.3	2.6	5.3		
L (mi)	45.07	43.38	45.07		
SH	7.03	10.6	4.3		
Q_2 (ft^3/s)	5,476	4,244	7,049	1,572	2,805
Q_{10} (ft^3/s)	16,639	8,179	27,043	10,404	18,864
Q_{25} (ft^3/s)	25,019	10,581	43,571	18,552	32,990
Q_{50} (ft^3/s)	32,672	12,547	59,668	26,996	47,121
Q_{100} (ft^3/s)	41,728	14,670	79,762	38,034	65,092

Step 3. Determine the design iso-probability line.

The combined watershed quantiles represent the exceedance probabilities at the confluence. The discharge at the confluence may occur based on numerous possibilities of contributing discharges from the main and tributary streams. These combinations represent the iso-probability line.

For example, Table E.1 shows that the 100-year discharges for the main, tributary, and combined watersheds are 11,637 ft^3/s, 6374 ft^3/s, and 14938 ft^3/s, respectively. The column in the table labeled "tributary residual" represents the flow occurring in the tributary during a 100-yr flow at the confluence if the main is simultaneously experiencing a 100-yr event. Similarly, the column labeled "main residual" represents the flow occurring in the main during a 100-yr event at the confluence if the tributary is simultaneously experiencing a 100-yr event. The iso-probability line is described by these two endpoints where 1) main flow = 11,637 ft^3/s and tributary flow = 3301 ft^3/s and 2) main flow = 8563 ft^3/s and tributary flow = 6374 ft^3/s. This 100-yr line is shown in Figure E.1. The figure also shows the iso-probability line for the 2-, 10-, 25-, and 50-year return periods.

Figures E.2 and E.3 illustrate the iso-probability lines for gage pairs 43 and 76, respectively.

Step 4. Determine the critical combination.

All combinations of events on an iso-probability line have the same exceedance probability. The critical combination is determined by modeling two or more of these events in the water surface profile model set up for the site. At a minimum, the two endpoints of the iso-probability line would likely be assessed.

For the 100-year event at gage pair 06, the iso-probability endpoints were previously identified as: 1) main flow = 11,637 ft^3/s and tributary flow = 3301 ft^3/s and 2) main flow = 8563 ft^3/s and tributary flow = 6374 ft^3/s. Both result in a 100-year discharge at the confluence. In case (1), the 100-year event on the main stream is combined with a discharge between the 10-year and 25-year event (3301 ft^3/s). In case (2), the 100-year tributary event is combined with an event between the 25-yr and 50-yr event (8,563 ft^3/s). As can be seen from Figure E.1, discharges of 10,000 ft^3/s on the main and 5,000 ft^3/s on the tributary also result in a 100-yr event at the confluence.

The critical combination, that is, the one producing the highest stage at the design location, is used for design. This stage represents the 100-yr stage at the design location.

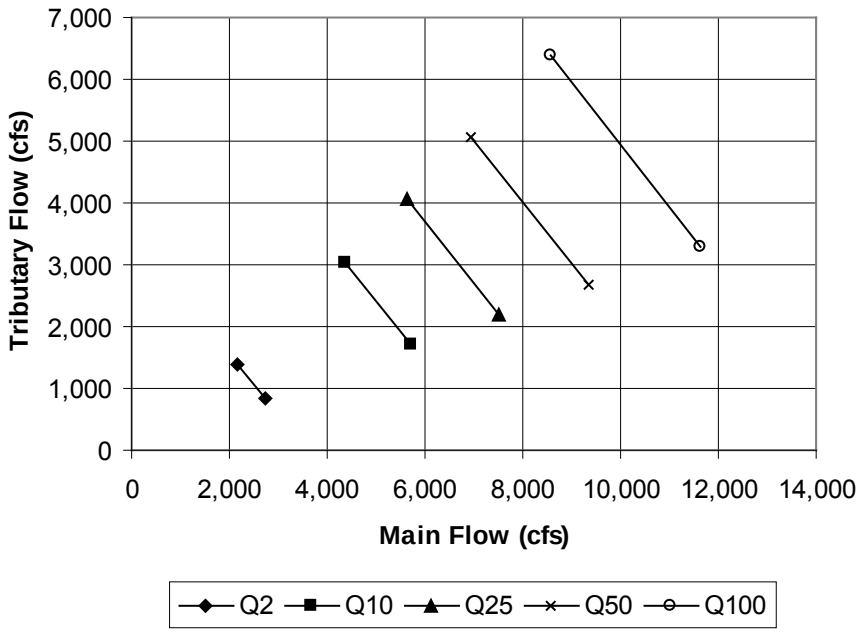

Figure E.1. Iso-Probability Lines for Gage Pair 06 (New Jersey).

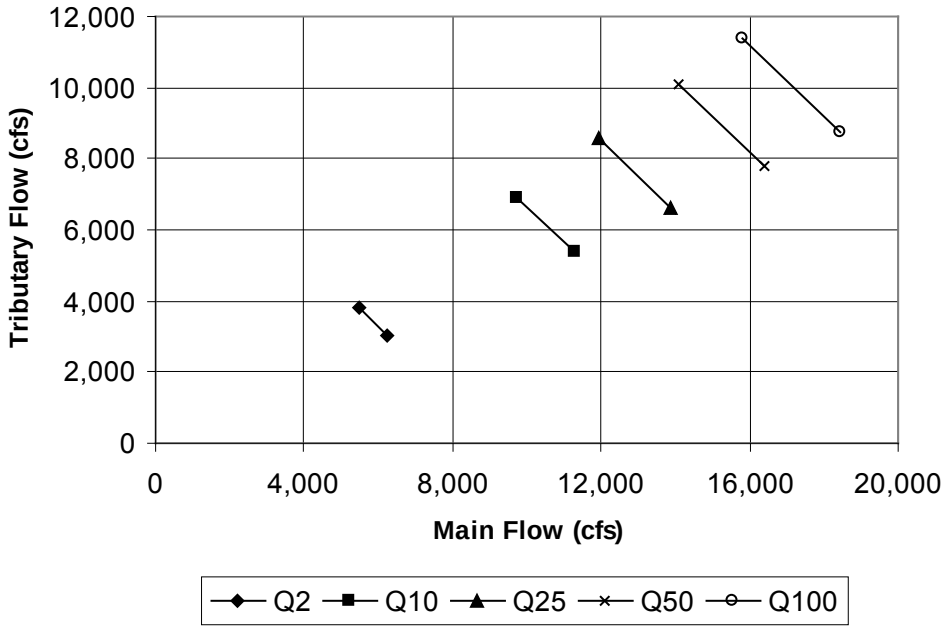

Figure E.2. Iso-Probability Lines for Gage Pair 43 (Washington).

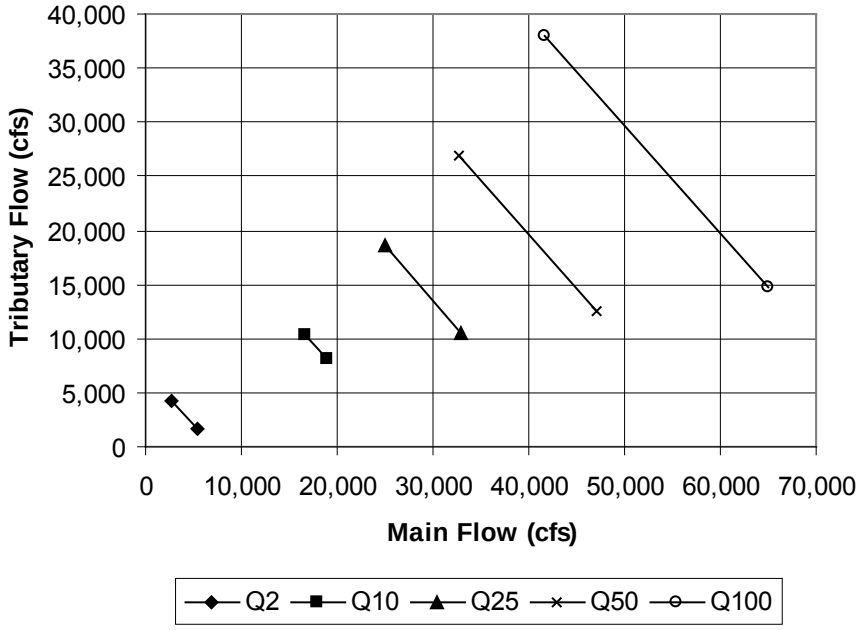

Figure E.3. Iso-Probability Lines for Gage Pair 76 (Texas).

E.2 Tabular/Graphical

Tabular and graphical presentations are not considered a distinct method, but rather an alternative means of presenting the joint probabilities and appropriate design events. In Section 2.2.7, Table 2.1 was provided as an example of a tabular method. Figures E.4 and E.5 represent graphical isolines for the 100:1 and 1:1 drainage area ratios from that table.

Application of the tabular/graphical presentation is illustrated with three gage pairs for a 10-yr design event. First, Gage Pair 08 has a drainage area ratio of approximately 1.0. According to Figure E.5, a 10-yr event for this drainage area ratio is simulated by using 10-yr events on both the main and tributary streams. This is equivalent to assuming perfect correlation between the two streams. For Gage Pair 21, the drainage area ratio is approximately 100. Consulting Figure E.4, a 10-yr event is simulated using any point on the 0.1 isoline. A third example is based on gage pair 76, which has a drainage area ratio of 1.6. Therefore, intermediate values between those shown in Figures E.4 and E.5 are appropriate.

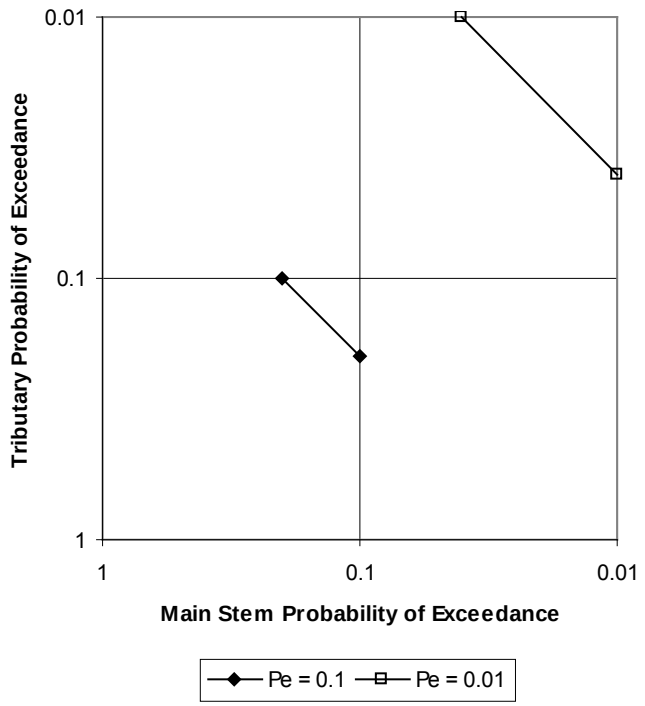

Figure E.4. Iso-Probability Lines for Area Ratio 100:1.

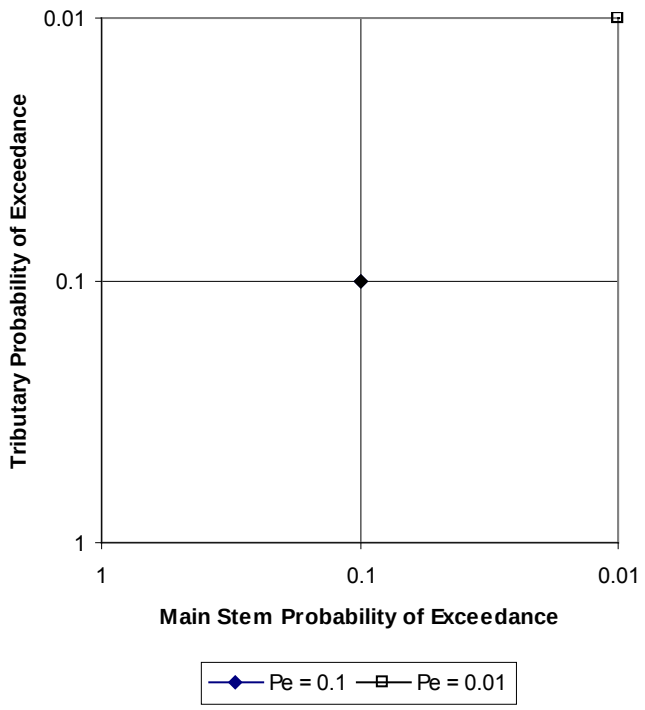

Figure E.5. Iso-Probability Lines for Area Ratio 1:1.

Appendix F. Data Analysis

Table of Contents

List of Tables ..... **F-i**

List of Figures ..... **F-ii**

F.1 Prototype Watershed Data F-1

F.1.1 Watershed Summaries F-1

F.1.2 Selected Events for Gage Pair 05 ..... F-2

F.1.3 Instantaneous Peaks and Mean Daily Flows F-4

F.1.3.1 Correlation Plots for Instantaneous Peaks and Mean Daily Flows F-4

F.1.3.2 Small Watersheds F-12

F.2 Watershed Regulation F-14

List of Tables

Table F.1. Data Summary for Gage Pair 05 (New Jersey) F-1

Table F.2. Data Summary for Gage Pair 06 (New Jersey) F-1

Table F.3. Data Summary for Gage Pair 08 (Maryland) F-1

Table F.4. Data Summary for Gage Pair 21 (Michigan) F-2

Table F.5. Data Summary for Gage Pair 43 (Washington) F-2

Table F.6. Data Summary for Gage Pair 76 (Texas) F-2

Table F.7. Summary of Gage Pairs with p-value > 0.05 ..... F-12

Table F.8. t-statistic; POM Dataset (R_A > 7 & A_{TOT} > 350 mi^2) F-13

Table F.9. t-statistic; POM Dataset (R_A > 7 & A_{TOT} > 350 mi^2); p-value < 0.05 ..... F-13

Table F.10. t-statistic; POT Dataset (R_A > 7 & A_{TOT} > 350 mi^2) F-13

Table F.11. Regulated Gage Pair Watershed Characteristics F-15

Table F.12. Regulation Characteristics F-15

Table F.13a. Unregulated/Regulated Pearson's ρ (POM) F-16

Table F.13b. Unregulated/Regulated Pearson's ρ (POT) F-16

Table F.14a. Unregulated/Regulated Kendall's τ (POM) F-16

Table F.14b. Unregulated/Regulated Kendall's τ (POM) F-16

List of Figures

Figure F.1a. Gage Pair 05, October 1989 Event F-3
Figure F.1b. Gage Pair 05, October 1989 Event (Expanded Scale) F-3
Figure F.2. Gage Pair 05, May 1990 Event F-4
Figure F.3a. Gage Pair 05, October 2005 Event F-5
Figure F.3b. Gage Pair 05, October 2005 Event (Expanded Scale) F-5
Figure F.4. Daily and Instantaneous Peaks for Gage Pair 05 (POM) F-6
Figure F.5. Daily and Instantaneous Peaks for Gage Pair 05 (POT) F-6
Figure F.6. Daily and Instantaneous Peaks for Gage Pair 06 (POM) F-7
Figure F.7. Daily and Instantaneous Peaks for Gage Pair 06 (POT) F-7
Figure F.8. Daily and Instantaneous Peaks for Gage Pair 08 (POM) F-8
Figure F.9. Daily and Instantaneous Peaks for Gage Pair 08 (POT) F-8
Figure F.10. Daily and Instantaneous Peaks for Gage Pair 21 (POM) F-9
Figure F.11. Daily and Instantaneous Peaks for Gage Pair 21 (POT) F-9
Figure F.12. Daily and Instantaneous Peaks for Gage Pair 43 (POM) F-10
Figure F.13. Daily and Instantaneous Peaks for Gage Pair 43 (POT) F-10
Figure F.14. Daily and Instantaneous Peaks for Gage Pair 76 (POM) F-11
Figure F.15. Daily and Instantaneous Peaks for Gage Pair 76 (POT) F-11

This appendix includes supplemental information referenced in the main report.

F.1 Prototype Watershed Data

Summaries of the available prototype watershed flow data, event plots, and correlations are provided in this section.

F.1.1 Watershed Summaries

Data availability for the prototype gage pairs are summarized in Tables F.1 through F.6.

Table F.1. Data Summary for Gage Pair 05 (New Jersey).

	Main (01403060)	Tributary (01403150)	Notes
Area (mi^2)	784	2	
Annual Duration Series	WY 1882, 1896, 1904-1909, 1936-1939, 1942, 1945-2006 (n=75)	WY 1980-2006 (n=27)	
Mean Daily	1904-1909, 1945-2006 (n=68)	1980-2006 (n=27)	
Instantaneous	1981-2006 (n=26)	1981-2006 (n=26)	15 min data

Table F.2. Data Summary for Gage Pair 06 (New Jersey).

	Main (01445500)	Tributary (01446000)	Coincident[1] Data	Notes
Area (mi^2)/T_c (h)	106/15.1	36.7/5.5		
Annual Duration Series	WY 1922-2006 (n=85)	WY 1923-2006 (nic 1996-2002) (n=77)	65% (n=50)	
Partial Duration Series	WY 1922-1996 (n=196 over 75 yrs, 2.6/yr)	WY 1923-1995 (n=141 over 73 yrs, 1.9/yr)	36% (n=71)	Some yrs in both gages only have annual peaks.
Mean Daily	10/1/1921-8/28/2007	10/1/1922-9/30/1961; 6/7/2003-8/28/2007		
Instantaneous	10/1/81-9/30/2006	6/7/2003-9/30/2006		15 min data

1. Peak on tributary occurs within plus or minus 1 day of the peak on the main.

Table F.3. Data Summary for Gage Pair 08 (Maryland).

	Main (01590000)	Tributary (01590500)	Notes
Area (mi^2)	8.5	6.9	
Annual Duration Series	WY 1932-1974 (n=43)	WY 1944-1990 nic 1953-1964 (n=35)	
Mean Daily	WY 1932-1974 (n=43)	WY 1943-1990 nic 1953-1974 (n=26)	
Instantaneous	None available	12/12/83 – 4/2/84	30 minute data

Table F.4. Data Summary for Gage Pair 21 (Michigan).

	Main (04140500)	Tributary (04141000)
Area (mi^2)	117	1.2
Annual Duration Series	WY 1951-2006 (n=56)	WY 1952-1987 (n=36)
Mean Daily	WY 1951-1982 (n=32)	WY 1952-1978 (n=27)
Instantaneous	None available	None available

Table F.5. Data Summary for Gage Pair 43 (Washington).

	Main (12082500)	Tributary (12083000)	Coincident[1] Data	Notes
Area (mi^2)/T_c (h)	133/3.8	70.3/3.0		
Annual Duration Series	WY 1943-2006 (n=64)	WY 1943-2006 (n=64)	53% (n=34)	
Partial Duration Series	WY 1943-1995 (n=187 over 53 yrs, 3.5/yr)	WY 1943-1995 (n=136 over 53 yrs, 2.6/yr)	56% (n=105)	
Mean Daily	6/1/1942-8/28/2007	6/1/1942-3/31/2007		
Instantaneous	10/1/1987-9/30/2006	10/1/1987-9/30/2006		15 min data

1. Peak on tributary occurs within plus or minus 1 day of the peak on the main.

Table F.6. Data Summary for Gage Pair 76 (Texas).

	Main (08164450)	Tributary (08164503)	Coincident[1] Data	Notes
Area (mi^2)/T_c (h)	289/25.6	178/32.6		
Annual Duration Series	WY 1978-2006 (n=29)	WY 1978-2006 (n=29)	45% (n=13)	
Partial Duration Series	WY 1980-1995 (nic 1992-1994) (n=26 over 13 yrs, 2/yr)	WY 1980-1995 (nic 1992-1994) (n=13 over 13 yrs, 1/yr)	19% (n=5)	Tributary does not appear to be a true PDS dataset.
Mean Daily	10/1/1977-8/28/2007	10/1/1977-8/28/2007		
Instantaneous	10/1/1997-9/30/2005	10/1/1991-9/30/2005		60 min data before 9/11/1997; 15 min data after

1. Peak on tributary occurs within plus or minus 1 day of the peak on the main.

F.1.2 Selected Events for Gage Pair 05

In addition to those presented in the main report, selected flood event hydrographs for Gage Pair 05 are presented to provide illustration of potential issues with using daily data especially for the small tributary watershed in this gage pair. Figures F.1a suggests that the daily and instantaneous peaks correspond to some degree, but the scale difference is the dominating feature of the hydrographs. Figure F.1b expands the scale to better show the tributary hydrograph. In this case, the daily value is much lower than the instantaneous peak because the hydrograph has a shorter duration as would be typical of the smaller drainage area and shorter time of concentration.

