— Technical Report Documentation Page —

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA/TX–05/0–4193–7		

4. Title and Subtitle	5. Report Date
An Initial-Abstraction, Constant-Loss Model for Unit Hydrograph Modeling for Applicable Watersheds in Texas	December 2007
7. Authors(s)	8. Performing Organization Report No.
William H. Asquith and Meghan C. Roussel	USGS SIR 2007–5243
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
U.S. Geological Survey 8027 Exchange Drive Austin, Texas 78754	11. Contract or Grant No. Project 0–4193
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Texas Department of Transportation Research and Technology Implementation Office P.O. Box 5080 Austin, Texas 78731	Technical report on research from 2001 to 2007 14. Sponsoring Agency Code

15. Supplementary Notes

Project conducted in cooperation with the Texas Department of Transportation and the Federal Highway Administration

16. Abstract

Estimation of representative hydrographs from design storms, which are known as design hydrographs, provides for cost-effective, risk-mitigated design of drainage structures such as bridges, culverts, roadways, and other infrastructure. During 2001–07, the U.S. Geological Survey (USGS), in cooperation with the Texas Department of Transportation, investigated runoff hydrographs, design storms, unit hydrographs, and watershed-loss models to enhance design hydrograph estimation in Texas. Design hydrographs ideally should mimic the general volume, peak, and shape of observed runoff hydrographs. Design hydrographs commonly are estimated in part by unit hydrographs. A unit hydrograph is defined as the runoff hydrograph that results from a unit pulse of excess rainfall uniformly distributed over the watershed at a constant rate for a specific duration. A time-distributed, watershed-loss model is required for modeling by unit hydrographs. This report develops a specific time-distributed, watershed-loss model known as an initial-abstraction, constant-loss model. For this watershed-loss model, a watershed is conceptualized to have the capacity to store or abstract an absolute depth of rainfall at and near the beginning of a storm. Depths of total rainfall less than this initial abstraction do not produce runoff. The watershed also is conceptualized to have the capacity to remove rainfall at a constant rate (loss) after the initial abstraction is satisfied. Additional rainfall inputs after the initial abstraction is satisfied contribute to runoff if the rainfall rate (intensity) is larger than the constant loss. The initial-abstraction, constant-loss model thus is a two-parameter model. The initial-abstraction, constant-loss model is investigated through detailed computational and statistical analysis of observed rainfall and runoff data for 92 USGS streamflow-gaging stations (watersheds) in Texas with contributing drainage areas from 0.26 to 166 square miles. The analysis is limited to a previously described, watershed-specific, gamma distribution model of the unit hydrograph. In particular, the initial-abstraction, constant-loss model is tuned to the gamma distribution model of the unit hydrograph. A complex computational analysis of observed rainfall and runoff for the 92 watersheds was done to determine, by storm, optimal values of initial abstraction and constant loss. Optimal parameter values for a given storm were defined as those values that produced a modeled runoff hydrograph with volume equal to the observed runoff hydrograph and also minimized the residual sum of squares of the two hydrographs. Subsequently, the means of the optimal parameters were computed on a watershed-specific basis. These means for each watershed are considered the most representative, are tabulated, and are used in further statistical analyses. Statistical analyses of watershed-specific, initial abstraction and constant loss include documentation of the distribution of each parameter using the generalized lambda distribution. The analyses show that watershed development has substantial influence on initial abstraction and limited influence on constant loss. The means and medians of the 92 watershed-specific parameters are tabulated with respect to watershed development; although they have considerable uncertainty, these parameters can be used for parameter prediction for ungaged watersheds. The statistical analyses of watershed-specific, initial abstraction and constant loss also include development of predictive procedures for estimation of each parameter for ungaged watersheds. Both regression equations and regression trees for estimation of initial abstraction and constant loss are provided. The watershed characteristics included in the regression analyses are (1) main-channel length, (2) a binary factor representing watershed development, (3) a binary factor representing watersheds with an abundance of rocky and thin-soiled terrain, and (4) curve number. Physical interpretations of the regression coefficients are made. Finally, an evaluation of an initial-abstraction, constant-loss model for general application is made through four techniques of parameter estimation: the mean and median watershed-specific values, regression equations, and regression trees. The results show that the four techniques have similar overall performance, but measurable differences exist. The four techniques, when combined, are shown to provide unbiased estimates of peak streamflow and reliably represent runoff volumes as well as times of peak streamflow occurrence with less uncertainty than any single technique. The combined initial-abstraction, constant-loss model used with the gamma distribution model of the unit hydrograph is suggested for general application in Texas, and discussion in the context of practical applications is provided.

17. Key Words	18. Distribution Statement
Rainfall, Design Storm, Unit Hydrograph, Loss Model, Design Hydrograph	No restrictions

19. Security Classif. (of report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	82	

Form DOT F 1700.7 (8-72) Reproduction of complete page authorized

Front cover:

Left: Position of gage shelter for 08055700 Bachman Branch at Dallas, Tex. after flood of April 28, 1966. Photograph taken by J.H. Montgomery, May 14, 1966.

Right: Newly established gage shelter for 08055700 Bachman Branch at Dallas, Tex., dated December 15, 1964 by FHR (initials).

Back cover:

Aerial photograph showing location of 08055700 Bachman Branch at Dallas, Tex., dated August 17, 1956.

Blank Page

An Initial-Abstraction, Constant-Loss Model for Unit Hydrograph Modeling for Applicable Watersheds in Texas

By William H. Asquith and Meghan C. Roussel

In cooperation with the Texas Department of Transportation

Scientific Investigations Report 2007–5243
(Texas Department of Transportation Research Report 0–4193–7)

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
DIRK KEMPTHORNE, Secretary

U.S. Geological Survey
Mark D. Myers, Director

U.S. Geological Survey, Reston, Virginia: 2007

For more information about the USGS and its products:
Telephone: 1-888-ASK-USGS
World Wide Web: http://www.usgs.gov/

Any use of trade, product, or firm names in this publication is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this report is in the public domain, permission must be secured from the individual copyright owners to reproduce any copyrighted materials contained within this report.

Suggested citation:
Asquith, W.H., and Roussel, M.C., 2007, An initial-abstraction, constant-loss model for unit hydrograph modeling for applicable watersheds in Texas: U.S. Geological Survey Scientific Investigations Report 2007–5243, 82 p. [http://pubs.usgs.gov/sir/2007/5243]

Contents

Abstract 1
Introduction 2
Purpose and Scope 3
Report Structure 3
Acknowledgments 3
Unit Hydrographs for Applicable Watersheds in Texas 3
Gamma Unit Hydrographs for Applicable Watersheds 5
Estimation of Gamma Unit Hydrograph Shape 7
Estimation of Gamma Unit Hydrograph Time to Peak 8
Analysis of Initial Abstraction and Constant Loss for 92 Watersheds in Texas 8
Analysis of Storm-Specific Initial Abstraction and Constant Loss 9
Analysis of Watershed-Specific Initial Abstraction and Constant Loss 10
Generalized Lambda Distribution Model of Initial Abstraction 12
Generalized Lambda Distribution Model of Constant Loss 12
Regression Equations and Regression Trees to Estimate Initial-Abstraction and Constant-Loss Values 14
Estimation of Initial Abstraction 17
Estimation of Constant Loss 19
General Discussion of the Regression Equations 20
Comparison of the Initial Abstraction Equation to the Curve Number Method 21
An Initial-Abstraction, Constant-Loss Model for Unit Hydrograph Modeling for Applicable Watersheds in Texas 22
The Combined I_A&C_L Model 23
The Combined I_A&C_L Model in Practical Applications 25
Further Discussion 26
Summary 28
References 30
Appendix 1—Selected Watershed Characteristics and Graphical and Statistical Output From the ϟI_A&ϟC_L Software for the 92 Watersheds 33
Examples of Graphical Output From the ϟI_A&ϟC_L Software 35
Selected Watershed Characteristics and $\overline{I_A}$ and $\overline{C_L}$ Values 38
Statistical Relation Between Main-Channel Length and Contributing Drainage Area for the 92 Watersheds 40
Appendix 2—Selected Code and Summary Output from R Environment Showing Final Steps of Regression Process 43
Summary of the I_A&C_L Equations 45
Computation of the I_A Equation 46
Computation of the C_L Equation 47
Abbreviated Residual Analysis for the I_A&C_L Equations 48
Computation of Selected Upper-Tail Values of t-Distribution 49
Summary of the I_A&C_L Trees 50
Example of Graphical Output From Modified ϟI_A&ϟC_L Software for Evaluation of the I_A&C_L Equations 53

Appendix 3—Algorithm Listings for Computations of Gamma Unit Hydrographs and the Initial-Abstraction, Constant-Loss Model . . . 57
Computation of q_p From T_p and K Values of Gamma Unit Hydrograph . . . 59
Computation of Excess Rainfall for I_A&C_L Model . . . 59
Computation of Gamma Unit Hydrograph Ordinates . . . 61
Appendix 4—Derivation of the Gamma Hydrograph Equation . . . 63
Appendix 5—Example Computations of Prediction Limits From Regression Equations . . . 67
Computation of Unit Hydrograph Shape and Prediction Limits . . . 69
Computation of Unit Hydrograph Time to Peak and Prediction Limits . . . 70
Computation of Initial Abstraction and Prediction Limits . . . 71
Computation of Constant Loss and Prediction Limits . . . 72
Glossary . . . 75
Technical Notes . . . 81

Figures

1. Map showing locations of U.S. Geological Survey streamflow-gaging stations in Texas used in analysis of unit hydrographs and the initial-abstraction, constant loss model . . . 4
2. Listing of R code for solution of equation 4 given q_p and T_p values . . . 6
3. Graph showing comparison of observed (Q_p^{obs}) and modeled (Q_p^{model}) peak streamflows estimated by the $^{\natural}I_A$&$^{\natural}C_L$ and K&T_p-GUH models for the 92 watersheds . . . 11
4. Graph showing distribution of $\overline{I_A}$ (watershed-specific I_A) and fitted generalized lambda distribution models for the 92 watersheds . . . 13
5. Graph showing distribution of $\overline{C_L}$ (watershed-specific C_L) and fitted generalized lambda distribution models for the 92 watersheds . . . 14
6. Listing of R code that represents the algorithm used for PRESS minimization in a weighted-least-squares context . . . 17
7. Diagram showing regression tree of $\overline{I_A}$ (the I_A tree) for estimation of I_A . . . 19
8. Diagram showing regression tree of $\overline{C_L}$ (the C_L tree) for estimation of C_L . . . 20
9. Whisker diagrams showing distribution of error statistics for each of the four techniques and the combined I_A&C_L model from numerical values listed in tables 6–8 . . . 24
10. Graph showing comparison of observed (Q_p^{obs}) and modeled ($\mathring{Q}_p$) peak streamflows from the combined I_A&C_L and K&T_p-GUH model for the 92 watersheds . . . 26

1.1. Graphs showing selected graphical output from $^{\natural}I_A$&$^{\natural}C_L$ software for storm of May 4, 1969 . . . 36
1.2. Graphs showing selected graphical output from $^{\natural}I_A$&$^{\natural}C_L$ software for storm of June 3, 1973 . . . 37
1.3. Listing of R code for computation equation 1.1 . . . 41

2.1. Listing of R code for computation of regression weights . . . 45
2.2. Listing of R code for summary of major regression computations shown in figures 2.3 and 2.4 . . . 45
2.3. Listing of R code for computation of I_A equation . . . 46
2.4. Listing of R code for computation of C_L equation . . . 47
2.5. Listing of R code for computations of the residuals for the I_A&C_L equations shown in figures 2.6 and 2.7 . . . 48
2.6. Graph showing residual plot for the 92 watersheds of I_A equation . . . 49
2.7. Graph showing residual plot for the 92 watersheds of C_L equation . . . 49

2.8. Listing of R code for computation of selected values of t-distribution for use in prediction-interval computations 49
2.9. Listing of R code for computation of the I_A tree and residual plot 50
2.10. Listing of R code for computation of the C_L tree and residual plot 51
2.11. Diagram showing raw I_A tree from R environment for the 92 watersheds used to generate figure 7 52
2.12. Diagram showing raw C_L tree from R environment for the 92 watersheds used to generate figure 8 52
2.13. Graph showing residual plot for the 92 watersheds of I_A tree 52
2.14. Graph showing residual plot for the 92 watersheds of C_L tree 52
2.15. Graphs showing selected worst-case graphical output showing multiple peaks from the modified ${}^{\natural}I_A{}^{\&\natural}C_L$ software using the $I_A{}^{\&}C_L$ equations 54
2.16. Graphs showing selected best-case graphical output from the modified ${}^{\natural}I_A{}^{\&\natural}C_L$ software using the $I_A{}^{\&}C_L$ equations 55

3.1. Listing of R code for solution of equation 4 given q_p and K values 59
3.2. Listing of R code for computation of excess rainfall using an initial-abstraction, constant-loss ($I_A{}^{\&}C_L$) model 60
3.3. Listing of R code for computation of an ordinate of a gamma unit hydrograph (GUH) as defined by equation 2 61

5.1. Listing of R code for computation of leverage for K equation 70
5.2. Listing of R code for computation of leverage for T_p equation 71
5.3. Listing of R code for computation of leverage for I_A equation 72
5.4. Listing of R code for computation of leverage for C_L equation 73

Tables

1. Common summary statistics of $\overline{I_A}$ and $\overline{C_L}$ (watershed-specific I_A and C_L) for the 92 watersheds 10
2. L-moment statistics of $\overline{I_A}$ and $\overline{C_L}$ (watershed-specific I_A and C_L) for the 92 watersheds 11
3. Parameters of generalized lambda distribution for I_A for the 92 watersheds 12
4. Parameters of generalized lambda distribution for C_L for the 92 watersheds 13
5. Comparison of I_A as predicted by the curve number (CN) method and the I_A equation 22
6. Common summary statistics of peak streamflow error (ε^{Q_p}) for the 92 watersheds 22
7. Common summary statistics of runoff hydrograph volume error (ε^{V}) for the 92 watersheds 23
8. Common summary statistics of time difference of peak streamflow ($\varepsilon^{T_{Q_p}}$) for the 92 watersheds 25

1.1. Selected watershed characteristics and $\overline{I_A}$ and $\overline{C_L}$ (watershed-specific) values for the 92 watersheds 38

2.1. Selected upper-tail values of t-distribution for three degrees of freedom 49

Conversion Factors

Inch/Pound to SI

Multiply	By	To obtain
	Length	
inch (in.)	25.4	millimeter (mm)
foot (ft)	.3048	meter (m)
mile (mi)	1.609	kilometer (km)
	Area	
square mile (mi^2)	2.590	square kilometer (km^2)
	Flow	
cubic feet per second (ft^3/s)	.02832	cubic meters per second (m^3/s)

SI to Inch/Pound

Multiply	By	To obtain
	Length	
millimeter (mm)	0.03937	inch (in.)
meter (m)	3.281	foot (ft)
	Area	
square kilometer (km^2)	.3861	square mile (mi^2)
	Flow	
cubic meters per second (m^3/s)	35.31	cubic feet per second (ft^3/s)

Horizontal and Vertical Datums

Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83).
Vertical coordinate information is referenced to the North American Vertical Datum of 1988 (NAVD 88).

Conventions

Acronyms and Computer Code Several acronyms are used in this report. With few exceptions these are typeset as a slightly larger, small caps version of the main document font (Times): gamma unit hydrograph (GUH). Computer code, specifically for the language R, which is shown both in figures or as in-line text, is typeset with variants of a monospaced font (Luximono) that changes in shape and weight with context. When code is listed in a figure, often it is necessary to break (wrap) lines of code. These breaks are shown by "↩" and "↪", for pre- and post-line breaks, respectively. General keywords of the computer code generally are typeset in bold: **`optimize`** or **`sum`**. Code comments are typeset in oblique face: *`# this is a code comment`*. All other code is typeset in regular, upright face: `Vfunc.K()`.

Footnotes and Technical Notes The purpose of footnotes, which are typeset at the bottom of the respective column, are to provide clarification or otherwise augment the context. In contrast, the purposes of the technical notes, which are provided at the back of the report, are to provide additional background information to the report or provide documention. Footnote numbers are shown in the text as a superscript; technical notes are shown as a lower-case alphabetical character in brackets.

Universal Resource Locations (URL) Links to online resources are typeset, like computer code examples, in the Luximono font with application of highly specialized-line breaking rules.

An Initial-Abstraction, Constant-Loss Model for Unit Hydrograph Modeling for Applicable Watersheds in Texas

By William H. Asquith and Meghan C. Roussel

Abstract

Estimation of representative hydrographs from design storms, which are known as design hydrographs, provides for cost-effective, risk-mitigated design of drainage structures such as bridges, culverts, roadways, and other infrastructure. During 2001–07, the U.S. Geological Survey (USGS), in cooperation with the Texas Department of Transportation, investigated runoff hydrographs, design storms, unit hydrographs, and watershed-loss models to enhance design hydrograph estimation in Texas. Design hydrographs ideally should mimic the general volume, peak, and shape of observed runoff hydrographs. Design hydrographs commonly are estimated in part by unit hydrographs. A unit hydrograph is defined as the runoff hydrograph that results from a unit pulse of excess rainfall uniformly distributed over the watershed at a constant rate for a specific duration. A time-distributed, watershed-loss model is required for modeling by unit hydrographs. This report develops a specific time-distributed, watershed-loss model known as an initial-abstraction, constant-loss model. For this watershed-loss model, a watershed is conceptualized to have the capacity to store or abstract an absolute depth of rainfall at and near the beginning of a storm. Depths of total rainfall less than this initial abstraction do not produce runoff. The watershed also is conceptualized to have the capacity to remove rainfall at a constant rate (loss) after the initial abstraction is satisfied. Additional rainfall inputs after the initial abstraction is satisfied contribute to runoff if the rainfall rate (intensity) is larger than the constant loss. The initial-abstraction, constant-loss model thus is a two-parameter model. The initial-abstraction, constant-loss model is investigated through detailed computational and statistical analysis of observed rainfall and runoff data for 92 USGS streamflow-gaging stations (watersheds) in Texas with contributing drainage areas from 0.26 to 166 square miles. The analysis is limited to a previously described, watershed-specific, gamma distribution model of the unit hydrograph. In particular, the initial-abstraction, constant-loss model is tuned to the gamma distribution model of the unit hydrograph. A complex computational analysis of observed rainfall and runoff for the 92 watersheds was done to determine, by storm, optimal values of initial abstraction and constant loss. Optimal parameter values for a given storm were defined as those values that produced a modeled runoff hydrograph with volume equal to the observed runoff hydrograph and also minimized the residual sum of squares of the two hydrographs. Subsequently, the means of the optimal parameters were computed on a watershed-specific basis. These means for each watershed are considered the most representative, are tabulated, and are used in further statistical analyses. Statistical analyses of watershed-specific, initial abstraction and constant loss include documentation of the distribution of each parameter using the generalized lambda distribution. The analyses show that watershed development has substantial influence on initial abstraction and limited influence on constant loss. The means and medians of the 92 watershed-specific parameters are tabulated with respect to watershed development; although they have considerable uncertainty, these parameters can

be used for parameter prediction for ungaged watersheds. The statistical analyses of watershed-specific, initial abstraction and constant loss also include development of predictive procedures for estimation of each parameter for ungaged watersheds. Both regression equations and regression trees for estimation of initial abstraction and constant loss are provided. The watershed characteristics included in the regression analyses are (1) main-channel length, (2) a binary factor representing watershed development, (3) a binary factor representing watersheds with an abundance of rocky and thin-soiled terrain, and (4) curve number. Physical interpretations of the regression coefficients are made. Finally, an evaluation of an initial-abstraction, constant-loss model for general application is made through four techniques of parameter estimation: the mean and median watershed-specific values, regression equations, and regression trees. The results show that the four techniques have similar overall performance, but measurable differences exist. The four techniques, when combined, are shown to provide unbiased estimates of peak streamflow and reliably represent runoff volumes as well as times of peak streamflow occurrence with less uncertainty than any single technique. The combined initial-abstraction, constant-loss model used with the gamma distribution model of the unit hydrograph is suggested for general application in Texas, and discussion in the context of practical applications is provided.

Introduction

Estimation of representative hydrographs[a] [1] from design storms[b] is an important goal of hydrologic engineering. The goal is important because these "design hydrographs," in general, and the peak streamflow of a design hydrograph, in particular, are critical for cost-effective, risk-mitigated design of drainage structures such as bridges, culverts, roadways, and other infrastructure.

During 2001–07, the U.S. Geological Survey (USGS), in cooperation with the Texas Department of Transportation (TxDOT) and in collaboration (TxDOT Research Project 0–4193) with researchers at Texas Tech University,[2] Lamar University,[3] and University of Houston,[4] investigated rainfall and runoff data from 92 USGS streamflow-gaging stations (watersheds) to enhance design hydrograph estimation for applicable watersheds[c] in Texas. The locations of the stations are shown in figure 1. Further background and results of the research program between the five agencies is in Asquith, Thompson, and others (2004), Asquith and others (2005), Asquith and others (2006), Thompson and others (2007), and references therein.

Design hydrographs ideally should mimic the general volume, peak streamflow, and shape of observed runoff hydrographs. Design hydrographs commonly are estimated in part by unit hydrographs. A unit hydrograph is defined as the runoff hydrograph that results from a unit pulse of excess rainfall uniformly distributed over a watershed at a constant rate for a specific duration. Among other concepts, runoff hydrograph estimation by the unit hydrograph method requires (1) a rainfall hyetograph that represents rainfall intensity of a storm over time on a watershed and (2) a method to convert this hyetograph into an appropriate excess rainfall hyetograph.[d] The conversion is made by a watershed-loss model.

There are two general classes of watershed-loss models: non-time distributed and time distributed. A non-time distributed model is appropriate when only total runoff volume and not an actual runoff hydrograph is required for a particular application. This class effectively is not considered further in this report.

Time-distributed, watershed-loss models are applicable when an analyst needs a design hydrograph. When the design storm hyetograph[e] is combined with a time-distributed, watershed-loss model, an excess rainfall hyetograph results. When this design excess rainfall hyetograph is convolved with the unit hydrograph, the runoff hydrograph of the design storm is produced. In particular, the time-distributed, watershed-loss model should be mathematically consistent in structure or general form with hypothesized processes

[1] Alphabetical superscripts [a] refer to Technical Notes located at back of report.

[2] David B. Thompson, Research Supervisor on TxDOT Research Project 0–4193 (2001–07), Texas Tech University, Lubbock, Tex. at the time this work was done.

[3] Xing Fang, Research Associate on TxDOT Research Project 0–4193 (2001–07), Lamar University, Beaumont, Tex.

[4] Theodore G. Cleveland, Research Associate on TxDOT Research Project 0–4193 (2001–07), University of Houston, Houston, Tex.

(depression and other macrostorage, infiltration, evaporation, and other physical processes), which abstract or otherwise prevent rainfall from becoming runoff.

Purpose and Scope

The primary purpose of this report is to develop a specific time-distributed, watershed-loss model known as an initial-abstraction, constant-loss model for general use for watersheds that are believed to be hydrologically similar to the 92 watersheds of this report. The development is made through detailed computational and statistical analysis of observed rainfall and runoff data for the 92 watersheds with contributing drainage areas from 0.26 to 166 square miles. The analysis is limited to a previously described, watershed-specific, gamma distribution model of the unit hydrograph based on the data for the watersheds (Asquith and others, 2005). A secondary purpose of this report is to explain how the initial-abstraction, constant-loss model and gamma distribution model of the unit hydrograph would be practically implemented in hydrologic-engineering circumstances. The explanation is limited to description of a combined initial-abstraction, constant-loss model, discussion of model bias, and discussion of the mathematical implementation of the unit hydrograph with discrete time steps.

Report Structure

This report contains several major, but codependent, elements:

1. A comprehensive summary of the unit hydrograph used for development of the watershed-loss model in the section titled "Unit Hydrographs for Applicable Watersheds in Texas."
2. A comprehensive statistical analysis of parameters for the watershed-loss model in the section titled "Analysis of Initial Abstraction and Constant Loss for 92 Watersheds in Texas."
3. A proposed watershed-loss model for general application in the section titled "An Initial-Abstraction, Constant-Loss Model for Unit Hydrograph Modeling for Applicable Watersheds in Texas."
4. Five appendixes that present numerous supporting details regarding the investigation.
   - Appendix 1 presents supporting computations and data from the section described for element 2.
   - Appendix 2 presents additional discussion and archival computations related to material presented in the section described for element 3.
   - Appendix 3 presents selected algorithms, which are intended to augment the application of the results reported here.
   - Appendix 4 presents mathematical derivation of a gamma hydrograph equation.
   - Appendix 5 presents example computations of prediction limits for four regression equations presented in this report.

Acknowledgments

The authors acknowledge the following colleagues of TxDOT for their considerable professional and technical support of this work and their critical feedback on project direction: George R. Herrmann, 0–4193 Project Director; Amy J. Ronnfeldt, 0–4193 Project Advisor; and David Stolpa (retired), 0–4193 Project Advisor. The authors are indebted to our research colleagues David B. Thompson (Texas Tech University), Theodore G. Cleveland (University of Houston), and Xing Fang (Lamar University) for nearly 7 years of intellectually stimulating and collaborative work. The breadth of their influence is difficult to describe; this report stands as a product of our partnership.

Unit Hydrographs for Applicable Watersheds in Texas

The unit hydrograph method is a well-known, textbook technique (Dingman, 2002) for estimating the runoff hydrograph given an excess rainfall hyetograph. Excess rainfall is a volume of rainfall per unit area (depth) after losses such as evaporation, infiltration,

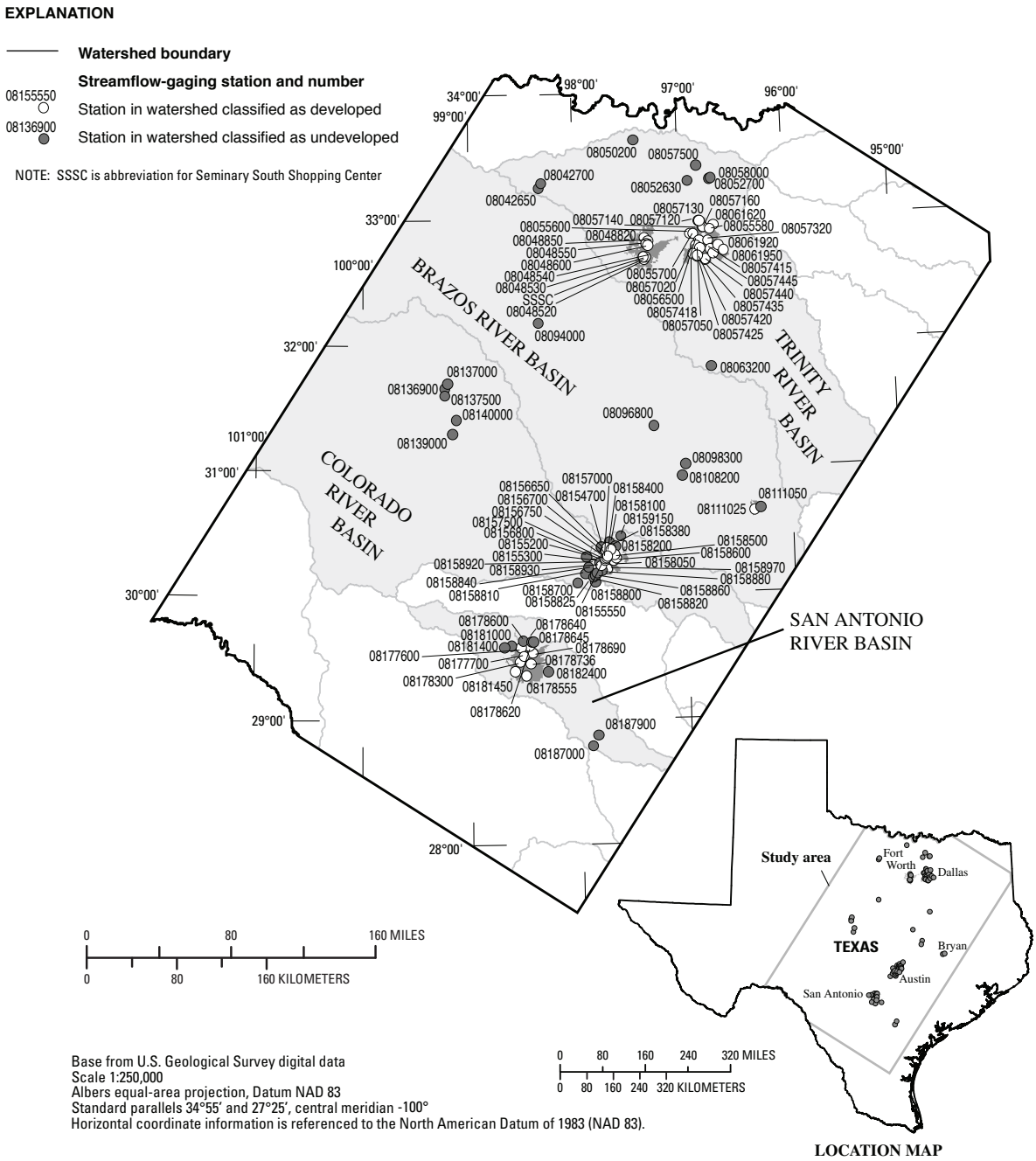


Figure 1. Map showing locations of U.S. Geological Survey streamflow-gaging stations in Texas used in analysis of unit hydrographs and the initial-abstraction, constant loss model.

depression storage, and others are subtracted. A unit hydrograph is defined as the runoff hydrograph that results from a unit pulse of excess rainfall uniformly distributed over a watershed at a constant rate for a specific duration (Chow and others, 1988, p. 213).