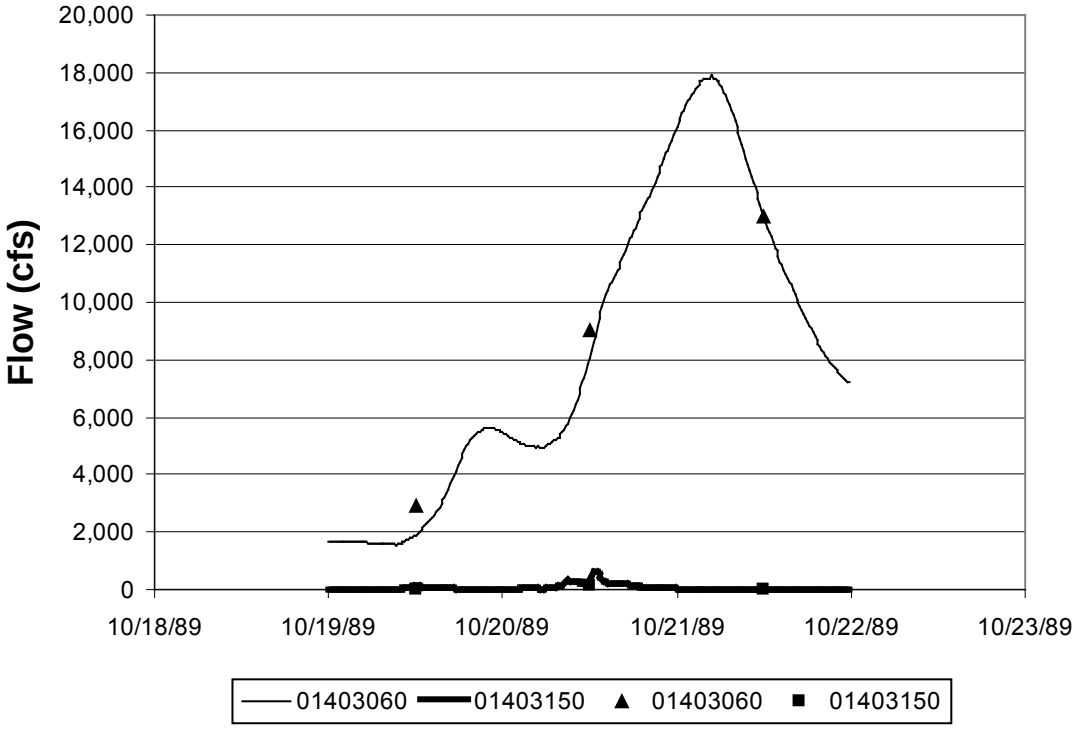

Figure F.1a. Gage Pair 05, October 1989 Event.

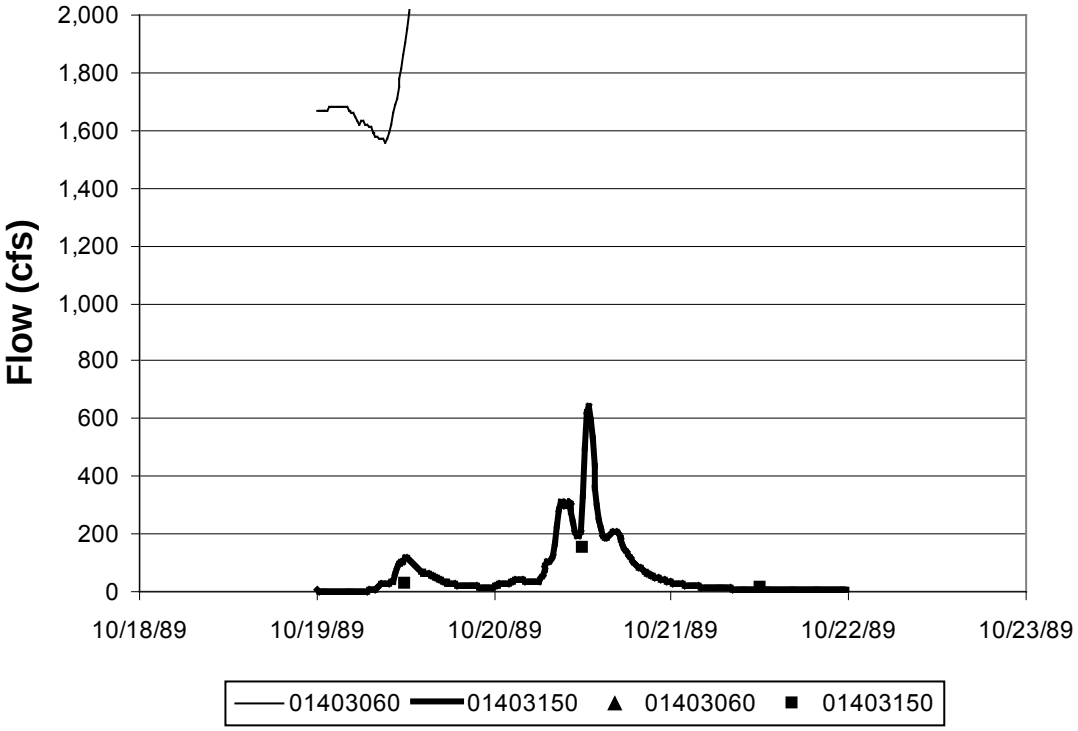

Figure F.1b. Gage Pair 05, October 1989 Event (Expanded Scale).

Figure F.2 describes an event where the same storm may have caused the peaks at both gages, but the tributary peak is receding before the rising limb of the main stream hydrograph is established. This event is an example where the daily average values completely miss the instantaneous peak for the tributary gage.

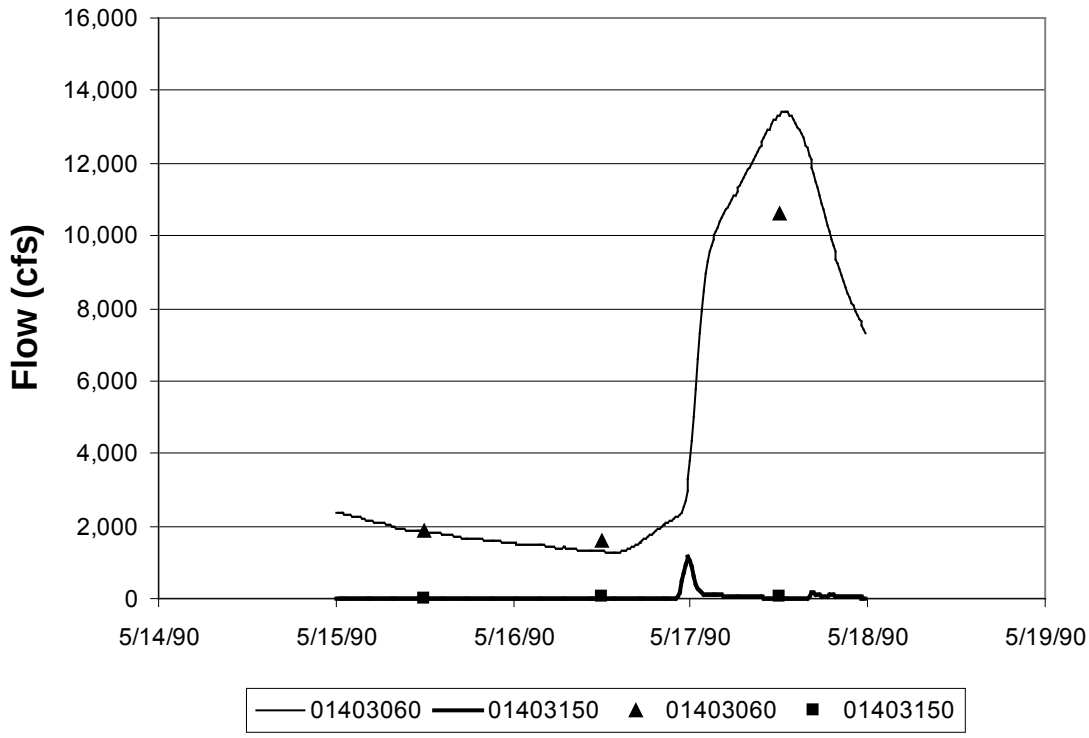

Figure F.2. Gage Pair 05, May 1990 Event.

Figure F.3a illustrates an earlier response for the tributary compared with the main for an October 2005 event. It shows that the daily flows are not highly representative of the instantaneous flows for the tributary gage in absolute terms.

However, the primary interest is how daily flows represent instantaneous flows in terms of their statistical nature. For example, the instantaneous and daily average peaks for the main shown in Figure F.3a both represent flows within the 0.5 to 0.1 exceedance probability quantile bins. The instantaneous and daily average peaks for the tributary are shown in Figure F.3b. Although the former falls within the 0.1 to 0.04-quantile bin, the daily average peak provides some evidence of an elevated condition. Its value falls within the 0.5 to 0.1-quantile bin. As expected for the smaller watershed, there is a loss of information, but not a total loss.

F.1.3 Instantaneous Peaks and Mean Daily Flows

Because of the wide availability of high quality mean daily flow data desirable for this research, it was important to determine whether mean daily data could reasonably be used as an alternative to instantaneous data. Some of the analyses supporting the conclusion that the use of mean daily data would allow the research to meet its goals are described in this section.

F.1.3.1 Correlation Plots for Instantaneous Peaks and Mean Daily Flows

The correlations between instantaneous and mean daily flow peaks are shown in Figures F.4 through F.15.

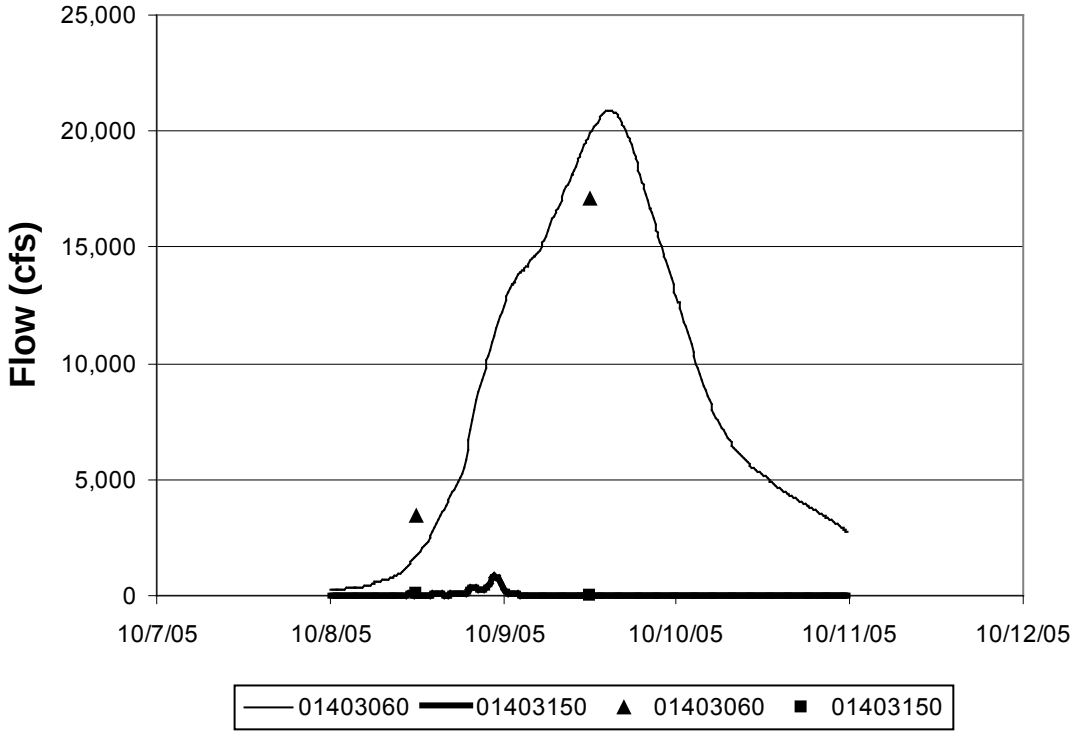

Figure F.3a. Gage Pair 05, October 2005 Event.

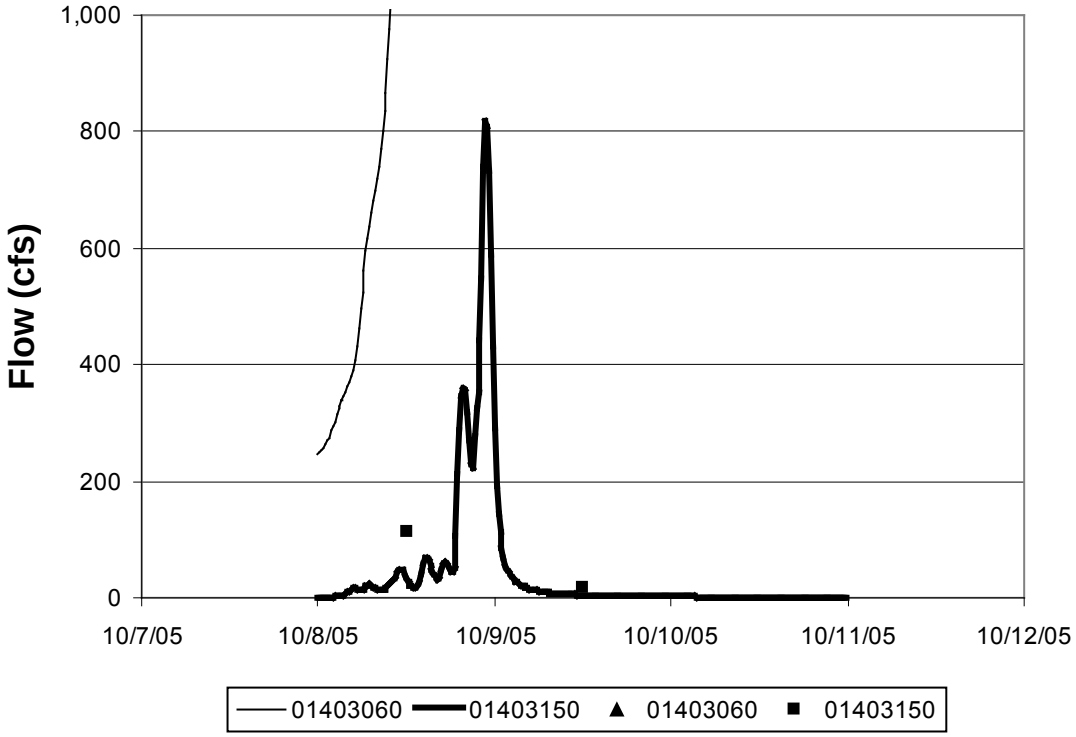

Figure F.3b. Gage Pair 05, October 2005 Event (Expanded Scale).

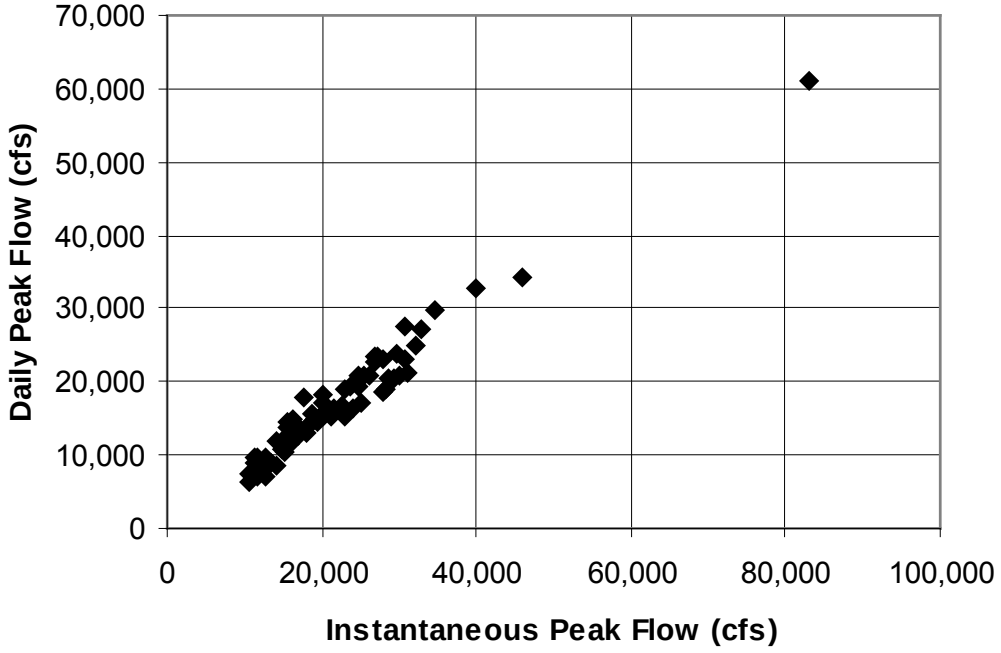

Figure F.4. Daily and Instantaneous Peaks for Gage Pair 05 (POM).

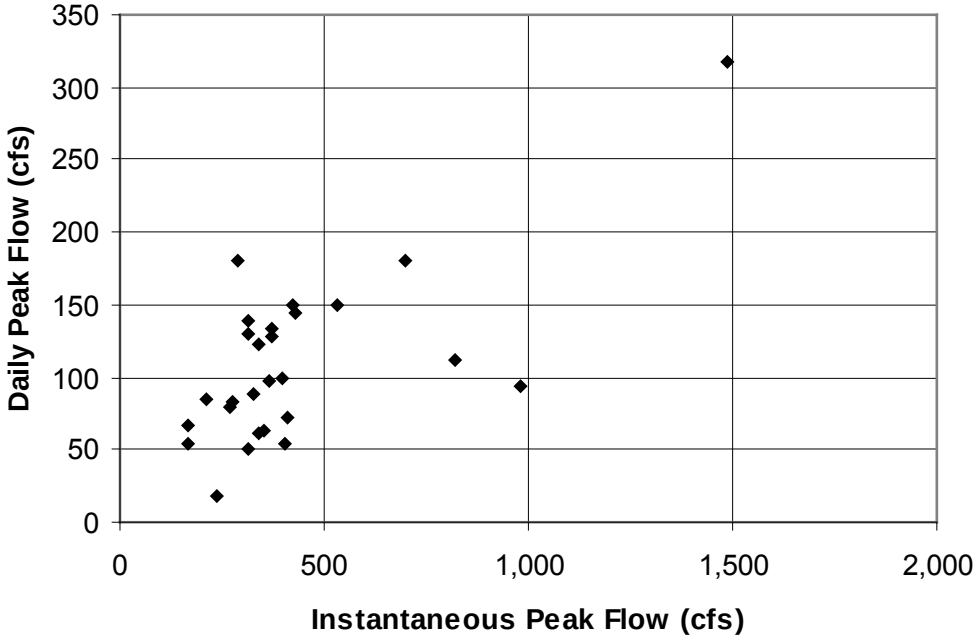

Figure F.5. Daily and Instantaneous Peaks for Gage Pair 05 (POT).

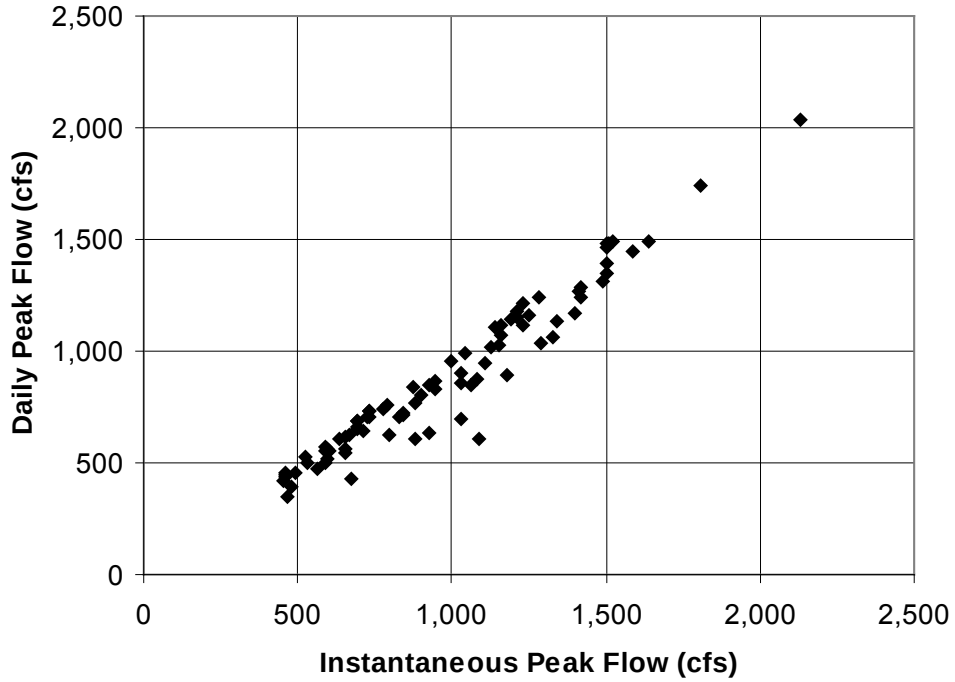

Figure F.6. Daily and Instantaneous Peaks for Gage Pair 06 (POM).