To use a unit hydrograph, a watershed is assumed to function as a linear system in which the concepts superposition, proportionality, and time invariance are appropriate.

SUPERPOSITION: The runoff hydrograph resulting from two consecutive pulses is computed using the addition of two unit hydrographs, separated by the effective duration of the first pulse. The duration of the unit hydrograph is equal to the time step of the rainfall pulses.

PROPORTIONALITY: The runoff hydrograph resulting from two simultaneous pulses of unit rainfall of a specific duration has ordinates that are twice as large as those resulting from a single unit pulse of rainfall of the same duration.

TIME INVARIANCE: A unit hydrograph has a specified time period (step), and this time step is the duration of the time-invariant (uniform-intensity) excess rainfall. As a result, the magnitude of runoff within each time step is uniform. The temporal input-response relation is fixed for a given watershed. The duration of the unit hydrograph should be specified in the name of the unit hydrograph. For this investigation, a 5-minute time step exclusively was used.

The relation between excess rainfall, the unit hydrograph, and the runoff hydrograph is algebraically straightforward. The discrete convolution for a linear system is used to generate the runoff hydrograph given an excess rainfall and a unit hydrograph. The convolution is

$$Q_n = \sum_{m=1}^{n \leq M} P_m U_{n-m+1}, \tag{1}$$

where Q_n is runoff, in cubic feet per second, which is estimated or modeled from the excess rainfall (P_m), in watershed inches; U_{n-m+1} are the ordinates of the unit hydrograph, in cubic feet per unit time; M is the number of excess rainfall pulses; and m and n are integers.

Gamma Unit Hydrographs for Applicable Watersheds

Asquith and others (2005) document four separate lines of inquiry into unit hydrographs for the 92 watersheds. These unit hydrographs were developed using an extensive database of USGS rainfall and runoff data for Texas. Almost all the database is summarized in Asquith, Thompson, and others (2004), and the remainder of the data are on file at the USGS Water Science Center, Austin, Tex. One of those lines of inquiry (Gamma Unit Hydrograph Analysis System, GUHAS) involved an algebraically straightforward, analyst-directed approach for 5-minute unit hydrograph estimation based on the unit hydrograph having the form of a gamma distribution (Evans and others, 2000). The results of GUHAS provide a starting point for this report. A unit hydrograph from GUHAS is referred to as the gamma unit hydrograph (GUH). The mathematics of the GUH are now described.

The structural form of the GUH used with GUHAS is discussed by Haan and others (1994). The GUH can attain shapes that mimic the general shape of many observed runoff hydrographs (unit or otherwise). The GUH has two unique parameters that can be variously expressed, but are expressed here in terms of watershed-depth peak streamflow (q_p) in inches over the watershed per hour (watershed inches per hour) and time to peak (T_p) in hours. The third parameter is a shape parameter (K) that is dependent on q_p and T_p. Expression and analysis of unit hydrographs in terms of q_p and T_p are important because the magnitude and timing of peak streamflow (Q_p) in cubic feet per second are critical for many designs.

Because of the importance of Q_p estimation in hydrologic-engineering practice, the optimal q_p and T_p values for GUHAS were selected for each storm by precisely matching the modeled peak streamflow (Q_p^{model}) in cubic feet per second and time of modeled peak streamflow ($T^{Q_p^{\text{model}}}$) to observed values of peak streamflow (Q_p^{obs}) in cubic feet per second and time of observed peak streamflow ($T^{Q_p^{\text{obs}}}$) for 1,984 individual storm peaks (Asquith and others, 2005, p. 29). The storm peaks come from datafile-by-datafile, graphical interpretation of the database described in Asquith, Thompson, and others (2004). GUHAS also permitted

individual (discrete) analysis of multiple Q_p values within a single storm if the storm had peaks that were substantially separate in time.

The equation defining a gamma hydrograph (unit or otherwise) is

$$\frac{q(t)}{q_p} = \left[\frac{t}{T_p} e^{1-(\frac{t}{T_p})}\right]^K, \tag{2}$$

where q_p is watershed-depth peak streamflow in watershed inches per hour, T_p is time to peak in hours, K is a shape parameter that is dependent on q_p and T_p, and $q(t)$ is streamflow in watershed inches per hour at time t in hours. This equation produces $q(t)$ ordinates of a GUH.[5] The relation between q_p and Q_p as well as $q(t)$ and Q_n is

$$Q = 645.33 \times q \times A, \tag{3}$$

where Q is streamflow in cubic feet per second, q is in watershed inches per hour, A is drainage area in square miles, and 645.33 is a units conversion factor.

Although three parameters (q_p, T_p, and K) are shown in equation 2, in practice, K is a function of q_p and T_p, and total runoff volume (V), or alternatively any two parameters will yield the third because $V = 1$ (unit volume) for a unit hydrograph. The V of a gamma hydrograph is computed by

$$V = q_p T_p \Gamma(K) \left(\frac{e^{(1)}}{K}\right)^K, \tag{4}$$

where $\Gamma(K)$ is the complete gamma function (Abramowitz and Stegun, 1964) for K. The time scale of the unit hydrograph is represented by T_p, but T_p does not represent the time base (T_b) or overall width in time of a runoff hydrograph. A numerical root solver or function minimizer[6] is required to compute one of the three parameters from the other two in equation 4. The R environment for statistical computing (R Development Core Team, 2006) can be used as well. A listing of minimal[7] code to solve equation 4 for K is shown in figure 2.

[5]The derivation of the gamma hydrograph equation from the gamma distribution is shown in appendix 4.

[6]Popular spreadsheets, such as NeoOffice-Calc (`http://www.neooffice.org`), generally provide a "Goal Seek" function, and the natural logarithm of the gamma function is the software function `gammaln()`.

[7]Minimal in this context assumes familiarity with the R environment.

Asquith and others (2005) provide equations to estimate the GUH for applicable Texas watersheds based on 92 of the 93 watersheds shown in figure 1. The streamflow-gaging station at Seminary South Shopping Center (fig. 1) was not used in Asquith and others (2005) and also was not used for this report because critical basin characteristics were indeterminate. The equations of K and T_p from Asquith and others (2005) are reproduced here in equations 5 and 11 and collectively are referred to as $K^{\&}T_p$ equations. A GUH set by these equations is referred to as a $K^{\&}T_p$-GUH. The $K^{\&}T_p$-GUHs are deemed appropriate for investigation of the watershed-loss model in this report on the basis of lines of reasoning in Asquith and others (2005, p. 14–15, 59–64) and supported by conclusions from independent modeling techniques described by Cleveland and others (2006).

To summarize, the $K^{\&}T_p$ equations were derived by Asquith and others (2005) from multiple-linear regression analysis (Helsel and Hirsch, 2002; Maindonald and Braun, 2003; Montgomery and others, 2001). For the equations, the mean shape parameters K and T_p for each watershed (Asquith and others, 2005, table 3) were used as regressor variables, and watershed char-

```
qp <- 0.3; # in watershed inches per hour
Tp <- 4.0; # in hours

# build an objective function for the ↩
    ↪shape parameter, which minimizes at ↩
    ↪zero when the gamma hydrograph has ↩
    ↪volume of 1 (a gamma unit hydrograph)
Vfunc.K <- function(K) { return( abs(qp * ↩
    ↪Tp * gamma(K) * (exp(1)/K)^K - 1) ) }

# minimize the objective function by ↩
    ↪specifying the function and the ↩
    ↪interval of shape parameter in which ↩
    ↪to search.
optimize( Vfunc.K, interval = c(0.01,10) )

$minimum
[1] 5.253242

$objective
[1] 4.339262e-07
# for qp and Tp, the K is about 5.25.
```

Figure 2. Listing of R code for solution of equation 4 given q_p and T_p values.

acteristics of main-channel length (L), dimensionless main-channel slope (S), and watershed development (D) were used as predictor variables. These characteristics are listed in Asquith and others (2005, table 2) as well as listed in appendix 1 (table 1.1).

Estimation of Gamma Unit Hydrograph Shape

The regression equation for the GUH shape parameter (Asquith and others, 2005, eq. 26) is

$$K = 10^{0.560-0.249D} L^{0.142}, \tag{5}$$

where K is shape parameter, $D = 0$ for undeveloped watersheds and $D = 1$ for developed watersheds,[8] and L is main-channel length of the watershed in miles. The equation has 88 degrees of freedom ($\mathrm{df}^{[K]} = 88$), a residual standard error ($\sigma^{[K]}$) of $0.2052 \log_{10}(K)$ units, and an adjusted R-squared of 0.292. The maximum leverage ($h^{[K]}_{\max}$) is about 0.132, which can be useful in assessing applicability of equation 5.

Weighted mean K (no $\log_{10}$ transformation) without regard to watershed development is about 3.9, and weighted mean values with regard to watershed development are 5.2 and 2.9 for undeveloped and developed watersheds, respectively. These values bracket the equivalent K value of 3.77 for a gamma dimensionless hydrograph equivalent to the Natural Resources Conservation Service (2006) dimensionless hydrograph. A larger shape parameter for undeveloped watersheds implies that undeveloped watersheds tend to have more-symmetrical dimensionless hydrographs than developed watersheds (Asquith and others, 2005, fig. 3), and the Natural Resources Conservation Service dimensionless hydrograph lies between them.

The prediction limits of K from equation 5 can be useful for expressing uncertainty when the equation is used. The limits for the $100 \times (1-\alpha)$ prediction interval are shown as

$$^{\downarrow}K^{\alpha/2}_{\oplus} \leq K_{\oplus} \leq {}^{\uparrow}K^{\alpha/2}_{\oplus}, \tag{6}$$

where down (up) arrow signifies the lower (upper) prediction limit for the prediction $K_{\oplus}$ at the α significance level. The symbol $\oplus$ in this report[9] designates the vector of the watershed characteristics or predictor variables for a given prediction.

The lower and upper limits are computed by

$$^{\downarrow}K^{\alpha/2}_{\oplus} = 10^{\log_{10}(K_{\oplus}) - \mathrm{t}_{[\alpha/2,\mathrm{df}]}\, \sigma^{[K]} \sqrt{1+h^{[K]}_{\oplus}}} \text{ and} \tag{7}$$

$$^{\uparrow}K^{\alpha/2}_{\oplus} = 10^{\log_{10}(K_{\oplus}) + \mathrm{t}_{[\alpha/2,\mathrm{df}]}\, \sigma^{[K]} \sqrt{1+h^{[K]}_{\oplus}}}, \tag{8}$$

where $K_{\oplus}$ is the prediction from equation 5, $\mathrm{t}_{[\alpha/2,\mathrm{df}]}$ is the upper tail of the t-distribution for $\mathrm{df}^{[K]}$ degrees of freedom at the α significance level, $\sigma^{[K]}$ is the residual standard error of equation 5, and $h^{[K]}_{\oplus}$ is the leverage of the prediction for the watershed.

The leverage $h^{[K]}_{\oplus}$ for a prediction for a watershed from equation 5 is important for estimation of the prediction limits of the equation. Leverage $h_{\oplus}$ is a measure of influence on the regression—in other words, $h_{\oplus}$ is a measure of distance that the predictor variables are from the center of the parameter space on which an equation is based. The $h^{[K]}_{\oplus}$ calculation requires matrix multiplication and the inverted covariance matrix of the linear-regression model. The computation of $h^{[K]}_{\oplus}$ is

$$h^{[K]}_{\oplus} = \begin{bmatrix} 1 & \log_{10}(L_{\oplus}) & D_{\oplus} \end{bmatrix} \times \begin{bmatrix} 0.10599 & -0.10349 & -0.03134 \\ -0.10349 & 0.13156 & 0.00859 \\ -0.03134 & 0.00859 & 0.04502 \end{bmatrix} \times \begin{bmatrix} 1 \\ \log_{10}(L_{\oplus}) \\ D_{\oplus} \end{bmatrix}, \tag{9}$$

where $L_{\oplus}$ and $D_{\oplus}$ are watershed-specific values, and the 3×3 matrix is the inverted covariance matrix of the regression. The prediction limits (interval) of $K_{\oplus}$ will increase (expand) as the watershed characteristics become farther away from those used (the parameter space) to develop equation 5. Equation applicability can be roughly assessed by the maximum leverage of the regression. In general, a regression equation is judged applicable for a given watershed if the leverage

[8] D as a concept and regressor variable is formally described on page 16.

[9] The literature commonly uses a subscripted symbol o, but the authors use the $\oplus$ symbol (similar to a "target") to visually reinforce the concept of prediction.

for a prediction is about

$$h_{\oplus}^{[K]} \leq h_{\max}^{[K]} = 0.132. \tag{10}$$

Estimation of Gamma Unit Hydrograph Time to Peak

The regression equation for the GUH timing parameter (Asquith and others, 2005, eq. 34) is

$$T_p = 10^{-1.49-0.354D} L^{0.602} S^{-0.672}, \tag{11}$$

where T_p is time to peak in hours, $D = 0$ for undeveloped watersheds and $D = 1$ for developed watersheds, L is main-channel length of the watershed in miles, and S is dimensionless main-channel slope. The equation has 87 degrees of freedom ($\mathrm{df}^{[T_p]} = 87$), a residual standard error ($\sigma^{[T_p]}$) of $0.1383 \log_{10}(T_p)$ units, and an adjusted R-squared of 0.858. The maximum leverage ($h_{\max}^{[T_p]}$) is about 0.136, which can be useful in assessing applicability of equation 11.

The prediction limits of T_p from equation 11 can be useful for expressing uncertainty when the equation is used. The limits for the $100 \times (1-\alpha)$ prediction interval are shown as

$$^{\downarrow}T_{p\oplus}^{\alpha/2} \leq T_{p\oplus} \leq {}^{\uparrow}T_{p\oplus}^{\alpha/2}, \tag{12}$$

where down (up) arrow signifies the lower (upper) prediction limit for T_p at the α significance level.

The lower and upper limits are computed by

$$^{\downarrow}T_{p\oplus}^{\alpha/2} = 10^{\log_{10}(T_{p\oplus}) - \mathrm{t}_{[\alpha/2,\mathrm{df}]}\, \sigma^{[T_p]} \sqrt{1+h_{\oplus}^{[T_p]}}} \text{ and} \tag{13}$$

$$^{\uparrow}T_{p\oplus}^{\alpha/2} = 10^{\log_{10}(T_{p\oplus}) + \mathrm{t}_{[\alpha/2,\mathrm{df}]}\, \sigma^{[T_p]} \sqrt{1+h_{\oplus}^{[T_p]}}}, \tag{14}$$

where $T_{p\oplus}$ is the prediction from equation 11, $\mathrm{t}_{[\alpha/2,\mathrm{df}]}$ is the upper tail of the t-distribution for $\mathrm{df}^{[T_p]}$ degrees of freedom at the α significance level, $\sigma^{[T_p]}$ is the residual standard error of equation 11, and $h_{\oplus}^{[T_p]}$ is the leverage of the prediction for the watershed.

The leverage $h_{\oplus}^{[T_p]}$ for a $T_{p\oplus}$ prediction for a watershed from equation 11 is important for estimation of the prediction limits of the equation. The computation of $h_{\oplus}^{[T_p]}$ is

$$h_{\oplus}^{[T_p]} = \begin{bmatrix} 1 & \log_{10}(L_{\oplus}) & \log_{10}(S_{\oplus}) & D_{\oplus} \end{bmatrix} \times \begin{bmatrix} 1.34775 & 0.16646 & 0.67709 & -0.01451 \\ 0.16646 & 0.19025 & 0.14719 & 0.01225 \\ 0.67709 & 0.14719 & 0.36919 & 0.00918 \\ -0.01451 & 0.01225 & 0.00918 & 0.04524 \end{bmatrix} \times \begin{bmatrix} 1 \\ \log_{10}(L_{\oplus}) \\ \log_{10}(S_{\oplus}) \\ D_{\oplus} \end{bmatrix}, \tag{15}$$

where $L_{\oplus}$, $S_{\oplus}$, and $D_{\oplus}$ are watershed-specific values, and the 4×4 matrix is the inverted covariance matrix of the regression. The prediction limits (interval) of $T_{p\oplus}$ will increase (expand) as the watershed characteristics become farther away from those used (the parameter space) to develop equation 11. Equation applicability can be roughly assessed by the maximum leverage of the regression. In general, a regression equation is judged applicable for a given watershed if the leverage for a prediction is about

$$h_{\oplus}^{[T_p]} \leq h_{\max}^{[T_p]} = 0.136. \tag{16}$$

For the $K^{\&}T_p$ equations, the general range of L for which the equations can be used is 1–50 miles, and the general range for S is 0.002–0.020 (Asquith and others, 2005, figs. 15 and 16).

Analysis of Initial Abstraction and Constant Loss for 92 Watersheds in Texas

The initial-abstraction, constant-loss model ($I_A{}^{\&}C_L$) for rainfall-runoff computations for the 92 watersheds is analyzed in this section. Initial abstraction (I_A) has units of length (watershed depth) or watershed inches. Constant loss (C_L) has units of length (again, watershed depth) per unit time or watershed inches per hour. For the $I_A{}^{\&}C_L$ model, a watershed is conceptualized to have the capacity to store or abstract an absolute depth of rainfall at and near the beginning of a storm. Depths of rainfall less than this initial abstraction do not produce runoff. The watershed also is conceptualized to have the

capacity to remove rainfall at a constant rate (loss) after the initial abstraction is satisfied. Additional rainfall inputs after the initial abstraction is satisfied contribute to runoff if the rainfall rate (intensity) is larger than the constant loss. The $I_A{}^{\&}C_L$ model thus is a two-parameter model of watershed losses.[10]

Values of C_L, which have units of length per unit time, can be compared to infiltration rates. The watershed-loss model is structured to be mathematically consistent with infiltration and other processes. The conceptualization of rainfall removal in this report primarily is as a (statistical) model of aggregate rainfall loss and not necessarily a model of infiltration that represents the actual physics of water movement in soil and other land-cover types.

Analysis of Storm-Specific Initial Abstraction and Constant Loss

A complex computational analysis of the database of observed rainfall and runoff for the 92 watersheds was done in summer 2006 using custom-built software (Joseph Vrabel, U.S. Geological Survey, written commun., 2006; source code is available from report web site). The purpose of the software (labeled as ${}^{\natural}I_A{}^{\&}{}^{\natural}C_L$) was to compute optimal (storm-specific, ${}^{\natural}$) parameter values (${}^{\natural}I_A$ and ${}^{\natural}C_L$) for each suitable storm[f] in the database. The analysis used the observed rainfall and runoff as computational input. The computations used a 5-minute time step; rainfall and runoff were linearly interpolated to 5-minute increments as needed. The GUH was estimated for each watershed by the $K^{\&}T_p$ equations. This single "watershed-specific GUH" was considered a representative unit hydrograph for the watershed. The analysis successfully processed 1,620 of about 1,660 storms inclusive of all watersheds and provided ${}^{\natural}I_A$ and ${}^{\natural}C_L$ values used in subsequent statistical analyses reported here.[11]

The assumption that the GUH set by the $K^{\&}T_p$ equations is representative for each watershed for the analysis of ${}^{\natural}I_A$ and ${}^{\natural}C_L$ is important. Although the "correctness" of this unit hydrograph is not ensured, the assumption implies that the $I_A{}^{\&}C_L$ model reported here is "tuned" against observed rainfall and runoff and $K^{\&}T_p$-GUH. Therefore, a complete framework in a conceptual and statistical sense is established. The $I_A{}^{\&}C_L$ model is linked to $K^{\&}T_p$-GUH, and in practice the two techniques are to be used together.

Values for ${}^{\natural}I_A$ and ${}^{\natural}C_L$ were computed by the ${}^{\natural}I_A{}^{\&}{}^{\natural}C_L$ software. These values are defined as those that generated an excess rainfall hyetograph, which when convolved with $K^{\&}T_p$-GUH, produced a modeled runoff hydrograph that has the same volume as the observed runoff hydrograph ($V^{\text{model}} = V^{\text{obs}}$) and a minimized residual sum of squares between the observed and modeled runoff hydrographs. ${}^{\natural}I_A$ and ${}^{\natural}C_L$ were further constrained to be physically meaningful—that is, values greater than or equal to zero. Negative values for either parameter imply that the "loss model" contributes to the total runoff volume of a storm, which is a physically meaningless scenerio.[g]

Values for ${}^{\natural}I_A$ and ${}^{\natural}C_L$ computationally were determined by an adaptive grid search algorithm in which an initial grid of I_A and C_L with coarse resolution was used. Successive grids at finer resolutions and volume-match tolerance were used to arrive at a subset of parameter space ("candidate pairs"). The residual sum of squares between the observed and modeled runoff hydrograph was computed for each candidate pair, and the candidate pair resulting in the smallest residual sum of squares was chosen as optimal. This approach capitalizes on the fact that it is possible to determine if the volume-matching constraint is met for a particular parameter pair without making computationally expensive model runs involving convolution of the excess rainfall hyetograph and unit hydrograph (eq. 1).

Subsequently, the means of ${}^{\natural}I_A$ and ${}^{\natural}C_L$ for each watershed were computed; the means ($\overline{I_A}$ and $\overline{C_L}$) are referred to as watershed-specific. $\overline{I_A}$ and $\overline{C_L}$ for each watershed are considered the most representative and are used in further statistical analyses reported here. Mathematically these values are computed by

$$\overline{I_A} = \frac{\sum_{i=1}^{92} w_i \times {}^{\natural}I_{A[i]}}{\sum_{i=1}^{92} w_i} \text{ and} \tag{17}$$

$$\overline{C_L} = \frac{\sum_{i=1}^{92} w_i \times {}^{\natural}C_{L[i]}}{\sum_{i=1}^{92} w_i}\ , \tag{18}$$

where $\overline{I_A}$ and $\overline{C_L}$ are watershed-specific values, and w_i

[10]The algorithm used to implement the $I_A{}^{\&}C_L$ model is described in appendix 3 (fig. 3.2).

[11]Two examples of the graphical output from the ${}^{\natural}I_A{}^{\&}{}^{\natural}C_L$ software are described and illustrated in appendix 1 (figs. 1.1 and 1.2).

is the number of suitable storms processed for watershed i for $1 \leq i \leq 92$, and ${}^{\natural}I_{A[i]}$ and ${}^{\natural}C_{L[i]}$ are optimal values for the ith storm. The $\overline{I_A}$, $\overline{C_L}$, and w_i for each watershed are listed in appendix 1 (table 1.1).

Analysis of Watershed-Specific Initial Abstraction and Constant Loss

Statistical analyses of $\overline{I_A}$ and $\overline{C_L}$ were done with the objectives of (1) documenting the parameter distribution and (2) developing predictive procedures for each parameter. Results of the analyses, which will influence the preferred ${}^{\natural}I_A{}^{\&}C_L$ model at the conclusion of this report, are described in this section.

Each successful model run for a storm produces a Q_p^{model} (the single largest streamflow for the storm). Each Q_p^{model} can be compared to Q_p^{obs} of the observed runoff hydrograph. The comparison between Q_p^{obs} and Q_p^{model} for each modeled runoff hydrograph using ${}^{\natural}I_A{}^{\&}{}^{\natural}C_L$ values is made in figure 3. The figure contains 1,620 data points (storms). An observed or modeled runoff hydrograph can contain more than one peak (a local maximum), but only one Q_p (the global maximum for the event) is used in the analysis. Also, the Q_p^{model} might not represent the same Q_p^{obs} in terms of timing or peak number in the case of multiple peaks.[12]

The data points have no apparent tendency to plot above or below the equal value line. The standard deviation of Q_p^{model} about the equal value line is about $0.20 \log_{10}$(cubic foot per second). It is concluded by interpretation of figure 3 that ${}^{\natural}I_A{}^{\&}{}^{\natural}C_L$ coupled with $K^{\&}T_p$-GUH produces unbiased estimates of Q_p. Values for Q_p often provide critical design criteria for drainage infrastructure, thus the observation of no bias is important.

The importance and lasting ramifications of the graph in figure 3 are stressed. The graph shows that ${}^{\natural}I_A{}^{\&}{}^{\natural}C_L$ and $K^{\&}T_p$-GUH produce reliable estimates of Q_p distributed over a wide range of almost 4 orders of magnitude with no hint of nonlinearity and with constant relative error. The lack of bias is remarkable considering that ${}^{\natural}I_A{}^{\&}{}^{\natural}C_L$ values were not optimized to match Q_p^{obs}. In particular, the functional performance of the complex ${}^{\natural}I_A{}^{\&}{}^{\natural}C_L$ software is confirmed.

[12]For example, see the observed and modeled (estimated) runoff hydrograph for the multi-peak storm represented in appendix 2 (fig. 2.15). This figure shows that the text file providing the data contains no less than eight distinctively individual peaks. For the processing of the ${}^{\natural}I_A{}^{\&}{}^{\natural}C_L$ software, only the Q_p^{obs} occurring at about 80 hours was matched. Note that the figure shows representative results for context of estimation at ungaged locations described later in the report and not the results from storm-specific analysis. Therefore, the observed and modeled hydrographs in the figure do not align on the global maximum seen at about 80 hours.

Table 1. Common summary statistics of $\overline{I_A}$ and $\overline{C_L}$ (watershed-specific I_A and C_L) for the 92 watersheds.

[Initial abstraction statistics in watershed inches; constant loss statistics in watershed inches per hour]

Minimum	First quartile	Median	Mean	Third quartile	Maximum	Standard deviation
Initial abstraction ($\overline{I_A}$)						
0.0157	0.504	0.821	0.876	1.14	1.910	0.4669
Initial abstraction—undeveloped watersheds						
.336	.787	1.111	1.106	1.317	1.902	.4044
Initial abstraction—developed watersheds						
.0157	.365	.564	.690	.955	1.910	.4331
Constant loss ($\overline{C_L}$)						
.0251	.287	.517	.559	.764	2.883	.4088
Constant loss—undeveloped watersheds						
.0251	.301	.481	.617	.823	2.883	.5087
Constant loss—developed watersheds						
.0322	.274	.520	.512	.644	1.218	.3035

Both the $\overline{I_A}$ and $\overline{C_L}$ for each of the 92 watersheds are listed in appendix 1 (table 1.1). Common summary statistics for $\overline{I_A}$ and $\overline{C_L}$ are listed in table 1. A distinction between undeveloped and developed watersheds is made in each table. The statistics were computed by first subselecting the 1,620 storms into those storms for which the number of modeled peaks (peaks produced by the model) for each storm was less than or equal to six; the total number of storms after this filtering is 1,603. For the 17 rejected storms, the modeled runoff hydrograph contained too many intermediate peaks for reliable assessment.

The mean value of $\overline{I_A}$ is about 1.11 watershed inches for an undeveloped watershed and about 0.69 watershed inch for developed watersheds. The $\overline{I_A}$ for undeveloped watersheds in the study typically are about 60 percent larger than $\overline{I_A}$ for developed watersheds. The mean value of $\overline{C_L}$ is about 0.62 watershed inch per hour for an undeveloped watershed and about 0.51 watershed inch per hour for developed watersheds.