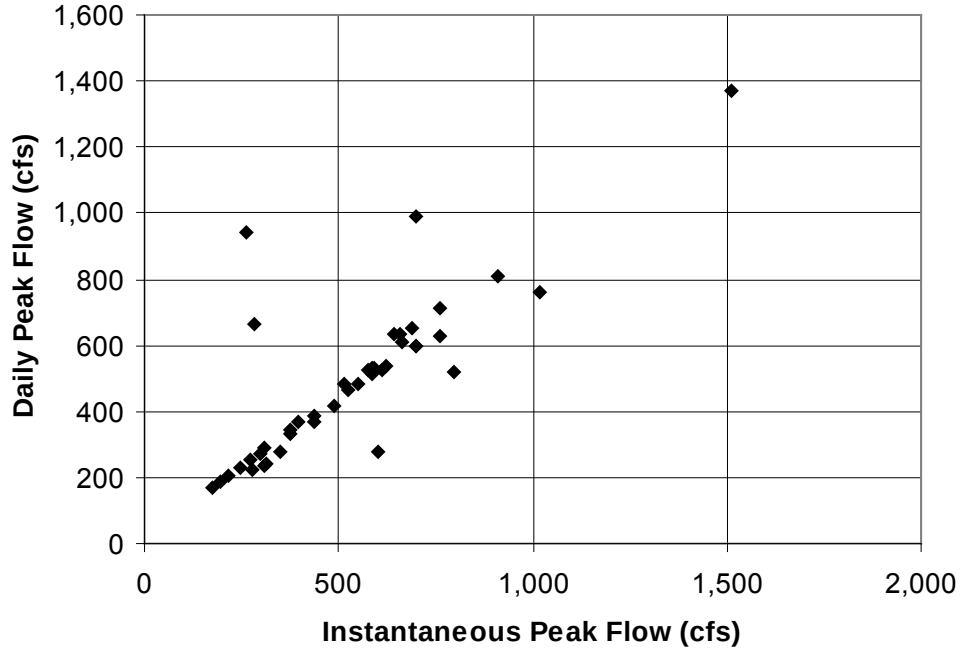

Figure F.7. Daily and Instantaneous Peaks for Gage Pair 06 (POT).

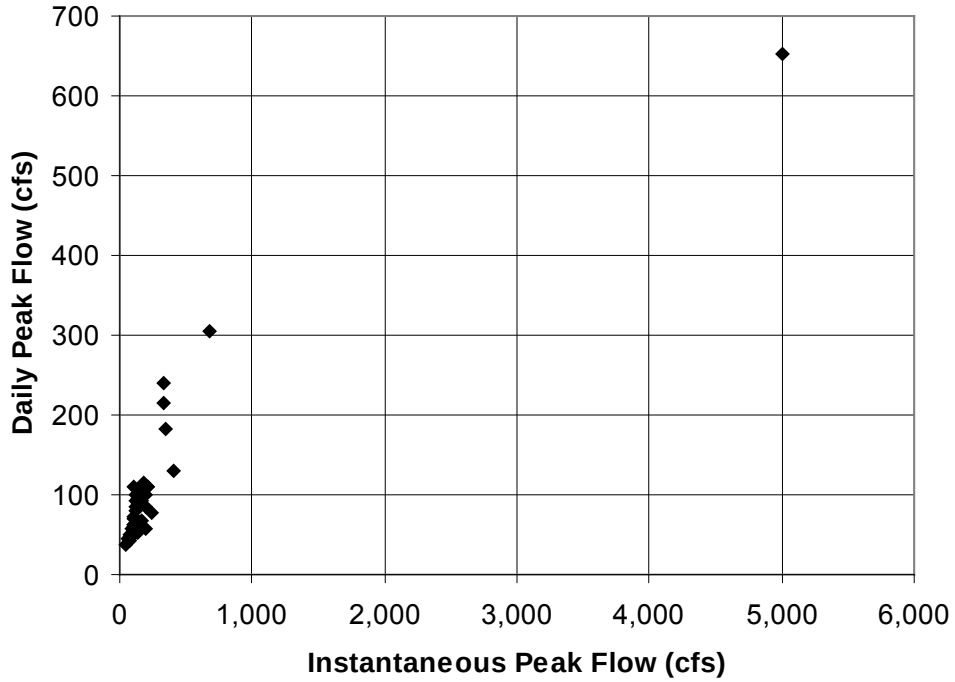

Figure F.8. Daily and Instantaneous Peaks for Gage Pair 08 (POM).

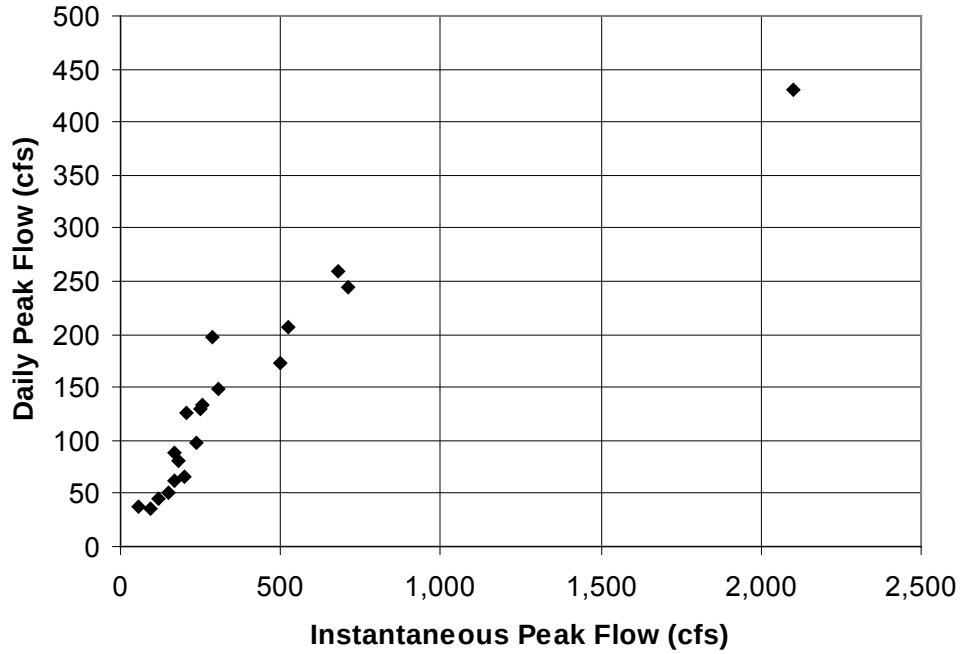

Figure F.9. Daily and Instantaneous Peaks for Gage Pair 08 (POT).

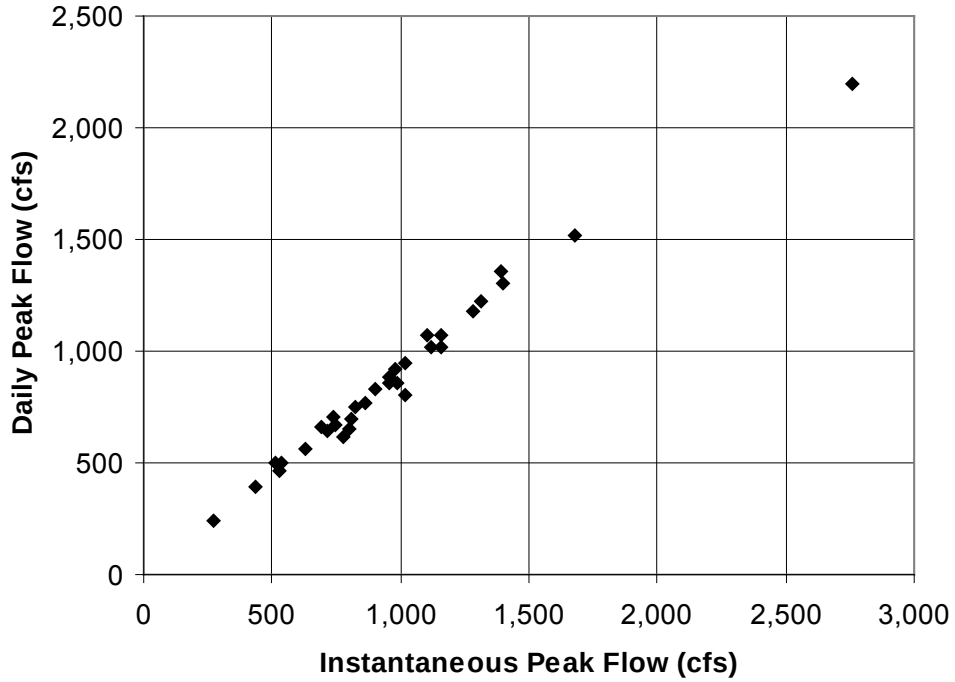

Figure F.10. Daily and Instantaneous Peaks for Gage Pair 21 (POM).

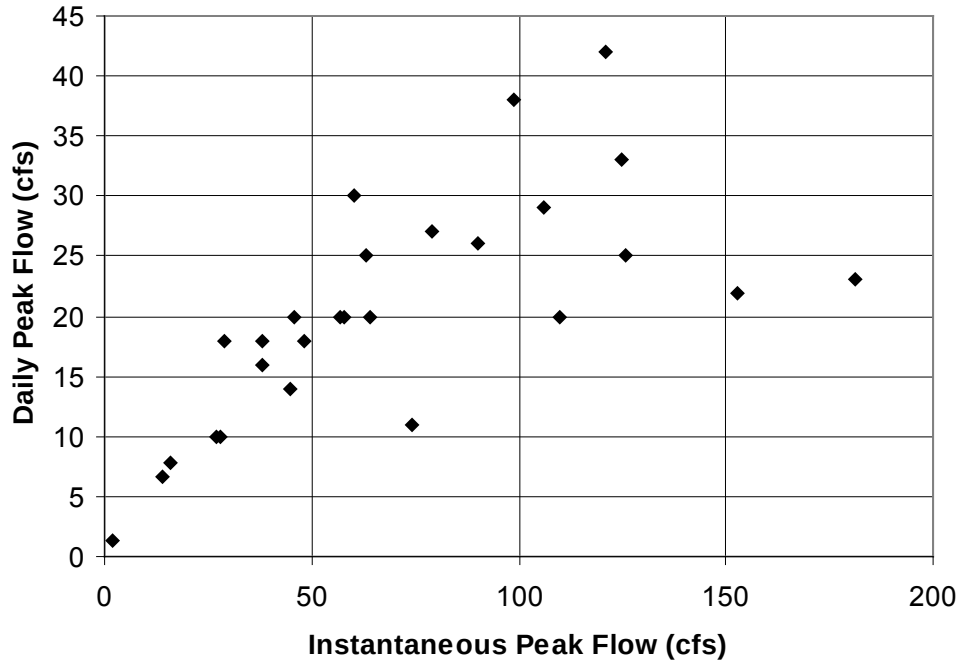

Figure F.11. Daily and Instantaneous Peaks for Gage Pair 21 (POT).

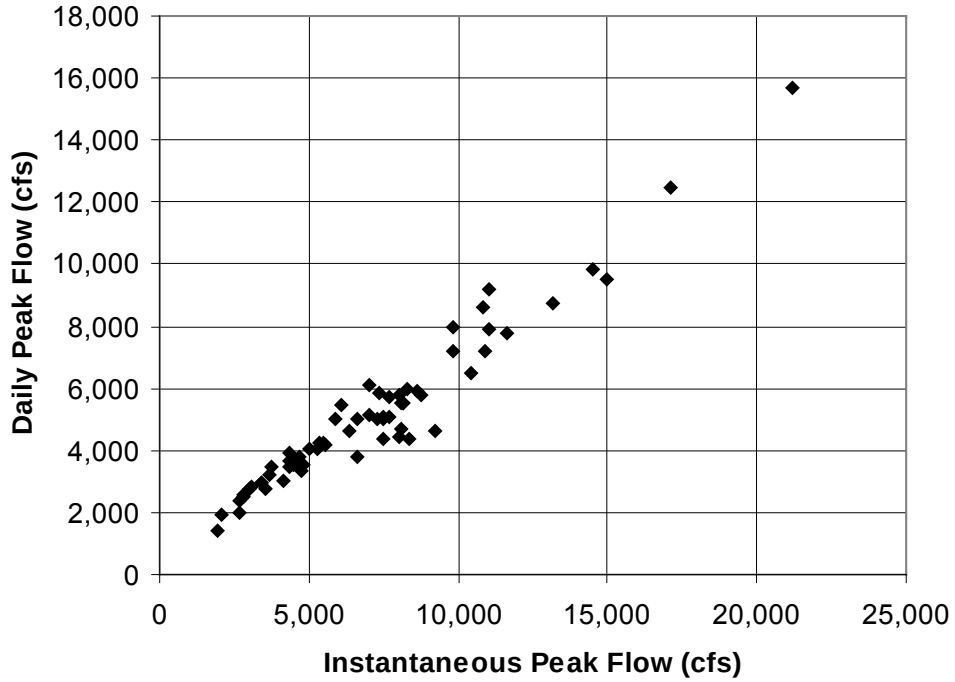

Figure F.12. Daily and Instantaneous Peaks for Gage Pair 43 (POM).

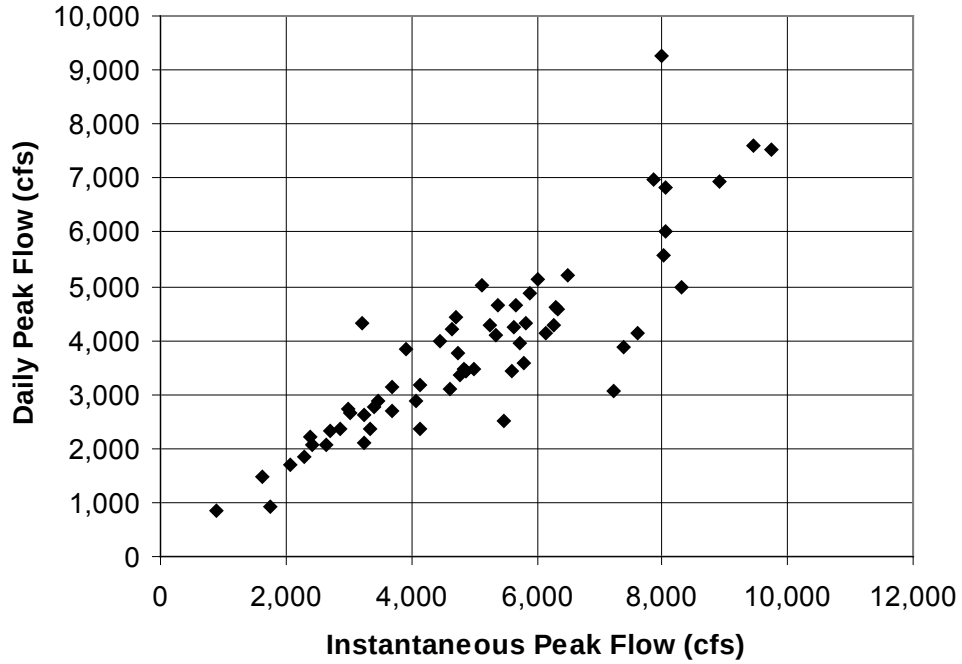

Figure F.13. Daily and Instantaneous Peaks for Gage Pair 43 (POT).

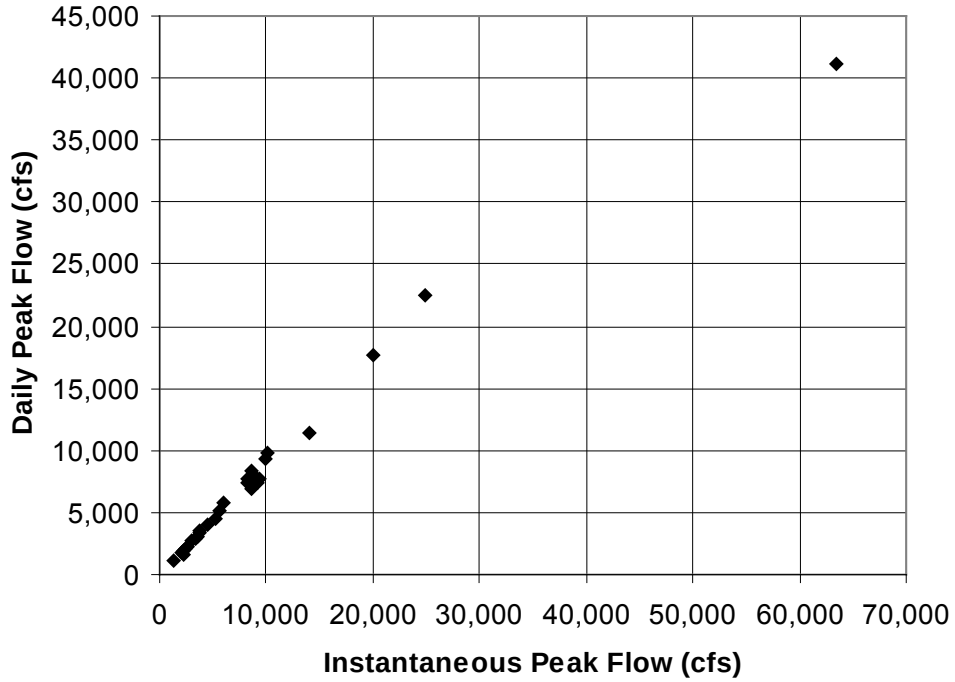

Figure F.14. Daily and Instantaneous Peaks for Gage Pair 76 (POM).

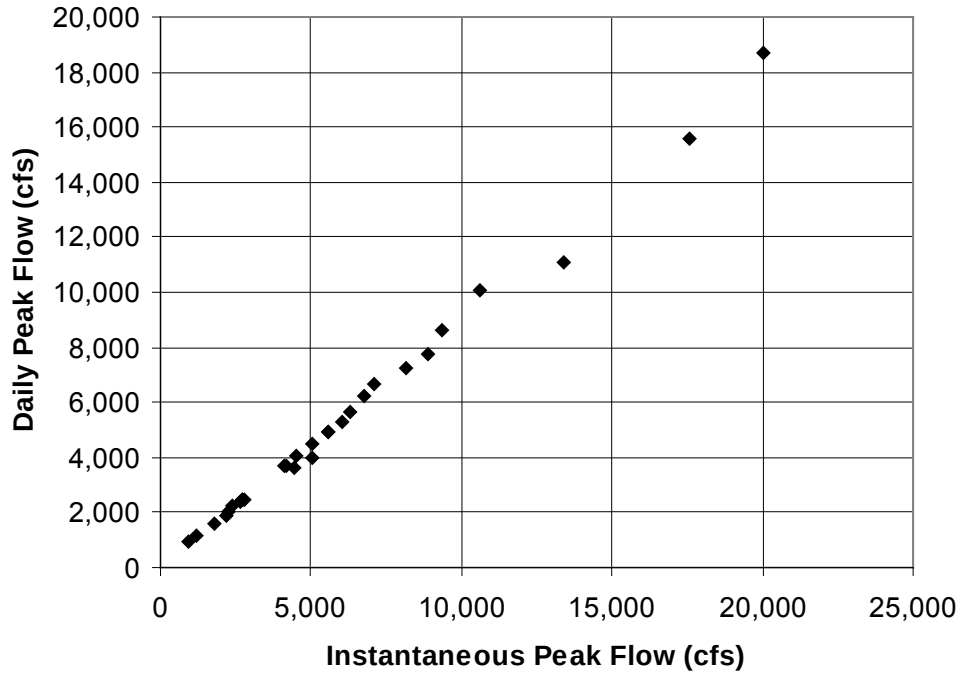

Figure F.15. Daily and Instantaneous Peaks for Gage Pair 76 (POT).

F.1.3.2 Small Watersheds

As described in Section 3.1.3 of the main report, there was particular concern that using daily data rather than instantaneous data would affect small watersheds (drainage area less than 6.5 mi^2). Seven gage pairs include a tributary watershed meeting this definition of small: GP02, GP03, GP05, GP11, GP21, GP33, and GP44. This represents less than 10 percent of the gage pairs in the 85 gage pair database.

The research team estimated Pearson's ρ and Kendall's τ for each gage pair used in the analyses for both the Peak on Main (POM) and Peak on Tributary (POT) dataset along with the corresponding p-values. If the use of daily data obscured a correlative relation between the two gages in a pair, one might expect that the p-value would show that the estimated correlations are not significant. A summary of the number of pairs with p-values greater than 0.05 (not significant) is provided in Table F.7. Depending on the dataset and parameter either 2 or 3 of the 7 gage pairs with a tributary area less than 6.5 mi^2 resulted in correlation parameters that were not significantly different from zero at the 5 percent significance level. In contrast, between 1 and 5 gage pairs with a larger tributary area were not statistically significant out of the remaining 78 gage pairs.

Table F.7. Summary of Gage Pairs with p-value > 0.05.

Dataset	Parameter	A_T < 6.5 mi^2 (Gage Pairs)	A_T > 6.5 mi^2 (Gage Pairs)	Total
POM	Pearson's ρ	3 (05,11,33)	5 (12,16,35,38,80)	8
	Kendall's τ	2 (05,11)	3 (16,35,80)	5
POT	Pearson's ρ	2 (05,21)	2 (66,84)	4
	Kendall's τ	2 (05,21)	1 (66)	3

Five of the seven gage pairs with small tributary watersheds have an area ratio, R_A, greater than 7 and a total watershed area, A_{TOT}, greater than 350 mi^2. According to the groupings discussed in Section 4.2.4 of the main report, these five gage pairs are a subset of the 19 gage pairs with these area ratio and total area characteristics. An analysis was conducted to determine if use of daily data obscured correlative relations among gage pairs to such a degree that one can say the 5 gage pairs with small tributary watersheds represent a different population than the 14 gage pairs with large tributary watersheds. The other three groupings could not be examined because there was an insufficient number of gage pairs with small watersheds in each of these groups to produce meaningful results.

Results of the t-test for significance on area ratio, total area, Pearson's ρ, and Kendall's τ are summarized in Table F.8. At a five percent level of significance, the values of Pearson's ρ and Kendall's τ in the two groups (A_T < 6.5 mi^2 and A_T > 6.5 mi^2) are considered to be from different populations. One might conclude that the use of daily values resulted in these pairs appearing to be from different populations when they should not be, especially, since the area ratios for the two groups are not considered to be from different populations statistically. However, the total area measurements from the two groups are considered to be from different populations. Therefore, the question becomes are the correlation parameters statistically different because of the use of daily data, because the total areas are different, or for some other reason?

Table F.8. t-statistic; POM Dataset (R_A > 7 & A_{TOT} > 350 mi^2).

Parameter	t	t' (at 5% significance)	Significant (Yes or No)
Area Ratio	1.53	1.92	No
Total Area	1.81	1.73	Yes
Pearson's ρ	2.99	1.95	Yes
Kendall's τ	2.07	1.87	Yes

To answer this question, the gage pairs with p-values for the correlation estimates greater than 0.05 were dropped and the data reanalyzed. Of the 19 gage pairs, 6 estimates of Pearson's ρ have associated p-values greater than 0.05. Of those, 3 gage pairs have a small tributary watershed and 3 gage pairs have a large tributary watershed. Similarly, of the 19 gage pairs, 5 estimates of Kendall's τ have associated p-values greater than 0.05. Of those, two gage pairs have a small tributary watershed and three gage pairs have a larger tributary watershed.

The results of the t-tests with the reduced dataset are summarized in Table F.9. With gage pairs where the p-value is greater than 0.05 removed, the conclusion is that the two groups are not from different populations. The only contrary indication is for total area as it relates to Pearson's ρ. The results in Table F.9 compared with the results in Table F.8 suggest that the use of daily data may have hindered the estimation of a statistically significant correlation parameter (ρ or τ), but the gage pairs with the small tributary watersheds can be said to be from the same population as those with the larger tributary watersheds.

Table F.9. t-statistic; POM Dataset (R_A > 7 & A_{TOT} > 350 mi^2); p-value < 0.05.

Assessment	Parameter	t	t' (at 5% significance)	Significant (Yes or No)
p-value < 0.05 for Pearson's ρ (n=13)	Area Ratio	0.54	1.81	No
	Total Area	2.71	1.81	Yes
	Pearson's ρ	1.60	2.10	No
p-value < 0.05 for Kendall's τ (n=14)	Area Ratio	0.70	2.22	No
	Total Area	1.29	1.81	No
	Kendall's τ	1.89	1.96	No

The POT dataset was examined in the same way. The resulting t-tests are summarized in Table F.10. For these data, the correlation parameter (ρ and τ) does not appear to be from different populations suggesting that the use of daily data did not introduce misleading data. Further analysis of these data after deleting gage pairs with a p-value greater than 0.05 was not pursued because only one gage pair of the 19 met that criterion (GP05).

Table F.10. t-statistic; POT Dataset (R_A > 7 & A_{TOT} > 350 mi^2).

Parameter	t	t' (at 5% significance)	Significant (Yes or No)
Area Ratio	1.53	1.92	No
Total Area	1.81	1.73	Yes
Pearson's ρ	1.75	1.89	No
Kendall's τ	0.78	2.07	No

F.2 Watershed Regulation

As stated in Section 2.3 of the main report, the gage pairs for this research were selected to be free of watershed regulation as identified by the USGS in the peak flow gage records. However, many watersheds are affected by regulation to some degree and it is of interest to know what limits there might be on the application of the methods recommended in this report when at least one of the watersheds in a pair is regulated. Two assessments are proposed.

The first level of assessment is for the designer to determine if it is appropriate to apply methods for estimating peak flow to a regulated watershed when those methods do not explicitly account for regulation. That is, if the regulation is sufficiently small in relation to the scope of the watershed, the designer can estimate peak flows using the preferred methods for the site. Defining “sufficiently small” requires application of professional judgment on the part of the designer. It follows that if the designer has determined that regulation on a pair of watersheds is sufficiently small that application of the same methodologies for estimating peak flows is acceptable, then the methods developed for this research are also considered valid.

The second level of assessment is required when the designer determines that regulation invalidates flow estimation techniques that do not address regulation explicitly. The literature is relatively quiet regarding the potential impact of regulation on watershed output. Benson examined flood frequency curves from relatively wet New England states (Benson, 1963) and relative dry western states of Texas, New Mexico, and Colorado (Benson, 1964). His threshold for a "significant" impact was a ten percent reduction in annual peak flows. (This would correspond to about a ten-percent reduction in the flood frequency curve). Benson reported that for watersheds in relatively wet regions, useable storage of 100 ac-ft/mi^2 was sufficient to cause the threshold impact in the annual peak series. For dryer climates, useable storage of 50 ac-ft/mi^2 was similarly sufficient to cause the threshold impact in the annual peak flow series.

Asquith (2002) examined the impact of regulation on Texas watersheds as measured in the peak discharge L-moments. He used maximum capacity, normal capacity, surface area, and the difference between maximum and normal reservoir storage as predictor variables in his research. His conclusion was that "regulation" is a complex interaction of watershed characteristics and that the impact is most apparent in the L-mean statistic. The remaining L-moments, L-CV and L-skew, might be affected, but no statistically significant effect was measured.

Asquith (2002) noted that the U.S. Geological Survey (USGS) standard practice is to consider "regulated" watersheds to be those in which at least ten percent of drainage area is affected by storage. Although his comments pertained to the Texas Water Science Center, this definition is generally used nationwide.

A set of watershed pairs was selected to test whether the impact of reservoirs on the resulting correlation between peak discharges from confluent watersheds could be observed, and if so, what implications might there be for applying the methods developed from this research. Four gage pairs were identified; one from the original database retrieved for this study (GP07) and three identified from the work of Asquith (2002). Watershed characteristics from these gage pairs are listed in Table F.11. These gage pairs are intended to inform the issue and are not considered statistically representative. The analysis performed here may suggest a more comprehensive analysis based on a larger and more representative dataset.

The objective was to determine if there is a measurable difference in the correlation parameter (Pearson’s ρ and Kendall’s τ) caused by regulation. The gage pairs in Table F.11 were assessed to identify the cause, period, and degree of regulation within the pair. An unregulated

period for the gage pair was identified as well as one or more regulated periods. A summary of these characteristics is provided in Table F.12.

Table F.11. Regulated Gage Pair Watershed Characteristics.

Characteristic	Pair 07	Pair R2	Pair R5	Pair R6
Main watershed				
Gage ID	01536500	07346000	08172000	08104700
Area (sq mi)	9,600	850	838	248
State	PA	TX	TX	TX
Tributary watershed				
Gage ID	01537500	07346070	08173000	08104900
Area (sq mi)	15.7	675	309	138
State	PA	TX	TX	TX
Joint Characteristics				
Total area (sq mi)	9976	1525	1147	386
Drainage area ratio	634	1.3	2.7	1.8

Table F.12. Regulation Characteristics.

Characteristic	Pair 07 (1950)	Pair 07 (1979)	Pair R2	Pair R5	Pair R6
Regulated Gage	01536500	01536500	07346000	08173000	08104700
Main or Tributary?	Main	Main	Main	Tributary	Main
Unregulated Period of Record (Water years)	1941-1949	1941-1949	1947-1957	1940-1963	1969-1979
Regulated Period of Record (Water years)	1950-1978	1979-1990	1958-1959, 1980-2009	1964-1983	1980-2009
Number of Unregulated Years	9	9	11	24	11
Number of Regulated Years	29	12	32	20	30
Regulation Source	4 flood control reservoirs	9 flood control reservoirs	1 flood control reservoir	27 flood retarding structures	1 multi-purpose reservoir
Percent of Watershed Affected by Regulation	4	12	100	39	99
Unit Storage (ac-ft/mi^2)	25	69	2052	225	804

Reservoirs were put into service at various times for Gage Pair 07. To represent this sequencing, two separate unregulated/regulated comparisons were determined. The unregulated period is the same for both: 1941-1949. In the first column for Gage Pair 07 the regulated period begins in 1950 based on the construction of 4 flood control reservoirs. Additional reservoirs were constructed later resulting in a second column for Gage Pair 07 with the regulated period beginning in 1979.

For gage pair R5, both watersheds became regulated in 1984. Therefore, 1964 to 1983, when only one watershed was regulated and the other was unregulated, was considered the regulated period. Gage pairs R2 and R6 required no additional considerations and their characteristics are summarized in the table.

Pearson's ρ and Kendall's τ were calculated for each unregulated and regulated period of record for both the peak on main (POM) and peak on tributary (POT) datasets. The p-value, indicating whether the result should be considered significantly different from zero, was also calculated for each estimate of Pearson's ρ and Kendall's τ. Elsewhere in this report, a p-value

less than 0.05 has been considered to represent a significant result; that criterion is also applied here. Summaries of the computations are presented in Tables F.13a, F.13b, F.14a, and F.14b.

Table F.13a. Unregulated/Regulated Pearson's ρ (POM)

		GP07 (1950)	GP07 (1979)	GPR2	GPR5	GPR6
Pearson's ρ	Unregulated	0.75	0.75	0.78	0.19	0.88
	Regulated	0.26	0.40	0.91	0.60	0.43
p-value	Unregulated	0.020	0.020	0.004	0.371	0.000
	Regulated	0.175	0.202	0.000	0.006	0.151
Regulated-Unregulated ρ		-0.49	-0.35	0.12	0.40	-0.45

Table F.13b. Unregulated/Regulated Pearson's ρ (POT)

		GP07 (1950)	GP07 (1979)	GPR2	GPR5	GPR6
Pearson's ρ	Unregulated	-0.55	-0.55	0.80	0.72	0.76
	Regulated	0.39	0.57	0.38	0.74	0.39
p-value	Unregulated	0.122	0.122	0.003	0.000	0.007
	Regulated	0.039	0.050	0.030	0.000	0.032
Regulated-Unregulated ρ		0.94	1.12	-0.42	0.02	-0.37

Table F.14a. Unregulated/Regulated Kendall's τ (POM)

		GP07 (1950)	GP07 (1979)	GPR2	GPR5	GPR6
Kendall's τ	Unregulated	0.63	0.63	0.53	0.25	0.67
	Regulated	0.42	0.47	0.13	0.64	0.21
p-value	Unregulated	0.020	0.020	0.026	0.087	0.003
	Regulated	0.001	0.033	0.284	0.000	0.101
Regulated-Unregulated τ		-0.21	-0.16	-0.39	0.39	-0.46

Table F.14b. Unregulated/Regulated Kendall's τ (POM)

		GP07 (1950)	GP07 (1979)	GPR2	GPR5	GPR6
Kendall's τ	Unregulated	-0.56	-0.56	0.75	0.67	0.67
	Regulated	0.26	0.21	0.16	0.60	0.19
p-value	Unregulated	0.045	0.045	0.001	0.000	0.003
	Regulated	0.045	0.381	0.200	0.000	0.143
Regulated-Unregulated τ		0.82	0.77	-0.59	-0.07	-0.48

The difference between the unregulated and regulated correlation parameter is included in each table. It was hypothesized that correlation between coincident watersheds is reduced with regulation because regulation attenuates peak flows. The presence of reservoirs on one watershed appears to interrupt the arrival of peak discharges at the confluence. For Pearson's ρ for the POM datasets (Table F.13a) 3 of the 5 cases show a decrease in Pearson's ρ in the regulated period of record compared to the unregulated period of record. For Pearson's ρ for the POT dataset (Table F.13b), Kendall's τ for the POM dataset (Table F.14a), and Kendall's τ for the POT dataset (Table F.14b), 2, 4, and 3, respectively, of the 5 cases show a decrease in

the correlation parameter. In the aggregate, 12 of 20 cases show the hypothesized decrease in the correlation parameter.

It is noted that gage pair 07 exhibits negative correlations for the unregulated period of record for the POT datasets (Tables F.13b and F.14b); negative correlation is not a reasonable result. For Pearson's ρ (Table F.13.b), it statistically insignificant as determined by the high p-value. Eliminating these four cases from consideration means that in the aggregate 12 of 16 cases show a decrease in correlation with regulation.

Because of the small sampling of regulated gage pairs in this analysis, few definitive conclusions can be drawn. The short record lengths of many of the datasets (several were less than or equal to 12 years), the potential for outliers in the subdivided periods of record, and the questionable levels of significance (p-values) for many of the computations all limit the prescriptive nature of the analysis.

However, two observations are offered:

1. Small levels of regulation might affect the estimated correlation. For example, the Gage Pair 07 (1950) comparison includes regulation that affects only 4 percent of the watershed with an average storage of 25 ac-ft/mi^2. Yet, the Kendall's τ (Table F.14a) was reduced from 0.63 to 0.42.
2. Regulation tends to reduce correlation. Therefore, if the methods of this research are applied to a regulated watershed, the application is likely to be conservative.

These observations are qualified by the small, non-representative set of data analyzed. The dataset also has variations in drainage area ratio and total area that may affect statistical behavior. In addition, Gage Pair R5 tends to show increased correlations with regulation. Possibly significant is that this gage pair includes a regulated tributary whereas the others include a regulated main stream.

Given the dataset comprised only four gage pairs (five unregulated/regulated comparisons), the results should not be considered a statistical analysis. However, the observations represent an initial review of a complex problem. A complete analysis would require at least another dozen watershed pairs.

If the standard protocol developed for this research were to be applied in the presence of regulation on one or both of the paired watersheds, then the greatest risk appears to be additional conservatism. That is, the presence of correlation between two watersheds will generally result in greater estimates of peak discharge for design. Therefore, the assumption of correlation when such correlation is not present should result in a conservative design (a somewhat larger structure), but should not result in a structure size that is insufficient to pass the required event at the selected risk level.

Appendix G. Application Guidelines

Table of Contents

List of Tables ..... **G-i**

List of Figures ..... **G-ii**

G.1 Specify the Design Condition G-1

G.2 Influence Reach Determination G-1

G.3 Select Joint Probability Method G-2

G.4 Copula-based Method G-3

G.4.1 Assess Site-Specific Dependence G-3

G.4.2 Determine Potential Combinations G-3

G.4.3 Estimate Discharges G-6

G.4.4 Analyze Combinations G-8

G.4.5 Combination Development Using Alternative Method G-9

G.4.5.1 Assess Site-Specific Dependence G-9

G.4.5.2 Determine Potential Combinations G-11

G.4.5.3 Estimate Discharges G-12

G.4.5.4 Analyze Combinations G-13

G.5 Total Probability Method G-14

G.5.1 Develop the Marginal Exceedance Frequency Curves G-14

G.5.2 Select the Conditional Probability Matrix G-15

G.5.3 Compute Stages G-18

G.5.4 Compute Total Probability G-18

G.5.5 Construct Total Probability Curve G-19

Appendix G provides more detailed application guidelines for situations of high sensitivity or vulnerability either for the infrastructure itself or for the traveling public.

Appendix H provides a step-by-step applications guide in a user-friendly format. It is intended to be applied to most design situations and objectives.

List of Tables

Table G.1. Example Watershed Pair Data. G-2

Table G.2. Kendall's τ Based on Grouping. G-3

Table G.3. 10-yr Joint Return Periods (Groupings). G-4

Table G.4. 25-yr Joint Return Periods (Groupings). G-4

Table G.5. 50-yr Joint Return Periods (Groupings). G-4
Table G.6. 100-yr Joint Return Periods (Groupings). G-5
Table G.7. 500-yr Joint Return Periods (Groupings). G-5
Table G.8. Example Regression Discharge Results. G-6
Table G.9. Potential Combination Discharges. G-8
Table G.10. Hypothetical Stage and Velocity at the Example Project Site. G-8
Table G.11. Kendall's τ Equation Selection Matrix. G-9
Table G.12. 10-yr Joint Return Periods. G-11
Table G.13. 25-yr Joint Return Periods. G-11
Table G.14. 50-yr Joint Return Periods. G-11
Table G.15. 100-yr Joint Return Periods. G-12
Table G.16. 500-yr Joint Return Periods. G-12
Table G.17. Potential Combination Discharges for Alternative Method. G-13
Table G.18. Hypothetical Stage and Velocity at the Example Project Site. G-13
Table G.19. Exceedance Frequency Curves Based on Bin AEPs. G-15
Table G.20. CPM Selection Table – Degree of Correlation. G-15
Table G.21a. CPM for Highly Correlated Flows (POM). G-16
Table G.21b. CPM for Highly Correlated Flows (POT). G-16
Table G.22a. CPM for Moderately Correlated Flows (POM). G-16
Table G.22b. CPM for Moderately Correlated Flows (POT). G-16
Table G.23a. CPM for Mixed Correlated Flows (POM). G-17
Table G.23b. CPM for Mixed Correlated Flows (POT). G-17
Table G.24a. CPM for Poorly Correlated Flows (POM). G-17
Table G.24b. CPM for Poorly Correlated Flows (POT). G-17
Table G.25. Summary of Stage Computations (ft). G-18
Table G.26. Total Probability Computation (POM). G-19
Table G.27. Total Probability Curve (POM). G-20

List of Figures

Figure G.1. Watershed Pair Discharge versus Return Period. G-7
Figure G.2. Total Probability Curve (POM) for Example Watershed. G-21

The purpose of this appendix is to provide guidelines for applying the methods developed from this research. The designer is encouraged to review the main report to understand the foundation of the methods.

Application of the methods is limited to watershed pairs satisfying three conditions: 1) neither watershed in the pair is substantially affected by regulation, 2) the sum of the watershed areas is less than 9,000 mi^2, and 3) the drainage area of the smaller watershed of the pair exceeds 1 mi^2. The watershed pairs used to develop the methods presented in this report satisfy these criteria.

The preliminary steps in the application of joint probability methods are implemented prior to selecting a method. They are:

1. Specify the design objective. The designer might be interested in determining tributary stage and/or velocity at the design location for a given annual exceedance probability (AEP). The desired design event or range of events is specified by the designer.
2. Determine if the project site is within the influence reach. If not, joint probability design techniques are not required.
3. Select the joint probability method.

Application of each method is described in subsequent sections of this appendix. Two methods are discussed: 1) a copula-based method and 2) the total probability method. See Appendix F.2 for a discussion of the implications of watershed regulation in the context of these methods.

G.1 Specify the Design Condition

The AEP and relevant design parameters are established. The design AEP might range from 0.1 to 0.002, corresponding to return periods from 10-yr to 500-yr. The relevant design parameters might include water-surface elevation (stage) for evaluating flooding potential and structure freeboard or velocity for evaluating erosion and scour. If both, the designer should recognize that different combinations of conditions on the two streams could be required. The designer might also wish to examine a range of design conditions as part of a cost benefit analysis.

G.2 Influence Reach Determination

Joint probability design techniques are necessary only if the project site on the tributary stream is sufficiently close to the confluence with the main stream such that the backwater from the main stream influences stage and velocity at the project site. (The converse of this is also true, but the following discussion is written assuming the project site is on the tributary.) To assess the potential interaction of flow and stage at the site of interest, the designer creates a hydraulic model for the site using HEC-RAS or other appropriate tool and runs two scenarios. Scenario A will use the design return period discharge on the tributary and the 2-yr return period discharge on the main stream. Scenario B will also use the design return period discharge on the tributary, but will use the design return period discharge on the main stream. Scenario B represents the conservative assumption that the two streams are completely dependent.

If the stage at the project site is the same between the two scenarios, within some tolerance, then the project site is outside the influence reach; joint probability techniques are not necessary. If the stage differs, the site is probably within the influence reach and one of the joint probability design procedures presented herein should be applied. In some situations where the project site is sufficiently far from the confluence or the statistical dependence

between the two watersheds is weak, application of a joint probability design procedure may ultimately show that the project site is not within the influence reach even though the initial assessment indicated that it was so located.

The professional engineer must define the acceptable tolerance on a case-by-case basis. If, for example, the purpose of the analysis is to redefine a FEMA floodplain, then a very small tolerance is appropriate. If the purpose is documentation of a hydraulics and scour analysis for a bridge with no encroachment and generous freeboard a higher tolerance is likely to be acceptable. Ambiguity favors selection of a small tolerance.

G.3 Select Joint Probability Method

Two methods were developed for use: 1) the copula-based method, and 2) the total probability method. Choice of the most appropriate method is dependent on the designer's goals, skills, experience, and preference. Both methods require the setup and calibration of a HEC-RAS (or other hydraulic model) to produce stage and velocity conditions for the analysis and design.