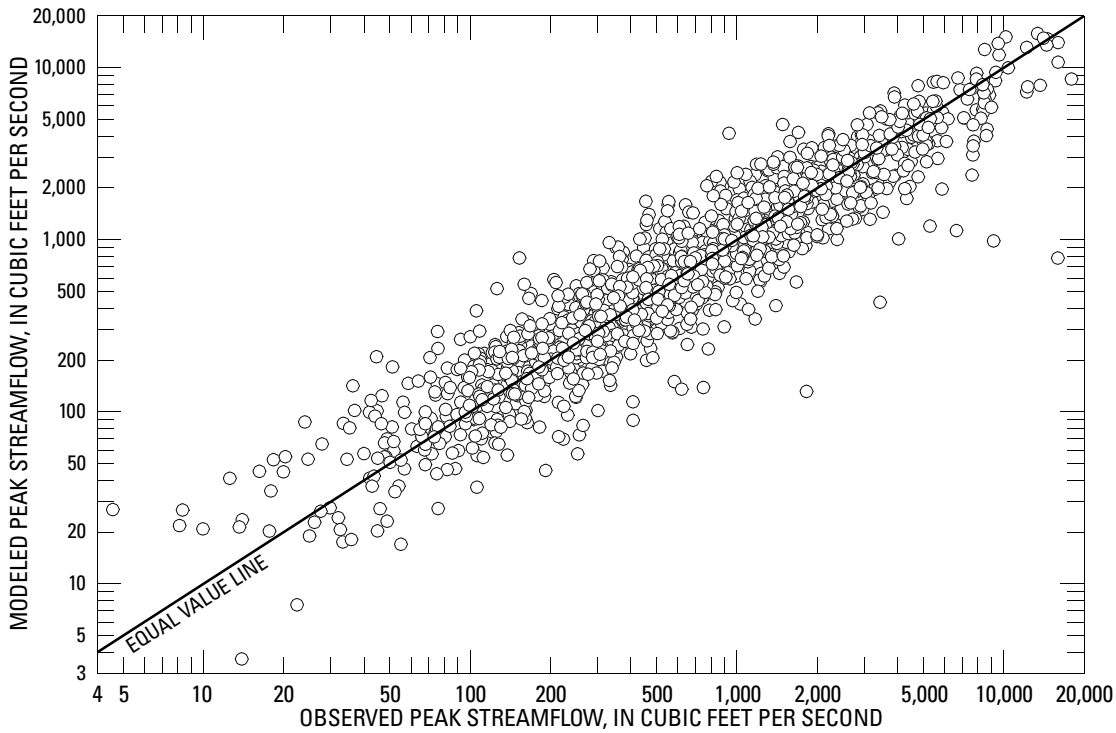


Figure 3. Graph showing comparison of observed (Q_p^{obs}) and modeled (Q_p^{model}) peak streamflows estimated by the $^{1}I_A{}^{\&}{}^{1}C_L$ and $K^{\&}T_p$-GUH models for the 92 watersheds.

The $\overline{C_L}$ for undeveloped watersheds in the study typically are about 21 percent larger than $\overline{C_L}$ for developed watersheds.

The distributions of $\overline{I_A}$ and $\overline{C_L}$ represent the variability associated with each parameter. The L-moment statistics Hosking (1990); Asquith (2007) of $\overline{I_A}$ and $\overline{C_L}$ of the two parameters are listed in table 2. A distinction between undeveloped and developed watersheds is made in the table.

The generalized lambda distribution (GLD) is a four-parameter distribution that can mimic a wide range of distributional forms (Karian and Dudewicz, 2000). The GLD was selected to model the distributions of $\overline{I_A}$ and $\overline{C_L}$ because the distribution has a simple and readily solved quantile function should the distributions be needed for end-user applications. Other than presenting two fitted GLDs in this section, further consideration of the GLD is not made in this report.

The quantile function of the GLD is

$$x(F) = \xi + \psi[F^{\kappa} - (1-F)^{\eta}], \qquad (19)$$

Table 2. L-moment statistics of $\overline{I_A}$ and $\overline{C_L}$ (watershed-specific I_A and C_L) for the 92 watersheds.

[Mean and L-scale are in inches for initial abstraction, and are in inches per hour for constant loss. --, dimensionless]

Mean (varies)	L-scale (varies)	τ_3 (--)	τ_4 (--)	τ_5 (--)
Initial abstraction ($\overline{I_A}$)				
0.875	0.267	0.0822	0.0795	0.0273
Initial abstraction—undeveloped watersheds				
1.106	.230	.0699	.108	−.0245
Initial abstraction—developed watersheds				
.690	.242	.182	.107	.0238
Constant loss ($\overline{C_L}$)				
.559	.206	.209	.188	.127
Constant loss—undeveloped watersheds				
.617	.248	.322	.225	.172
Constant loss—developed watersheds				
.512	.173	.0668	.117	.0653

where $x(F)$ is the quantile for nonexceedance probability (cumulative probability) F, ξ is a location parameter with the same units as x, ψ is a scale parameter with the same units as x, κ is a dimensionless shape parameter, and η is another dimensionless shape parameter.

The parameters of the GLD can be fit by several techniques including the method of moments or method of percentiles (Karian and Dudewicz, 2000). The parameters also can be estimated by the method of L-moments, which uses the L-moments listed in table 2. The L-moments of the GLD in the context of application of the method of L-moments are summarized by Asquith (2007).

Finally, the GLD models for distributions of $\overline{I_A}$ and $\overline{C_L}$, respectively, are described in the next two sections. The primary objectives for providing the distribution models are (1) to comprehensively document the results of the ${}^{\natural}I_A{}^{\&\natural}C_L$ software by watershed and (2) to accommodate potential usage of the two GLD models for advanced applications, such as simulation studies, that are outside of the scope of this report.

Generalized Lambda Distribution Model of Initial Abstraction

The GLD model for the distribution of $\overline{I_A}$ is

$$I_A(F) = \xi^{[I_A]} + \psi^{[I_A]}[F^{\kappa^{[I_A]}} - (1-F)^{\eta^{[I_A]}}], \quad (20)$$

where $I_A(F)$ is I_A in watershed inches for nonexceedance probability F, and the four parameters are defined the same as for equation 19. The lack of mean notation (overline) on $I_A(F)$ is by design—the watershed mean is implicit in the notation. The values for the parameters are listed in table 3. A distinction between undeveloped and developed watersheds is made in the table. The three GLD distributions of $\overline{I_A}$ that correspond to the three parameter sets listed in table 3 are depicted in figure 4. These fitted distributions are acquired by substitution of the parameters listed in the table into equation 20.

Two interpretations of the $\overline{I_A}$ distributions shown in figure 4 are made. First, the GLD provides a reliable parametric model of the distribution of the data. Second, the range or variation of $\overline{I_A}$ estimates is moderately large and slightly asymmetrical (skewed). Critically important for application, the distinction between undeveloped and developed watersheds has a large influence on I_A magnitude.

Table 3. Parameters of generalized lambda distribution for I_A for the 92 watersheds.
[--, dimensionless]

Location $\xi^{[I_A]}$ (inches)	Scale $\psi^{[I_A]}$ (inches)	Shape $\kappa^{[I_A]}$ (--)	Shape $\eta^{[I_A]}$ (--)
All 92 watersheds			
1.028	0.954	5.331	2.151
Undeveloped watersheds (41 watersheds)			
.894	1.805	.0892	.248
Developed watersheds (51 watersheds)			
1.015	.946	8.454	1.225

Specifically, $\overline{I_A}$ generally is smaller by about 0.3–0.5 watershed inch when a watershed is classified as developed. This observation is consistent with generally accepted understanding. A developed (urbanized) watershed is expected to have less capacity to store an absolute depth of rainfall than an undeveloped watershed when other factors are equal. This conclusion is reinforced by analysis in the section titled "Estimation of Initial Abstraction."

Generalized Lambda Distribution Model of Constant Loss

The GLD model for the distribution of $\overline{C_L}$ is

$$C_L(F) = \xi^{[C_L]} + \psi^{[C_L]}[F^{\kappa^{[C_L]}} - (1-F)^{\eta^{[C_L]}}], \quad (21)$$

where $C_L(F)$ is C_L in watershed inches per hour for nonexceedance probability F, and the four parameters are defined as for equation 19. As with the model for the distribution of $\overline{I_A}$, the lack of overline on $C_L(F)$ is by design because the watershed mean is implicit in the notation. The values for the parameters are listed in table 4. A distinction between undeveloped and developed watersheds is made in the table. The three GLD distributions of $\overline{C_L}$ that correspond to the three parameter sets listed in table 4 are depicted in figure 5. These fitted distributions are acquired by substitution of the parameters listed in the table into equation 21.

Two interpretations of the $\overline{C_L}$ distributions shown in figure 5 are made. First, the GLD provides a reli-

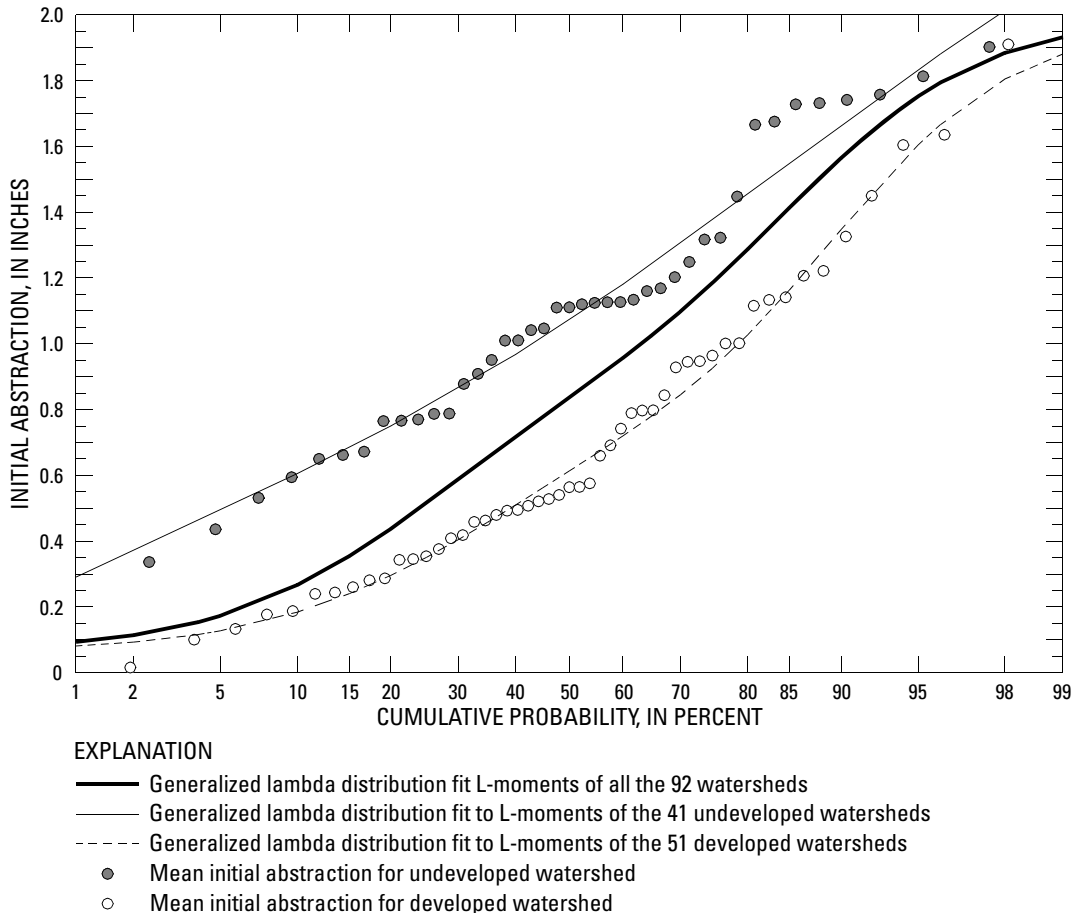


Figure 4. Graph showing distribution of $\overline{I_A}$ (watershed-specific I_A) and fitted generalized lambda distribution models for the 92 watersheds.

Table 4. Parameters of generalized lambda distribution for C_L for the 92 watersheds.
[--, dimensionless]

Location $\xi^{[C_L]}$ (inches)	Scale $\psi^{[C_L]}$ (inches)	Shape $\kappa^{[C_L]}$ (--)	Shape $\eta^{[C_L]}$ (--)
All 92 watersheds			
0.337	−4.237	−0.0207	−0.0683
Undeveloped watersheds (41 watersheds)			
.272	−2.051	−.0270	−.164
Developed watersheds (51 watersheds)			
.636	.681	7.538	2.338

able parametric model of the distribution of the data. Second, the range or variation of the $\overline{C_L}$ estimates is moderately large and slightly asymmetrical (skewed).

Potentially important for application, the distinction between undeveloped and developed watersheds has limited influence on C_L magnitude. However, for the right tail of the distributions (cumulative probabilities greater than about 60 percent), undeveloped watersheds have larger C_L. The right tail of the distribution for developed watersheds also curves to the right, which might indicate an upper limit on C_L of about 1.3 watershed inches per hour or slightly larger for the types of developed watersheds represented in the database. This observation is consistent with generally accepted understanding of urbanization of a watershed. Developed (urbanized) watersheds should generally have smaller C_L values than undeveloped watersheds when other factors are equal. However, the effect appar-

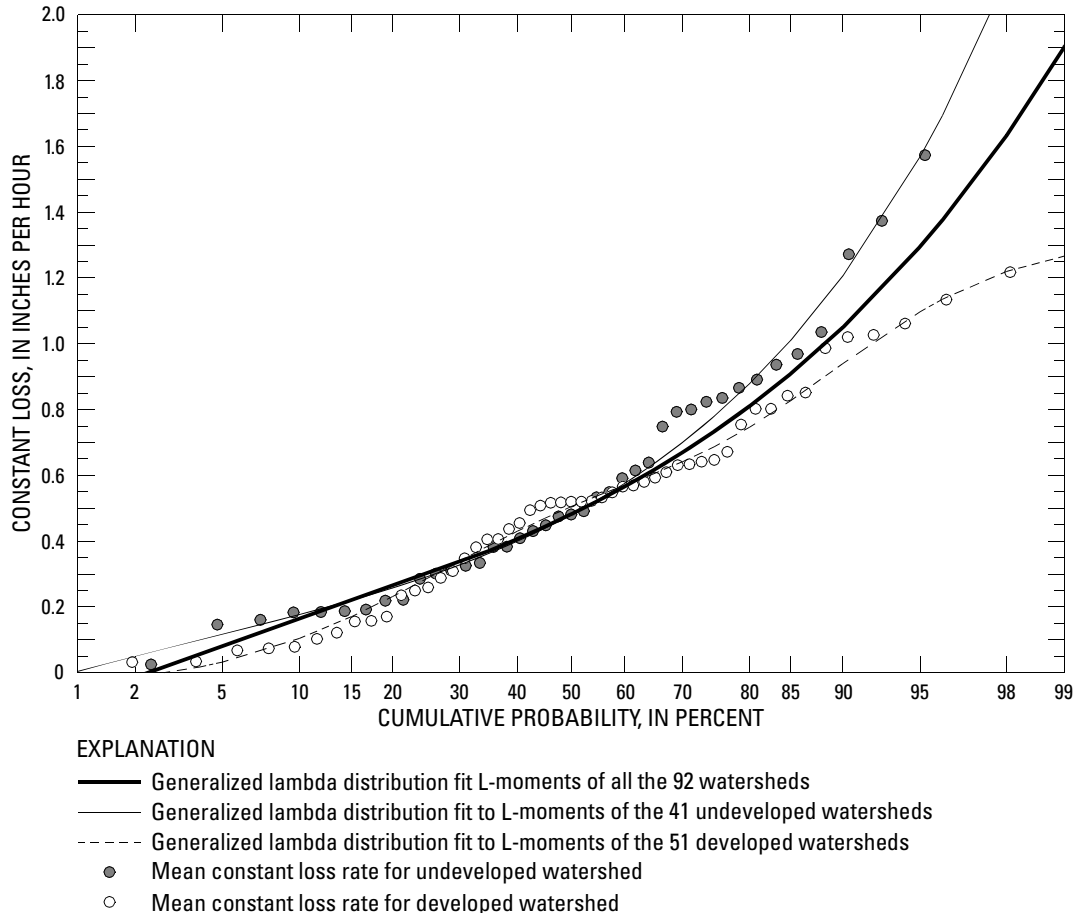


Figure 5. Graph showing distribution of $\overline{C_L}$ (watershed-specific C_L) and fitted generalized lambda distribution models for the 92 watersheds.

ently is slight. This conclusion is further discussed in the section titled "Estimation of Constant Loss."

A final comment about the GLD model for the distribution of $\overline{C_L}$ is needed because there are limitations on the fitted GLD. The $C_L(F)$ distributions defined in table 4 are negative in the far-left tail (cumulative probabilities less than about 2 percent). In applications involving these C_L distributions and predictions resulting in negative values, C_L should be truncated at zero.

Regression Equations and Regression Trees to Estimate Initial-Abstraction and Constant-Loss Values

Regression equations and regression trees to estimate I_A and C_L for applicable watersheds in Texas are developed in this section. For practical applications, estimation of I_A and C_L for ungaged watersheds is necessary. This estimation can be provided by regression between either $\overline{I_A}$ or $\overline{C_L}$ (each a regressor variable) and selected watershed characteristics (predictor variables). Two types of regression are described, and the resulting equations and trees for the I_A&C_L model are shown. As part of the regression process, two separate, but parallel, analyses of $\overline{I_A}$ and $\overline{C_L}$ were done.

Several distinct watershed characteristics were evaluated as predictor variables for estimation of $\overline{I_A}$ or $\overline{C_L}$ for ungaged watersheds. Favorable characteristics are judged in part by their statistical significance in regression computations; characteristics are considered significant at the 0.05 level (p-value[13] ≤ 0.05). The watershed characteristics for each of the 92 water-

[13]P-values are shown in appendix 2 (figs. 2.3 and 2.4), under the label `Pr(>|t|)`.

sheds were obtained from Asquith and others (2005) or determined (or computed) as necessary, as described as follows:

1. Values for contributing drainage area A for each watershed were obtained from Asquith and others (2005, table 2) and are listed in appendix 1 (column 6 of table 1.1). Preliminary analysis showed that A has predictive properties on I_A and C_L similar to those of main-channel length L. One reason for the similarity is the statistically significant relation between L and A (see appendix 1, eq. 1.1 and fig. 1.3). The authors decided not to use A for final regression computations in part because a dimensionless quantity (I_A/L) can be formulated to directly express proportionality, or the time quantity (L/C_L) can be formulated to express a duration. However, these two ratios are not discussed further in this report.

2. Values for main-channel length L were obtained from Asquith and others (2005, table 2) and are listed in appendix 1 (column 7 of table 1.1). The L is defined as the length in stream-course miles of the longest defined channel shown in a 30-meter digital elevation model from the approximate watershed headwaters to the outlet. L is statistically significant for estimation of $\overline{I_A}$ and $\overline{C_L}$ and is used in the final regression computations.

3. Values for dimensionless main-channel slope S were obtained from Asquith and others (2005, table 2) and are listed in appendix 1 (column 8 of table 1.1). The S is defined as the change in elevation in feet (ΔE) between the two end points of L divided by L in feet: $S = \Delta E/(5{,}280 \times L)$. A 30-meter digital elevation model was used to compute S for this report. Preliminary analysis showed that S has statistically insignificant predictive capabilities after the stronger prediction properties of L are accounted for by regression. The authors decided not to use S for final regression computations.

4. Natural Resources Conservation Service (2006) curve numbers (CN) were estimated. The CNs for each of the 92 watersheds are listed in appendix 1 (column 5 of table 1.1). The table lists CN to the nearest tenth, which reflects the value used in the analyses. However, the authors acknowledge that for a typical watershed, CN likely has measurable resolution no better than to the nearest integer. The CN is a parameter used in the Curve Number (CN) method to estimate the maximum potential retention of rainfall in a watershed (Natural Resources Conservation Service, 2006) and reportedly accounts for differences between soil types, land-cover classifications, and other hydrologic conditions of the land surface that affect watershed storage of rainfall. The CN method generally represents a non-time-distributed, watershed-loss model. The CN method likely is one of the most common rainfall and runoff models for typical hydrologic-engineering applications. Because CN is well known, represents an overall capacity of a watershed to absorb rainfall, and is statistically significant for estimation of $\overline{I_A}$ and $\overline{C_L}$, CN is used for final regression computations.

5. Soil types and textures were obtained from U.S. Geological Survey (2006). Exploratory analysis of the relations between $\overline{I_A}$ and $\overline{C_L}$ and soil types and textures indicated that various measures (sand, silt, clay, loam, hydraulic conductivity, and others) were inferior to the CN in terms of predictive capabilities. The authors decided not to use soil types and textures for final regression computations. The soil data are not listed in this report, and further details are not provided.

6. Values of percentage impervious cover were obtained[h] from U.S. Geological Survey (2006). In a suburban to urban area, impervious cover includes roads, parking lots, and rooftops. Exploratory analysis of the relations between $\overline{I_A}$ and $\overline{C_L}$ indicates that percentage impervious cover was inferior to the combined predictive capabilities of CN and D. The authors acknowledge that differences between the time period of the percentage impervious cover data and the general period of data for the 92 watersheds exist.

The authors decided not to use percentage impervious cover for final regression computations. The values for percentage impervious cover are not listed in this report, and further details are not provided.

7. The binary classification of D was considered. Values for D were obtained from Asquith and others (2005, table 1) and are listed in appendix 1 (column 4 of table 1.1). The D factor (Asquith and others, 2005, table 1) is a state variable representing a binary classification of the state of development in a watershed (undeveloped and developed). The classification scheme parallels and accommodates the disparate discussion and conceptualization in more than 220 reports that provided the original data for the rainfall and runoff database (Asquith, Thompson, and others, 2004). The classification scheme also is used in this report. The authors elected to use D, as it represents a highly generalized treatment of complex and potentially temporally changing watershed development. D is statistically significant for estimation of $\overline{I_A}$ and $\overline{C_L}$ and is used in final regression computations.

8. Another binary classification considered was whether the watershed is characteristized by rock-dominated terrain and thin soils, rock channels, and karstic features,[i] such as the limestone in and around the Austin and San Antonio areas. The rock-classification factor (R) was obtained through local knowledge and geologic and soil data for the watersheds. The R for each of the 92 watersheds is listed in appendix 1 (column 3 of table 1.1). R is statistically significant for estimation of $\overline{I_A}$ and $\overline{C_L}$ and is used in final regression computations.

In summary, of the eight candidate watershed characteristics, the four characteristics of L, D, R, and CN were used in the final stages of linear regression (equation) analysis. Further, these same characteristics were used for regression tree analysis.

Regression trees (Faraway, 2006, chapter 13, "Trees") result from an alternative method of regression (sometimes termed recursive partitioning) when compared to regression that produces equations. The tree is constructed such that partitions (branches) are determined by an algorithm that seeks to split and minimize residual sum of squares. The R environment (R Development Core Team, 2006) coupled with the `rpart` package (Therneau and Atkinson, 2006) was used for regression tree construction. Modest "pruning" of the limbs was made to mitigate for over-partitioning to small sample sizes (small number of watersheds per terminal branch). Specifically, a weighted partition was used in which the number of suitable storms provided the basis for the weights. For the regression trees presented in this report, the vertical extent (length) of the branches is proportional to the amount of variability explained by each partition.

The regression equation computations were done using the R environment (R Development Core Team, 2006). The general form of the $I_A{}^{\&}C_L$ equations reported here is

$$Y = a + bL^{\varphi} + cD + dR + eCN, \tag{22}$$

where Y is the predictor variable ($\overline{I_A}$ or $\overline{C_L}$), and the five coefficients (a, b, c, d, and e) are estimated by weighted-least-squares regression in which the weights were chosen on the basis of the number of suitable storms listed in appendix 1 (table 1.1) for each watershed. The power transformation factor (φ) is determined independently from the least-squares procedure by a process referred to as "PRESS minimization." The process entails the minimization of the PRESS statistic of a candidate regression by iterative refinement of φ values.

The PRESS statistic is a validation statistic that measures the performance of the equation in predicting new values. Small PRESS values are desirable. Helsel and Hirsch (2002, p. 247) state that, "Minimizing PRESS means that the equation produces the least error when making new predictions." An example of PRESS minimization in an annual peak streamflow context with source code is available in Asquith and Thompson (2005).

An algorithm for PRESS minimization is shown in figure 6 using the I_A equation for illustration. The code listing assumes that I_A (`IA`), L (`L`), D (`DU`), R (`RK`),

```
PRESS <- function(model) {
  res <- weighted.residuals(model)
  hats <- hatvalues(model) # "leverages"
  return( sum((res/(1-hats))^2 ) /
              length(hats)) # PRESS
}
makeIAeq <- function(phi.b, phi.e,
                     phi.stp) {
  minpress <- 10000; minphi <- 10000
  minwls <- 0 # the model of min PRESS
  for(phi in seq(phi.b, phi.e,
                 by=phi.stp)) {
    if(phi == 0) next # skip power=0
    my.L <- L^phi # transform
    aWLS <- lm(IA~my.L+DU+RK+CN,
               weights=W) # fit a linear ↩
                  ↪model, see eq. 22
    press <- PRESS(aWLS.OUT) # compute it
    if(press < minpress) {
      minpress <- press; minphi <- phi
      minwls <- aWLS # a better model
    }
  }
  return(list(PHI=minphi, PRESS=minpress,
              EQUATION=aWLS))
}
```

Figure 6. Listing of R code that represents the algorithm used for PRESS minimization in a weighted-least-squares context.

CN (CN), and the regression weights *w* (W), already are available. The values phi.b, phi.e, and phi.stp represent the starting, ending, and step interval for candidate values of φ.

Intermediate results of the exploratory transformation research are not reported here. However, these results indicated that one watershed (08158825 Little Bear Creek at Farm to Market 1826, Manchaca, Tex.) should be dropped, in part, from further analysis because only one storm was available. This watershed is not used for the final regression equations and trees reported here. Therefore, it is implicit in subsequent discussion that this watershed is not used; although for consistency, the text will continue to refer to 92 (not 91) watersheds.[14] Finally, for the C_L regression, D was not used because watershed development was not statistically significant. This conclusion is reinforced by figure 5—for the 92 watersheds, watershed development apparently has no substantial influence on $\overline{C_L}$.

The final results of the PRESS minimization show that L^{φ} transformation provides marginally better performance (smaller PRESS) than a more-common $\log_{10}(L)$ transformation. The φ coefficients are each labeled as $\varphi^{[I_A]}$ and $\varphi^{[C_L]}$ for the respective I_A and C_L equations. The results are

$$I_A : \log_{10}(L) \text{ PRESS} = 0.1052 \qquad \varphi^{[I_A]} \text{ PRESS} = 0.1036$$
$$C_L : \log_{10}(L) \text{ PRESS} = 0.0744 \qquad \varphi^{[C_L]} \text{ PRESS} = 0.0743$$

Additional evaluation of model preference is made with the Akaike Information Criterion (AIC) in which smaller values for I_A and C_L indicate a preferable model:

$$I_A : \log_{10}(L) \text{ AIC} = 74.81 \qquad \varphi^{[I_A]} \text{ AIC} = 73.18$$
$$C_L : \log_{10}(L) \text{ AIC} = 48.17 \qquad \varphi^{[C_L]} \text{ AIC} = 48.12$$

The PRESS and AIC results indicate that the power transformation on L has a larger benefit for I_A estimation because the PRESS and AIC statistics show more reduction relative to any potential benefit for C_L. This conclusion was confirmed by visual comparison of the residual plots (fitted values on horizontal axis and residual on vertical axis). Some curvature existed in the $I_A : \log_{10}(L)$ residual plots (not shown in this report), and this curvature was removed by the $I_A : \varphi^{[I_A]}$ transformation (residual plots are shown in appendix 2, figs. 2.6 and 2.7).

Estimation of Initial Abstraction

The regression equation[15] for estimation of I_A has $\varphi^{[I_A]} = -0.9041$ and is

$$I_A = 2.045 - 0.5497L^{-0.9041} - 0.1943D + 0.2414R - 0.01354CN, \tag{23}$$

where I_A is initial abstraction in watershed inches, L is main-channel length of the watershed in miles, $D = 0$ for undeveloped watersheds and $D = 1$ for developed watersheds, $R = 0$ for non-rocky watersheds and $R = 1$ for rocky watersheds, and CN is the curve number.

[14]Exceptions to this practice are the residual plots in appendix 2, figures 2.6 and 2.7 (regression equations) and figures 2.13 and 2.14 (regression trees) in which the residual for station 08158825 is computed and shown.

[15]The equation and other ancillary information about the I_A equation are shown in appendix 2 (fig. 2.3).

The equation has 86 degrees of freedom ($\mathrm{df}^{[I_A]} = 86$), a residual standard error ($\sigma^{[I_A]}$) of 0.3025 watershed inch, and an adjusted R-squared of 0.345. The maximum leverage ($h_{\max}^{[I_A]}$) is about 0.272, which can be useful in assessing applicability of equation 23.

The prediction limits of I_A from equation 23 can be useful for expressing uncertainty when the equation is used. The limits for the $100 \times (1-\alpha)$ prediction interval are shown as

$$^{\downarrow}I_{A\oplus}^{\alpha/2} \leq I_{A\oplus} \leq {}^{\uparrow}I_{A\oplus}^{\alpha/2}, \tag{24}$$

where down (up) arrow signifies the lower (upper) prediction limit for I_A at the α significance level.

The lower and upper limits are computed by

$$^{\downarrow}I_{A\oplus}^{\alpha/2} = I_{A\oplus} - \mathrm{t}_{[\alpha/2,\mathrm{df}]}\, \sigma^{[I_A]} \sqrt{1 + h_{\oplus}^{[I_A]}} \text{ and} \tag{25}$$

$$^{\uparrow}I_{A\oplus}^{\alpha/2} = I_{A\oplus} + \mathrm{t}_{[\alpha/2,\mathrm{df}]}\, \sigma^{[I_A]} \sqrt{1 + h_{\oplus}^{[I_A]}}, \tag{26}$$

where $I_{A\oplus}$ is the prediction from equation 23, $\mathrm{t}_{[\alpha/2,\mathrm{df}]}$ is the upper tail of the t-distribution for $\mathrm{df}^{[I_A]}$ degrees of freedom at the α significance level, $\sigma^{[I_A]}$ is the residual standard error of equation 23, and $h_{\oplus}^{[I_A]}$ is the leverage of the prediction for the watershed.

The leverage $h_{\oplus}^{[I_A]}$ for an $I_{A\oplus}$ prediction for a watershed from equation 23 is important for estimation of the prediction limits of the equation. The computation of $h_{\oplus}^{[I_A]}$ is

$$h_{\oplus}^{[I_A]} = \begin{bmatrix} 1 & L_{\oplus}^{-0.9041} & D_{\oplus} & R_{\oplus} & CN_{\oplus} \end{bmatrix} \times \begin{bmatrix} 2.38000 & 0.17571 & 0.21256 & -0.04512 & -0.03106 \\ 0.17571 & 0.40497 & 0.00579 & -0.01074 & -0.00348 \\ 0.21256 & 0.00579 & 0.06893 & -0.01141 & -0.00304 \\ -0.04512 & -0.01074 & -0.01141 & 0.04742 & 0.00042 \\ -0.03106 & -0.00348 & -0.00304 & 0.00042 & 0.00041 \end{bmatrix} \times \begin{bmatrix} 1 \\ L_{\oplus}^{-0.9041} \\ D_{\oplus} \\ R_{\oplus} \\ CN_{\oplus} \end{bmatrix}, \tag{27}$$

where $L_{\oplus}$, $D_{\oplus}$, $R_{\oplus}$, and $CN_{\oplus}$ are watershed-specific values, and the 5×5 matrix is the inverted covariance matrix of the regression. The prediction limits (interval) of $I_{A\oplus}$ will increase (expand) as the watershed characteristics become farther away from those used (the parameter space) to develop equation 23. Equation applicability can be roughly assessed by the maximum leverage of the regression. In general, a regression equation is judged applicable for a given watershed if the leverage for a prediction is about

$$h_{\oplus}^{[I_A]} \leq h_{\max}^{[I_A]} = 0.272. \tag{28}$$

However, this rule of thumb might be insufficient in the context here, and more discussion is provided in the section titled "General Discussion of the I_A&C_L Equations."

The I_A regression tree (I_A tree) for estimation of I_A is shown in figure 7. Each value for I_A is shown in bold type, and the number of samples n and residual standard error σ also are listed. For example, a hypothetical watershed with $CN = 83$ and $L = 4$ miles has an I_A of 0.75 ± 0.0830 watershed inch. The database has 12 watersheds meeting the criteria of the hypothetical watershed. Unlike for the I_A equation, D apparently does not have substantial predictive properties for I_A in a regression-tree context. Therefore, conclusions based on which parameters are important can be influenced by model structure; the I_A equation and I_A tree are structurally distinct.

With respect to results of the I_A tree, it might appear counterintuitive that, for the right branch of the first test ($CN > 80$), watersheds with $R = 1$ have larger I_A ($I_A = 1.17$ watershed inches) than watersheds with $R = 0$ ($I_A = 0.924$ watershed inch). The authors hypothesize that the nature of the limestone (karst) watersheds of the Austin and San Antonio areas is the cause. These watersheds during periods lacking insufficient or abundant rainfall have substantial capacity to store an absolute depth of rainfall early in a storm event. As previously described, the assignment of the R value is specific to rock-dominated watersheds in the Central Texas area; extrapolation of the rock classification to other "rocky" watersheds, such as watersheds dominated by outcrops of granite, might not be appropriate.

Finally, the residuals of the I_A tree are shown in appendix 2 (fig. 2.13). The residuals are evenly distributed, although with only six unique fitted values, to the residuals for the I_A equation in appendix 2 (fig. 2.6).

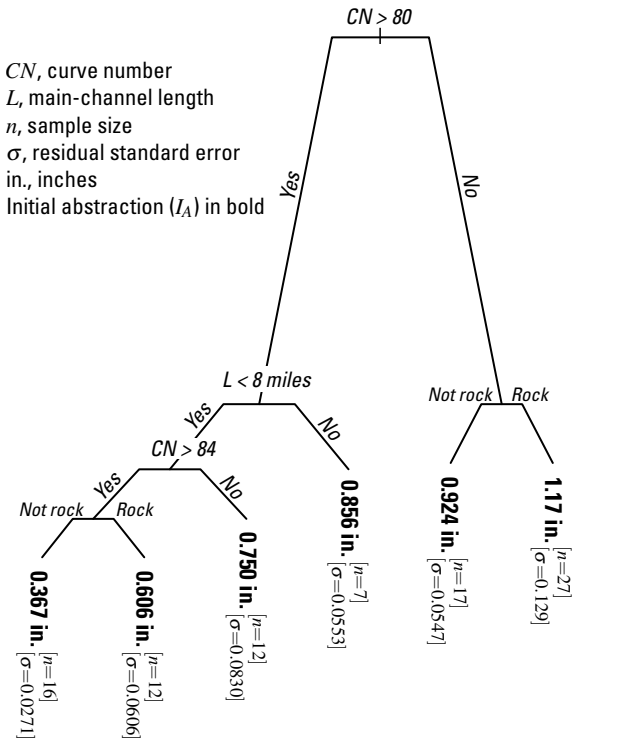


Figure 7. Diagram showing regression tree of $\overline{I_A}$ (the I_A tree) for estimation of I_A.

Estimation of Constant Loss

The regression equation[16] for estimation of C_L has $\varphi^{[C_L]} = 0.2312$ and is

$$C_L = 2.535 - 0.4820L^{0.2312} + 0.2271R - 0.01676CN, \tag{29}$$

where C_L is constant loss in watershed inches per hour, L is main-channel length of the watershed in miles, $R = 0$ for non-rocky watersheds and $R = 1$ for rocky watersheds, and CN is curve number. The equation has 87 degrees of freedom ($\mathrm{df}^{[C_L]} = 87$), a residual standard error ($\sigma^{[C_L]}$) of 0.2649 watershed inch per hour, and an adjusted R-squared of 0.301. The maximum leverage ($h_{\max}^{[C_L]}$) is about 0.183, which can be useful in assessing applicability of equation 29.

The prediction limits of C_L from equation 29 can be useful for expressing uncertainty when the equation is used. The limits for the $100 \times (1-\alpha)$ prediction interval are shown as

$$^{\downarrow}C_{L\oplus}^{\alpha/2} \le C_{L\oplus} \le {}^{\uparrow}C_{L\oplus}^{\alpha/2}, \tag{30}$$

where down (up) arrow signifies the lower (upper) prediction limit for I_A at the α significance level.

The lower and upper limits are computed by

$$^{\downarrow}C_{L\oplus}^{\alpha/2} = C_{L\oplus} - \mathrm{t}_{[\alpha/2,\mathrm{df}]}\,\sigma^{[C_L]}\sqrt{1+h_{\oplus}^{[C_L]}} \text{ and} \tag{31}$$

$$^{\uparrow}C_{L\oplus}^{\alpha/2} = C_{L\oplus} + \mathrm{t}_{[\alpha/2,\mathrm{df}]}\,\sigma^{[C_L]}\sqrt{1+h_{\oplus}^{[C_L]}}, \tag{32}$$

where $C_{L\oplus}$ is the prediction from equation 29, $\mathrm{t}_{[\alpha/2,\mathrm{df}]}$ is the upper tail of the t-distribution for $\mathrm{df}^{[C_L]}$ degrees of freedom at the α significance level, $\sigma^{[C_L]}$ is the residual standard error of equation 29, and $h_{\oplus}^{[C_L]}$ is the leverage of the prediction for the watershed.

The leverage $h_{\oplus}^{[C_L]}$ for a $C_{L\oplus}$ prediction for a watershed from equation 29 is important for estimation of the prediction limits of the equation. The computation of $h_{\oplus}^{[C_L]}$ is

$$h_{\oplus}^{[C_L]} = \begin{bmatrix} 1 & L_{\oplus}^{0.2312} & R_{\oplus} & CN_{\oplus} \end{bmatrix} \times \begin{bmatrix} 2.79336 & -0.45089 & -0.01554 & -0.02600 \\ -0.45089 & 0.17984 & 0.00375 & 0.00222 \\ -0.01554 & 0.00375 & 0.04538 & -0.00011 \\ -0.02600 & 0.00222 & -0.00011 & 0.00028 \end{bmatrix} \times \begin{bmatrix} 1 \\ L_{\oplus}^{0.2312} \\ R_{\oplus} \\ CN_{\oplus} \end{bmatrix}, \tag{33}$$

where $L_{\oplus}$, $R_{\oplus}$, and $CN_{\oplus}$ are the watershed-specific values, and the 4×4 matrix is the inverted covariance matrix of the regression. The prediction limits (interval) of $C_{L\oplus}$ will increase (expand) as the watershed characteristics become farther away from those used (the parameter space) to develop equation 29. Equation applicability can be roughly assessed by the maximum leverage of the regression. In general, a regression equation is judged applicable for a given watershed if the leverage for a prediction is about

$$h_{\oplus}^{[C_L]} \le h_{\max}^{[C_L]} = 0.183. \tag{34}$$

[16]The equation and other ancillary information about the C_L equation are shown in appendix 2 (fig. 2.4).

However, this criteria might be insufficient in the context here, and more discussion is provided in the next section.

The C_L regression tree (C_L tree) for estimation of C_L is shown in figure 8. Each value for C_L is shown in bold type, and the number of samples n and residual standard errors σ also are listed. For example, a hypothetical watershed with $CN = 63$ and $L = 4$ miles has a C_L of 0.862 ± 0.101 watershed inch per hour. The database has 12 watersheds meeting the criteria of the hypothetical watershed. As shown for the I_A tree (fig. 7), D does not have substantial predictive properties for the C_L tree, which is consistent with the C_L equation.

The residuals of the C_L tree are shown in appendix 2 (fig. 2.14). The residuals are evenly distributed, although with only five unique fitted values, to the residuals for the C_L equation in appendix 2 (fig. 2.7).

Visual comparison of the I_A and C_L trees is informative. The vertical length of the branches for the C_L tree (fig. 8) are more similar than those for the I_A tree (fig. 7). In particular, the initial branch for the I_A tree ($CN > 80$) is much longer than the other branches. Thus, CN has an appreciably longer influence on I_A than subsequent branches of the I_A tree.

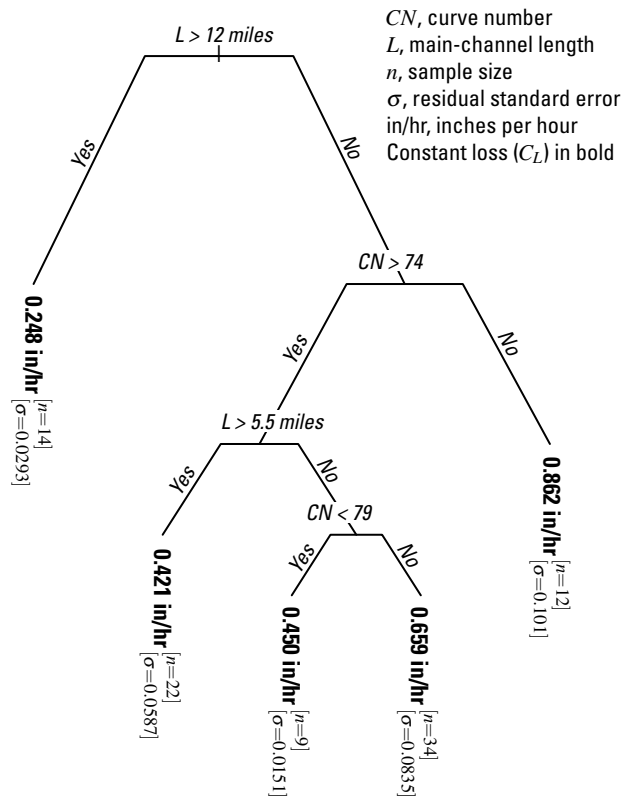


Figure 8. Diagram showing regression tree of $\overline{C_L}$ (the C_L tree) for estimation of C_L.

General Discussion of the Regression Equations

Discussion of the $I_A{}^{\&}C_L$ equations (eqs. 23 and 29) in the context of physical interpretations is useful for reliable implementation of the results reported here. The $\varphi^{[I_A]}$ and $\varphi^{[C_L]}$ coefficients were each derived by PRESS minimization. The authors are uncertain whether φ is intrinsically important or whether the fact that $\varphi^{[I_A]}$ and $\varphi^{[C_L]}$ differ in magnitude and sign is important for further interpretation. In particular for the I_A equation, the $\varphi^{[I_A]}$ transformation removes apparent curvature in the regression residuals relative to a more common $\log_{10}$ transformation.

For the I_A equation, the coefficient on D is -0.1943 watershed inch, which means that developed watersheds generally have about 1/5-inch less initial rainfall storage than undeveloped watersheds.

The coefficients on R for I_A and C_L are $+0.2414$ and $+0.2271$ watershed inch and watershed inch per hour, respectively. The positive signs mean that rock-dominated, thin-soiled watersheds tend to have larger rainfall losses. This observation appears logically consistent with many watersheds in the Austin and San Antonio areas. In general, rock-dominated, thin-soiled, karst watersheds represented by $R = 1$ in the database have about 1/4-inch larger I_A or 1/4-inch per hour larger C_L than other watersheds.

The coefficients on CN for I_A and C_L are -0.01354 and -0.01676 watershed inch and watershed inch per hour, respectively. The negative signs show that rainfall losses decrease as CN increases. This is consistent with the broadly understood definition of CN. In general, an increase of 10 units of CN represents about $-1/7$-inch of I_A and represents about $-1/6$-inch per hour of C_L.

Description and interpretation of the influence of L on I_A and C_L is more complex than for the other three regressor variables. Further, L values are less constrained than values for the other characteristics, which can result in physically inconsistent predictions from the $I_A{}^{\&}C_L$ equations. Specifically, mathemati-

cal analysis of the I_A equation for a "worst-case scenario" ($D = 1$, $R = 0$, and $CN = 100$) shows that $L \gtrapprox 1.12$ miles to maintain $I_A \geq 0$ (physical consistency). Further mathematical analysis of the C_L equation for another worst-case scenario ($R = 0$ and $CN = 100$) shows that $L \lessapprox 12.5$ miles to maintain $C_L \geq 0$. Therefore, the most restrictive limits on L are about $1.12 \lessapprox L \lessapprox 12.5$; however, less restrictive limits of L exist for other combinations of D, R, and CN.

Some users might want to compare the L restrictions to a typical value of A for a given watershed, given L. Equation 1.1 (appendix 1) can be used to estimate a typical value of A, which is denoted as A^{pred}. The approximate A range of the most restrictive L values is $0.44 \lessapprox A^{\text{pred}} \lessapprox 24$ square miles. This length and area information, used in conjunction with the $h_{\oplus}^{[I_A]}$ and $h_{\oplus}^{[C_L]}$ values, can be used to assess equation applicability for a given watershed. In general, the $I_A{}^{\&}C_L$ equations likely are not applicable, or at least have questionable applicability, for very small watersheds ($A \lessapprox 0.4$) or for watersheds greater than about several tens of square miles, unless other combinations of L and remaining parameters are used.

Now that generalized limits of L are established, interpretation of effects of L on I_A and C_L can be made. For the I_A equation, $\varphi^{[I_A]}$ is negative so as L increases, its influence on the prediction of I_A decreases. The coefficient on L is negative, therefore, I_A increases as L (or watershed size) increases. The increase of I_A approaches a limit as L becomes large. Conversely, for the C_L equation, $\varphi^{[C_L]}$ is positive so as L increases, its influence on the prediction of C_L increases. The coefficient on L is negative, therefore, C_L decreases as L increases. Although it seems logical that larger watersheds should have larger I_A, it is not clear what effect watershed size, as expressed by L (and A; see eq. 1.1), should have on C_L.

Mathematical analysis of the approximate global limits for the $I_A{}^{\&}C_L$ equations is informative. The most restrictive limits on L ($1.12 \lessapprox L \lessapprox 12.5$ miles) provide a convenient starting point. CN in the range of $10 \leq CN \leq 100$ are considered in this analysis. For this analysis the following combinations of watershed characteristics were used to compute smallest and highest loss-parameter values: $L = 1.12|12.5$, $D = 0|1$, $R = 0|1$, and $CN = 10|100$, where the $|$ symbol represents the distinction between watershed development ($D = 0 \mid D = 1$). The analysis shows that the approximate global limits of I_A are $0 \lessapprox I_A \lessapprox 2$ watershed inches, and the approximate limits for C_L are $0 \lessapprox C_L \lessapprox 1.9$ watershed inches per hour. For both I_A and C_L equations these limits are consistent with the general limits of the actual data shown in figures 4 and 5.

Comparison of the Initial Abstraction Equation to the Curve Number Method

The CN method involves a computation of initial abstraction. The equation is

$$I_A^{[CN]} = 0.2 S_{\max}, \tag{35}$$

where $I_A^{[CN]}$ is the I_A of the CN method, $S_{\max}$ is known as maximum potential retention, and 0.2 represents an empirical coefficient known as the "best approximation from observed data" (Pilgrim and Cordery, 1993, p. 9.22). The relation between $S_{\max}$ and CN is

$$S_{\max} = \frac{1000}{CN} - 10. \tag{36}$$

The CN method is widely used by hydrologic engineers and others. Therefore, comparison of the I_A predicted from the CN method to those predicted from a representative solution (representative watershed) of the I_A equation likely is of interest to practitioners. Listed in table 5 are I_A values predicted by each method. The $I_A^{[CN]}$ ranges from 0 to 18 watershed inches; whereas, the I_A equation yields a much smaller range from 0.56 to 1.8 watershed inches. These ranges are distinctively different. The range of CN for the 92 watersheds used for the I_A equation is between about 65 and 90; therefore, considerable CN space is not represented by the watersheds.

By this relative comparison listed in table 5, the I_A equation and, by association, the I_A tree do not produce very large I_A values for small CN—larger V^{model} values and larger Q_p^{model} values are expected when either the I_A equation or I_A tree is used. It is not known how reliable the I_A equation or, by association, the I_A tree is for watersheds with small CN. However, the authors argue that the I_A estimation methods described in this report are expected to provide reliable estimates of

Table 5. Comparison of I_A as predicted by the curve number (CN) method and the I_A equation.

[For this table and the I_A equation (eq. 23) and the curve number (CN, $I_A^{[CN]}$) equation (eq. 35), the authors have assumed $L = 5$ miles, $R = 0$, and $D = 0$. --, dimensionless]

CN (--)	$I_A^{[CN]}$ (inches)	I_A equation (inches)	CN (--)	$I_A^{[CN]}$ (inches)	I_A equation (inches)
10	18	1.8	60	1.3	1.1
20	8.0	1.7	70	.86	.97
30	4.7	1.5	80	.50	.83
40	3.0	1.4	90	.22	.70
50	2.0	1.2	100	0	.56

I_A (as a concept) because the methods were explicitly developed to predict I_A. It is unknown how reliable the $0.2S_{\max}$ formulation implicit in the CN method is for the 92 watersheds.

Table 6. Common summary statistics of peak streamflow error (ε^{Q_p}) for the 92 watersheds.

[All values are in $\log_{10}$(cubic feet per second).]

Minimum	First quartile	Median	Mean	Third quartile	Maximum	Standard deviation
$\overline{I_A{}^{\&}C_L}$ model (1,276 storms)						
−2.31	−0.204	0.047	0.100	0.340	3.06	0.536
$\check{I}_A{}^{\&}\check{C}_L$ model (1,319 storms)						
−2.38	−.249	.007	.049	.287	5.04	.546
$I_A{}^{\&}C_L$ equations (1,342 storms)						
−2.28	−.187	.045	.093	.305	4.09	.533
$I_A{}^{\&}C_L$ trees (1,350 storms)						
−2.39	−.203	.049	.079	.311	2.35	.494
Combined $I_A{}^{\&}C_L$ model (1,232 storms)						
−2.34	−.226	.013	.001	.253	1.68	.408

An Initial-Abstraction, Constant-Loss Model for Unit Hydrograph Modeling for Applicable Watersheds in Texas

An $I_A{}^{\&}C_L$ model for watershed-loss estimation for applicable watersheds in Texas is evaluated and discussed in the context of practical applications in this section. The $I_A{}^{\&}C_L$ equations (eqs. 23 and 29) are shown with statistically significant variables, but those variables only explain about 30–34 percent of the variation in the $\overline{I_A}$ and $\overline{C_L}$ values. The $I_A{}^{\&}C_L$ trees explain a similar percentage. Therefore, the mean or median of $\overline{I_A}$ and $\overline{C_L}$ (table 1), with respect to D, also could be reasonable estimates for a watershed. Thus, four techniques are identified for general estimation for ungaged watersheds. The four techniques are abbreviated as

$\overline{I_A{}^{\&}C_L}$ model (means) $\quad \begin{matrix}\overline{I_A} = 1.106 \mid 0.690 \\ \overline{C_L} = 0.617 \mid 0.512\end{matrix}$,

$\check{I}_A{}^{\&}\check{C}_L$ model (medians) $\quad \begin{matrix}\check{I}_A = 1.111 \mid 0.564 \\ \check{C}_L = 0.481 \mid 0.520\end{matrix}$,

$I_A{}^{\&}C_L$ equations (eqs. 23 and 29), and

$I_A{}^{\&}C_L$ trees (figs. 7 and 8),

where the | symbol represents the distinction between watershed development ($D = 0 \mid D = 1$). The units of I_A and C_L are watershed inches and watershed inches per hour, respectively.

As part of evaluation of the four techniques, three types of error are defined. The error between the observed and modeled Q_p is defined as

$$\varepsilon^{Q_p} = \log_{10}(Q_p^{\text{obs}}) - \log_{10}(Q_p^{\text{model}}), \tag{37}$$

where ε^{Q_p} is the peak streamflow error in $\log_{10}$(cubic feet per second), Q_p^{obs} is the observed peak streamflow, and Q_p^{model} is the peak streamflow from a context-specific model. The error for runoff volume (ε^V) in watershed inches and the time difference of peak streamflow ($\varepsilon^{T_{Q_p}}$) in hours are similarly defined, but $\log_{10}$ transformations are not used.

The ${}^{\natural}I_A{}^{\&}{}^{\natural}C_L$ software was modified to implement the four techniques with the primary purpose of evaluating the $I_A{}^{\&}C_L$ equations and $I_A{}^{\&}C_L$ trees, but the other two techniques (the $\overline{I_A{}^{\&}C_L}$ and $\check{I}_A{}^{\&}\check{C}_L$ models) also were included. As a result, all storms in the database for the 92 watersheds were reprocessed four separate times.[17] The ε^{Q_p}, ε^V, and $\varepsilon^{T_{Q_p}}$ values for each storm were computed. Common summary statistics of ε^{Q_p}, ε^V, and $\varepsilon^{T_{Q_p}}$, respectively, are listed in tables 6–8.

[17] A storm-specific processing example using the $I_A{}^{\&}C_L$ equations is shown in appendix 2 (fig. 2.15).

Table 7. Common summary statistics of runoff hydrograph volume error (ε^V) for the 92 watersheds.

[All values are in watershed inches.]

Minimum	First quartile	Median	Mean	Third quartile	Maximum	Standard deviation
$\overline{I_A{}^{\&}C_L}$ model (1,276 storms)						
−2.77	−0.156	0.148	0.293	0.525	8.00	0.882
$\check{I}_A{}^{\&}\check{C}_L$ model (1,319 storms)						
−2.83	−.252	.090	.214	.465	7.10	.863
$I_A{}^{\&}C_L$ equations (1,342 storms)						
−2.27	−.158	.135	.253	.538	9.17	.874
$I_A{}^{\&}C_L$ trees (1,350 storms)						
−2.33	−.167	.109	.243	.526	9.27	.850
Combined $I_A{}^{\&}C_L$ model (1,232 storms)						
−2.55	−.210	.100	.237	.489	8.38	.864

Specifically, the values are listed in the first four entry lines of the three tables. (The last entry is the topic of the section titled "The Combined $I_A{}^{\&}C_L$ Model.") Storms with no modeled runoff, which occasionally occurred because the estimated losses were greater than observed rainfall, were not used for the statistical summary. The reported numbers of storms in the three tables thus are less than 1,620 and also are not equal among the four techniques.

To augment presentation of the statistics listed in tables 6–8, selected statistics are graphically depicted in figure 9. The statistics in tables 6–8 require further discussion. The minimum and maximum errors are large; however, all storms processed in the statistical computations were retained without regard to the specific nature of the rainfall and runoff data. Highly complex patterns of rainfall and runoff are in the database, which contribute to the magnitude of the smallest and largest errors. Of more general interest for the discussion here is the interquartile range (third quartile minus first quartile). The interquartile range is small compared to the minimum and maximum range, which indicates that the errors have much smaller variation. The central tendency, as indicated by the median and mean, require residual-specific discussion.

For ε^{Q_p} the medians are slightly greater than zero, and the means are between about $0.05\log_{10}$ and $0.10\log_{10}$(cubic foot per second). The authors conclude that nearly unbiased estimates of Q_p result from each of the four techniques. The standard deviations of ε^{Q_p} all are about $0.54\log_{10}$(cubic foot per second), much larger than the $0.20\log_{10}$(cubic foot per second) in figure 3. The variations in ε^{Q_p} for each technique are about 2.5 times larger than the Q_p variation. Although Q_p estimation results are nearly unbiased, the approximate equality of the standard deviations of ε^{Q_p} contributes to considerable ambiguity regarding which technique is preferable for watershed-loss estimation. The preferable technique could be that with the smallest standard deviation of ε^{Q_p}.

Both the medians and means for ε^V are greater than zero and large enough to conclude that somewhat biased estimates of V result from each of the four techniques. The bias is between about 0.1 and 0.3 inch. The positive sign on the medians and means in table 7 indicates that V is being underestimated. Therefore, the authors suggest that between about 1/10 and 1/4 inch be added to V^{model} after application of the watershed losses and the unit hydrograph. If done in a practical application, the addition only occurs to the total runoff produced by the excess rainfall and unit hydrograph computations. The additional volume is not to be distributed across the excess rainfall hyetograph because that would result in an increase in the modeled Q_p; such an increase is contrary to the conclusion that nearly unbiased estimates of Q_p already occur.

For $\varepsilon^{T_{Q_p}}$ the medians and means are less than zero but are small enough to conclude that generally unbiased estimates of T^{Q_p} result. However, specific interpretation of the medians and means indicates that T^{Q_p} is overestimated by about 15 minutes.