The copula-based method is a new method that allows for direct estimation of combinations of design discharges for a given exceedance probability, which are evaluated through a series of applications of a hydraulic model to determine the critical combination. Application of the copula method is described in Section G.4.

The total probability method (TPM) is well established, however development of conditional probability relations is a challenge for ungaged watersheds. One of the products of this research is a series of conditional probability matrices to meet that challenge. A stage-probability relation that conveniently facilitates comparison of a range of alternatives and events and requires several iterations of the hydraulic model is produced by application of the total probability method. Application is described in Section G.5.

If a designer needs to examine a single exceedance probability, the copula method might be preferred because it can likely be implemented in less time. If, however, the designer wishes to analyze a range of exceedance probabilities for a cost-benefit analysis of design alternatives, for example, the TPM might be preferred. It can be noted that a complete stage-frequency curve can also be generated using the copula method by repeating the copula method over a range of exceedance probabilities.

An illustrative example of each method is provided along with the description of the required steps. The sample watershed pair is described in Table G.1 and is gage pair 43 from the state of Washington. Although gage information is available, those data are not used so that application of the technology to an ungaged watershed pair is demonstrated. It is assumed that the design site is on the tributary stream: Mineral Creek.

Table G.1. Example Watershed Pair Data

	Nisqually River	Mineral Creek
Gage ID	12082500	12083000
A (mi^2)	133	70.3
P_M (in)	94	98

G.4 Copula-Based Method

If the copula-based method is selected, the design process continues with the following steps:

1. Assess site-specific dependence (correlation) between hydrologic events occurring on the main and tributary streams.
2. Determine possible AEP/return period combinations for the site for a selected joint exceedance probability and dependence parameter.
3. Estimate the discharges for the identified combinations.
4. Analyze the hydraulics of the system using the discharge combinations to determine the critical combination, that is, the one with the most severe (conservative) conditions. This result is the appropriate design condition.

G.4.1 Assess Site-Specific Dependence

Kendall's τ is used to quantify dependence between discharges from the two watersheds in the copula-based method. Two methods were developed and are described in more detail in Chapter 4: regression equations and data groupings. For the present discussion, the recommended data groupings method is illustrated.

With the data groupings method, the estimate of Kendall's τ depends on whether the drainage area ratio, R_A, is greater than or less than 7 and whether the total area, A_{TOT}, is greater than or less than 350 mi^2. It also depends on whether the best-fit or envelope estimates are appropriate for the design situation. The estimates of τ for each of these conditions are summarized in Table G.2. (Watershed pairs where $R_A > 7$ and $A_{TOT} < 350$ mi^2 are treated with the group where $R_A < 7$ and $A_{TOT} < 350$ mi^2.)

Table G.2. Kendall's τ Based on Grouping

Method	Drainage Area Ratio	$A_{TOT} < 350$ mi^2	$A_{TOT} > 350$ mi^2
Representative	$R_A < 7$	0.68	0.58
	$R_A > 7$	-	0.45
Envelope	$R_A < 7$	0.81	0.77
	$R_A > 7$	-	0.57

For the example, the representative method is selected. With the drainage area ratio, R_A, equal to 1.9 and A_{TOT} equal to 203 mi^2, Kendall's τ is taken as 0.68. (The estimated Kendall's τ values from the gage data are 0.50 and 0.53 for the POM and POT datasets, respectively.)

G.4.2 Determine Potential Combinations

Potential combinations for the 10-, 25-, 50-, 100-, and 500-yr joint design frequency events are provided in Chapter 4 and are repeated here in Tables G.3 through G.7. Within each table, the row labeled with the appropriate values of R_A and A_{TOT} for the project site contains combinations of events the designer might analyze with the hydraulic model to determine which is most critical for the design objective. The column labeled "equal" represents the return period on both the main and tributary streams equivalent to the target joint return period. All combinations in each row correspond to the joint return period given in the table title.

Table G.3. 10-yr Joint Return Periods (Groupings)

				Return Period on One Stream			
			Kendall's τ	1.25	2	5	Equal
Representative	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.68	-	10	9	8
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.59	-	10	9	7
	$R_A > 7$	$A_{TOT} < 350$ mi^2					
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.45	10	9	7	6
Envelope	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.81	-	-	10	9
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.77	-	-	10	8
	$R_A > 7$	$A_{TOT} < 350$ mi^2					
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.57	-	10	8	7

Table G.4. 25-yr Joint Return Periods (Groupings)

				Return Period on One Stream				
			Kendall's τ	1.25	2	5	10	Equal
Representative	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.68	-	-	25	24	19
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.59	-	25	24	22	17
	$R_A > 7$	$A_{TOT} < 350$ mi^2						
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.45	25	24	22	18	14
Envelope	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.81	-	-		25	22
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.77	-	-	-	25	21
	$R_A > 7$	$A_{TOT} < 350$ mi^2						
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.57		25	24	22	17

Table G.5. 50-yr Joint Return Periods (Groupings)

				Return Period on One Stream					
			Kendall's τ	1.25	2	5	10	25	Equal
Representative	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.68	-	-	-	50	48	38
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.59	-	50	49	48	40	33
	$R_A > 7$	$A_{TOT} < 350$ mi^2							
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.45	50	49	47	43	30	27
Envelope	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.81	-	-	-		50	43
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.77	-	-	-	50	49	41
	$R_A > 7$	$A_{TOT} < 350$ mi^2							
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.57	-	50	49	48	40	33

Table G.6. 100-yr Joint Return Periods (Groupings)

				Return Period on One Stream						
			Kendall's τ	1.25	2	5	10	25	50	Equal
Representative	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.68	-	-	-	100	98	92	75
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.59	-	-	100	98	94	80	67
	$R_A > 7$	$A_{TOT} < 350$ mi^2								
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.45	100	99	96	92	80	58	54
Envelope	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.81	-	-	-	-	100	99	86
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.77	-	-	-		100	98	83
	$R_A > 7$	$A_{TOT} < 350$ mi^2								
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.57	-	100	99	98	93	79	66

Table G.7. 500-yr Joint Return Periods (Groupings)

				Return Period on One Stream							
			Kendall's τ	1.25	2	5	10	25	50	100	Equal
Representative	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.68	-	-	-	-	500	499	495	376
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.59	-	-	500	499	497	491	476	331
	$R_A > 7$	$A_{TOT} < 350$ mi^2									
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.45	500	499	495	490	477	456	417	269
Envelope	$R_A < 7$	$A_{TOT} < 350$ mi^2	0.81	-	-	-	-	-	-	500	430
	$R_A < 7$	$A_{TOT} > 350$ mi^2	0.77	-	-	-	-	-	500	499	414
	$R_A > 7$	$A_{TOT} < 350$ mi^2									
	$R_A > 7$	$A_{TOT} > 350$ mi^2	0.57	-	-	500	499	496	490	473	327

For the example, the 50-yr joint design frequency event, and therefore Table G.5, is relevant. Within the table, the row labeled with the grouping for representative evaluation (and value of Kendall's τ) contains combinations of events the designer should analyze with the hydraulic model to determine which is most critical for the design objective.

The grouping is $R_A < 7$ and $A_{TOT} < 350$ mi^2 (Kendall's $\tau = 0.68$) for the example. By inspection of Table G.5, the following return period pairs are equivalent to the coincident 50-yr event (main stream return period, tributary return period): (10,50), (25,48), and (38,38). A 10-yr event on the main stream combined with a 50-yr event on the tributary stream represents a 50-yr coincident event for this watershed pair, as does the simultaneous occurrence of 38-yr events. The intermediate event also represents a coincident 50-yr event.

The transpose pairings provide two additional events: (50,10) and (48,25). Therefore, a 50-yr event on the main stream combined with a 10-yr event on the tributary also represents a 50-yr coincident event. In total, five potential combinations are identified.

Depending on the location of the site in the influence reach and the design objective, some combinations might be discarded before proceeding with the remainder of the analyses. For example, if the site is close to the confluence of the two streams, the main stream is substantially larger than tributary, and stage is the design variable, the (10-yr, 50-yr) combination might be discarded as its use is unlikely to produce the highest stage. Combinations might also be discarded if they are not substantially different from other combinations.

G.4.3 Estimate Discharges

The designer determines the discharges corresponding to the potential combination return periods using any appropriate method for the watersheds. Because the set of potential combinations will likely include return periods not generally estimated using standard methods, interpolation will be necessary for some return periods. Regression equations, and other methods, do not always produce "well-behaved" flood frequency curves. For this reason, it is recommended that the designer plot the estimated discharges prior to interpolating at the required exceedance probabilities.

In the illustrative example based on the Nisqually River and Mineral Creek, United States Geological Survey (USGS) regression equations are applied (Sumioka, et al., 1998). The watershed pair is located in USGS region 2 for which the following regression equation applies:

$$Q_T = aA^{b_1}P_M^{b_2} \tag{G.1}$$

where,

Q_T = design flow for return period, T, ft^3/s

A = drainage area, mi^2

P_M = mean annual precipitation, in

a, b_1, b_2 = regression constants (See Table G.8)

The input variables for the main and tributary watersheds were listed in Table G.1. The typical return period estimates of discharge are summarized in Table G.8. The discharge estimates resulting from application of Equation G.1 are displayed in Figure G.1. Discharges for return periods not shown in Table G.8 can be interpolated graphically from Figure G.1 or estimated algebraically.

Table G.8. Example Regression Discharge Results

	Regression Constants			Discharge (ft^3/s)	
Return Period	a	b_1	b_2	Nisqually River	Mineral Creek
2	0.817	0.877	1.02	6,260	3,810
10	0.845	0.875	1.14	11,300	6,920
25	0.912	0.874	1.17	13,900	8,550
50	0.808	0.872	1.23	16,400	10,100
100	0.801	0.871	1.26	18,400	11,400

The 1.25-yr return period discharge can be calculated by extrapolating from the 2-yr and 10-yr discharges as follows:

$$Q_{1.25} = 10^{1.292\log Q_2 - 0.292\log Q_{10}} \tag{G.2}$$

where,

$Q_{1.25}$ = 1.25-yr discharge, ft^3/s

Q_2 = 2-yr discharge, ft^3/s

Q_{10} = 10-yr discharge, ft^3/s

Other discharges between values summarized in Table G.8 may be interpolated in log-space using Equations G.3 and G.4.

$$\log Q_x = \frac{\log T_x - \log T_L}{\log T_H - \log T_L}\left(\log Q_H - \log Q_L\right) + \log Q_L \qquad \text{(G.3)}$$

$$Q_x = 10^{\log Q_x} \qquad \text{(G.4)}$$

where,

Q_x = Discharge at return period x, ft^3/s

Q_H = Known discharge at return period higher than x, ft^3/s

Q_L = Known discharge at return period lower than x, ft^3/s

T_x = Return period of interest, years

T_H = Return period of Q_H, years

T_L = Return period of Q_L, years

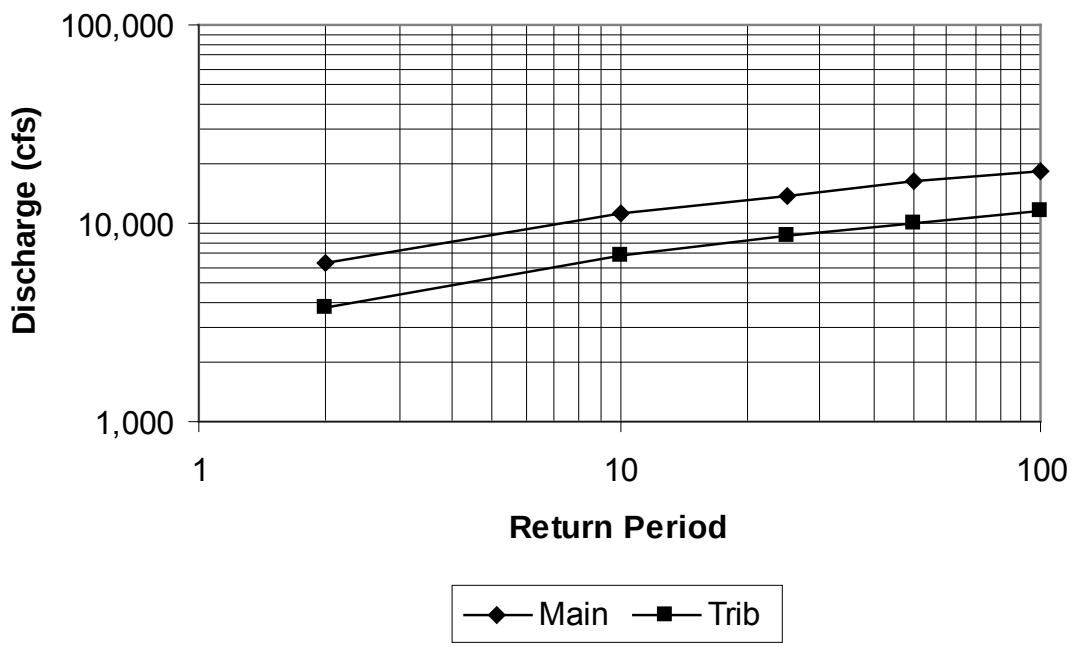

Figure G.1. Watershed Pair Discharge versus Return Period

For example, the discharge corresponding to a 38-yr return period for the Nisqually River is calculated as follows. The resulting discharge is rounded to three significant figures to be consistent with the values in Table G.8.

$$\log Q_x = \frac{\log(38) - \log(25)}{\log(50) - \log(25)}\left(\log(16{,}400) - \log(13{,}900)\right) + \log(13{,}900)$$

$$\log Q_x = \frac{1.580 - 1.398}{1.699 - 1.398}\left(4.215 - 4.143)\right) + 4.143 = 4.186$$

$$Q_x = 10^{\log Q_x} = 10^{4.186} = 15{,}346 = 15{,}300$$

Discharges for the combinations identified previously are graphically or algebraically interpolated and summarized in Table G.9. Each of the combinations in the table represents a joint exceedance probability of 0.02 or the joint 50-yr return period event for the discharges.

Table G.9. Potential Combination Discharges

Combination	Nisqually River Return Period (yrs)	Mineral Creek Return Period (yrs)	Nisqually River Discharge (ft^3/s)	Mineral Creek Discharge (ft^3/s)
1	10	50	11,300	10,100
2	25	48	13,900	10,000
3	38	38	15,300	9,420
4	48	25	16,200	8,550
5	50	10	16,400	6,920

G.4.4 Analyze Combinations

After development of discharge pairs for the potential combinations, a hydraulic model such as HEC-RAS is applied to site conditions using those discharge combinations. The initial data collection, setup, and calibration of HEC-RAS are completed. Then, in the case of this example, the five combinations in Table G.9 are executed changing only the main and tributary discharges. Although HEC-RAS was not executed for this example, the hypothetical results in Table G.10 illustrate the process.

Table G.10. Hypothetical Stage and Velocity at the Example Project Site

Combination	Stage (ft)	Velocity (ft/s)
1	28.7	14.2
2	29.6	11.1
3	30.2	10.2
4	30.4	9.5
5	29.8	11.7

Depending on the design objective (stage or velocity) and site location within the influence reach, one of the combinations will yield the extreme condition for that design objective. In this example, combination 4 results in the highest stage; 30.4 ft is the 50-yr water surface elevation at the site. Similarly, combination 1 yields the highest velocity; 14.2 ft/s is the 50-yr velocity.

Results from different combinations are not combined. If the design objective is most concerned about stage, the maximum stage is used. If the design objective is most concerned about velocity, the maximum velocity is used. For scour analyses, which depend on both stage and velocity, scour computations using each combination are completed, and the maximum scour is used. The extreme condition for the appropriate design objective is considered the

design condition corresponding to that return period. As in this example, a different combination might be the design condition for stage and velocity.

The designer might analyze all combinations identified in the previous step. However, inspection of the combinations might reveal that some combinations will result in computations that do not differ substantially from another and can be dropped from consideration. Similarly, the location of the design site within the influence reach might be such that certain combinations will not require analysis to determine the critical combination and the design event. If there is uncertainty as to which combinations, all combinations should be tested.

G.4.5 Combination Development Using Alternative Method

The designer might choose to use the regression equations rather than the data groupings to estimate Kendall's τ and to subsequently develop the alternative combinations. The same example is used to illustrate the process.

G.4.5.1 Assess Site-Specific Dependence

Best-fit equations were developed using the least-squares method for determining equation parameters. The envelope equations were developed to cover (envelope) 90 percent of the data in the fitting process. Therefore, the envelope equations provide a more conservative estimate of the correlation between discharges of the confluent watersheds. An equation selection matrix is provided in Table G.11.

Table G.11. Kendall's τ Equation Selection Matrix

		Group 1. Smaller watershed less than or equal to 400 mi^2.	Group 2. Smaller watershed greater than 400 mi^2.
Best-fit	POM	Eq. 4.3	Eq. 4.7
Best-fit	POT	Eq. 4.4	Eq. 4.8
Envelope	POM	Eq. 4.11	Eq. 4.15
Envelope	POT	Eq. 4.12	Eq. 4.16

The equations and their limitations are repeated here. Chapter 4 in the main report may be consulted for more information. The equations for Group 1 are presented first followed by the equations for Group 2.

For Kendall's τ, the POM analysis resulted in the best-fit equation of:

$$\tau = 0.69(R_A)^{-0.15} \quad \text{(Equation 4.3)}$$

where,

τ = Kendall's τ, dimensionless

The best-fit equation for estimating Kendall's τ for the POT case is:

$$\tau = 0.66(R_A)^{-0.11} \quad \text{(Equation 4.4)}$$

When applying Equations 4.3 and 4.4, the mainstream drainage area should be adjusted to 2,000 square miles if it exceeds that value before the area ratio is computed. If R_A is estimated to be greater than 400, then 400 should be used in the equations.

The database used for this study did not include data pairs where the smaller watershed drainage area was less than one square mile. Therefore, the results of this research may not apply for watershed pairs with this characteristic. However, a professional engineer must exercise judgment on a site-specific basis to determine if Equations 4.3 and 4.4 are appropriate or if better design alternatives exist.

For Group 2, the best-fit equation for estimating Kendall's τ for the POM dataset is:

$$\tau = 0.035(P_M)^{0.76} \quad \text{(Equation 4.7)}$$

For the POT dataset, the best-fit equation for Kendall's τ is:

$$\tau = 0.12(P_M)^{0.46} \quad \text{(Equation 4.8)}$$

Equations 4.7 and 4.8 are limited to a range of P_M between 15 and 52 inches. If P_M is estimated to be greater than 52 inches, then 52 should be used in the equation. If P_M is estimated to be less than 15 inches, then 15 should be used.

In addition to the best-fit equations, envelope equations were also developed. The envelope equations provide a design option where the designer could be confident that the correlation would not be underestimated.

Group 1, Kendall's τ, POM:

$$\tau = 0.69(R_A)^{-0.15} + 0.23 \quad \text{(Equation 4.11)}$$

Group 1, Kendall's τ, POT:

$$\tau = 0.66(R_A)^{-0.11} + 0.23 \quad \text{(Equation 4.12)}$$

Group 2, Kendall's τ, POM:

$$\tau = 0.035(P_M)^{0.76} + 0.21 \quad \text{(Equation 4.15)}$$

Group 2, Kendall's τ, POT:

$$\tau = 0.12(P_M)^{0.46} + 0.09 \quad \text{(Equation 4.16)}$$

If the design site is more likely to be influenced by the main stream because it is closer to the main stream than the upstream end of the influence reach, then use of the Peaks on Main (POM) equations might be justified. If the converse is true, the Peaks on Tributary (POT) equations might be justified. If the designer is not confident which (POM or POT) is applicable, then both should be used and the maximum value selected for the design.

The computed value of τ should be rounded to the nearest tenth. The value of τ is limited to a maximum value of 0.9, which indicates substantial dependence between the two watersheds; a value of 0.2 is the practical minimum value.

For the example, the best-fit equations are selected. With the drainage area ratio, R_A, equal to 1.9, Equations 4.3 and 4.4 are applied as follows:

$$\tau = 0.69(R_A)^{-0.15} = 0.69(1.9)^{-0.15} = 0.63 \quad \text{(Equation 4.3)}$$

$$\tau = 0.66(R_A)^{-0.11} = 0.66(1.9)^{-0.11} = 0.62 \quad \text{(Equation 4.4)}$$

In this example, the Kendall's τ value is nearly identical for the POM and POT cases. It is rounded to the nearest tenth resulting in an estimate of 0.6.

G.4.5.2 Determine Potential Combinations

Potential combinations for the 10-, 25-, 50-, 100-, and 500-yr joint design frequency events are provided in Section 4.3 and are repeated here as Tables G.12 through G.16. Within each table, the row labeled with the value of Kendall's τ for the project site contains combinations of events the designer may analyze with a hydraulic model to determine which is most critical for the design objective. The column labeled "equal" represents the return period on both the main stream and tributary that is equivalent to the joint return period.