The Combined $I_A{}^{\&}C_L$ Model

There is ambiguity as to which of the four techniques is preferable for watershed-loss estimation. In particular for Q_p, the results in table 6 indicate that each of the four techniques has similar bias and similar standard deviations of error. Therefore, a judgement was made to combine all four techniques—the combined $I_A{}^{\&}C_L$ model. The combined Q_p, denoted as $\mathring{Q}_p$,

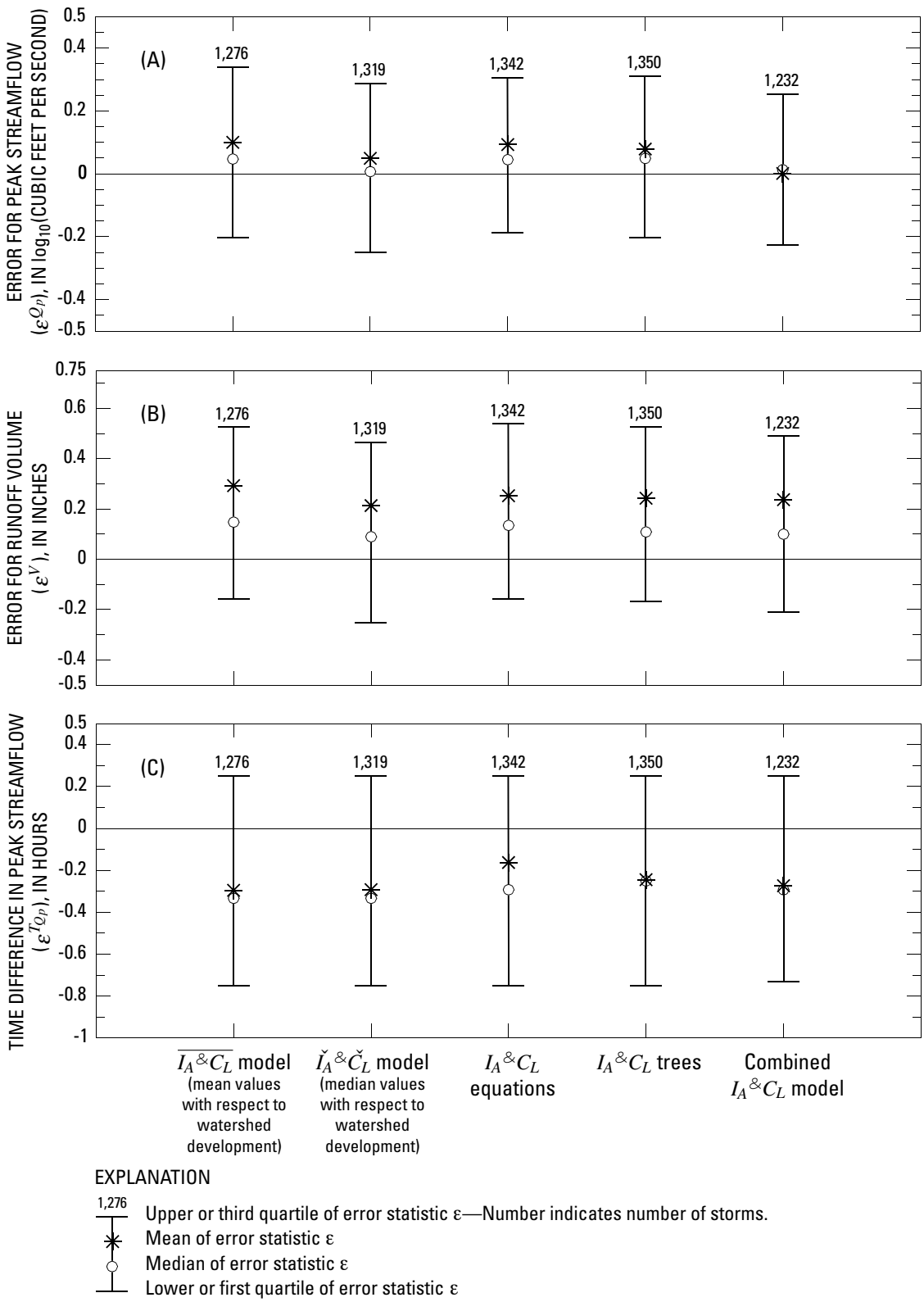


Figure 9. Whisker diagrams showing distribution of error statistics for each of the four techniques and the combined $I_A{}^{\&}C_L$ model from numerical values listed in tables 6–8.

Table 8. Common summary statistics of time difference of peak streamflow ($\varepsilon^{T_{Q_p}}$) for the 92 watersheds.

[All values are in hours. The equalities of the minimum, first quartile, third quartile, and maximum columns have been verified.]

Minimum	First quartile	Median	Mean	Third quartile	Maximum	Standard deviation
$\overline{I_A{}^{\&}C_L}$ model (1,276 storms)						
−70.4	−0.750	−0.333	−0.295	0.250	55.8	6.44
$\check{I}_A{}^{\&}\check{C}_L$ model (1,319 storms)						
−70.4	−.750	−.333	−.292	.250	55.8	6.24
$I_A{}^{\&}C_L$ equations (1,342 storms)						
−70.4	−.750	−.292	−.162	.250	55.8	6.39
$I_A{}^{\&}C_L$ trees (1,350 storms)						
−70.4	−.750	−.250	−.244	.250	55.5	6.40
Combined $I_A{}^{\&}C_L$ model (1,232 storms)						
−70.4	−.729	−.292	−.273	.250	55.8	6.18

is defined as the arithmetic mean

$$\mathring{Q}_p = \frac{1}{4}\left(Q_p^{\overline{I_A{}^{\&}C_L}\,\text{model}} + Q_p^{\check{I}_A{}^{\&}\check{C}_L\,\text{model}} + Q_p^{I_A{}^{\&}C_L\,\text{equations}} + Q_p^{I_A{}^{\&}C_L\,\text{trees}}\right). \quad (38)$$

Other quantities representing combined V, denoted as $\mathring{V}$, and combined T^{Q_p}, denoted as $\mathring{T}^{Q_p}$, are defined as

$$\mathring{V} = \frac{1}{4}\left(V^{\overline{I_A{}^{\&}C_L}\,\text{model}} + V^{\check{I}_A{}^{\&}\check{C}_L\,\text{model}} + V^{I_A{}^{\&}C_L\,\text{equations}} + V^{I_A{}^{\&}C_L\,\text{trees}}\right) \text{ and} \quad (39)$$

$$\mathring{T}^{Q_p} = \frac{1}{4}\left(T^{\overline{I_A{}^{\&}C_L}\,\text{model}} + T^{\check{I}_A{}^{\&}\check{C}_L\,\text{model}} + T^{I_A{}^{\&}C_L\,\text{equations}} + T^{I_A{}^{\&}C_L\,\text{trees}}\right). \quad (40)$$

The common summary statistics of the residuals for the combined $I_A{}^{\&}C_L$ model are listed in the last row of tables 6–8. In particular the Q_p estimation clearly is unbiased, and the standard deviation is reduced substantially. Specifically, these two statistics are much closer to zero and unbiased (mean) or smaller (standard deviation) than those listed for the four techniques. The approximate $0.41\log_{10}$(cubic foot per second) standard deviation (last row, last column of table 6) is about 105 percent larger than the $0.20\log_{10}$(cubic foot per second) for the Q_p^{model} for $^{\natural}I_A{}^{\&}{}^{\natural}C_L$ and $K^{\&}T_p$-GUH in the graph of figure 3. The bias and standard deviations of ε^V and $\varepsilon^{T_{Q_p}}$ for the combined $I_A{}^{\&}C_L$ model remain similar to those for the other four techniques. The authors recognize that sample size differences could influence the computations and thus the conclusions, but the sample sizes are large enough that concern is mitigated.

To demonstrate the application of the combined $I_A{}^{\&}C_L$ model, a comparison between $\mathring{Q}_p$ and Q_p^{obs} is shown in figure 10. The figure contains 1,232 data points. The relation demonstrates that $\mathring{Q}_p$ effectively are unbiased estimates of Q_p^{obs}. The variability of the data in figure 10 clearly is larger than the variability in figure 3; the larger variability is reflected by the larger standard deviation about the equal value line. There is evidence of larger variance for the smallest Q_p^{obs} values, the source of which is not identified.

As with the graph in figure 3, the importance and lasting ramifications of the graph in figure 10 are stressed. The graph shows that the combined $I_A{}^{\&}C_L$ model and $K^{\&}T_p$-GUH likely will perform reliably for the 92 watersheds. Therefore, the combined $I_A{}^{\&}C_L$ model and $K^{\&}T_p$-GUH likely will perform reliably for similar, that is, applicable ungaged watersheds.

The Combined $I_A{}^{\&}C_L$ Model in Practical Applications

The combined $I_A{}^{\&}C_L$ model is preferred over individual methods of the four techniques for circumstances involving design of drainage infrastructure. For applicable watersheds, the analyst makes four separate computations using a single $K^{\&}T_p$-GUH as the unit hydrograph. The computations will use the $\overline{I_A{}^{\&}C_L}$ model, $\check{I}_A{}^{\&}\check{C}_L$ model, $I_A{}^{\&}C_L$ equations, and $I_A{}^{\&}C_L$ trees as four individual—although not mutually independent—watershed-loss models. The arithmetic mean (eqs. 38–40) of the specific values of $\mathring{Q}_p$, $\mathring{V}$, and $\mathring{T}^{Q_p}$ represent the best available estimates of these three important factors in hydrologic-engineering design. Analysts might consider adding between about

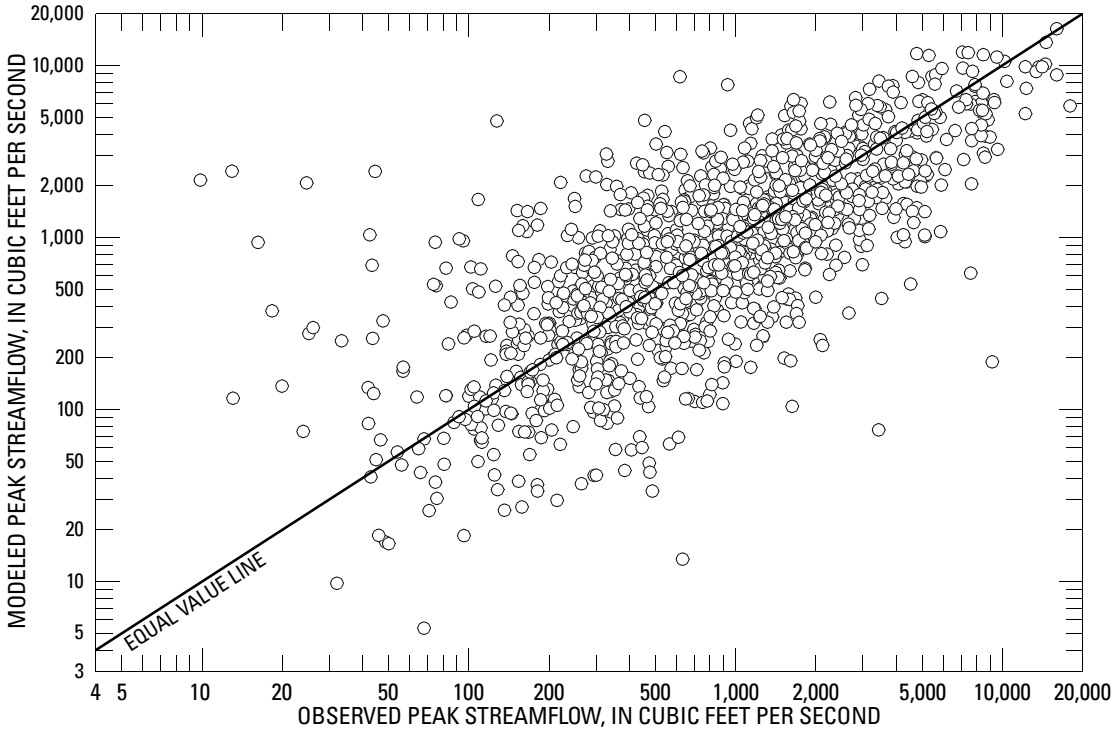


Figure 10. Graph showing comparison of observed (Q_p^{obs}) and modeled ($\mathring{Q}_p$) peak streamflows from the combined $I_A{}^{\&}C_L$ and $K^{\&}T_p$-GUH model for the 92 watersheds.

1/10 and 1/4 inch to the $\mathring{V}$ to compensate for expected underestimation of V by the combined $I_A{}^{\&}C_L$ model and $K^{\&}T_p$-GUH. Finally, analysts should expect $\mathring{T}^{Q_p}$ to be on average about 15 minutes too late.[j]

For circumstances in which no runoff results from one, two, or three of the four models, then each of those models should be dropped from equations 38–40, and the integer in the denominators accordingly changed. In other words, zero runoff conditions are ignored. If the models of all four techniques produce no runoff, then it follows that no runoff for the design storm hyetograph is produced by the combined $I_A{}^{\&}C_L$ model.

Finally, if a single $I_A{}^{\&}C_L$ model must be identified as the most favorable, the authors prefer the $I_A{}^{\&}C_L$ trees. The $I_A{}^{\&}C_L$ trees are preferred because the trees produce the lowest standard deviation of ε^{Q_p} (see column 7, table 6). The authors recognize, however, that the other residual statistics for each of the four techniques (tables 6–8) are similar. In general, the $I_A{}^{\&}C_L$ trees are easier to use than the $I_A{}^{\&}C_L$ equations and have six or fewer unique outcomes—the trees provide unambiguous I_A and C_L estimates. On the other hand, the I_A and C_L equations offer a wider range of outcomes and, as a result, offer interpolation (or extrapolation) to watershed types well (or not well) represented by the characteristics of the 92 watersheds. The $I_A{}^{\&}C_L$ equations also provide a means to estimate prediction limits for an arbitrary watershed, and the prediction limits in turn provide a measure of equation applicability to an arbitrary watershed.

Further Discussion

INHERENT MODEL BIAS: The combined $I_A{}^{\&}C_L$ model and $K^{\&}T_p$-GUH are based on rainfall and runoff data for the 92 watersheds. Storms for which no runoff occurred are not represented in the database. At times some storms with a given depth and duration produce no runoff; however, at other times an apparently similar storm will produce runoff. The difference in watershed response can be attributed in part to errors in areal and temporal distribution of the rainfall as represented by the rainfall data for the 92 watersheds and temporal dif-

ferences in antecedent moisture condition. Antecedent moisture condition is a measure of the potential for runoff based on history of rainfall in the watershed. During periods of abundant rainfall, the potential for runoff is greater than during periods of below-average rainfall. Regardless of the sources of storm-to-storm variations in rainfall and runoff relations, from an overall perspective of runoff potential, the combined $I_A{}^{\&}C_L$ model underestimates watershed losses. The underestimation occurs because the computational analysis is biased by the fact that only runoff-producing storms were analyzed, or could be analyzed.

MODELING TIME STEP: The combined $I_A{}^{\&}C_L$ model and $K^{\&}T_p$-GUH are each derived from computer models using 5-minute time steps. This is a small time increment. However, in hydrologic-engineering practice the size of rainfall time steps, and hence rainfall-runoff models using unit hydrographs, commonly is tens of minutes to several hours. Attention is required by analysts in the specification of q_p, T_p, and K of the GUH so that the Q_p^{model} is not underestimated because T_p obtained from equation 11 is unattainable. This condition occurs because the modeling time steps are not integer multiples of the T_p. For example, suppose $T_p = 7.5$ hours from equation 11 and the analyst requires 1-hour increments for modeling. The 1-hour unit hydrograph potentially peaks at 7 or 8 hours, but not 7.5 hours. The magnitude of Q_p^{model} will be underestimated. This is not a problem specific with the GUH, but an artifact of the mathematical implementation of a unit hydrograph with discrete time steps. Therefore, the authors suggest the following algorithm for parameter computation to compensate for the influence of modeling time step. The boxed symbols for q_p, T_p, and K represent final values.

1. Compute K (dimensionless) by equation 5 and T_p in hours by equation 11.

2. Compute $\boxed{q_p}$ by equation 4 using K and T_p from step 1. (This q_p is regarded as the "true peak" streamflow, which this algorithm seeks to preserve.)

3. Determine a convenient time step δ (such as 15, 30, 60, 120, . . .) for which 5 minutes is an integer divisor and for which 5 to about 16 time steps prior to the peak are obtained. This computation is

$$5 \leq \text{int}(60T_p/\delta) \leq 16,$$

where $\text{int}(x)$ is a function that returns an integer of the argument x, and 60 accounts for an hours-to-minutes conversion. The quantity between the inequalities is labeled $\aleph$; this value is to be determined by the analyst.

4. Compute $\boxed{T_p} = \delta \times \aleph/60$.

5. Compute $\boxed{K}$ by equation 4 using $\boxed{q_p}$ and $\boxed{T_p}$ from steps 2 and 4. This step adjusts the shape of the GUH so that $\boxed{q_p}$ is honored yet occurs at $\boxed{T_p}$, which is a time that is consistent with the modeling time step δ.

6. Apply the combined $I_A{}^{\&}C_L$ model with a GUH defined by equation 2 using $\boxed{q_p}$, $\boxed{T_p}$, and $\boxed{K}$.

This algorithm is suggested for general application of the results reported here and also is applicable to the other unit hydrographs. For example, the algorithm could be used for the unit hydrographs described by Asquith and others (2005).

FINAL REMARKS: In conclusion, after 7 years of research of rainfall and runoff data in the context of design storms and resulting runoff context, a reliable framework for estimation of design hydrographs using a unit hydrograph with the structure of a gamma distribution and time-distributed, watershed-loss model is available. Both the unit hydrograph and the watershed-loss models are tuned to Texas hydrology. The authors suggest the term "Texas Unitgraph Method Type G" to collectively encompass the combined $I_A{}^{\&}C_L$ model and $K^{\&}T_p$-GUH set in accordance with the algorithm to ensure proper estimation of Q_p because of subtle mathematical characteristics of finite time-step modeling. The "Type G" is appended because alternative unit hydrographs, coupled with alternative watershed-loss models could emerge by other studies of the subject. The Texas Unitgraph Method Type G contributes to the state-of-the-practice (2007) for applicable watersheds in Texas.

Summary

Estimation of representative hydrographs from design storms, which are known as design hydrographs, provides for cost-effective, risk-mitigated design of drainage structures such as bridges, culverts, roadways, and other infrastructure. During 2001–07, the U.S. Geological Survey (USGS), in cooperation with the Texas Department of Transportation, investigated runoff hydrographs, design storms, unit hydrographs, and watershed-loss models to enhance design hydrograph estimation.

Design hydrographs ideally should mimic the general volume, peak streamflow (Q_p), and shape of observed runoff hydrographs. Design hydrographs commonly are estimated in part by unit hydrographs. A unit hydrograph is defined as the runoff hydrograph that results from a unit pulse of excess rainfall uniformly distributed over a watershed at a constant rate for a specific duration. A time-distributed, watershed-loss model is required for modeling by unit hydrographs. In particular, the time-distributed, watershed-loss model should be mathematically consistent in structure or general form with hypothesized processes (depression and other macrostorage, infiltration, evaporation, and other physical processes), which prevent rainfall from becoming runoff. This report develops a specific time-distributed, watershed-loss model known as an initial-abstraction, constant-loss model ($I_A{}^{\&}C_L$).

For the $I_A{}^{\&}C_L$ model, a watershed is conceptualized to have the capacity to store or abstract an absolute depth of rainfall at and near the beginning of a storm. Depths of rainfall less than this initial abstraction do not produce runoff. The watershed also is conceptualized to have the capacity to remove rainfall at a constant rate after the initial abstraction is satisfied. Additional rainfall inputs after the initial abstraction is satisfied contribute to runoff if the rainfall rate (intensity) is larger than the constant loss. The $I_A{}^{\&}C_L$ model thus is a two-parameter model of watershed losses. The $I_A{}^{\&}C_L$ model is developed through detailed computational and statistical analysis of observed rainfall and runoff data for 92 USGS streamflow-gaging stations (watersheds) in Texas with contributing drainage areas from 0.26 to 166 square miles.

For the development of the $I_A{}^{\&}C_L$ model, the unit hydrograph is limited to a previously described, watershed-specific, gamma-distribution, unit-hydrograph model (GUH). The procedures for computation of GUH are reproduced and extensively discussed. The GUH was taken from a previous report, which documents a unit hydrograph investigation for the same 92 watersheds. The GUH has two unique parameters that can be variously expressed, but are considered in terms of watershed-depth peak streamflow (q_p) and time to peak (T_p). The third parameter is a shape parameter (K) that is dependent on q_p and T_p.

The equations to estimate K and T_p of GUH for applicable Texas watersheds are collectively referred to as the $K^{\&}T_p$ equations, and the GUH set by these equations is referred to as $K^{\&}T_p$-GUH. The $K^{\&}T_p$-GUH is deemed appropriate for development of the $I_A{}^{\&}C_L$ model.

To initiate $I_A{}^{\&}C_L$ model development, a complex computational analysis of the database of observed rainfall and runoff for the 92 watersheds was done using custom-built software (labeled as ${}^{\lightning}I_A{}^{\&}{}^{\lightning}C_L$). The computations used a 5-minute time step; rainfall and runoff values were linearly interpolated to 5-minute increments as needed. The purpose of the software was to compute optimal (storm-specific, $\lightning$) parameter values (${}^{\lightning}I_A$ and ${}^{\lightning}C_L$) for each suitable storm in the database. The GUH was estimated for each watershed by the $K^{\&}T_p$ equations. This "watershed-specific GUH" was considered a representative unit hydrograph for the watershed. The analysis successfully processed 1,620 of about 1,660 storms and provided ${}^{\lightning}I_A$ and ${}^{\lightning}C_L$ values used in later statistical analyses.

The assumption that the GUH set by the $K^{\&}T_p$ equations is representative for each watershed for the analysis of ${}^{\lightning}I_A$ and ${}^{\lightning}C_L$ is important. Although the "correctness" of $K^{\&}T_p$-GUH is not ensured, the assumption implies that the $I_A{}^{\&}C_L$ model reported is "tuned" against observed rainfall and runoff and $K^{\&}T_p$-GUH. Therefore, the $I_A{}^{\&}C_L$ model is linked to $K^{\&}T_p$-GUH, and in practice the two techniques are to be used together.

Values for ${}^{\lightning}I_A$ and ${}^{\lightning}C_L$ were computed by the ${}^{\lightning}I_A{}^{\&}{}^{\lightning}C_L$ software. Those values were used to generate an excess rainfall hyetograph (time series of rainfall

intensity) and when convolved with the GUH, produced a modeled runoff hydrograph that has the same volume as the observed runoff hydrograph and a minimized residual sum of squares between the observed and modeled runoff hydrographs. The means of ${}^{\natural}I_A$ and ${}^{\natural}C_L$ for each watershed were computed; the means ($\overline{I_A}$ and $\overline{C_L}$) are referred to as watershed-specific. $\overline{I_A}$ and $\overline{C_L}$ for each watershed are considered the most representative and are used in further statistical analyses.

Statistical analyses of $\overline{I_A}$ and $\overline{C_L}$ were done with the objectives of (1) documenting the parameter distribution and (2) developing predictive procedures of each parameter for ungaged watersheds. The statistical analyses of $\overline{I_A}$ and $\overline{C_L}$ document the distribution for each parameter. The four-parameter generalized lambda distribution (GLD) was used as a parametric model, which was fit by the method of L-moments. The L-moments and corresponding GLD parameters for both $\overline{I_A}$ and $\overline{C_L}$ are tabulated. The analyses show that watershed development has substantial influence on I_A and limited influence on C_L. The mean and median watershed-specific values are tabulated with respect to watershed development. Although considerable variability exists, these values are used in later analyses to develop procedures for I_A and C_L estimation for ungaged watersheds.

The statistical analyses also document predictive procedures for estimation of $\overline{I_A}$ and $\overline{C_L}$ for ungaged watersheds with respect to watershed development. Both regression equations and regression trees for estimation of I_A and C_L are provided. The watershed characteristics included in the regressions are main-channel length (L), watershed development (a binary factor $D = 0 \mid 1$), abundance of rocky terrain with thin soils and limestone and karst features (a binary factor $R = 0 \mid 1$), and curve number (CN). Other characteristics assessed were dimensionless main-channel slope, soil types and textures, and percentage impervious cover.

The regression equations for I_A and C_L are accompanied by mathematical results to assess equation applicability and prediction limits. Physical interpretations of the regression coefficients are made. In summary, the coefficient on D for the I_A equation implies that developed watersheds generally have about 1/5-inch less initial rainfall storage than undeveloped watersheds. The coefficients on R imply that rock-dominated, thin-soiled watersheds as represented by some of the 92 watersheds have about 1/4-inch larger I_A and about 1/4-inch per hour larger C_L than other watersheds. The coefficients on CN are consistent with the broadly understood definition of CN. An increase of 10 units of CN represents about $-1/7$-inch of I_A and represents about $-1/6$-inch per hour of C_L. Description and interpretation of the influence of L on I_A and C_L is more complex than for D, R, and CN.

Regression trees result from an alternative method of regression (sometimes termed recursive partitioning) when compared to regression that produces equations. A tree is constructed such that partitions (branches) are determined by an algorithm that seeks to split and minimize residual sum of squares. A tree lists at each terminal branch the value for I_A or C_L, the number of samples, and residual standard error. Unlike for the I_A equation, D apparently does not have substantial predictive properties for I_A in a regression-tree context. Therefore, conclusions based on which parameters are important can be influenced by model structure; the I_A equation and I_A tree are structurally distinct.

Subsequent to the regression analyses, an $I_A{}^{\&}C_L$ model for watershed-loss estimation for applicable watersheds in Texas is evaluated. The $I_A{}^{\&}C_L$ equations with statistically significant variables explain about 30–34 percent of the variation in the $\overline{I_A}$ and $\overline{C_L}$ values. The $I_A{}^{\&}C_L$ trees explain a similar amount. Therefore, the mean or median of $\overline{I_A}$ and $\overline{C_L}$, with consideration of watershed development, also could be reasonable estimates for a watershed. Four techniques are identified for general estimation for ungaged watersheds. The four techniques are abbreviated as $\overline{I_A{}^{\&}C_L}$ (mean values), $\check{I}_A{}^{\&}\check{C}_L$ (median values), $I_A{}^{\&}C_L$ equations, and $I_A{}^{\&}C_L$ trees. The units on I_A and C_L are watershed inches and watershed inches per hour, respectively.

The ${}^{\natural}I_A{}^{\&}{}^{\natural}C_L$ software was modified to implement the four techniques for the 92 watersheds. As a result, all storms in the database for each of the 92 watersheds were reprocessed four separate times. With each reprocessing, the (1) Q_p error, (2) runoff volume error, and (3) time difference of Q_p were computed for each suitable storm. Analysis of the three error types was

made. The analysis of the Q_p error indicates that nearly unbiased estimates of Q_p result from each of the four techniques. The variations of the Q_p errors for each technique are about 2.5 times larger than the Q_p variation. The analysis of the runoff volume error indicates that generally biased estimates result from each of the four techniques. The bias is between about 0.1 and 0.3 inch. The positive values indicate that runoff volume is being underestimated. A method to compensate for the underestimation is suggested. The analysis of the time difference of Q_p indicates that generally unbiased estimates result. However, specific interpretation indicates that time of Q_p occurrence is overestimated by about 15 minutes.

The analysis of the three error types shows that there is ambiguity as to which single technique is preferable for watershed-loss estimation. In particular for Q_p, the results indicate that each of the four techniques has similar bias and approximately equal standard deviations of error. Therefore, a judgement was made to combine all four techniques—the combined $I_A{}^{\&}C_L$ model.

The combined $I_A{}^{\&}C_L$ model uses the arithmetic mean of Q_p, volume, and time of Q_p occurrence. These averages provide better estimates than each technique individually. In particular the Q_p estimation clearly is unbiased, and the standard deviation is reduced substantially. Specifically, the mean Q_p error for the combined model is closer to zero, and the standard deviation is smaller than the corresponding statistics for the four techniques. The approximate $0.41\log_{10}$(cubic foot per second) standard deviation of the Q_p error is about 105 percent larger than the $0.20\log_{10}$(cubic foot per second) for optimal $^{\natural}I_A$ and $^{\natural}C_L$. The bias and standard deviations of the volume and time difference of Q_p for the combined model remain similar to those from the four techniques.

The combined $I_A{}^{\&}C_L$ model and $K^{\&}T_p$-GUH are based on paired rainfall and runoff data. Therefore storms for which no runoff occurred are not represented in the database. At times some storms with a given depth and duration produce no runoff; however, at other times an apparently similar storm will produce runoff. The difference in watershed response can be attributed in part to temporal differences in antecedent moisture condition. The combined $I_A{}^{\&}C_L$ model is conservative, which means that from an overall perspective of runoff potential, watershed losses are underestimated. The underestimation occurs because the computational analysis is biased by the fact that only runoff-producing storms were analyzed, or could be analyzed. An algorithm to compensate for subtle mathematical characteristics of the unit hydrograph method in regard to the influence of modeling time step is suggested for general application of the combined $I_A{}^{\&}C_L$ model and $K^{\&}T_p$-GUH.

The authors suggest the term "Texas Unitgraph Method Type G" to collectively encompass the combined $I_A{}^{\&}C_L$ model and $K^{\&}T_p$-GUH set in accordance with the algorithm to ensure proper estimation of Q_p because of subtle mathematical characteristics of finite time-step modeling. The Texas Unitgraph Method Type G contributes to the state-of-the-practice (2007) for applicable watersheds in Texas.

References

Abramowitz, Milton, and Stegun, I.A., 1964 [reprinted 1972], Handbook of mathematical functions: National Bureau of Standards Applied Mathematics Series, v. 55, 1,046 p.

Asquith, W.H., 2007, L-moments of the generalized lambda distribution: Computational Statistics and Data Analysis, v. 51, p. 4,484–4,496.

Asquith, W.H., and Roussel, M.C., 2004, Atlas of depth-duration frequency of precipitation annual maxima for Texas: U.S. Geological Survey Scientific Investigations Report 2004–5041, 106 p. [http://pubs.usgs.gov/sir/2004/5041/]

Asquith, W.H., Roussel, M.C., Cleveland, T.G., Fang, Xing, and Thompson, D.B., 2006, Statistical characteristics of storm interevent time, depth, and duration for eastern New Mexico, Oklahoma, and Texas: U.S. Geological Survey Professional Paper 1725, 299 p. [http://pubs.usgs.gov/pp/pp1725/]

Asquith, W.H., Roussel, M.C., Thompson, D.B., Cleveland, T.G., and Fang, Xing, 2004, Summary of dimensionless Texas hyetographs and distribution of storm depth developed for Texas Department of Transportation Research Project 0–4194:

Austin, Center for Transportation Research, University of Texas at Austin, Texas Department of Transportation Research Report 0–4194–4, 68 p.