Table G.12. 10-yr Joint Return Periods

	Return Period on One Stream			
Kendall's τ	1.25	2	5	Equal
0.2	9	7	3	4
0.3	10	8	5	5
0.4	-	9	6	6
0.5	-	10	8	6
0.6	-	-	9	7
0.7	-	-	10	8
0.8	-	-	-	9
0.9	-	-	-	9

Table G.13. 25-yr Joint Return Periods

	Return Period on One Stream				
Kendall's τ	1.25	2	5	10	Equal
0.2	23	20	13	7	8
0.3	24	22	17	11	11
0.4	25	24	21	16	13
0.5	-	25	23	20	15
0.6	-	-	24	22	17
0.7	-	-	25	24	19
0.8	-	-	-	25	21
0.9	-	-	-	-	23

Table G.14. 50-yr Joint Return Periods

	Return Period on One Stream					
Kendall's τ	1.25	2	5	10	25	Equal
0.2	47	41	30	21	7	15
0.3	49	46	39	31	15	20
0.4	50	49	45	40	25	25
0.5	-	50	48	45	34	30
0.6	-	-	50	48	42	34
0.7	-	-	-	50	47	39
0.8	-	-	-	-	49	43
0.9	-	-	-	-	50	46

Table G.15. 100-yr Joint Return Periods

	Return Period on One Stream						
Kendall's τ	1.25	2	5	10	25	50	Equal
0.2	95	86	67	53	32	12	28
0.3	98	94	84	74	54	28	39
0.4	100	98	94	88	73	48	49
0.5	-	100	98	95	87	68	59
0.6	-	-	100	99	95	83	68
0.7	-	-	-	100	99	93	77
0.8	-	-	-	-	100	99	85
0.9	-	-	-	-	-	100	93

Table G.16. 500-yr Joint Return Periods

	Return Period on One Stream							
Kendall's τ	1.25	2	5	10	25	50	100	Equal
0.2	485	452	395	354	294	239	167	132
0.3	496	486	462	439	398	353	286	189
0.4	499	497	489	480	459	431	381	243
0.5	-	500	498	496	488	474	445	293
0.6	-	-	-	500	498	494	481	341
0.7	-	-	-	-	500	499	497	385
0.8	-	-	-	-	-	-	500	426
0.9	-	-	-	-	-	-	500	464

For the example application, the table for the 50-yr joint design frequency event is required and Kendall's τ is estimated to be 0.6. By inspection of Table G.14, the following pairs of coincident events are equivalent to the coincident 50-yr event (main stream return period, tributary return period): (5,50), (10,48), (25,42), and (34,34). A 5-yr event on the main stream combined with a 50-yr event on the tributary represents a 50-yr coincident event for this watershed pair, as does the simultaneous occurrence of 34-yr events. All of these combinations represent a coincident 50-yr event.

In addition, the transpose pairings provide three additional events: (50,5), (48,10), and (42,25). Therefore, a 50-yr event on the main stream combined with a 5-yr event on the tributary represents a 50-yr coincident event. In total, seven potential combinations are identified.

Depending on the location of the site in the influence reach and the design objective, some combinations might be discarded at this point. For example, if the site is close to the confluence of the two streams, the main stream is substantially larger than tributary, and stage is the design variable, the (5,50) combination may be discarded as its use is unlikely to produce the highest stage. Combinations may also be discarded if they are not substantially different from other combinations.

G.4.5.3 Estimate Discharges

The designer determines the discharges corresponding to the potential combination return periods using any appropriate method for the watersheds. Because the set of potential combinations will likely include return periods not generally estimated using standard methods, interpolation will be necessary for some return periods. Regression equations, and other

methods, do not always produce “well-behaved” flood frequency curves. For this reason, it is recommended that the designer plot the estimated discharges prior to performing any interpolations.

Flows are estimated using the information and techniques previously applied in Section G.4.3. The resulting combination discharges are summarized in Table G.17.

Table G.17. Potential Combination Discharges for Alternative Method

Combination	Nisqually River Return Period (yrs)	Mineral Creek Return Period (yrs)	Nisqually River Discharge (ft^3/s)	Mineral Creek Discharge (ft^3/s)
1	5	50	8,760	10,100
2	10	48	11,300	10,000
3	25	42	13,900	9,680
4	34	34	15,000	9,210
5	42	25	15,700	8,550
6	48	10	16,200	6,920
7	50	5	16,400	5,350

G.4.5.4 Analyze Combinations

After development of discharge pairs for the potential combinations, a hydraulic model such as HEC-RAS is applied to site conditions using those discharge combinations. The initial data collection, setup, and calibration of HEC-RAS are completed. Then, in the case of this example, the seven combinations in Table G.17 are executed changing only the main and tributary discharges. Although HEC-RAS was not executed for this example, the hypothetical results in Table G.18 illustrate the process.

Table G.18. Hypothetical Stage and Velocity at the Example Project Site

Combination	Stage (ft)	Velocity (ft/s)
1	28.0	14.0
2	28.6	12.8
3	29.5	11.0
4	30.0	10.0
5	30.3	9.4
6	29.7	10.6
7	29.2	11.6

Depending on the design objective (stage or velocity) and location within the influence reach, one of the combinations will yield the extreme condition for that design objective. In this example, combination 5 results in the highest stage such that 30.3 ft is taken as the 50-yr water surface elevation at the site. Similarly, combination 1 yields the highest velocity such that 14.0 ft/s is the 50-yr velocity.

Results from different combinations are not combined. If the design objective is most concerned about stage, the maximum stage is used. If the design objective is most concerned about velocity, the maximum velocity is used. For scour analyses, which depend on both stage and velocity, scour computations using each combination are completed, and the maximum scour is used. The extreme condition for the appropriate design objective is considered the design condition corresponding to that return period. As in this example, a different combination might be the design condition for stage and velocity.

The designer might analyze all combinations identified in the previous step. However, inspection of the combinations might reveal that some combinations will result in computations that do not differ substantially from others and can be dropped from consideration. Similarly, the location of the design site within the influence reach might be such that certain combinations will not require analysis to determine the critical combination and the design event. If there is uncertainty as to which combinations of potential design events should be used, all combinations should be tested.

G.5 Total Probability Method

If the total probability method has been selected, the design process continues with the following steps:

1. Construct marginal exceedance frequency curves for the main and tributary streams. Index values of $Q_{M,i}$ may be used, that is discreet values of Q_M that represent a range of Q and a corresponding probability of occurrence $P(Q_{M,i})$.
2. Select the conditional probability matrix representing the probabilities of y_T given an occurrence for each value of $Q_{Mi,i}$. The result is $P[y_T|Q_{Mi,i}]$.
3. Compute stages, y_T, for various combinations of Q_T and Q_M. A HEC-RAS model of the confluence and design location is a likely tool.
4. Compute the total probability of occurrence $P[y_T]$ for each stage.
5. Construct the total probability curve. Sort the results of Step 4 by stage and plot the exceedance probability curve. Select the stage corresponding to the selected design event.

The total probability method is applied twice. The conditional probability matrices for both the POM and POT cases are applied in Step 2. Steps 3 through 5 are completed separately for both cases and the highest stage computed from Step 5 from the POM and POT analyses is selected for design. The following sections illustrate the application of this method using the watershed pair from Washington State described in Table G.1.

G.5.1 Develop the Marginal Exceedance Frequency Curves

The marginal exceedance frequency curves provide a relation between discharge on a given stream and the probability of occurrence of that discharge. To simplify the process, the complete range of exceedance probabilities are divided into several groups or “bins.” Each bin is represented by an index value that is centrally located within the bin. In this example, the marginal exceedance values are divided into seven bins as summarized in Table G.19. These bins were selected so that the annual exceedance probability (AEP) index values of 0.5 through 0.002 represent common design events (2-yr through 500-yr, respectively). The index value is approximately the geometric mean of the range for each bin as shown in Table G.19. Using bin number 3 for illustration, the index value of 0.1 represents a range of events with AEPs from 0.063 to 0.141. In terms of return period, the index value of the 10-yr event represents all events between the 16-yr and 7.1-yr events.

The probability that a flow will be in the given bin is summarized in the fourth column of Table G.19 and is determined from the AEP range. The difference between the upper and lower bin values represents the probability of a flow occurring in the range defined by the bin. Using bin number 3 for illustration, P(Q) = 0.141-0.063 = 0.078. Therefore, the probability of an event occurring within this bin is 0.078 or 7.8 percent. The sum of the values in the P(Q) column must equal 1 to account for all possible outcomes.

The corresponding flow rates are computed from the regression equations used in Section G.4.3, and summarized in Table G.8, including an interpolation for the 0.2 AEP and an extrapolation for the 0.002 AEP. The flow rates (Q_M and Q_T) in conjunction with the probability of occurrence, P(Q), represent the marginal exceedance frequency curves for the main and tributary watersheds.

Table G.19. Exceedance Frequency Curves Based on Bin AEPs

Bin number	Bin AEP (Index Value)	Bin AEP range	$P(Q_i)$	Nisqually River Q_{Mi} (ft^3/s)	Mineral Creek Q_T (ft^3/s)
1	0.5	0.316 > AEP >= 1.000	0.684	6,260	3,810
2	0.2	0.141 > AEP >= 0.316	0.175	8,760	5,350
3	0.1	0.063 > AEP >= 0.141	0.078	11,300	6,920
4	0.04	0.028 > AEP >= 0.063	0.035	13,900	8,550
5	0.02	0.014 > AEP >= 0.028	0.014	16,400	10,100
6	0.01	0.004 > AEP >= 0.014	0.010	18,400	11,400
7	0.002	0.000 > AEP >= 0.004	0.004	23,900	14,800

G.5.2 Select the Conditional Probability Matrix

As described in Chapter 5, the appropriate conditional probability matrix (CPM) is based on the drainage area ratio and the total area. The selection table from that section is repeated below as Table G.20.

Table G.20. CPM Selection Table – Degree of Correlation

Area ratio	Total area (sq. mi.)			
	1 to 50	50 to 100	100 to 1000	1000 or greater
1 to 10	High	Moderate	Mixed	Mixed
10 to 100	High	Moderate	Mixed	Poor
100 or greater	High	Moderate	Poor	Poor

For the subject watershed pair, the drainage areas of the main and tributary watersheds are 133 mi^2 and 70.3 mi^2, respectively. Consequently, the total area is 203 mi^2 and the area ratio is 1.9. Consulting Table G.20 reveals that the appropriate selection is the “mixed” CPM.

There are two CPMs for each correlation selection. One is for the POM and one for the POT. If the designer is certain that one or the other is applicable, based on the location of the design site in the influence reach, that matrix should be selected. However, if the designer is uncertain which governs, both should be analyzed and the highest stage selected for design. The CPM tables from Chapter 5 are repeated below as Tables G.21a through G.24b.

For this illustration, the POM for the “mixed” flow condition is selected first (Table G.23a). Because Q_M and Q_T are dependent (to some degree), each row of the table represents a member of the family of curves for the conditional probability of Q_T given Q_M. For example, if a flow on the main stream is represented by the 0.10 AEP bin (10-yr event), there is a 0.3 probability of coincident occurrence on the tributary of a 2-yr event (AEP=0.50), a 0.3 probability of a 5-yr event (AEP=0.20), a 0.3 probability of a 10-yr event (AEP=0.10), and a 0.1 probability of a 25-yr event (AEP=0.04).

Table G.21a. CPM for Highly Correlated Flows (POM)

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Sum
Main Stream AEP	0.500	1							1
	0.200		1						1
	0.100			1					1
	0.040				1				1
	0.020					1			1
	0.010						1		1
	0.002							1	1

Table G.21b. CPM for Highly Correlated Flows (POT)

		Tributary Stream AEP						
		0.500	0.200	0.100	0.040	0.020	0.010	0.002
Main Stream AEP	0.500	1						
	0.200		1					
	0.100			1				
	0.040				1			
	0.020					1		
	0.010						1	
	0.002							1
	Sum	1	1	1	1	1	1	1

Table G.22a. CPM for Moderately Correlated Flows (POM)

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Sum
Main Stream AEP	0.500	0.9	0.1						1
	0.200	0.4	0.5	0.1					1
	0.100	0.2	0.3	0.4	0.1				1
	0.040		0.2	0.3	0.4	0.1			1
	0.020			0.2	0.3	0.4	0.1		1
	0.010				0.2	0.3	0.4	0.1	1
	0.002					0.2	0.4	0.4	1

Table G.22b. CPM for Moderately Correlated Flows (POT)

		Tributary Stream AEP						
		0.500	0.200	0.100	0.040	0.020	0.010	0.002
Main Stream AEP	0.500	0.9	0.4	0.2				
	0.200	0.1	0.5	0.3	0.2			
	0.100		0.1	0.4	0.3	0.2		
	0.040			0.1	0.4	0.3	0.2	
	0.020				0.1	0.4	0.3	0.2
	0.010					0.1	0.4	0.4
	0.002						0.1	0.4
	Sum	1	1	1	1	1	1	1

Table G.23a. CPM for Mixed Correlated Flows (POM)

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Sum
Main Stream AEP	0.500	0.9	0.1						1
	0.200	0.5	0.4	0.1					1
	0.100	0.3	0.3	0.3	0.1				1
	0.040	0.3	0.1	0.2	0.3	0.1			1
	0.020	0.3		0.1	0.2	0.3	0.1		1
	0.010	0.3			0.1	0.2	0.3	0.1	1
	0.002	0.3				0.1	0.3	0.3	1

Table G.23b. CPM for Mixed Correlated Flows (POT)

		Tributary Stream AEP						
		0.500	0.200	0.100	0.040	0.020	0.010	0.002
Main Stream AEP	0.500	0.9	0.5	0.3	0.3	0.3	0.3	0.3
	0.200	0.1	0.4	0.3	0.1			
	0.100		0.1	0.3	0.2	0.1		
	0.040			0.1	0.3	0.2	0.1	
	0.020				0.1	0.3	0.2	0.1
	0.010					0.1	0.3	0.3
	0.002						0.1	0.3
	Sum	1	1	1	1	1	1	1

Table G.24a. CPM for Poorly Correlated Flows (POM)

		Tributary Stream AEP							
		0.500	0.200	0.100	0.040	0.020	0.010	0.002	Sum
Main Stream AEP	0.500	0.684	0.175	0.078	0.035	0.014	0.010	0.004	1
	0.200	0.684	0.175	0.078	0.035	0.014	0.010	0.004	1
	0.100	0.684	0.175	0.078	0.035	0.014	0.010	0.004	1
	0.040	0.684	0.175	0.078	0.035	0.014	0.010	0.004	1
	0.020	0.684	0.175	0.078	0.035	0.014	0.010	0.004	1
	0.010	0.684	0.175	0.078	0.035	0.014	0.010	0.004	1
	0.002	0.684	0.175	0.078	0.035	0.014	0.010	0.004	1

Table G.24b. CPM for Poorly Correlated Flows (POT)

		Tributary Stream AEP						
		0.500	0.200	0.100	0.040	0.020	0.010	0.002
Main Stream AEP	0.500	0.684	0.684	0.684	0.684	0.684	0.684	0.684
	0.200	0.175	0.175	0.175	0.175	0.175	0.175	0.175
	0.100	0.078	0.078	0.078	0.078	0.078	0.078	0.078
	0.040	0.035	0.035	0.035	0.035	0.035	0.035	0.035
	0.020	0.014	0.014	0.014	0.014	0.014	0.014	0.014
	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010
	0.002	0.004	0.004	0.004	0.004	0.004	0.004	0.004
	Sum	1	1	1	1	1	1	1

G.5.3 Compute Stages

The tributary stage is computed using a HEC-RAS model or other appropriate tool. For each pair of main and tributary flows of interest from the CPM, a stage at the design location is computed. For the CPM in Table G.23a there are 28 non-zero entries each requiring a HEC-RAS run. Generally, the majority of effort preparing HEC-RAS analyses is the initial data collection, setup, and calibration. Executing multiple scenarios, as is anticipated here, requires only changing the main and tributary discharges within the calibrated model

The stage computations for each pair are summarized in Table G.25. (The stage values are hypothetical to illustrate the computation process as no HEC-RAS runs were actually performed.) For example, the 24 ft stage in the upper left box results from using the 0.5 AEP flow on the main stream and the 0.5 AEP flow on the tributary stream, which from Table G.19 are 6,260 ft^3/s and 3,810 ft^3/s, respectively.

Table G.25. Summary of Stage Computations (ft)

		Tributary Stream AEP						
		0.500	0.200	0.100	0.040	0.020	0.010	0.002
Main Stream AEP	0.500	24.0	24.6					
	0.200	25.2	26.0	26.6				
	0.100	26.2	27.2	28.0	28.6			
	0.040	27.3	28.2	29.2	30.0	30.6		
	0.020	28.3		30.2	31.2	32.0	32.6	
	0.010	29.3			32.2	33.2	34.0	34.6
	0.002	30.3				34.2	35.2	36.0

G.5.4 Compute Total Probability

The total probability is computed using Equation 2.2 for each computed stage. Equation 2.2 is repeated here as follows:

$$P[y_T] = \sum_{i=1}^{n} P[y_T \mid Q_{M,i}]P[Q_{M,i}] \quad \text{(Equation 2.2)}$$

where,

$P[y_T]$ = probability that y_T will occur

$P[y_T|Q_{M,i}]$ = probability that y_T will occur given $Q_{M,i}$

$P[Q_{M,i}]$ = probability that $Q_{M,i}$ will occur

The computations use the marginal probabilities, $P(Q_{M,i})$, from Step 1, the conditional probabilities, $P(y_T|Q_{M,i})$, from Step 2, and the stages from Step 3. The results are summarized in Table G.26. Each row in the table represents an entry in Table G.25 starting at the upper left then moving from left to right and then from top to bottom. For example, the first row in Table G.26 represents the probability of occurrence of the 24.0 ft stage in the upper left corner of Table G.25. Using Equation 2.2:

$$P[y_T = 24.0 \text{ ft}] = \sum_{i=1}^{n} P[y_T = 24.0 \text{ ft} \mid Q_{M,i} = 6260 \text{ ft}^3]P[Q_{M,i} = 6260 \text{ ft}^3] = 0.9\text{x}0.684 = 0.6156$$

The remaining rows in Table G.26 represent the analogous computation for the other stage entries in Table G.25.

Table G.26. Total Probability Computation (POM)

y_T	$P(Q_M)$	$P(y_T\|Q_M)$	$P(y_T)$
24.0	0.684	0.9	0.6156
24.6	0.684	0.1	0.0684
25.2	0.175	0.5	0.0875
26.0	0.175	0.4	0.0700
26.6	0.175	0.1	0.0175
26.2	0.078	0.3	0.0234
27.2	0.078	0.3	0.0234
28.0	0.078	0.3	0.0234
28.6	0.078	0.1	0.0078
27.3	0.035	0.3	0.0105
28.2	0.035	0.1	0.0035
29.2	0.035	0.2	0.0070
30.0	0.035	0.3	0.0105
30.6	0.035	0.1	0.0035
28.3	0.014	0.3	0.0042
30.2	0.014	0.1	0.0014
31.2	0.014	0.2	0.0028
32.0	0.014	0.3	0.0042
32.6	0.014	0.1	0.0014
29.3	0.010	0.3	0.0030
32.2	0.010	0.1	0.0010
33.2	0.010	0.2	0.0020
34.0	0.010	0.3	0.0030
34.6	0.010	0.1	0.0010
30.3	0.004	0.3	0.0012
34.2	0.004	0.1	0.0004
35.2	0.004	0.3	0.0012
36.0	0.004	0.3	0.0012

G.5.5 Construct Total Probability Curve

The computations from Step 4 (Table G.26) are sorted by stage as illustrated in Table G.27. If any values of stage, y_T, are identical, the rows should be combined so that each row represents a unique value of y_T. The final step is to create the probability of exceedance estimates.

For the first row in the table, the probability of occurrence of a stage of 24 ft is represented as 0.6156. Therefore, the probability of exceedance is 1-0.6156 = 0.3844. For the next row, the probability of exceedance is 1-0.6156-0.0684 = 0.3160. The computation of exceedance probability for each stage value is summarized in the table. The result is then plotted as shown in Figure G.2.

The stage for a design exceedance probability can be taken either from Figure G.2 or Table G.27. For the design example, the 50-year stage (AEP = 0.02) is approximately 30.4 feet.

Table G.27. Total Probability Curve (POM)

y_T	$P(y_T)$	P_e
24.0	0.6156	0.3844
24.6	0.0684	0.3160
25.2	0.0875	0.2285
26.0	0.0700	0.1585
26.2	0.0234	0.1351
26.6	0.0175	0.1176
27.2	0.0234	0.0942
27.3	0.0105	0.0837
28.0	0.0234	0.0603
28.2	0.0035	0.0568
28.3	0.0042	0.0526
28.6	0.0078	0.0448
29.2	0.0070	0.0378
29.3	0.0030	0.0348
30.0	0.0105	0.0243
30.2	0.0014	0.0229
30.3	0.0012	0.0217
30.6	0.0035	0.0182
31.2	0.0028	0.0154
32.0	0.0042	0.0112
32.2	0.0010	0.0102
32.6	0.0014	0.0088
33.2	0.0020	0.0068
34.0	0.0030	0.0038
34.2	0.0004	0.0034
34.6	0.0010	0.0024
35.2	0.0012	0.0012
36.0	0.0012	0.0000

The total probability curve shown in Table G.27 and Figure G.2 is based on the POM conditional probability matrix. The process is repeated beginning at Step 2 for the POT scenario. The POT conditional probability matrix for the "mixed" scenario is selected. Stages are then computed in Step 3, however, many will already be computed for the POM case. Many of the same combinations of flows remain relevant; it is their probability of occurrence that changes. In Step 4, the total probability is computed. Equation 2.2 is used again, but is rewritten with Q_T as the dominant variable:

$$P[y_T] = \sum_{i=1}^{n} P[y_T \mid Q_{T,i}] P[Q_{T,i}]$$

where,

$P[y_T]$ = probability that y_T will occur

$P[y_T|Q_{T,i}]$ = probability that y_T will occur given $Q_{T,i}$

$P[Q_{T,i}]$ = probability that $Q_{T,i}$ will occur

The total probability curve is then constructed in Step 5 as it was for the POM scenario. The highest stage from the two computations (POM or POT) is used for design.

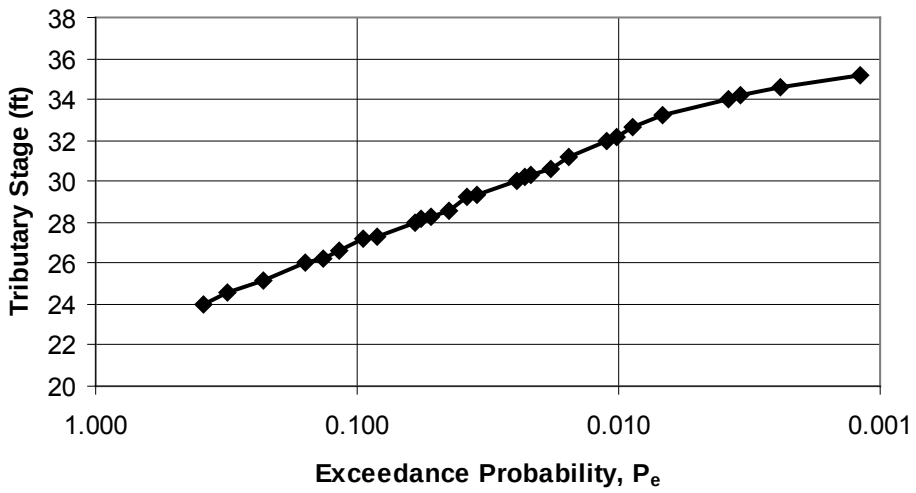

Figure G.2. Total Probability Curve (POM) for Example Watershed

Appendix H. Step-by-Step Application Guide

Table of Contents

List of Tables ..... **H-ii**
List of Figures ..... **H-ii**
H.1 Introduction **H-1**
H.2 Design Steps **H-3**
H.2.1 Step 1: Specify the Design Condition H-3
H.2.2 Step 2. Compute Discharge Range of Interest H-3
H.2.3 Step 3: Determine Influence Reach H-4
H.2.4 Step 4: Determine Hydrologic Combinations H-5
H.2.5 Step 5: Estimate Discharges for Combinations H-8
H.2.6 Step 6: Analyze Combinations H-9
H.3 Design Examples **H-10**
H.3.1 Nisqually River and Mineral Creek H-10
H.3.2 Raritan River and Middle Brook H-14

Appendix H provides a step-by-step applications guide in a user-friendly format. It is intended to be applied to most design situations and objectives.

Appendix G provides more detailed application guidelines for situations of high sensitivity or vulnerability either for the infrastructure itself or for the traveling public.

List of Tables

Table H.1. Influence Reach Determination Options H-4
Table H.2. Watershed Categories H-5
Table H.3. Return Period Combinations for the 10-yr Joint Return Period H-6
Table H.4. Return Period Combinations for the 25-yr Joint Return Period H-6
Table H.5. Return Period Combinations for the 50-yr Joint Return Period H-7
Table H.6 Return Period Combinations for the 100-yr Joint Return Period H-7
Table H.7. Return Period Combinations for the 500-yr Joint Return Period H-8
Table H.8. Nisqually/Mineral Regression Constants H-11
Table H.9. Nisqually/Mineral Discharge Range H-11
Table H.10. Nisqually/Mineral Influence Reach Determination H-12
Table H.11. Nisqually/Mineral Discharges H-12
Table H.12. Nisqually/Mineral Discharge Combinations H-13
Table H.13. Stage and Velocity For Nisqually/Mineral Combinations H-13
Table H.14. Raritan/Middle Brook Regression Constants H-15
Table H.15. Raritan/Middle Brook Watershed Data H-15
Table H.16. Raritan/Middle Brook Discharge Range H-15
Table H.17. Raritan/Middle Brook Influence Reach Determination H-16
Table H.18. Raritan/Middle Brook Discharges H-16
Table H.19. Raritan/Middle Brook Discharge Combinations H-17
Table H.20. Stage and Velocity For Raritan/Middle Brook Combinations H-17

List of Figures

Figure H.1. Influence Reach Schematic H-1
Figure H.2. Hypothetical Flow-Frequency Curves H-8

H.1 Introduction

Highway drainage structures are often located near the confluence of two streams where they may be subject to inundation by high flows from either stream. These structures should be designed to meet specified performance objectives for floods of a specified return period, e.g., the 100-year flood, as specified by the applicable design criteria. Because the flooding of structures on one stream can be affected by high flows on the other stream, it is necessary to know the relationship between the joint exceedance probability of the confluent stream pair, i.e., the ***joint probability*** of the coincident flows, with the individual exceedance probability on each stream.

The joint probability question arises when a structure is located where the hydraulic behavior of some combination of the main and tributary streams may result in critical hydraulic design conditions. Specifically, there is a portion of the tributary stream that is influenced by both the discharge of the tributary stream and the backwater caused by the main stream. This is referred to as the ***influence reach***. The location of the structure within the influence reach, as well as the joint hydrologic behavior of the confluent streams, will determine the importance of the confluent streams on the appropriate design conditions for the structure.

The concept of an influence reach is shown schematically in Figure H.1 along with three possible alternative structure locations relative to the influence reach. For any given confluent pair of streams, there is a distance upstream of the confluence within which the flow from both streams will influence the hydraulic conditions at the structure. This distance is referred to as X_{max} in the figure and will vary with the magnitude of the discharge. If the distance of the structure location above the confluence, x, is less than X_{max}, then the structure is said to be located in the influence reach and joint probability analysis may be necessary to design a cost-effective and reliable structure. Although the figure shows the structure location and influence reach on the tributary stream, the analysis is the same if the structure and influence reach is on the main stream.

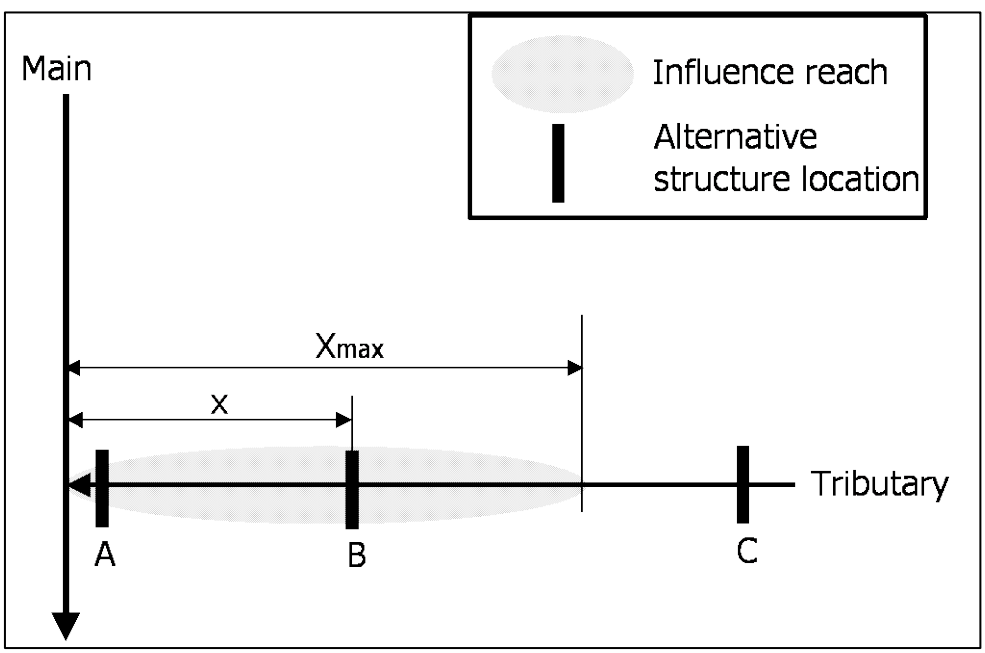

Figure H.1. Influence Reach Schematic

For a structure located very close to the confluence of the main and tributary streams (location A in Figure H.1), the structure is sufficiently close to the confluence such that the critical design condition is determined by the backwater at the confluence. This case essentially reduces to an analysis of the flow at the confluence and joint probability analysis may not be necessary. If the designer is unsure if the site is sufficiently close, the joint probability design procedure should be applied.

For a structure located a significant distance from (or at a much higher elevation than) the confluence (location C in Figure H.1), the structure is beyond the influence reach and joint probability analysis is not needed. This case represents the typical riverine hydraulic design condition.

Alternative location B in Figure H.1 represents the joint probability problem where the distance of the structure from the confluence, x, is less than X_{max} and the structure is within the influence reach. The hydraulic conditions, e.g. depth and velocity for a structure at location B is a function of the flow at the confluence, which establishes the downstream control elevation, and the flow in the tributary, which determines the water surface profile from the downstream control to the design location. More simply put, the hydraulic conditions at location B can be considered a function of the main and tributary stream discharges.

In a typical riverine situation (one stream) we have a three-step design procedure:

1. Select an appropriate design return period.
2. Estimate the flow rate associated with that return period.
3. Apply that flow rate to a hydraulic analysis of the structure.

For confluent streams, we will still use the appropriate design return period for the site, but refer to it as the joint probability. Then, we require additional analyses to determine the appropriate return periods to use on the main stream and the tributary stream such that the joint probability is as we require it to be. This design procedure provides the necessary tools.

The purpose of this appendix is to provide a step-by-step applications guide in a user friendly format. The approach used in this guide is intended to be applied to most design situations and objectives. However, for situations of high sensitivity or vulnerability either for the infrastructure itself or for the traveling public, the designer is encouraged to review the main report to understand the foundation of this method as well as to review the more detailed application guidelines in Appendix G.

Application of this design procedure is limited to watershed pairs satisfying three conditions: 1) neither watershed in the pair is substantially affected by regulation, 2) the sum of the watershed areas is less than 9,000 mi^2, and 3) the drainage area of the smaller watershed of the pair exceeds 1 mi^2. (Refer to Appendix F.2 for a detailed discussion of regulation. The key point to remember is that the application of this confluent stream procedure requires use of a hydrologic method to generate peak discharges on the main and tributary streams. If either of the watersheds is regulated such that a designer would not use the hydrologic procedure on the watershed in a typical single stream analysis because of regulation, then the designer should not use that hydrologic procedure in the context of confluent stream analyses either.]

The next section provides an overview of the design steps. The final section provides two example applications.

H.2 Design Steps

Application of joint probability analysis to a site potentially affected by confluent streams includes six basic steps:

1. Specify the design condition.
2. Compute the discharge range of interest.
3. Determine if the site is within the influence reach.
4. Determine potential hydrologic combinations.
5. Estimate discharges for each potential hydrologic combination.
6. Perform hydraulic analyses for each hydrologic combination.

H.2.1 Step 1: Specify the Design Condition

The designer must choose the appropriate annual design condition appropriate for the site and type of structure, e.g. bridge, culvert, etc. This step must be made for any waterway crossing or other channel work and is no different for the joint probability problem. The annual exccedance probability (AEP) and relevant design parameters are established first. The design AEP might range from 0.1 to 0.002, corresponding to return periods from 10-yr to 500-yr. The designer may also wish to examine a range of design conditions as part of a cost benefit analysis.

The designer must also choose whether a representative (best-fit) or envelope analysis is appropriate for the site and design objectives. For this methodology, the representative approach results in the best estimate, while the envelope approach provides a more conservative estimate based on applying a design envelope around the underlying data. In most design situations, the representative approach is appropriate as it will result in the best estimate of conditions at the site. If, however, there is reason to be more conservative, the envelope approach is recommended.

This distinction can be compared to application of USGS regression equations for estimating discharges. In most situations we use the number from the equation as the best representative estimate. However, the equation also comes with information on the standard error of the equation that provides a basis for increasing the discharge if there is reason to be more conservative.

H.2.2 Step 2: Compute Discharge Range of Interest

Four discharges – two on the main stream and two on the tributary stream – establish the range of discharges that are of interest. For both the main and tributary streams the 2-yr and design condition discharges are calculated. The 2-yr discharge is a widely available low flood flow value used to define the lower end of the range. The design condition is the return period selected in Step 1 and defines the higher end of the range. For example, if a 100-yr design condition has been chosen, we calculate the 2-yr and 100-yr events for the two confluent streams.

Joint probability analysis does not require that a designer use special hydrological techniques to estimate discharge from a watershed. Depending on the watershed and applicable guidance, any appropriate peak flow or hydrograph technique may be used. For example, if USGS regression peak flow techniques are acceptable in a given situation for a typical riverine application, that technique would also be acceptable in a joint probability context.

H.2.3 Step 3: Determine Influence Reach

Joint probability design techniques are necessary only if the project site is within the influence reach. To determine this, the designer creates a hydraulic model for the site using HEC-RAS or other appropriate tool. The hydraulic model must extend downstream of the confluence to properly capture the interaction between the confluent streams. For the stream with the hydraulic structure, it must extend upstream of the hydraulic structure to the same extent as we would extend it in a typical riverine situation. For the other confluent stream, the model needs only to be extended upstream sufficiently so that the boundary condition does not influence the water surface elevation at the confluence.

Once the hydraulic model is setup, two hydrologic scenarios are run to determine if our structure is within the influence reach. Both scenarios, summarized in Table H.1, will use the 2-yr return period discharge for the stream with the hydraulic structure. Scenario I will include the 2-yr return period discharge on the other stream and Scenario II will include the design return period discharge on the other stream.

Table H.1. Influence Reach Determination Options

Scenario	Stream Where Hydraulic Structure is Located	Confluent Stream
I	2-yr	2-yr
II	2-yr	design return period

The two water surface elevations at our hydraulic structure site are compared. If the water surface elevations at the project site are the same between the two scenarios the project site is outside the influence reach (not affected by backwater from the confluent stream) and joint probability techniques are not necessary. If not, the site is within the influence reach and the joint probability design procedure should be applied.

Determining whether hydraulic conditions are different between Scenario I and Scenario II requires a definition of how close the two conditions must be to be considered the same. Defining this must consider many factors including the purpose of the analysis, the potential for error in the discharge estimates, and the sensitivity of the hydraulic conditions to the discharge estimates. Equation H.1 provides a guideline for making this determination. If the equation is true, the site is considered within the influence reach.

$$WSE_{II} - WSE_{I} \geq \max\left(H, WSE_{I+10\%} - WSE_{I}\right) \qquad \text{(H.1)}$$

where,

WSE_{I} = water surface elevation at site for Scenario I, ft

WSE_{II} = water surface elevation at site for Scenario II, ft

$WSE_{I+10\%}$ = water surface elevation at site for Scenario I with discharge rates on both streams increased by 10 percent, ft

H = reference elevation change set at 0.2 ft

$WSE_{I+10\%}$ represents the water surface elevation at the site under Scenario I when the discharges on both confluent streams are increased by 10 percent. This provides an indicator of the sensitivity of the water surface elevation to the discharge estimates. The reference elevation change, H, is a constant that provides a minimum guideline value for those situations where the

system is less sensitive to discharge. When in doubt about the use of the guideline in Equation H.1, the designer should assume that the site is within the influence reach and proceed with the joint probability analysis.

When setting up the HEC-RAS, or other, hydraulic model, the respective discharges for the main and tributary are applied to those reaches above the confluence. The flow below the confluence is the sum of the flows on the confluent streams.

H.2.4 Step 4: Determine Hydrologic Combinations

Appropriate combinations of flow to be considered for a pair of confluent streams depend on the degree to which their flood flow patterns are correlated. Based on the research data base and analyses on which this method is based, this may be determined by the ratio of the watershed areas and the combined watershed area. The drainage area ratio, R_A, is computed by dividing the drainage area of the larger watershed by the drainage area of the smaller watershed. The total watershed area, A_{TOT}, is computed by summing the two drainage areas. Table H.2 summarizes the categories.

Table H.2. Watershed Categories

		Total Watershed Area	
		$A_{TOT} < 350$ mi^2	$A_{TOT} \geq 350$ mi^2
Drainage Area Ratio	$R_A < 7$	SS	SL
	$R_A \geq 7$	LS	LL

The category label is a two-letter code. The first letter references the drainage area ratio. If it is less than 7, the ratio is considered to be small and is indicated by an “S.” If it is greater than or equal to 7 then the ratio is considered to be large and is indicated by an “L.” Similarly, the second letter in the code references the total watershed area. If the total watershed area is less than 350 mi^2, the letter “S” is assigned for small. If not, the letter “L” is assigned for large.

Combinations of individual return period flows on the tributary and main streams for the 10-, 25-, 50-, 100-, and 500-yr joint design frequency events are summarized in Table H.3 through Table H.7. The designer selects the table corresponding to the joint design return period and finds the watershed category within that table. Then, depending on whether the representative approach or envelope approach has been adopted, a series of return period combinations are listed. For the 10- and 25-year joint return periods there are 3 combinations while for the 50-, 100-, and 500-year joint return periods 5 combinations are listed.

Table H.3. Return Period Combinations for the 10-yr Joint Return Period

Category	Approach	Location	Combination		
			1	2	3
SS	Representative	Tributary	3	8	10
		Main	10	8	3
	Envelope	Tributary	5	9	10
		Main	10	9	5
SL	Representative	Tributary	2	7	10
		Main	10	7	2
	Envelope	Tributary	4	8	10
		Main	10	8	4
LS	Representative	Tributary	3	8	10
		Main	10	8	3
	Envelope	Tributary	5	9	10
		Main	10	9	5
LL	Representative	Tributary	1.25	6	10
		Main	10	6	1.25
	Envelope	Tributary	2	7	10
		Main	10	7	10

Table H.4. Return Period Combinations for the 25-yr Joint Return Period

Category	Approach	Location	Combination		
			1	2	3
SS	Representative	Tributary	6	19	25
		Main	25	19	6
	Envelope	Tributary	13	22	25
		Main	25	22	13
SL	Representative	Tributary	3	17	25
		Main	25	17	3
	Envelope	Tributary	10	21	25
		Main	25	21	10
LS	Representative	Tributary	6	19	25
		Main	25	19	6
	Envelope	Tributary	13	22	25
		Main	25	22	13
LL	Representative	Tributary	1.25	14	25
		Main	25	14	1.25
	Envelope	Tributary	3	17	25
		Main	25	17	3

Table H.5. Return Period Combinations for the 50-yr Joint Return Period

Category	Approach	Location	Combination				
			1	2	3	4	5
SS	Representative	Tributary	11	31	38	43	50
		Main	50	43	38	31	11
	Envelope	Tributary	25	39	43	46	50
		Main	50	46	43	39	25
SL	Representative	Tributary	5	26	34	40	50
		Main	50	40	34	26	5
	Envelope	Tributary	20	36	41	45	50
		Main	50	45	41	36	20
LS	Representative	Tributary	11	31	38	43	50
		Main	50	43	38	31	11
	Envelope	Tributary	25	39	43	46	50
		Main	50	46	43	39	25
LL	Representative	Tributary	2	19	27	35	50
		Main	50	35	27	19	2
	Envelope	Tributary	4	25	33	40	50
		Main	50	40	33	25	4

Table H.6 Return Period Combinations for the 100-yr Joint Return Period

Category	Approach	Location	Combination				
			1	2	3	4	5
SS	Representative	Tributary	20	62	75	86	100
		Main	100	86	75	62	20
	Envelope	Tributary	50	77	86	92	100
		Main	100	92	86	77	50
SL	Representative	Tributary	9	52	67	81	100
		Main	100	81	67	52	9
	Envelope	Tributary	39	72	83	91	100
		Main	100	91	83	72	39
LS	Representative	Tributary	20	62	75	86	100
		Main	100	86	75	62	20
	Envelope	Tributary	50	77	86	92	100
		Main	100	92	86	77	50
LL	Representative	Tributary	2	37	54	70	100
		Main	100	70	54	37	2
	Envelope	Tributary	7	50	66	79	100
		Main	100	79	66	50	7

Table H.7. Return Period Combinations for the 500-yr Joint Return Period

Category	Approach	Location	Combination				
			1	2	3	4	5
SS	Representative	Tributary	98	308	376	429	500
		Main	500	429	376	308	98
	Envelope	Tributary	249	385	430	462	500
		Main	500	462	430	385	249
SL	Representative	Tributary	39	257	336	402	500
		Main	500	402	336	257	39
	Envelope	Tributary	200	361	414	453	500
		Main	500	453	414	361	200
LS	Representative	Tributary	98	308	376	429	500
		Main	500	429	376	308	98
	Envelope	Tributary	249	385	430	462	500
		Main	500	462	430	385	249
LL	Representative	Tributary	5	182	269	349	500
		Main	500	349	269	182	5
	Envelope	Tributary	30	246	327	395	500
		Main	500	395	327	246	5

H.2.5 Step 5: Estimate Discharges for Combinations

The next step is to estimate the discharges corresponding to the combination return periods using the hydrologic method applied in Step 2. Because the set of combinations will include individual return periods not directly estimated using standard methods, interpolation will be necessary in those cases. This is illustrated by example flow-frequency curves for hypothetical main and tributary streams shown in Figure H.2. The plotted points represent the return periods generally available from various hydrologic techniques: 2-, 10-, 25-, 50-, and 100-yr. The curves represent the range of values in between the plotted points.