Asquith, W.H., Roussel, M.C., Thompson, D.B., Cleveland, T.G., and Fang, Xing, 2005, Unit hydrograph estimation for applicable Texas watersheds: Lubbock, Texas Tech Center for Multidisciplinary Research in Transportation, Texas Tech University, Texas Department of Transportation Research Report 0–4193–4, 71 p.

Asquith, W.H., and Thompson, D.B., 2005, Alternative regression equations for estimation of annual peak-streamflow frequency for undeveloped watersheds in Texas using PRESS minimization: Lubbock, Texas Tech Center for Multidisciplinary Research in Transportation, Texas Tech University, Texas Department of Transportation Research Report 0–4405–2, 27 p.

Asquith, W.H., Thompson, D.B., Cleveland, T.G., and Fang, Xing, 2004, Synthesis of rainfall and runoff data used for Texas Department of Transportation Research Projects 0–4193 and 0–4194: U.S. Geological Survey Open-File Report 2004–1035, 1,050 p. [http://pubs.usgs.gov/of/2004/1035/]

Bates, R.L., 1987, Glossary of geology (3rd ed.): Alexandria, Va., American Geological Institute, 754 p.

Chow, V.T., Maidment, D.R., and Mays, L.W., 1988, Applied hydrology: New York, McGraw-Hill, 572 p.

Cleveland, T.G., He, Xin, Asquith, W.H., Fang, Xing, and Thompson, D.B., 2006, Instantaneous unit hydrograph selection for rainfall-runoff modeling of small watersheds in north and south central Texas: Journal of Irrigation and Drainage Engineering, v. 132, p. 479–485.

Dingman, S.L., 2002, Physical hydrology (2nd ed.): Upper Saddle River, N.J., Prentice Hall, 646 p.

Evans M., Hastings, N.A.J., and Peacock, J.B., 2000, Statistical distributions (3rd ed.): New York, Wiley, 221 p.

Faraway, J.J., 2006, Extending the linear model with R—Generalized linear, mixed effects and nonparametric regression models: Boca Raton, Fla., Chapman and Hall/CRC, 301 p.

Good, P.I., and Hardin, J.W., 2003, Common errors in statistics (and how to avoid them): Hoboken, N.J., Wiley, 222 p.

Haan, C.T., Barfield, B.J., and Hayes, J.C., 1994, Design hydrology and sedimentology for small catchments: San Diego, Academic Press, 588 p.

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. A3, 510 p. [http://pubs.usgs.gov/twri/twri4a3/]

Hosking, J.R.M., 1990, L-moments—Analysis and estimation of distributions using linear combinations of order statistics: Journal Royal Statistical Society B, v. 52, no. 1, p. 105–124.

Karian, Z.A., and Dudewicz, E.J., 2000, Fitting statistical distributions—The generalized lambda distribution and generalized bootstrap methods: Boca Raton, Fla., CRC Press, 438 p.

Maindonald, J., and Braun, J., 2003, Data analysis and graphics using R—An example-based approach: Cambridge, United Kingdom, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 362 p.

Montgomery, D.C., Peck, E.A., and Vining, G.G., 2001, Introduction of linear regression analysis—Wiley Series in Probability and Statistics: New York, Wiley, 641 p.

Natural Resources Conservation Service, 2006, National engineering handbook: Part 630, chap. 15, accessed November 1, 2007, at http://www.info.usda.gov/CED/Default.cfm?xSbj=59&xAud=24#National%20Engineering%20Handbook

Pilgrim, D.H., and Cordery, Ian, 1993, Flood runoff, chapter 9, *in* Maidment, D.R., ed., Handbook of hydrology: New York, McGraw-Hill, p. 9.1–9.42.

R Development Core Team, 2006, R—A language and environment for statistical comput-

ing: Vienna, Austria, R Foundation for Statistical Computing, ISBN 3–900051–07–0, http://www.R-project.org.

Therneau, T.M., and Atkinson, Beth, 2006, rpart package—Recursive partitioning: R port by Brian Ripley, R package version 3.1-29, S-PLUS 6.x original at http://mayoresearch.mayo.edu/mayo/research/biostat/splusfunctions.cfm

Thompson, D.B., Cleveland, T.G., and Fang, Xing, 2007, Loss-rate functions for selected Texas watersheds: Lubbock, Texas Tech Center for Multidisciplinary Research in Transportation, Texas Tech University, Texas Department of Transportation Research Report 0–4193–5, [in press].

U.S. Geological Survey, 2006, Land cover characterization program—National land cover/MRLC: Earth Resources Observation Systems (EROS) Data Center, accessed various dates in December 2006, at http://landcover.usgs.gov/index.asp

Williams-Sether, Tara, Asquith, W.H., Thompson, D.B., Cleveland, T.G., and Fang, Xing, 2004, Empirical, dimensionless, cumulative-rainfall hyetographs developed from 1959–86 storm data for selected small watersheds in Texas: U.S. Geological Survey Scientific Investigations Report 2004–5075, 125 p. [http://pubs.usgs.gov/sir/2004/5075/]

Appendix 1—Selected Watershed Characteristics and Graphical and Statistical Output From the $^{\lightning}I_A$ & $^{\lightning}C_L$ Software for the 92 Watersheds

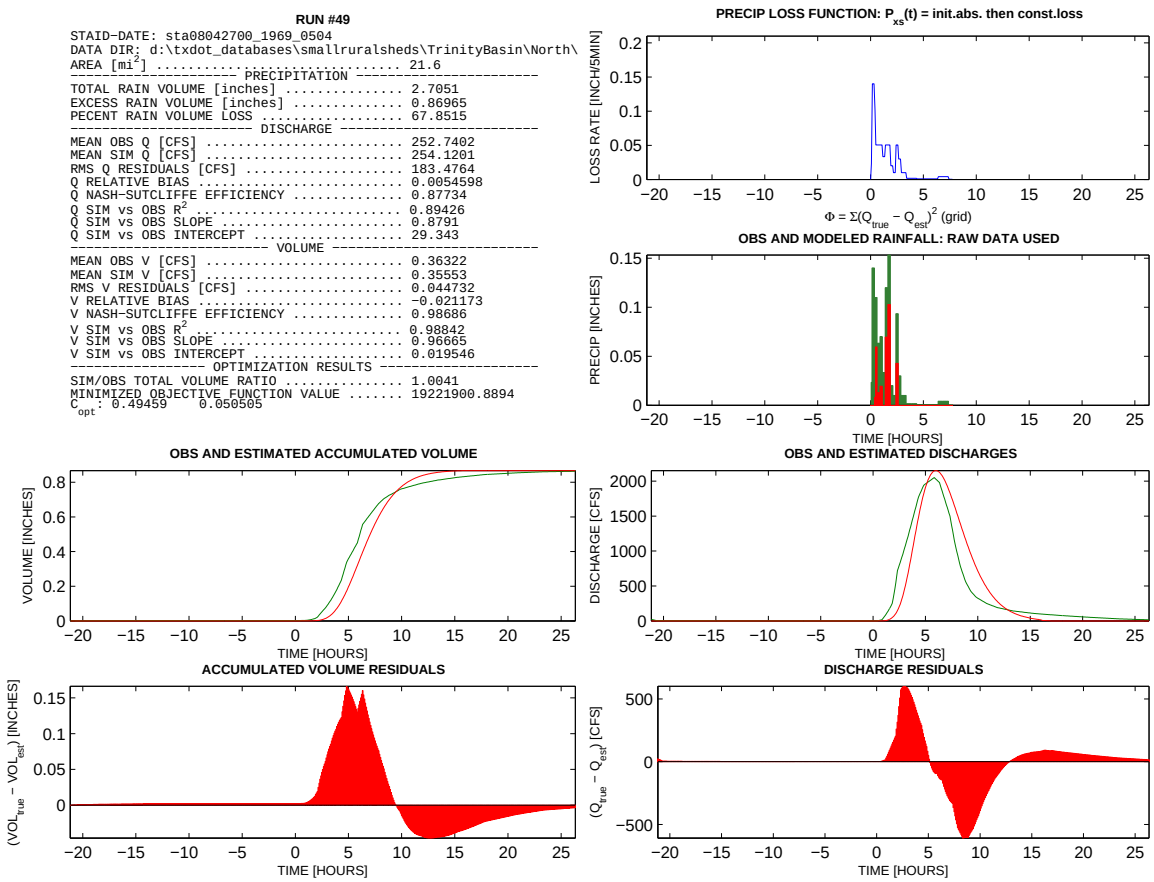
Blank Page

Examples of Graphical Output From the $^{\natural}I_A{}^{\&\natural}C_L$ Software

Two examples of direct graphical output from the $^{\natural}I_A{}^{\&\natural}C_L$ software (Joseph Vrabel, U.S. Geological Survey, written commun., 2006; source code is available from report web site) are shown in this section. The output for a storm that occurred on May 4, 1969, for station 08042700 is shown in figure 1.1, and the output for a storm on June 3, 1973, for station 08048530 is shown in figure 1.2. The numerous intermediate results are shown in the upper-left corner of each figure; these results served the purposes of quality control and quality assurance of the software development. The negative time values in the graphs represent a need specific to the output routines of the software to accommodate the database. Of interest for this report are the $^{\natural}I_A{}^{\&\natural}C_L$ values; these are shown in the $\mathtt{C_{opt}}$ field at the bottom of the tabular material on the upper-left corner of the figures. Red lines represent the results of the modeled runoff hydrograph, and green lines represent observed data.

The May 4, 1969, storm in figure 1.1 (shown to right of $\mathtt{C_{opt}}$) has $^{\natural}I_A = 0.495$ watershed inch and $^{\natural}C_L = 0.606$ watershed inch per hour (the figure shows $^{\natural}C_L = 0.0505$ in units of inch per 5 minutes). This storm is characterized as a single-peaked event (green line, second plot from bottom on right). The modeled runoff hydrograph (red line) using $K^{\&}T_p$-GUH and the $^{\natural}I_A{}^{\&\natural}C_L$ values is in remarkably good agreement with the observed data given that the unit hydrograph is constrained as a gamma distribution, the I_A and C_L values are physically meaningful, and storm volume is matched (see top plot on left). In general, this figure represents an excellent demonstration of software operation for a well-configured storm. Although not considered in the computations, the $T^{Q_p^{\text{model}}}$ approximates $T^{Q_p^{\text{obs}}}$.

The June 3, 1973, storm (fig. 1.2) has $^{\natural}I_A = 1.24$ watershed inches and $^{\natural}C_L = 0.588$ watershed inch per hour. This storm is characterized by about four temporal clusters of rainfall; hence, about four substantial Q_p^{obs} occurred. The $K^{\&}T_p$-GUH and the $^{\natural}I_A{}^{\&\natural}C_L$ modeling can produce a reasonable overall fit to complex observed runoff hydrographs. The $K^{\&}T_p$-GUH was defined by data analysis of generally discrete peaks (Asquith and others, 2005); in contrast the $^{\natural}I_A{}^{\&\natural}C_L$ software, as demonstrated by the second example, operated on the actual file structure of the database (Asquith, Thompson, and others, 2004).



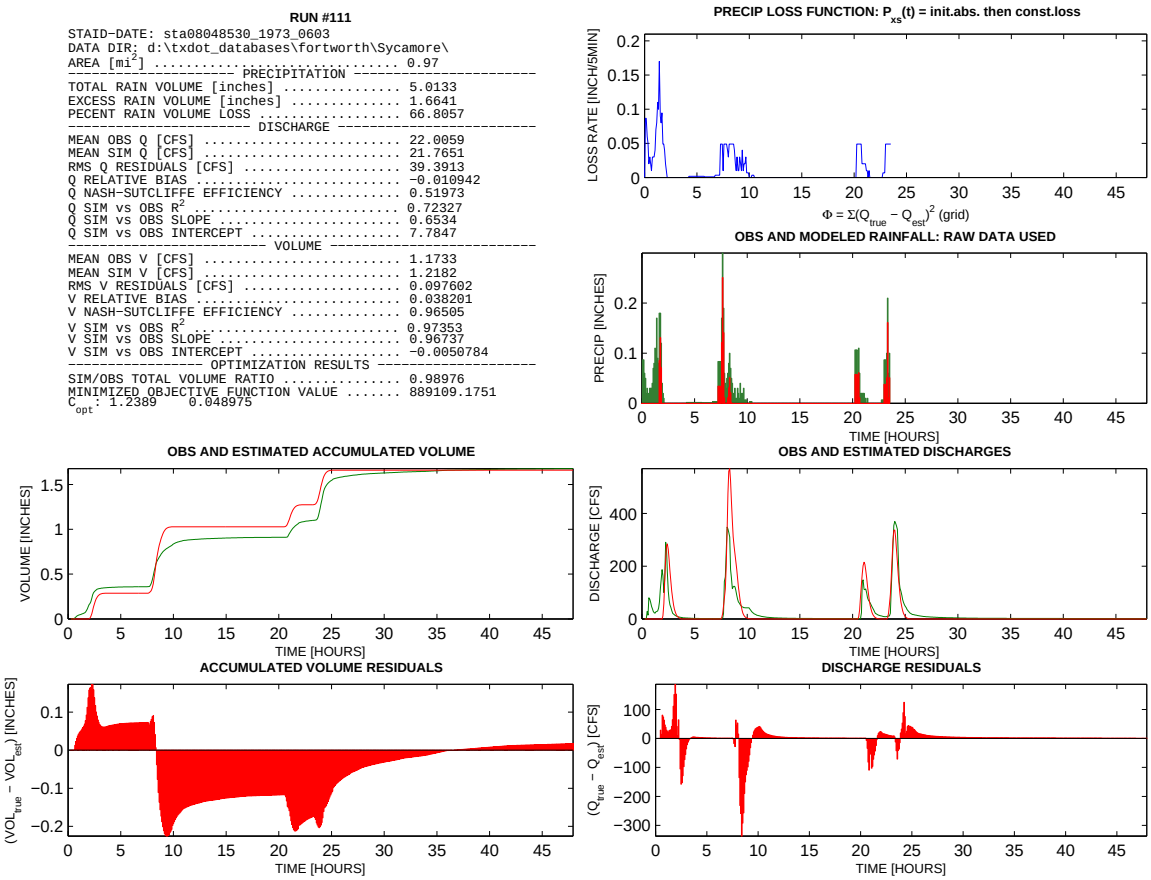
EXPLANATION

- Rainfall data
- Modeled data (ESTIMATED)
- Observed data (OBS)
- C_{opt} Storm-specific, initial-abstraction ($^{\ell}I_A$) and constant-loss ($^{\ell}C_L$) values

ABBREVIATIONS

CFS	cubic feet per second
OBS	observed
Q	streamflow (discharge)
SIM	simulated (modeled)
V	volume (watershed depth)

Figure 1.1. Graphs showing selected graphical output from $^{\ell}I_A$ & $^{\ell}C_L$ software for storm of May 4, 1969.



EXPLANATION

- Rainfall data
- Modeled data (ESTIMATED)
- Observed data (OBS)
- C_{opt} Storm-specific, initial-abstraction ($^{\natural}I_A$) and constant-loss ($^{\natural}C_L$) values

ABBREVIATIONS

CFS	cubic feet per second
OBS	observed
Q	streamflow (discharge)
SIM	simulated (modeled)
V	volume (watershed depth)

Figure 1.2. Graphs showing selected graphical output from $^{\natural}I_A$ & $^{\natural}C_L$ software for storm of June 3, 1973.

Selected Watershed Characteristics and $\overline{I_A}$ and $\overline{C_L}$ Values

For each of the 92 watersheds the results of $K^{\&}T_p$-GUH and the $^{\natural}I_A{}^{\&\natural}C_L$ estimation are listed in table 1.1 on a watershed-specific basis. Listed in the table are the number of suitable storms processed by the $^{\natural}I_A{}^{\&\natural}C_L$ software. This number provided a basis for weight factors for the regressions described in the text. The most salient watershed characteristics also are listed. The *R* and *CN* values are specific to this report. The *D* values derive from Asquith and others (2005, table 1), and the remaining three watershed characteristics (*A*, *L*, and *S*) derive from Asquith and others (2005, table 2). Finally, the two columns on the far right of the table list the watershed-specific values ($\overline{I_A}$ and $\overline{C_L}$, respectively; see eqs. 17 and 18) of the $^{\natural}I_A$ and $^{\natural}C_L$ values. The $\overline{I_A}$ and $\overline{C_L}$ values provide the basis for the four techniques and the combined $I_A{}^{\&}C_L$ model described in the report.

Table 1.1. Selected watershed characteristics and $\overline{I_A}$ and $\overline{C_L}$ values for the 92 watersheds.

[*R*, rock classification (no = 0, yes = 1); *D*, development classification (no = 0, yes = 1); *CN*, curve number; *A*, contributing drainage area; *L*, main-channel length; *S*, dimensionless main-channel slope; $\overline{I_A}$, watershed-mean initial abstraction; $\overline{C_L}$, watershed-mean constant loss; --, dimensionless; mi^2, square miles; mi, miles; in., watershed inches; in./hr, watershed inches per hour]

Station number	Number of suitable storms	*R* (yes/no)	*D* (yes/no)	*CN* (--)	*A* (mi^2)	*L* (mi)	*S* (--)	$\overline{I_A}$ (in.)	$\overline{C_L}$ (in./hr)
08042650	14	0	0	63.4	6.82	4.632	0.01378	1.322	0.866
08042700	57	0	0	62.5	21.6	11.57	.006025	.766	.800
08048520	24	0	1	82.3	17.7	7.530	.005081	.494	.381
08048530	27	0	1	86.7	.97	1.700	.01181	.243	1.134
08048540	23	0	1	88.0	1.35	2.370	.01119	.100	.802
08048550	25	0	1	91.2	1.08	2.017	.004507	.563	.548
08048600	27	0	1	84.3	2.15	3.845	.004729	.742	.522
08048820	19	0	1	83.4	5.64	6.027	.005970	1.115	.067
08048850	25	0	1	83.0	12.3	9.397	.005059	.928	.249
08050200	33	0	0	79.6	.77	2.643	.01068	.661	.301
08052630	29	0	0	85.4	2.10	3.298	.006489	.435	.183
08052700	57	0	0	84.1	75.5	23.23	.002201	.877	.146
08055580	7	0	1	85.2	1.94	2.997	.007204	.015	.754
08055600	10	0	1	86.1	7.51	6.742	.006012	.345	1.020
08055700	39	0	1	85.5	10.0	7.766	.005048	.492	.520
08056500	41	0	1	85.8	7.98	6.365	.006338	.408	.259
08057020	7	0	1	85.5	4.75	5.092	.009707	.564	.569
08057050	3	0	1	85.7	9.42	6.206	.007812	.176	.532
08057120	5	0	0	80.2	6.77	5.187	.007412	.650	.383
08057130	7	0	1	82.9	1.22	2.629	.009077	.343	.235
08057140	6	0	1	86.8	8.50	7.466	.005758	.286	.630
08057160	8	0	1	90.3	4.17	5.343	.006380	.259	.405
08057320	5	0	1	85.7	6.92	5.416	.005595	.280	.121
08057415	8	0	1	87.8	1.25	1.884	.006333	.187	.580
08057418	7	0	1	79.1	7.65	5.649	.007879	.375	.593
08057420	10	0	1	81.0	13.2	8.334	.006454	.354	.519
08057425	10	0	1	82.9	11.5	6.155	.007877	.458	.455
08057435	4	0	1	81.1	5.91	4.122	.008684	.843	.078

Table 1.1. Selected watershed characteristics and $\overline{I_A}$ and $\overline{C_L}$ values for the 92 watersheds—Continued.

Station number	Number of suitable storms	R (yes/no)	D (yes/no)	CN (--)	A (mi^2)	L (mi)	S (--)	$\overline{I_A}$ (in.)	$\overline{C_L}$ (in./hr)
08057440	3	0	1	79.1	2.53	3.517	.008347	1.221	.073
08057445	8	0	1	86.5	9.03	8.416	.003623	1.604	.170
08057500	29	0	0	78.2	2.14	2.070	.01061	.769	.448
08058000	26	0	0	80.1	1.26	2.087	.01025	.672	.548
08061620	8	0	1	85.0	8.05	5.522	.003876	.133	.436
08061920	9	0	1	86.0	13.4	7.645	.003890	.691	.032
08061950	31	0	1	85.3	23.0	12.65	.003070	.963	.034
08063200	33	0	0	79.4	17.6	8.730	.004013	1.109	.184
08094000	29	0	0	78.4	3.34	3.350	.008705	1.124	.408
08096800	47	0	0	80.0	5.25	4.493	.01117	.764	.890
08098300	18	0	0	80.5	22.2	13.73	.002635	.531	.191
08108200	20	0	0	79.9	48.60	19.96	.002524	.594	.187
08111025	7	0	1	70.0	1.33	2.548	.007061	.520	.102
08111050	7	0	0	70.0	1.94	2.453	.005792	1.133	.025
08136900	22	0	0	75.8	21.8	12.42	.007657	1.126	.311
08137000	37	0	0	74.5	4.02	4.404	.004730	1.046	.480
08137500	5	0	0	76.5	70.4	19.39	.005549	1.665	.221
08139000	29	0	0	74.6	3.42	3.357	.01518	1.160	.614
08140000	29	0	0	74.4	5.41	5.908	.009265	1.041	.823
08154700	14	1	0	68.9	22.3	10.04	.01069	1.248	.748
08155200	6	1	0	70.7	89.7	28.50	.004844	1.009	.793
08155300	8	1	0	69.8	116	45.07	.004030	.908	.490
08155550	10	1	1	87.3	3.12	3.660	.01258	1.001	1.217
08156650	13	1	1	83.6	2.79	2.999	.01150	.797	.671
08156700	16	1	1	86.6	7.03	4.527	.009245	.575	.986
08156750	14	1	1	86.8	7.56	5.130	.008750	.528	.851
08156800	23	1	1	87.0	12.3	10.58	.007481	.789	.493
08157000	40	1	1	88.3	2.31	4.119	.009794	.659	.516
08157500	39	1	1	89.1	4.13	5.164	.009425	.540	.646
08158050	10	1	1	83.9	13.1	7.361	.007925	1.635	.157
08158100	15	1	0	72.6	12.6	5.669	.009120	1.317	1.573
08158200	17	1	0	75.6	26.2	10.92	.006628	.950	1.035
08158380	2	1	1	88.9	5.22	4.015	.006982	.946	.155
08158400	10	1	1	85.6	5.57	4.477	.006726	.507	.288
08158500	14	1	1	76.7	12.1	8.590	.006769	.479	.565
08158600	21	1	1	74.5	51.3	19.47	.004951	1.000	.508
08158700	6	1	0	73.3	124	33.28	.004513	1.119	.430
08158800	2	1	0	69.8	166	48.94	.003916	1.731	.218
08158810	7	1	0	67.9	12.2	6.287	.01109	1.447	.638
08158820	2	1	0	67.2	24.0	14.85	.007462	1.757	.285
08158825	1	1	0	69.8	21.0	12.53	.006649	1.813	2.883

Table 1.1. Selected watershed characteristics and $\overline{I_A}$ and $\overline{C_L}$ values for the 92 watersheds—Continued.

Station number	Number of suitable storms	R (yes/no)	D (yes/no)	CN (--)	A (mi^2)	L (mi)	S (--)	$\overline{I_A}$ (in.)	$\overline{C_L}$ (in./hr)
08158840	10	1	0	68.0	8.24	4.960	.01191	.787	.968
08158860	2	1	0	79.4	23.1	12.79	.007875	1.902	.533
08158880	13	1	0	79.4	3.58	4.404	.01127	.336	1.373
08158920	13	1	1	77.5	6.30	4.974	.01173	1.449	.633
08158930	17	1	1	75.2	19.0	10.40	.008850	1.133	.609
08158970	16	1	1	77.7	27.6	17.61	.006454	1.910	.308
08159150	27	1	0	78.8	4.61	3.739	.008156	1.110	.333
08177600	13	1	1	84.8	.33	1.305	.01437	1.206	.641
08177700	22	1	1	72.0	21.2	10.96	.006584	1.141	.841
08178300	30	1	1	85.7	3.26	3.584	.01665	.462	.802
08178555	10	1	1	84.2	2.43	4.052	.002431	.418	.517
08178600	12	1	0	79.7	9.54	7.051	.01254	1.168	.475
08178620	3	1	1	60.0	4.05	3.608	.01197	1.325	1.061
08178640	8	1	0	78.4	2.45	3.044	.01960	1.675	.325
08178645	6	1	0	78.2	2.33	3.958	.01627	1.727	.160
08178690	41	1	1	84.4	.26	1.172	.004040	.239	1.026
08178736	12	1	1	92.3	.45	1.670	.009415	.944	.347
08181000	9	1	0	79.2	5.57	5.421	.01569	1.126	.936
08181400	15	1	0	79.8	15.0	9.821	.01215	1.741	.591
08181450	30	1	1	87.3	1.19	3.130	.003207	.796	.406
08182400	23	1	0	80.0	7.01	4.867	.005721	1.202	.382
08187000	30	1	0	83.8	3.29	2.780	.009742	.786	.835
08187900	17	1	0	73.3	8.43	4.869	.005251	1.010	1.272

Statistical Relation Between Main-Channel Length and Contributing Drainage Area for the 92 Watersheds

Several of the watershed characteristics listed in table 1.1 would be expected to be correlated such as S and L or A and L. In particular, the correlation, or more precisely, the relation between L and A might be of interest because only L is used in the I_A&C_L equations (eqs. 23 and 29). Some users might want to compare the limits of equations 23 and 29 in terms of L to the expected size of the watershed given L. The relation between L and A is defined by equation 1.1:

$$A^{\text{pred}} = 10^{-0.439} L^{1.66} = 0.364 L^{1.66}, \tag{1.1}$$

where L is main-channel length in miles and A^{pred} is predicted contributing drainage area in square miles. Regression statistics for the relation are shown in figure 1.3. The values −0.439 and 1.66 in the equation are the regression coefficients listed in the figure.

```
# The usual syntax highlighting in this report is purposefully suppressed in the figure.
summary(lm(log10(A)~log10(L)))

Call:
lm(formula = log10(A) ~ log10(L))

Residuals:
      Min        1Q    Median        3Q       Max
-0.375570 -0.090104  0.006886  0.100678  0.287949

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.43902    0.03748  -11.71   <2e-16 ***
log10(L)     1.66076    0.04510   36.82   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1402 on 90 degrees of freedom
Multiple R-Squared: 0.9378,     Adjusted R-squared: 0.9371
F-statistic:  1356 on 1 and 90 DF,  p-value: < 2.2e-16
```

Figure 1.3. Listing of R code for computation equation 1.1.

Blank Page

Appendix 2—Selected Code and Summary Output from R Environment Showing Final Steps of Regression Process

Blank Page

For purposes of archival and verification of report content, this appendix documents some of the final steps used in the development of the $I_A{}^{\&}C_L$ equations and $I_A{}^{\&}C_L$ trees. This appendix also includes a summary of critical values for the t-distribution pertinent to $I_A{}^{\&}C_L$ equations. Finally, the appendix ends with an example evaluation of $I_A{}^{\&}C_L$ equations. The data used for regression analysis are listed in table 1.1 of appendix 1.

Summary of the $I_A{}^{\&}C_L$ Equations

The development of the $I_A{}^{\&}C_L$ equations required considerable exploratory analysis and judgement. In turn this knowledge influenced $I_A{}^{\&}C_L$ tree development. This section highlights selected code and summary output from the R environment as a means to archive the final regression process. For the discussion here, assume that the regression weights (`W`) are available and are based on the number of suitable storms per watershed (column 2 of table 1.1). The data objects are the `IA.final` and `CL.final` **`list`** objects, and each object contains the final regression models. Figure 6 provides an example of a **`list`** for the I_A and C_L **`list`** objects, respectively.

The R environment requires rescaling of regression weights so their sum equals the number of data points for appropriate computation of the residual standard error (σ). This computation was made using the code in figure 2.1. The code was used to compute `W`. This vector of weights also is shown in figure 6.

```
MLRweights <- function(avector) {
  tmp = length(avector)/sum(avector)
  return (tmp*avector)
}
```

Figure 2.1. Listing of R code for computation of regression weights.