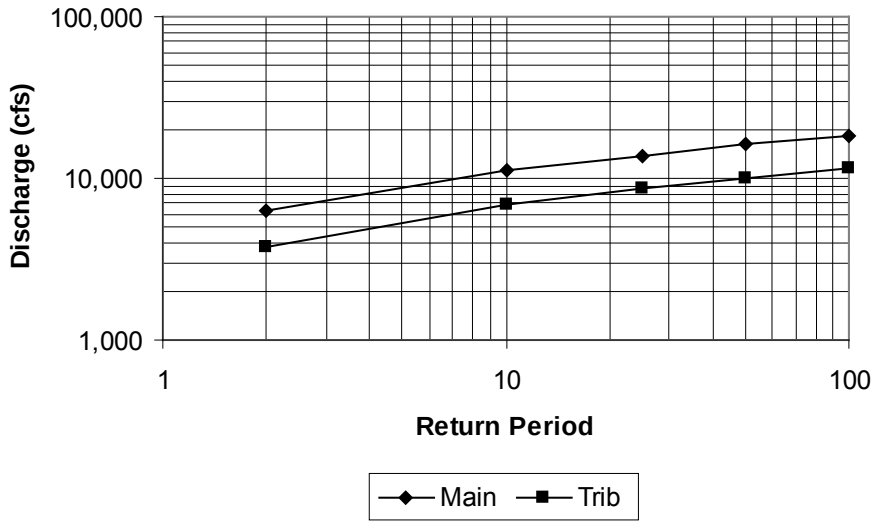

Figure H.2. Hypothetical Flow-Frequency Curves

When other return periods are required, for example, a 77-year return period discharge, it is obtained by interpolating between the 50-year and 100-year values. This interpolation is accomplished using Equations H.2 and H.3.

$$\log Q_x = \frac{\log T_x - \log T_L}{\log T_H - \log T_L}\left(\log Q_H - \log Q_L\right) + \log Q_L \tag{H.2}$$

$$Q_x = 10^{\log Q_x} \tag{H.3}$$

where,

Q_x	=	Discharge at return period x, ft^3/s
Q_H	=	Known discharge at return period higher than x, ft^3/s
Q_L	=	Known discharge at return period lower than x, ft^3/s
T_x	=	Return period of interest, years
T_H	=	Return period of Q_H, years
T_L	=	Return period of Q_L, years

For the 10- and 25-year return periods in the LL watershed category using the representative approach an estimate of the 1.25-yr return period is needed. The 1.25-yr return period discharge can be calculated by extrapolating from the 2-yr and 10-yr discharges as follows:

$$Q_{1.25} = 10^{1.292 \log Q_2 - 0.292 \log Q_{10}} \tag{H.4}$$

where,

$Q_{1.25}$	=	1.25-yr discharge, ft^3/s
Q_2	=	2-yr discharge, ft^3/s
Q_{10}	=	10-yr discharge, ft^3/s

H.2.6 Step 6: Analyze Combinations

After development of discharge pairs for the potential combinations, the same hydraulic model applied in Step 3 for each discharge combination is used to analyze the discharge combinations. Depending on the design objective (stage or velocity) and site location within the influence reach, one of the combinations will yield the extreme condition for a given design objective.

Results from different combinations are not combined. If the design objective is most concerned about stage, the maximum stage is used. If the design objective is most concerned about velocity, the maximum velocity is used. For scour analyses, which depend on stage and velocity, scour computations using each combination are completed, and the maximum scour is used. The extreme condition for the appropriate design objective is considered the design condition corresponding to that joint return period. A different combination might be the design condition for stage, velocity, and scour.

H.3 Design Examples

Two design examples are provided to illustrate the joint probability analysis procedure. They are based on actual watersheds in the states of Washington and New Jersey. In addition to the hydrologic analyses, application of HEC-RAS, or another hydraulic model, is required to complete a joint probability analysis. However, the implementation of riverine hydraulic models in the context of joint probability analyses is not unique. Therefore, these examples do not feature actual HEC-RAS analyses. Rather, they provide realistic, but synthetic, data representative of hydraulic model outputs to illustrate the procedures.

H.3.1 Nisqually River and Mineral Creek

Mineral Creek flows into the Nisqually River in the state of Washington. The drainage areas for the two watersheds at the confluence are 70.3 and 133 square miles, respectively. A bridge is being proposed over Mineral Creek upstream of the confluence.

Step 1. Specify the Design Condition.

The appropriate joint probability return period is the same as the return period for the type of structure being evaluated in a typical riverine situation. For this example, the 50-year return period (with certain freeboard requirements) is required. Therefore, we select the 50-year return period as our specified design condition for the joint probability.

We also need to decide whether we will choose the representative approach or the more conservative envelope approach. Since there are no features of the proposed project that suggest a more conservative approach is justified, the representative approach is selected.

Step 2. Compute the Discharge Range of Interest.

The discharge range of interest is from the 2-year to the 50-year (design condition) events. These values are estimated using an acceptable hydrologic methodology for the site. One acceptable method for this location is the United States Geological Survey (USGS) regression equations for the area. (Sumioka, S.S., D.L. Kresch, and K.D. Kasnick, "Magnitude and Frequency of Floods in Washington, USGS Water Resources Investigations Report 97-4277, 1998). The watershed pair is located in USGS region 2 for which the following regression equation applies:

$$Q_T = aA^{b_1}P_M^{b_2} \tag{H.5}$$

where,

Q_T = design flow for return period, T, ft^3/s

A = drainage area, mi^2

P_M = mean annual precipitation, in

a, b_1, b_2 = regression constants

The regression constants are provided in Table H.8.

Table H.8. Nisqually/Mineral Regression Constants

Return Period	Regression Constants		
	a	b_1	b_2
2	0.090	0.877	1.51
10	0.129	0.868	1.57
25	0.148	0.864	1.59
50	0.161	0.862	1.61
100	0.174	0.861	1.62

The drainage areas of the two watersheds were previously noted as 70.3 and 133 square miles for Mineral Creek and the Nisqually River, respectively. The mean annual precipitation for these watersheds is 98 and 94 inches, respectively. Applying USGS equations to the site-specific data yields the discharge range summarized in Table H.9.

Table H.9. Nisqually/Mineral Discharge Range

Return Period	Discharge (ft^3/s)	
	Nisqually River	Mineral Creek
2	6,260	3,810
50	16,400	10,100

Step 3. Determine if Site is Within Influence Reach.

A hydraulic model appropriate for the site is setup. HEC-RAS is a common choice when one-dimensional analysis is acceptable. Since the proposed bridge is to be located on Mineral Creek, the following two scenarios are analyzed in the hydraulic model:

- Scenario I includes the 2-year discharge on the Nisqually River and the 2-year discharge on Mineral Creek.
- Scenario II includes the 50-year discharge on the Nisqually River and the 2-year discharge on Mineral Creek.

The water surface elevations at the proposed bridge site are extracted from the two scenario runs and the difference is computed ($WSE_{II} - WSE_{I}$). These data are summarized in Table H.10. The difference in water surface elevation between scenarios I and II is 0.5 feet. Referring to equation H.1, this difference is greater than the reference elevation change, H, which is 0.2 ft, therefore we know that equation H.1 is true and our site is within the reference reach.

Based on this conclusion, it is not necessary to run scenario I again with the discharges increased by 10 percent. However, for illustrative purposes scenario I+10% is run with the results summarized in Table H.10. Equation H.1 remains true for this site; our site is within the influence reach for the discharge range of interest. (There is no harm in performing a joint

probability analysis for a site beyond the influence reach in terms of the results as one of the discharge combinations is equivalent to the typical riverine situation.)

Table H.10. Nisqually/Mineral Influence Reach Determination

Scenario	Mineral Creek Return Period	Nisqually River Return Period	Mineral Creek Flow (ft^3/s)	Nisqually River Flow (ft^3/s)	Water Surface Elevation (ft)
I	2-year	2-year	3,810	6,260	28.0
II	2-year	50-year	3,810	16,400	28.5
I+10%	2-year + 10%	2-year + 10%	4,190	6,890	28.1

Step 4. Determine Hydrologic Combinations.

For this example, the drainage area ratio, R_A, is equal to 133/70.3 = 1.9 and A_{TOT} is equal to 133 + 70.3 = 203 mi^2. From Table H.2 we determine that is site falls within category SS.

For our joint design return period of 50-years, the relevant potential hydrologic combinations are found in Table H.5. With a watershed category of SS and using the representative approach we see that there are five potential combinations. Combination 1 is to use the 11-year return period on Mineral Creek and the 50-year return period on the Nisqually River. Combinations 2 through 5 progressively increase the return period on Mineral Creek while decreasing the return period on the Nisqually River. All combinations represent the 50-yr joint return period.

Step 5. Estimate Discharges for Hydrologic Combinations.

For our example, we need to estimate the 11-, 31-, 38-, 43-, and 50-year discharges on both Mineral Creek and the Nisqually River. We previously estimated the 50-year discharges using the applicable USGS equations in Step 2. To estimate the others we use the same set of USGS equations to estimate the 10- and 25-year discharges and then interpolate to derive the 11-, 31-, 38-, and 43-yr discharges. Table H.11 provides the results of applying equation H.5 for the return periods available in the USGS equations.

Table H.11. Nisqually/Mineral Discharges

Return Period	Discharge (ft^3/s)	
	Nisqually River	Mineral Creek
10	11,300	6,920
25	13,900	8,550
50	16,400	10,100

Discharges for the intermediate values are computed using Equations H.2 and H.3. For example, the discharge corresponding to a 38-yr return period for the Nisqually River required for combination 3 is calculated as follows:

$$\log Q_x = \frac{\log(38) - \log(25)}{\log(50) - \log(25)}\left(\log(16{,}400) - \log(13{,}900)\right) + \log(13{,}900)$$

$$\log Q_x = \frac{1.580 - 1.398}{1.699 - 1.398}\left(4.215 - 4.143)\right) + 4.143 = 4.186$$

$$Q_x = 10^{\log Q_x} = 10^{4.186} = 15{,}346 = 15{,}300$$

Using the same procedure, the remaining required discharges are computed and summarized in Table H.12.

Table H.12. Nisqually/Mineral Discharge Combinations

Combination	Mineral Creek Return Period (yrs.)	Nisqually River Return Period (yrs.)	Mineral Creek Discharge (ft^3/s)	Nisqually River Discharge (ft^3/s)
1	11	50	7,070	16,400
2	31	43	9,000	15,800
3	38	38	9,460	15,300
4	43	31	9,740	14,600
5	50	11	10,100	11,500

Step 6. Perform Hydraulic Analyses for Each Combination.

After development of discharge pairs for the hydrologic combinations, each pair is applied to the hydraulic model developed for Step 3. Table H.13 summarizes the stage and velocity at the structure site for each combination of discharges.

Table H.13. Stage and Velocity For Nisqually/Mineral Combinations

Combination	Stage (ft)	Velocity (ft/s)
1	28.7	14.2
2	29.6	11.1
3	30.2	10.2
4	30.4	9.5
5	29.8	11.7

The most extreme results for the variables of interest are taken as the 50-year joint probability design conditions. For freeboard, stage is the driving variable; maximum stage is derived from combination 4. For scour, the designer takes the results from all five combinations and performs a scour computation for each. The maximum scour is used for design.

It is important to remember that the most extreme values for depth, velocity, or scour may not come from the same combination. Results from different combinations are not combined. In this example, the designer should not take the maximum depth from combination 4 and the maximum velocity from combination 1 to compute scour. These do not occur under the same conditions and should not be evaluated as if they did. More information on the development and application of this joint probability method is found in the main report and appendix G.

H.3.2 Raritan River and Middle Brook

The West Branch of Middle Brook flows into the Raritan River in the state of New Jersey. The drainage areas for the two watersheds at the confluence are 2 and 784 square miles, respectively. Approximately 11 square miles of the Raritan River watershed is noncontributing making the effective drainage area 773 square miles. A culvert installation is being proposed on the West Branch of Middle Brook upstream of the confluence.

Step 1. Specify the Design Condition.

The appropriate joint probability return period is the same as the return period for the type of structure being evaluated in a typical riverine situation. For this example, the 25-year return period is required. Therefore, we select the 25-year return period as our specified design condition for the joint probability.

We also need to decide whether we will choose the representative approach or the more conservative envelope approach. Since there are no features of the proposed project that suggest a more conservative approach is justified, the representative approach is selected.

Step 2. Compute the Discharge Range of Interest.

The discharge range of interest is from the 2-year to the 25-year (design condition) events. These values are estimated using an acceptable hydrologic methodology for the site. One acceptable method for this location is the United States Geological Survey (USGS) regression equations for the area. (Watson, Kara M. and Robert D. Schopp, "Methodology for Estimation of Flood Magnitude and Frequency for New Jersey Streams," Scientific Investigations Report 2009-5167, 2009). The watershed pair is located in the unglaciated piedmont region of New Jersey for which the following regression equation applies:

$$Q_T = aA^{b_1}(ST+1)^{b_2}S^{b_3}(D+1)^{b_4} \qquad \text{(H.6)}$$

where,

Q_T = design flow for return period, T, ft^3/s

A = drainage area, mi^2

ST = percentage of basin covered by streams, lakes, and wetlands

S = basin slope, ft/mile

D = population density, $persons/mi^2$

a, b_1, b_2, b_3, b_4 = regression constants

The regression constants are provided in Table H.14.

The drainage areas of the two watersheds were previously noted as 2 and 773 square miles for Middle Brook and the Raritan River, respectively. Additional data for the watersheds is summarized in Table H.15.

Applying USGS equations to the site-specific data yields the discharge range summarized in Table H.16.

Table H.14. Raritan/Middle Brook Regression Constants

	A	b_1	b_2	b_3	b_4
2	37.8	0.753	-0.054	0.251	0.127
5	75.3	0.741	-0.084	0.254	0.104
10	108	0.736	-0.104	0.258	0.092
25	159	0.732	-0.127	0.263	0.079
50	204	0.729	-0.144	0.267	0.070
100	256	0.728	-0.158	0.271	0.062

Table H.15. Raritan/Middle Brook Watershed Data

Data Type	Raritan River	Middle Brook
Area (mi^2)	773	2
Storage (percent)	14.6	22.3
Slope (ft/mi)	11.0	70.2
Pop. Density (persons/ mi^2)	523	765

Table H.16. Raritan/Middle Brook Discharge Range

	Discharge (ft^3/s)	
Return Period	Raritan River	Middle Brook
2	19,700	360
25	45,000	920

Step 3. Determine if Site is Within Influence Reach.

A hydraulic model appropriate for the site is setup. HEC-RAS is a common choice when one dimensional analysis is acceptable. Since the proposed bridge is to be located on the West Branch of Middle Brook, the following two scenarios are analyzed in the hydraulic model:

- Scenario I includes the 2-year discharge on the Raritan River and the 2-year discharge on Middle Brook.
- Scenario II includes the 25-year discharge on the Raritan River and the 2-year discharge on Middle Brook.

The water surface elevations at the proposed bridge site are extracted from the two scenario runs and the difference is computed ($WSE_{II} – WSE_{I}$). These data are summarized in Table

H.17. The difference in water surface elevation between scenarios I and II is 0.4 feet. Referring to equation H.1, this difference is greater than the reference elevation change, H, which is 0.2 ft, therefore we know that equation H.1 is true and our site is within the reference reach.

Table H.17. Raritan/Middle Brook Influence Reach Determination

Scenario	Middle Brook Return Period	Raritan River Return Period	Middle Brook Flow (ft^3/s)	Raritan River Flow (ft^3/s)	Water Surface Elevation (ft)
I	2-year	2-year	360	19,700	43.8
II	2-year	25-year	360	45,000	44.2
I+10%	2-year + 10%	2-year + 10%	400	21,700	43.9

Based on this conclusion, it is not necessary to run scenario I again with the discharges increased by 10 percent. However, for illustrative purposes scenario I+10% is run with the results summarized in Table H.17. Equation H.1 remains true for this site; our site is within the influence reach for the discharge range of interest. (There is no harm in performing a joint probability analysis for a site beyond the influence reach in terms of the results as one of the discharge combinations is equivalent to the typical riverine situation.)

Step 4. Determine Hydrologic Combinations.

For this example, the drainage area ratio, R_A, is equal to 773/2 = 386 and A_{TOT} is equal to 773 + 2 = 775 mi^2. From Table H.2 we determine that is site falls within category LL.

For our joint design return period of 25-years, the relevant hydrologic combinations are found in Table H.4. With a watershed category of LL and using the representative approach we see that there are three combinations. Combination 1 is to use the 1.25-year discharge on Middle Brook and the 25-year discharge on the Raritan River. Combination 2 employs the 14-yr event on both streams and combination 3 uses the 25-year event on the Middle Brook and the 1.25-year event on the Raritan River. All combinations represent the 25-yr joint return period.

Step 5. Estimate Discharges for Hydrologic Combinations.

For our example, we need to estimate the 1.25-, 14-, and 25-year discharges on both Middle Brook and the Raritan River. We previously estimated the 25-year discharges using the applicable USGS equation in Step 2. To estimate the 14-yr discharge we need to use the same set of USGS equations to estimate the 10- and 25-year discharges and then interpolate. Table H.18 provides the results of applying equation H.5 for the return periods available in the USGS equations.

Table H.18. Raritan/Middle Brook Discharges

Return Period	Discharge (ft^3/s)	
	Raritan River	Middle Brook
2	19,700	360
10	35,800	720
25	45,000	920

Discharges for the 14-year event are computed using Equations H.2 and H.3. For example, the discharge corresponding to a 14-yr return period for the Raritan River needed for combination 2 is calculated as follows.

$$\log Q_x = \frac{\log(14) - \log(10)}{\log(25) - \log(10)}\left(\log(45{,}000) - \log(35{,}800)\right) + \log(35{,}800)$$

$$\log Q_x = \frac{1.146 - 1.000}{1.398 - 1.000}\left(4.653 - 4.554)\right) + 4.554 = 4.590$$

$$Q_x = 10^{\log Q_x} = 10^{4.590} = 38{,}905 = 38{,}900$$

Using the same procedure, the 14-year discharge on Middle Brook is also computed.

The 1.25-year event is estimated using Equation H.4. For the Raritan River, the calculation is as follows:

$$Q_{1.25} = 10^{1.292\log Q_2 - 0.292\log Q_{10}} = 10^{1.292\log(19{,}700) - 0.292\log(35{,}800)} = 16{,}500$$

All needed discharges are summarized in Table H.19.

Table H.19. Raritan/Middle Brook Discharge Combinations

Combination	Middle Brook Return Period (yrs.)	Raritan River Return Period (yrs.)	Middle Brook Discharge (ft^3/s)	Raritan River Discharge (ft^3/s)
1	1.25	25	290	45,000
2	14	14	790	38,900
3	25	1.25	920	16,500

Step 6. Perform Hydraulic Analyses for Each Combination.

After development of discharge pairs for the hydrologic combinations, each pair is applied to the hydraulic model developed for Step 3. Table H.20 summarizes the stage and velocity at the structure site for each combination of discharges.

Table H.20. Stage and Velocity For Raritan/Middle Brook Combinations

Combination	Stage (ft)	Velocity (ft/s)
1	44.2	4.9
2	44.8	10.8
3	45.0	11.8

The most extreme results for the variables of interest are taken as the 25-year joint probability design conditions. For consideration of freeboard, stage is the driving variable; maximum stage is derived from combination 3. For consideration of scour, the designer takes the results from all three combinations and performs a scour computation for each. The maximum scour is used for design. However, in this case the maximum stage and velocity are both from combination 3, therefore, that combination should be the governing case for scour.

Although not true in this example, it is important to remember that the most extreme values for depth, velocity, or scour may come from different combinations. If that is the case, as in the first example in this document, results from different combinations are not combined. More information on the development and application of this joint probability method is found in the main report and appendix G.