The R function **`summary`** provides a compact and complete summary of an arbitrary regression model. However, this function lacks a report of the inverted covariance matrix and the maximum leverage, which are needed to compute prediction limits or evaluate equation applicability, respectively. These computations require matrix operations. Therefore, the `PostRegressComps` function was written to report this information. The `PostRegressComps` function is shown in figure 2.2. The code shown in figure 2.1 was used to compute the `W` shown in figure 2.2. The next two sections summarize the regression output for the $I_A{}^{\&}C_L$ equations (eqs. 23 and 29). The last section of this appendix describes some analysis of the regression residuals.

```
PostRegressComps <- function(themodel) {
   print(summary(themodel))
   WGT  <- diag(W) # convert weight vector to diagonal of matrix
   X  <- model.matrix(themodel)
   Xt <- t(X) # transposition of matrix X
   invcov  <- chol2inv(chol(Xt %*% WGT %*% X))
   hat.max <- max(diag(X %*% invcov %*% Xt)) # the maximum leverage
   cat(c("Inverted␣Covariance␣Matrix␣=\n"))
   print(invcov)
   cat(c("Max␣hat␣=",hat.max,"\n"))
}
```

Figure 2.2. Listing of R code for summary of major regression computations shown in figures 2.3 and 2.4.

Computation of the I_A Equation

The R environment output used for the I_A equation and ancillary results in the text is listed in figure 2.3. The `IA.final$EQUATION` was provided by the `makeIAeq` function shown in figure 6. For the context here, the vectors `W`, `RK`, `DU`, `CN`, `L`, and `IA` were derived from columns 2–5, 7, and 9 of table 1.1 in appendix 1.

```
# The usual syntax highlighting in this report is purposefully suppressed in the figure.
PostRegressComps(IA.final$EQUATION)

Call:
lm(formula = IA ~ L + DU + RK + CN, weights = W)

Residuals:
      Min        1Q    Median        3Q       Max
-0.669438 -0.181769 -0.006319  0.190959  0.869198

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.044544   0.466631   4.382 3.31e-05 ***
L           -0.549662   0.192485  -2.856 0.005383 **
D           -0.194257   0.079413  -2.446 0.016476 *
R            0.241418   0.065870   3.665 0.000427 ***
CN          -0.013542   0.006151  -2.202 0.030359 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3025 on 86 degrees of freedom
Multiple R-Squared: 0.374,      Adjusted R-squared: 0.3448
F-statistic: 12.84 on 4 and 86 DF,  p-value: 3.064e-08

Inverted Covariance Matrix =
            [,1]         [,2]         [,3]          [,4]          [,5]
[1,]  2.37999804  0.175708102  0.212563891 -0.0451175500 -0.0310582490
[2,]  0.17570810  0.404970626  0.005791072 -0.0107367190 -0.0034822895
[3,]  0.21256389  0.005791072  0.068930682 -0.0114055443 -0.0030351045
[4,] -0.04511755 -0.010736719 -0.011405544  0.0474242713  0.0004221009
[5,] -0.03105825 -0.003482289 -0.003035104  0.0004221009  0.0004134964
Max hat = 0.272324981146127
```

Figure 2.3. Listing of R code for computation of I_A equation (eq. 23).

Computation of the C_L Equation

The R environment output used for the C_L equation and ancillary results in the text is listed in figure 2.4. The `CL.final$EQUATION` was provided by a `makeCLeq` function (not shown in this report) that is analogous to `makeIAeq` shown in figure 6. The difference between the two functions is the absence of D in the `makeCLeq`. For the context here, the vectors `W`, `RK`, `CN`, `L`, and `CL` were derived from columns 2–3, 5, 7, and 10 of table 1.1 in appendix 1.

```
# The usual syntax highlighting in this report is purposefully suppressed in the figure.
PostRegressComps(CL.final$EQUATION)

Call:
lm(formula = CL ~ L + RK + CN, weights = W)

Residuals:
     Min       1Q   Median       3Q      Max
-0.56178 -0.17150 -0.04103  0.11943  0.73977

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.534955   0.442787   5.725 1.45e-07 ***
L           -0.482038   0.112350  -4.290 4.61e-05 ***
R            0.227130   0.056436   4.025 0.000121 ***
CN          -0.016763   0.004444  -3.772 0.000295 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2649 on 87 degrees of freedom
Multiple R-Squared: 0.3241,     Adjusted R-squared: 0.3008
F-statistic:  13.9 on 3 and 87 DF,  p-value: 1.742e-07

Inverted Covariance Matrix =
            [,1]         [,2]          [,3]          [,4]
[1,]  2.79335601 -0.450888349 -0.0155384759 -0.0259975911
[2,] -0.45088835  0.179840242  0.0037527131  0.0022169510
[3,] -0.01553848  0.003752713  0.0453789744 -0.0001118594
[4,] -0.02599759  0.002216951 -0.0001118594  0.0002814360
Max hat = 0.183227906877118
```

Figure 2.4. Listing of R code for computation of C_L equation (eq. 29).

Abbreviated Residual Analysis for the I_A&C_L Equations

An abbreviated analysis of the residuals of the I_A&C_L equations is shown in this section. A listing of R code for computation of the residuals for the I_A&C_L equations is shown in figure 2.5, and the residual plots for I_A and C_L, which were produced directly from the R environment, are shown in figures 2.6 and 2.7. Figure 2.6 shows that some asymmetry in the I_A residuals is present; there is a tendency for more large positive residuals, which indicates some chance of underestimation of I_A. Figure 2.7 shows some evidence for increasing variance with increasing C_L. The extreme outlier in the upper right of figure 2.7 represents station 08158825, which was not used to develop either of the I_A&C_L equations. It is concluded that the residual plots are acceptable.

```
getFittedIAvalues <- function(model) {
  thepower <- model$POWER; coes <- model$WLS$coefficients
  myfitted <- coes[1] + coes[2]*L^thepower + coes[3]*DU + coes[4]*RK + coes[5]*CN
  return(myfitted)
}
getFittedCLvalues <- function(model) {
  thepower <- model$POWER; coes <- model$WLS$coefficients
  myfitted <- coes[1] + coes[2]*L^thepower + coes[3]*RK + coes[4]*CN
  return(myfitted)
}
IA.fit <- getFittedIAvalues(IA.outlier);
CL.fit <- getFittedCLvalues(CL.outlier)
pdf("ia_residual_plot.pdf")
plot(IA.fit,ia-IA.fit,
     xlim=c(0.3,1.3),
     ylim=c(-1,1),
     xlab="FITTED␣INITIAL␣ABSTRACTION,␣IN␣INCHES",
     ylab="RESIDUAL␣(Observed␣-␣Modeled),␣IN␣INCHES")
abline(0,0)
dev.off()
pdf("cl_residual_plot.pdf")
plot(CL.fit,cl-CL.fit,
     xlim=c(0.1,1.1),
     ylim=c(-2,2)
     xlab="FITTED␣CONSTANT␣LOSS,␣IN␣INCHES␣PER␣HOUR",
     ylab="RESIDUAL␣(Observed␣-␣Modeled),␣IN␣INCHES␣PER␣HOUR")
abline(0,0)
dev.off()
```

Figure 2.5. Listing of R code for computations of the residuals for the I_A&C_L equations shown in figures 2.6 and 2.7.

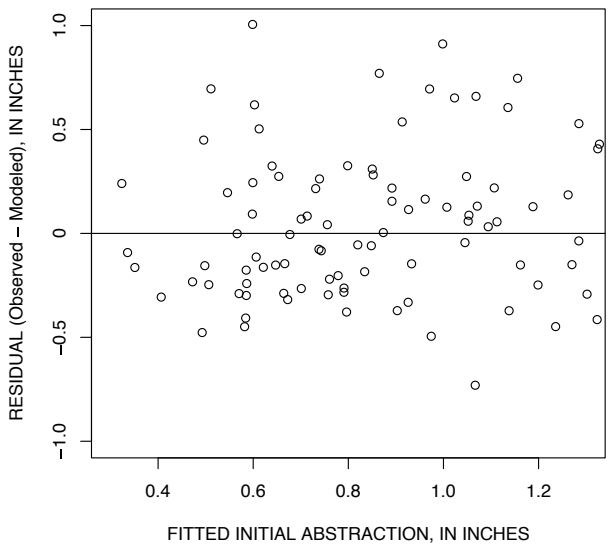


Figure 2.6. Residual plot for the 92 watersheds of I_A equation.

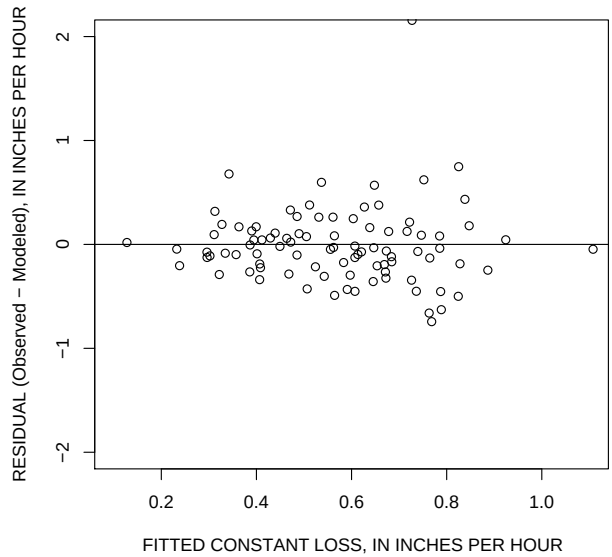


Figure 2.7. Residual plot for the 92 watersheds of C_L equation.

Computation of Selected Upper-Tail Values of t-Distribution

The upper-tail values of the t-distribution are used to compute the prediction intervals for the $K^{\&}T_p$ equations (eqs. 5 and 11) and for the $I_A{}^{\&}C_L$ equations (eqs. 23 and 29). The prediction intervals are shown in equations 7 and 8 (for K), equations 13 and 14 (for T_p), equations 25 and 26 (for I_A), and equations 31 and 32 (for C_L). Three unique degrees of freedom (df) are required: 86, 87, and 88. Prediction intervals in percent were selected, and the upper-tail probabilities ($\alpha/2$) were computed. The upper-tail values for the t-distribution are listed in table 2.1; the values were derived from computations made by R environment code shown in figure 2.8.

Table 2.1. Selected upper-tail values of t-distribution for three degrees of freedom.

[α, $100 \times (1-\alpha)$ prediction interval of a regression equation; $t_{[\alpha/2,df]}$, upper-tail values of t-distribution for the degree of freedom (df)]

Prediction interval (percent)	$\alpha/2$	$t_{[\alpha/2,86]}$	$t_{[\alpha/2,87]}$	$t_{[\alpha/2,88]}$
99	0.005	2.6342	2.6335	2.6329
98	.010	2.3705	2.3700	2.3695
95	.025	1.9879	1.9876	1.9873
90	.050	1.6628	1.6626	1.6624
80	.100	1.2915	1.2914	1.2912
75	.125	1.1582	1.1581	1.1580
50	.250	.6774	.6773	.6773

```
options(digits=4) # four decimals on output ↩
    ↪table
pinterval <- c(99, 98, 95, 90, 80, 75, 50) #↩
    ↪ in percent
zero.p5.alpha <- (1-(pinterval/100))/2
data.frame(INTERVAL=pinterval,ALPHAdiv2=zero↩
    ↪.p5.alpha, T86=qt(zero.p5.alpha,86,↩
    ↪lower.tail=FALSE), T87=qt(zero.p5.↩
    ↪alpha,87,lower.tail=FALSE), T88=qt(↩
    ↪zero.p5.alpha,88,lower.tail=FALSE))
  INTERVAL ALPHAdiv2    T86    T87    T88
1       99     0.005 2.6342 2.6335 2.6329
2       98     0.010 2.3705 2.3700 2.3695
3       95     0.025 1.9879 1.9876 1.9873
4       90     0.050 1.6628 1.6626 1.6624
5       80     0.100 1.2915 1.2914 1.2912
6       75     0.125 1.1582 1.1581 1.1580
7       50     0.250 0.6774 0.6773 0.6773
```

Figure 2.8. Listing of R code for computation of selected values of t-distribution for use in prediction-interval computations.

Summary of the $I_A{}^{\&}C_L$ Trees

The purpose of this section is to archive some of the final steps used in the development of the $I_A{}^{\&}C_L$ trees (figs. 7 and 8). A listing of R code for computation of the $I_A{}^{\&}C_L$ trees and the residuals of the trees is shown in figures 2.9 and 2.10 for I_A and C_L, respectively. The same watershed characteristics as used for the respective regression equation were used. For example, L, D, R, and CN were statistically significant for the I_A equation and therefore used for the tree construction. However, the tree regression did not identify D as useful for I_A prediction. Thus, D was not used for the C_L tree construction.

For purposes of archival and verification of report content, the raw output of the R environment of the trees is shown in figures 2.11 and 2.12. These figures subsequently were redrafted for the report to become figures 7 and 8 with the addition of the residual standard error σ and the sample size n. The residual standard error and sample size are produced by the **summary**(ia.tree.prune) and **summary**(cl.tree.prune). The residuals are shown in figures 2.13 and 2.14. For these examples, the vectors W, RK, DU, CN, L, IA, and CL were derived from columns 2–5, 7, and 9–10 of table 1.1. Additional directives and details regarding the R code are listed throughout figures 2.9 and 2.10.

```
library(rpart) # load library for tree building
Outliers <- c(69) # remove station 08158825
ia.tree <- rpart(IA[-Outliers]~L[-Outliers]+
                 DU[-Outliers]+RK[-Outliers]+CN[-Outliers],
                 weights=MLRweights(W[-Outliers]))
#plot(ia.tree,branch=.4) # hide this plot
#text(ia.tree)
#mtext("Initial Abstraction--no prune")

# printcp(ia.tree) # commented out to save space
ia.tree.prune <- prune.rpart(ia.tree,0.032) # prune a limb
# summary(ia.tree.prune) # commented out to save space

pdf("ia_tree.pdf")
plot(ia.tree.prune,branch=.4)
text(ia.tree.prune)
mtext("Initial␣Abstraction")
dev.off()

pdf("ia_tree_residual_plot.pdf")
plot(IA[-Outliers]-residuals(ia.tree.prune),
     residuals(ia.tree.prune),
     xlab="FITTED␣VALUES,␣IN␣INCHES",
     ylab="RESIDUAL␣(Observed␣-␣Model),␣IN␣INCHES",
     xlim=c(0.3,1.3),
     ylim=c(-1,1))
points(1.17,IA[69]-1.17) # add marker for station 08158825
abline(0,0)
dev.off()
```

Figure 2.9. Listing of R code for computation of the I_A tree and residual plot.

```
library(rpart) # load library for tree building
Outliers <- c(69) # remove station 08158825
cl.tree <- rpart(CL[-Outliers]~L[-Outliers]+
                 RK[-Outliers]+CN[-Outliers],
                 weights=MLRweights(W[-Outliers]))
#plot(cl.tree,branch=.4)
#text(cl.tree)
#mtext("Constant Loss--no prune")

printcp(cl.tree) # commented out to save space
cl.tree.prune <- prune.rpart(cl.tree,0.03) # prune a limb
summary(cl.tree.prune) # commented out to save space

pdf("cl_tree.pdf")
plot(cl.tree.prune,branch=.4)
text(cl.tree.prune)
mtext("Constant␣Loss")
dev.off()

pdf("cl_tree_residual_plot.pdf")
plot(CL[-Outliers]-residuals(cl.tree.prune),
     residuals(cl.tree.prune),
     xlab="FITTED␣VALUES,␣IN␣INCHES␣PER␣HOUR",
     ylab="RESIDUAL␣(Observed␣-␣Model),␣IN␣INCHES␣PER␣HOUR",
     xlim=c(0.1,1.1),
     ylim=c(-2,2))
points(0.248,CL[69]-0.248,col=2) # The residual equals 2.64 inches per hour for station ↩
    ↪08158825 and is beyond graph limits, but the limits are set to match those of the ↩
    ↪residual plot for the CL equation
abline(0,0)
dev.off()
```

Figure 2.10. Listing of R code for computation of the C_L tree and residual plot.

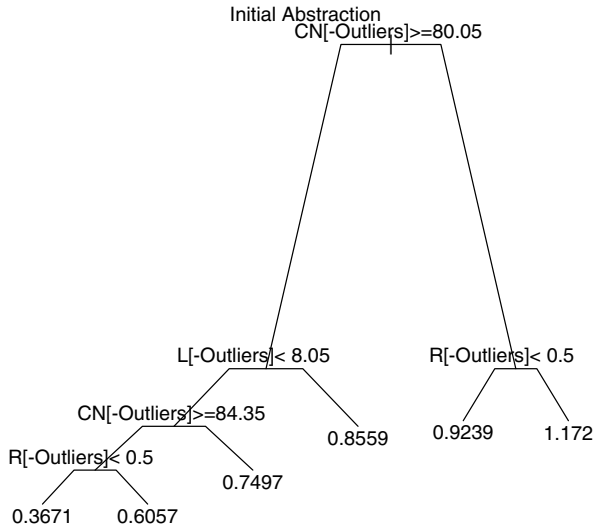


Figure 2.11. Diagram showing raw I_A tree from R environment for the 92 watersheds used to generate figure 7.

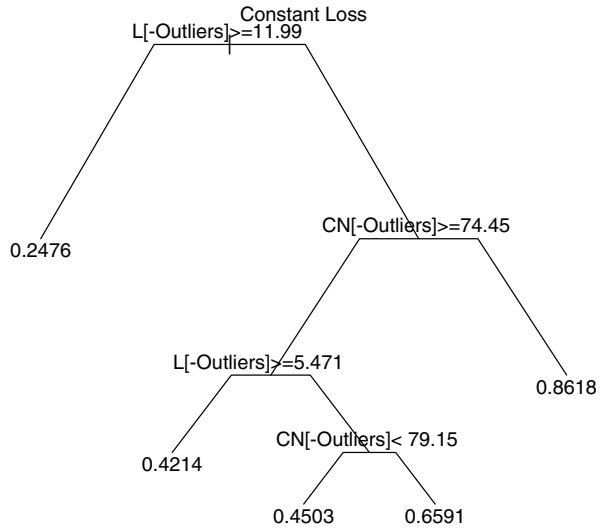


Figure 2.12. Diagram showing raw C_L tree from R environment for the 92 watersheds used to generate figure 8.

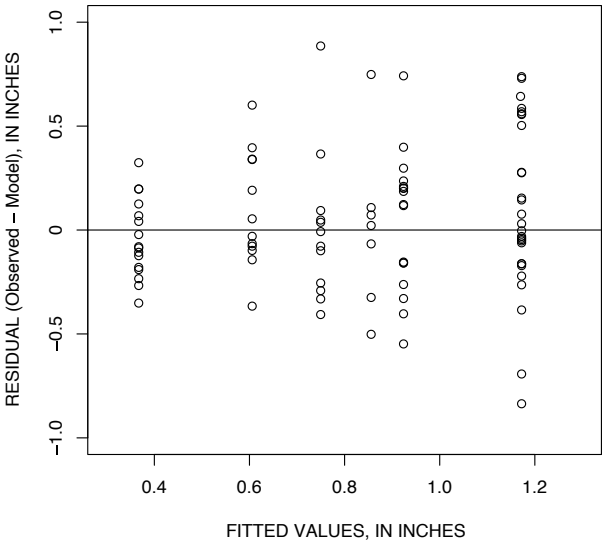


Figure 2.13. Graph showing residual plot for the 92 watersheds of I_A tree.

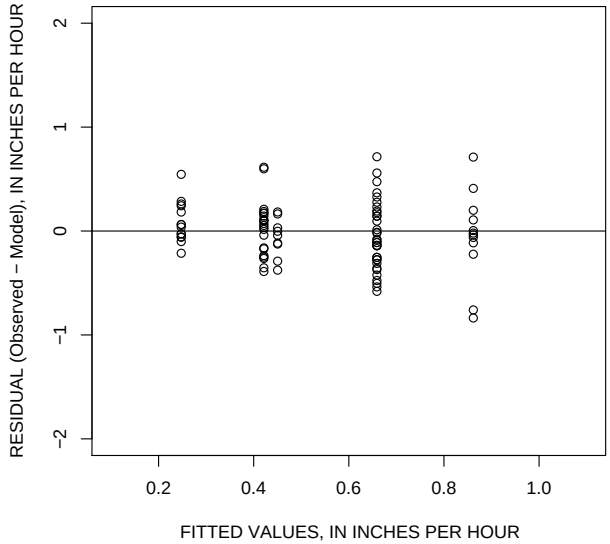
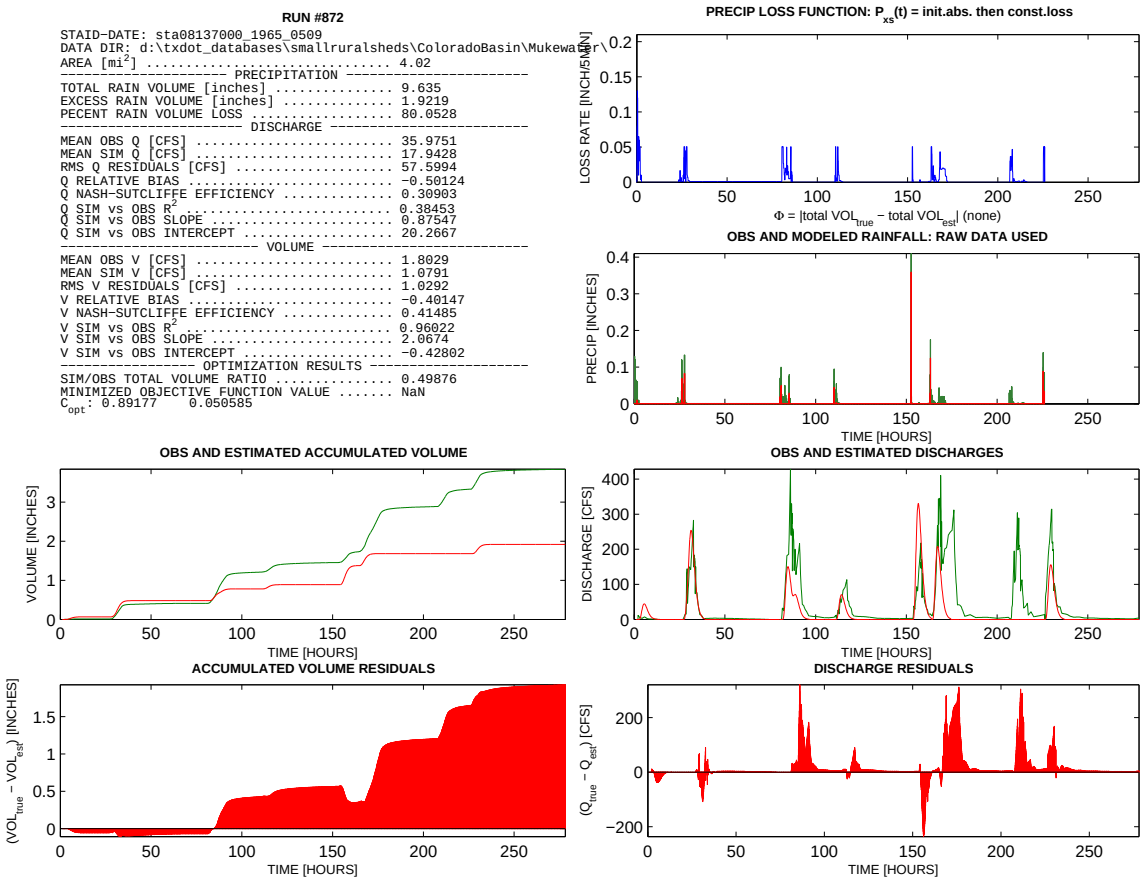


Figure 2.14. Graph showing residual plot for the 92 watersheds of C_L tree.

Example of Graphical Output From Modified ${}^{\natural}I_A{}^{\&}{}^{\natural}C_L$ Software for Evaluation of the $I_A{}^{\&}C_L$ Equations

An example of direct graphical output from the modified ${}^{\natural}I_A{}^{\&}{}^{\natural}C_L$ software (Joseph Vrabel, U.S. Geological Survey, written commun., 2007; source code is available from report web site) is shown in this section. The "original" ${}^{\natural}I_A{}^{\&}{}^{\natural}C_L$ software was modified so that I_A and C_L values from each of the four techniques (introduced on page 22) and the combined $I_A{}^{\&}C_L$ model (introduced on page 23) could be evaluated. The modified software computed the Q_p^{model}, V^{model}, and $T^{Q_p^{\text{model}}}$ so that the residuals ε^{Q_p}, ε^V, and $\varepsilon^{T_{Q_p}}$ could be computed.

This section provides an example output from the modified software in which the I_A and C_L for one of the 92 watersheds were estimated by the $I_A{}^{\&}C_L$ equations (one of the four techniques). The software then computed a model of the runoff hydrograph from the input rainfall. It is important to emphasize the modeled runoff hydrograph has not been constrained to match Q_p^{obs}, V^{obs}, and $T^{Q_p^{\text{obs}}}$. The output for a storm that occurred on May 9, 1965, for station 08137000 is shown in figure 2.15. The numerous intermediate results are shown in the upper-left corner of the figure. These results are legacy output from the ${}^{\natural}I_A{}^{\&}{}^{\natural}C_L$ software. Red lines represent the results of the modeled runoff hydrograph, and green lines represent observed data. For this example, the $I_A{}^{\&}C_L$ equations and $K^{\&}T_p$-GUH produce six distinct peaks. Some peaks (for example, second from left) are well modeled; whereas other peaks (for example, second from right) are entirely absent in the modeled runoff hydrograph. The example, although unusual in the database because of complexity and length (temporal duration), provides an essentially worst-case example. For comparative purposes, an essentially best-case example for a storm that occurred on May 30, 1976, for station 08048550 is shown in figure 2.16.



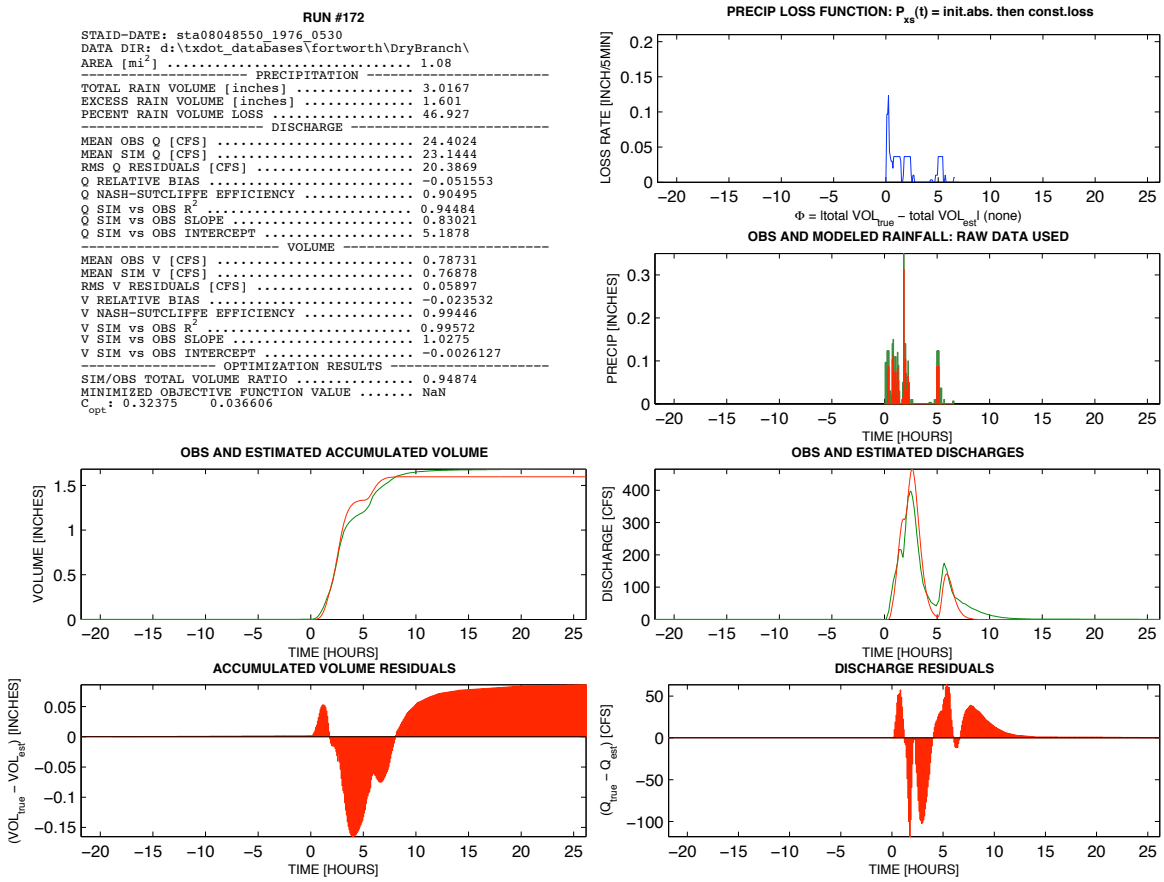
EXPLANATION

- Rainfall data
- Modeled data (ESTIMATED)
- Observed data (OBS)
- C_{opt} Storm-specific, initial-abstraction ($^{\ell}I_A$) and constant-loss ($^{\ell}C_L$) values

ABBREVIATIONS

CFS	cubic feet per second
OBS	observed
Q	streamflow (discharge)
SIM	simulated (modeled)
V	volume (watershed depth)

Figure 2.15. Graphs showing selected worst-case graphical output showing multiple peaks from the modified $^{\ell}I_A$ & $^{\ell}C_L$ software using the I_A & C_L equations.



EXPLANATION

- Rainfall data
- Modeled data (ESTIMATED)
- Observed data (OBS)
- C_{opt} Storm-specific, initial-abstraction ($^{\&}I_A$) and constant-loss ($^{\&}C_L$) values

ABBREVIATIONS

CFS	cubic feet per second
OBS	observed
Q	streamflow (discharge)
SIM	simulated (modeled)
V	volume (watershed depth)

Figure 2.16. Graphs showing selected best-case graphical output from the modified $^{\&}I_A{}^{\&}C_L$ software using the $I_A{}^{\&}C_L$ equations.

Blank Page

Appendix 3—Algorithm Listings for Computations of Gamma Unit Hydrographs and the Initial-Abstraction, Constant-Loss Model

Blank Page

The R environment (R Development Core Team, 2006) is a multi-platform, open-source, statistical, and programming environment that is well suited for demonstration of computational topics discussed in this report. Therefore, this appendix contains several topical and useful code listings to aid in implementation of this report.

Computation of q_p From T_p and K Values of Gamma Unit Hydrograph

The $K^{\&}T_p$ equations define the GUH; however, the value for q_p requires computation. A code listing for computation of q_p from T_p and K values using equation 4 is provided in figure 3.1. The code example shows that q_p for $K = 3.56$ and $T_p = 6.5$ hours is about 0.113 watershed inch per hour.

```
getqp <- function(K,Tp) {
   Vfunc.qp <- function(qp) { return( abs(qp * Tp * gamma(K) * (exp(1)/K)^K - 1) ) }
   root <- optimize( Vfunc.qp, interval = c(0.001,20) )
   return(root$minimum)
}
K  <- 3.56; Tp <- 6.5; # hours
qp <- getqp(K,Tp)
print(qp)
[1] 0.113133
```

Figure 3.1. Listing of R code for solution of equation 4 given q_p and K values.

Computation of Excess Rainfall for $I_A{}^{\&}C_L$ Model

Computation of excess rainfall from rainfall using the $I_A{}^{\&}C_L$ model is shown by the algorithm listed in figure 3.2. The code provides the `useIACL` function that takes a vector of rainfall (`rain`), the interval in seconds of time (`delT`), the I_A (`IA`), and C_L (`CL`). Example output from the `useIACL` function for arbitrary rainfall and selected input parameters is shown at the bottom of figure 3.2. Algorithmic description for a 1-hour `delT=3600` follows:

1. Compute the cumulative sum of `rain`.	If `rain = c(1,2,3)`, then `sumrain = c(1,3,6)`.
2. Substract `IA` from `sumrain` (`tmp.rain = sumrain-IA`)	If `IA = 1.25`, then `tmp.rain = c(-0.25,1.75,4.75)`.
3. Correct `tmp.rain` to non-negative values.	Then `tmp.rain = c(0,1.75,4.75)`.
4. Compute `excess.rain[i] = tmp.rain[i]-tmp.rain[i-1]`.	Then `excess.rain = c(0,1.75,3)`.
5. Compute excess rainfall `excess.rain` by subtraction of `CL`	If `CL = 1.85`, then `excess.rain = excess.rain-CL` and `excess.rain = c(0,-0.10,1.15)`.
6. Correct `excess.rain` to non-negative values	Then `excess.rain = c(0,0,1.15)`.

```
useIACL <- function(rain,delT,IA,CL) {
  # rain is a vector of incremental rainfall in inches.
  # delT is the time step in seconds; Ia is in inches; Cl is in watershed inches per hour.
  CL.per.unittime <- CL*delT/3600 # unit conversion
  n <- length(rain) # how many items in the vector.

  # Push a zero onto the vector and recompute n; a "seq(2,n)" is used later.
  if(n == 1) {
     rain <- c(rain,0)
     n <- length(rain)
  }

  # Compute a vector of the cumulative sum of rainfall.
  sumrain <- cumsum(rain)

  # If length 0 (not a vector?) just return whatever was passed to the function
  if(length(sumrain) == 0) { return(rain) }

  # If the total rainfall (last index) is zero return the vector of no rainfall
  if(sumrain[length(sumrain)] == 0) { return(rain) }

  # From the cumulative sum, substract (offset) the vector by the initial abstraction.
  tmp.rain <- sumrain - IA

  # Now make tmp.rain physically meaningful; < 0 become 0.
  tmp.rain <- sapply(tmp.rain, function(x) { if(x < 0) { return(0) }
                                                 else { return(x) } })

  excess.rain    <- vector(mode = "numeric") # declare new vector
  excess.rain[1] <- tmp.rain[1] # initialize first index

  # The excess.rain is to become incremental rainfall
  # (recall that tmp.rain is cumulative).
  for(i in seq(2,n)) { excess.rain[i] <- tmp.rain[i] - tmp.rain[i-1] }

  # Now apply the constant loss in the correct units
  for(i in seq(1,n)) { excess.rain[i] <- excess.rain[i] - CL.per.unittime }

  # Now make the excess vector physically meaningful; < 0 become 0.
  excess.rain <- sapply(excess.rain, function(x) { if(x < 0) { return(0) }
                                                       else { return(x) } })
  return(excess.rain) # return the excess rainfall vector to the user
}
useIACL(c(1,1,1,1),900,1.5,0)
# [1] 0.0 0.5 1.0 1.0

useIACL(c(1,1,1,1),900,0,2)
# [1] 0.5 0.5 0.5 0.5

useIACL(c(1,2,4,0,.4,0),300,.76,.45)
# [1] 0.2025 1.9625 3.9625 0.0000 0.3625 0.0000
```

Figure 3.2. Listing of R code for computation of excess rainfall using an initial-abstraction, constant-loss ($I_A{}^{\&}C_L$) model.

Computation of Gamma Unit Hydrograph Ordinates

A listing of code for computation of individual ordinates of the GUH (eq. 2) when values for q_p, T_p, and K are available, or when only q_p and T_p are available, is provided in figure 3.3. A demonstration of code usage also is provided. The code produces a GUHt object, which retains the t, $q(t)$, q_p, T_p, and K values. The GUH is constrained to unit volume ($V = 1$) according to equation 4 when q_p and T_p are given (as done in the listing). However, the GUH function does not verify that $V = 1$ when q_p, T_p, and K are given. This is done purposefully and with the intent of the GUH fitting into a larger algorithmic framework. Processing efficiency is increased because the **uniroot** function is bypassed. The code example shows that for $t = 2.5$ hours into the event, the unit hydrograph ordinate $q(2.5) = 0.276$ watershed inch per hour.

```
GUH <- function(tt, qp, Tp, K=NULL) {
  if(is.null(K)) {
    Vfunc.K <- function(K) { return(1 - qp * Tp * gamma(K) * (exp(1)/K)^K ) }
    root <- uniroot(Vfunc.K, c(0.001,20)) # K on range [0.001,20]
    K    <- root$root # extract K
  }
  ttp  <- tt/Tp # compute once, then use twice
  guh  <- qp * ( ttp * exp(1 - ttp) )^K; # definitive equation for gamma hydrograph
  return( list(Time=tt, qT=guh,
               qp=qp, Tp=Tp, K=K) )
}
# demonstrate usage
qp <- 0.3 # watershed inches per hour
Tp <- 3 # hours
GUHt <- GUH(2.5,qp,Tp) # 2.5 hours into event
str(GUHt) # nice listing follows
List of 5
 $ Time: num 2.5
 $ qT  : num 0.276
 $ qp  : num 0.3
 $ Tp  : num 3
 $ K   : num 5.25
```

Figure 3.3. Listing of R code for computation of an ordinate of a gamma unit hydrograph (GUH) as defined by equation 2.

Blank Page

Appendix 4—Derivation of the Gamma Hydrograph Equation

Blank Page

At the request of the cooperator, the Texas Department of Transportation, a derivation of the gamma hydrograph equation (eq. 2) is provided in this appendix. The probability density function of the gamma distribution is

$$f(x) = \frac{1}{b\Gamma(c)}\left(\frac{x}{b}\right)^{c-1}\exp\left(\frac{-x}{b}\right), \tag{4.1}$$

where b is a scale parameter greater than zero, c is a shape parameter greater than zero, $\Gamma(x)$ is the gamma function with argument c, and x is a whole number. An alternative formulation of the gamma distribution in the context of a runoff hydrograph with a watershed functioning as a system of cascading reservoirs of water storage is

$$q(t) = \frac{1}{\tau\Gamma(n)}\left(\frac{t}{\tau}\right)^{n-1}\exp\left(\frac{-t}{\tau}\right), \tag{4.2}$$

where n is the number of reservoirs modeling the watershed, τ is a storage coefficient (or the residence time of a reservoir), $\Gamma(u)$ is the gamma function of n, t represents time, and $q(t)$ is the streamflow at time t. Because of the importance of peak streamflow Q_p and time of Q_p occurrence T^{Q_p}, it is useful to express $q(t)$ and t in terms of $q_p(T_P)$ and T_p. The gamma hydrograph can then be written as

$$q_p(T_p) = \frac{1}{\tau\Gamma(n)}\left(\frac{T_p}{\tau}\right)^{n-1}\exp\left(\frac{-T_p}{\tau}\right). \tag{4.3}$$

The first step in deriving the gamma hydrograph equation is to find τ in terms of t and n. The derivation of T_p in terms of t and n is made by taking the first derivative of $q(t)$ and setting the derivative equal to zero. The maximum of the function is produced. The derivative is

$$\frac{\mathrm{d}\,q_p(T_p)}{\mathrm{d}\,T_p} = \frac{1}{\tau\Gamma(n)}\left[(n-1)T_p^{n-2}\exp\left(\frac{-T_p}{\tau}\right) + T_p^{n-1}\left(\frac{-1}{\tau}\right)\exp\left(\frac{-T_p}{\tau}\right)\right], \tag{4.4}$$

$$\frac{\mathrm{d}\,q_p(T_p)}{\mathrm{d}\,T_p} = 0 = (n-1)T_p^{n-2}\exp\left(\frac{-T_p}{\tau}\right) - \frac{1}{\tau}T_p^{n-1}\exp\left(\frac{-T_p}{\tau}\right), \tag{4.5}$$

$$\frac{1}{\tau}T_p^{n-1}\exp\left(\frac{-T_p}{\tau}\right) = (n-1)T_p^{n-2}\exp\left(\frac{-T_p}{\tau}\right), \tag{4.6}$$

$$\frac{1}{\tau}T_p^{n-1} = (n-1)T_p^{n-2}, \tag{4.7}$$

$$\frac{T_p^{n-1}}{T_p^{n-2}} = \tau(n-1), \tag{4.8}$$

$$T_p = \tau(n-1), \tag{4.9}$$

$$\tau = \frac{T_p}{n-1}. \tag{4.10}$$

The gamma hydrograph equation is produced by the ratio

$$\frac{q(t)}{q_p(T_p)} = \frac{1}{\tau\Gamma(n)}\left(\frac{t}{\tau}\right)^{n-1}\exp\left(\frac{-t}{\tau}\right) \Big/ \frac{1}{\tau\Gamma(n)}\left(\frac{T_p}{\tau}\right)^{n-1}\exp\left(\frac{-T_p}{\tau}\right), \tag{4.11}$$

$$= \frac{\left(\frac{t}{\tau}\right)\exp\left(\frac{-t}{\tau}\right)}{\left(\frac{T_p}{\tau}\right)\exp\left(\frac{-T_p}{\tau}\right)}. \tag{4.12}$$

Next, τ is substituted in terms of T_p and n

$$\frac{q(t)}{q_p(T_p)} = \left(\frac{t}{T_p}\right)^{n-1}\frac{\exp\left(\frac{-t}{T_p}(n-1)\right)}{\exp\left(\frac{-T_p}{T_p}(n-1)\right)}, \tag{4.13}$$

$$= \left(\frac{t}{T_p}\right)^{n-1}\frac{\exp\left(\frac{-t}{T_p}(n-1)\right)}{\exp(n-1)}, \tag{4.14}$$

$$= \left(\frac{t}{T_p}\right)^{n-1}\left[\exp\left(\frac{-t}{T_p}(n-1)-(1-n)\right)\right], \tag{4.15}$$

$$= \left(\frac{t}{T_p}\right)^{n-1}\left[\exp\left((n-1)\left(\frac{-t}{T_p}+1\right)\right)\right], \tag{4.16}$$

$$= \left[\frac{t}{T_p}\exp\left(1-\frac{t}{T_p}\right)\right]^{n-1}. \tag{4.17}$$

Let $K = n-1$, and finally the gamma hydrograph equation results after dropping the (T_p) term for q_p because q_p occurs at T_p and "exp" is replaced with the more conventional "e":

$$\frac{q(t)}{q_p} = \left[\frac{t}{T_p}\mathrm{e}^{1-\left(\frac{t}{T_p}\right)}\right]^{K}. \tag{4.18}$$

Appendix 5—Example Computations of Prediction Limits From Regression Equations

Blank Page

This appendix shows example computations of the prediction limits for the $K^{\&}T_p$ and $I_A{}^{\&}C_L$ equations. Suppose that a hypothetical watershed has the following characteristics: $L_\oplus = 7.8$ miles, $S_\oplus = 0.003$, $D_\oplus = 0$, $R_\oplus = 1$, and $CN_\oplus = 75$. Suppose that the analyst needs the 90-percent prediction limits of the K (eq. 5), T_p (eq. 11), I_A (eq. 23), and C_L (eq. 29) equations.

The 90-percent prediction limits require the $100 \times (1-\alpha)$ limits and thus $\alpha = 0.10$. The leverage $h_\oplus$ for each of the four equations requires matrix computations, and the $h_\oplus$ are shown in the following sections. The residual standard error σ of the four equations are $\sigma^{[K]} = 0.2052$ ($\log_{10}$), $\sigma^{[T_p]} = 0.1383$ ($\log_{10}$), $\sigma^{[I_A]} = 0.3025$ watershed inch, and $\sigma^{[C_L]} = 0.2649$ watershed inch per hour. The degrees of freedom (df) of the four equations are $\mathrm{df}^{[K]} = 88$, $\mathrm{df}^{[T_p]} = 87$, $\mathrm{df}^{[I_A]} = 86$, and $\mathrm{df}^{[C_L]} = 87$. These values along with α are used to compute the value $\mathrm{t}_{[\alpha/2,\mathrm{df}]}$ for the t-distribution (see table 2.1). The three unique values for the t-distribution for the four prediction limits given $\alpha = 0.10$ and df are $\mathrm{t}_{[0.050,86]} = 1.6628$ (I_A equation), $\mathrm{t}_{[0.050,87]} = 1.6626$ (T_p and C_L equations), and $\mathrm{t}_{[0.050,88]} = 1.6624$ (K equation).

The $100 \times (1-\alpha)$ prediction limits have a specific interpretation. The prediction limits define an interval for which an expression of probability is made—in the examples here, the 90-percent prediction interval. This interval is interpreted as follows (Good and Hardin, 2003, p. 101): For a large number of prediction limits, there is a 90-percent chance that the true value of K, T_p, I_A, or C_L for a given watershed is between the limits. It is not correct to conclude that the probability of the true value being included in the 90-percent interval is 90 percent for any single prediction. As a result, for 10 percent of these limits, the true value is either less than or greater than the lower and upper limits, respectively.

A summary of the results of the computations for the hypothetical watershed shown in the following four sections follows: (1) The 90-percent prediction limits for the $K_\oplus$ are 2.19 and 10.8 with a prediction of 4.86. (2) The limits for the $T_{P\oplus}$ are 3.2 and 9.56 hours with a prediction of 5.53 hours. (3) The limits for $I_{A\oplus}$ are 0.68 and 1.70 watershed inches with a prediction of 1.19 watershed inches. (4) The limits for $C_{L\oplus}$ are 0.28 and 1.18 watershed inches per hour with a prediction of 0.73 watershed inch per hour.

Computation of Unit Hydrograph Shape and Prediction Limits

For the hypothetical watershed, $K_\oplus$ is computed by substitution of $D_\oplus$ and $L_\oplus$ into equation 5:

$$K_\oplus = 10^{0.560-0.249D_\oplus} L_\oplus^{0.142},$$
$$K_\oplus = 10^{0.560-0.249(0)}(7.8)^{0.142}, \text{ and}$$
$$K_\oplus = 4.86 \text{ dimensionless.}$$

The leverage $h_\oplus^{[K]}$ is computed by substitution of $D_\oplus$ and $L_\oplus$ into equation 9:

$$h_\oplus^{[K]} = \begin{bmatrix} 1 & \log_{10}(7.8) & 0 \end{bmatrix} \times \begin{bmatrix} 0.10599 & -0.10349 & -0.03134 \\ -0.10349 & 0.13156 & 0.00859 \\ -0.03134 & 0.00859 & 0.04502 \end{bmatrix} \times \begin{bmatrix} 1 \\ \log_{10}(7.8) \\ 0 \end{bmatrix},$$

which results in $h_{\oplus}^{[K]} = 0.02604$. R code demonstrating the matrix multiplications and the resulting $h_{\oplus}^{[K]}$ is shown in figure 5.1.

```
cv <- c(1, log10(7.8), 0) # a column vector
rv <- t(cv) # transpose the vector into a row vector
h <- matrix(c( 0.10599, -0.10349, -0.03134,
              -0.10349,  0.13156,  0.00859,
              -0.03134,  0.00859,  0.04502),nrow=3,ncol=3)
rv %*% h %*% cv # perform matrix multiplications
          [,1]
[1,] 0.02604402
```

Figure 5.1. Listing of R code for computation of leverage for K equation.

Finally, the lower (eq. 7) and upper (eq. 8) 90-percent prediction limits are computed by

$$
\begin{aligned}
{}^{\downarrow}K_{\oplus} &= 10^{\log_{10}(4.86) - 1.6624 \times 0.2052\sqrt{1+0.02604}} = 2.19 \text{ and} \\
{}^{\uparrow}K_{\oplus} &= 10^{\log_{10}(4.86) + 1.6624 \times 0.2052\sqrt{1+0.02604}} = 10.8\ .
\end{aligned}
$$

Computation of Unit Hydrograph Time to Peak and Prediction Limits

For the hypothetical watershed, $T_{p\oplus}$ is computed by substitution of $D_{\oplus}$, $L_{\oplus}$, and $S_{\oplus}$ into equation 11:

$$
\begin{aligned}
T_{p\oplus} &= 10^{-1.49 - 0.354 D_{\oplus}} L_{\oplus}^{0.602} S_{\oplus}^{-0.672}, \\
T_{p\oplus} &= 10^{-1.49 - 0.354(0)} (7.8)^{0.602} (0.003)^{-0.672}, \text{ and} \\
T_{p\oplus} &= 5.53 \text{ hours.}
\end{aligned}
$$

The leverage $h_{\oplus}^{[T_p]}$ is computed by substitution of $D_{\oplus}$, $L_{\oplus}$, and $S_{\oplus}$ into equation 15:

$$
\begin{aligned}
h_{\oplus}^{[T_p]} = &\begin{bmatrix} 1 & \log_{10}(7.8) & \log_{10}(0.003) & 0 \end{bmatrix} \times \\
&\begin{bmatrix} 1.34775 & 0.16646 & 0.67709 & -0.01451 \\ 0.16646 & 0.19025 & 0.14719 & 0.01225 \\ 0.67709 & 0.14719 & 0.36919 & 0.00918 \\ -0.01451 & 0.01225 & 0.00918 & 0.04524 \end{bmatrix} \times \\
&\begin{bmatrix} 1 \\ \log_{10}(7.8) \\ \log_{10}(0.003) \\ 0 \end{bmatrix},
\end{aligned}
$$

which results in $h_{\oplus}^{[T_P]} = 0.06704$. R code demonstrating the matrix multiplications and the resulting $h_{\oplus}^{[T_p]}$ is shown in figure 5.2.

```
cv <- c(1, log10(7.8), log10(0.003), 0) # a column vector
rv <- t(cv) # transpose the vector into a row vector
h <- matrix(c( 1.34775,  0.16646,  0.67709, -0.01451,
               0.16646,  0.19025,  0.14719,  0.01225,
               0.67709,  0.14719,  0.36919,  0.00918,
              -0.01451,  0.01225,  0.00918,  0.04524),nrow=4,ncol=4)
rv %*% h %*% cv # perform matrix multiplications
          [,1]
[1,] 0.06703983
```

Figure 5.2. Listing of R code for computation of leverage for T_p equation.

Finally, the lower (eq. 13) and upper (eq. 14) prediction limits are computed by

$$^{\downarrow}T_{p\oplus} = 10^{\log_{10}(5.53)-1.6626\times 0.1383\sqrt{1+0.06704}} = 3.20 \text{ hours and}$$
$$^{\uparrow}T_{p\oplus} = 10^{\log_{10}(5.53)+1.6626\times 0.1383\sqrt{1+0.06704}} = 9.56 \text{ hours.}$$

Computation of Initial Abstraction and Prediction Limits

For the hypothetical watershed, $I_{A\oplus}$ is computed by substitution of $L_\oplus$, $D_\oplus$, $R_\oplus$, and $CN_\oplus$ into equation 23:

$$I_A = 2.045 - 0.5497L_\oplus^{-0.9041} - 0.1943D_\oplus + 0.2414R_\oplus - 0.01354CN_\oplus,$$
$$I_A = 2.045 - 0.5497(7.8)^{-0.9041} - 0.1943(0) + 0.2414(1) - 0.01354(75), \text{ and}$$
$$I_A = 1.19 \text{ inches.}$$

The leverage $h_\oplus^{[I_A]}$ is computed by substitution of $L_\oplus$, $D_\oplus$, $R_\oplus$, and $CN_\oplus$ into equation 27:

$$h_\oplus^{[I_A]} = \begin{bmatrix} 1 & (7.8)^{-0.9041} & 0 & 1 & 75 \end{bmatrix} \times \begin{bmatrix} 2.38000 & 0.17571 & 0.21256 & -0.04512 & -0.03106 \\ 0.17571 & 0.40497 & 0.00579 & -0.01074 & -0.00348 \\ 0.21256 & 0.00579 & 0.06893 & -0.01141 & -0.00304 \\ -0.04512 & -0.01074 & -0.01141 & 0.04742 & 0.00042 \\ -0.03106 & -0.00348 & -0.00304 & 0.00042 & 0.00041 \end{bmatrix} \times \begin{bmatrix} 1 \\ (7.8)^{-0.9041} \\ 0 \\ 1 \\ 75 \end{bmatrix},$$

which results in $h_\oplus^{[I_A]} = 0.02732$. R code demonstrating the matrix multiplications and the resulting $h_\oplus^{[I_A]}$ is shown in figure 5.3.

```
cv <- c(1, (7.8)^-0.9041, 0, 1, 75) # a column vector
rv <- t(cv) # transpose the vector into a row vector
h <- matrix(c( 2.38000,  0.17571,  0.21256, -0.04512, -0.03106,
               0.17571,  0.40497,  0.00579, -0.01074, -0.00348,
               0.21256,  0.00579,  0.06893, -0.01141, -0.00304,
              -0.04512, -0.01074, -0.01141,  0.04742,  0.00042,
              -0.03106, -0.00348, -0.00304,  0.00042,  0.00041),nrow=5,ncol=5)
rv %*% h %*% cv # perform matrix multiplications
           [,1]
[1,] 0.02731614
```

Figure 5.3. Listing of R code for computation of leverage for I_A equation.

Finally, the lower (eq. 25) and upper (eq. 26) prediction limits are computed by

$$^{\downarrow}I_{A\oplus} = 1.19 - 1.6628 \times 0.3025\sqrt{1+0.02732} = 0.68 \text{ inch, and}$$
$$^{\uparrow}I_{A\oplus} = 1.19 + 1.6628 \times 0.3025\sqrt{1+0.02732} = 1.70 \text{ inches.}$$

Computation of Constant Loss and Prediction Limits

For the hypothetical watershed, $C_{L\oplus}$ is computed by substitution of $L_\oplus$, $R_\oplus$, and $CN_\oplus$ into equation 29:

$$C_L = 2.535 - 0.4820L_\oplus^{0.2312} + 0.2271R_\oplus - 0.01676CN_\oplus,$$
$$C_L = 2.535 - 0.4820(7.8)^{0.2312} + 0.2271(1) - 0.01676(75), \text{ and}$$
$$C_L = 0.73 \text{ watershed inch per hour.}$$

The leverage $h_\oplus^{[C_L]}$ is computed by substitution of $L_\oplus$, $R_\oplus$, and $CN_\oplus$ into equation 33:

$$h_\oplus^{[C_L]} = \begin{bmatrix} 1 & (7.8)^{0.2312} & 1 & 75 \end{bmatrix} \times \begin{bmatrix} 2.79336 & -0.45089 & -0.01554 & -0.02600 \\ -0.45089 & 0.17984 & 0.00375 & 0.00222 \\ -0.01554 & 0.00375 & 0.04538 & -0.00011 \\ -0.02600 & 0.00222 & -0.00011 & 0.00028 \end{bmatrix} \times \begin{bmatrix} 1 \\ (7.8)^{0.2312} \\ 1 \\ 75 \end{bmatrix},$$

which results in $h_\oplus^{[C_L]} = 0.02863$. R code demonstrating the matrix multiplications and the resulting $h_\oplus^{[C_L]}$ is shown in figure 5.4.

```
cv <- c(1, (7.8)^0.2312, 1, 75) # a column vector
rv <- t(cv) # transpose the vector into a row vector
h <- matrix(c( 2.79336, -0.45089, -0.01554, -0.02600,
              -0.45089,  0.17984,  0.00375,  0.00222,
              -0.01554,  0.00375,  0.04538, -0.00011,
              -0.02600,  0.00222, -0.00011,  0.00028),nrow=4,ncol=4)
rv %*% h %*% cv # perform matrix multiplications
          [,1]
[1,] 0.02862544
```

Figure 5.4. Listing of R code for computation of leverage for C_L equation.

Finally, the lower (eq. 31) and upper (eq. 32) prediction limits are computed by

$$^{\downarrow}C_{L\oplus} = 0.73 - 1.6626 \times 0.2649\sqrt{1+0.02863} = 0.28 \text{ watershed inch per hour, and}$$

$$^{\uparrow}C_{L\oplus} = 0.73 + 1.6626 \times 0.2649\sqrt{1+0.02863} = 1.18 \text{ watershed inches per hour.}$$

Blank Page

Glossary

Acronyms

AIC — Akaike Information Criterion is a measure of information content of a regression model in which a tradeoff between number of parameters and the fit of the model is made.

GLD — The generalized lambda distribution defined in equation 19.

GUH — The gamma unit hydrograph as generally specified by equation 2 and particularly defined by $K^{\&}T_p$ equations.

GUHAS — The Gamma Unit Hydrograph Analysis System: Custom-built, analyst-directed software (Asquith and others, 2005) for estimation of K and T_p values from observed rainfall and runoff data.

PRESS — The PREdiction Sum of Squares statistic for a regression, which is regarded as a measure of performance when a regression equation is used to predict new data (see fig. 6).

Symbols

$|$ — Represents "or" and for this report represents the distinction between watershed development ($D = 0 \mid D = 1$).

$\oplus$ — Represents the $\oplus$-vector of the watershed characteristics or regressor variables used to make a prediction for a given regression equation.

$\gtrapprox$ — Represents an approximate "greater than" inequality for the quantities on the left of the symbol.

$\lessapprox$ — Represents an approximate "less than" inequality for the quantities on the left of the symbol.

Hebrew and Greek Alphabet

$\aleph$ — A quantity defined in item 3 on page 27 for the algorithm to compensate for the influence of modeling time step that begins on page 27.

α — The $100 \times (1 - \alpha)$ prediction interval of a regression equation and also the scale parameter of the GLD.

δ — A quantity defined in item 3 on page 27 for the algorithm to compensate for the influence of modeling time step that begins on page 27.

ΔE — The change in elevation in feet between the two end points of L.

ε^{Q_p} — The Q_p error (eq. 37) in $\log_{10}$(cubic feet per second) that is defined as $\varepsilon^{Q_p} = \log_{10}(Q_p^{\text{obs}}) - \log_{10}(Q_p^{\text{model}})$.

ε^{V} — The V error in watershed inches that is defined as $\varepsilon^{V} = V^{\text{obs}} - V^{\text{model}}$, where V^{obs} is observed V in watershed inches and V^{model} is modeled V in watershed inches.

$\varepsilon^{T_{Q_p}}$ — The T^{Q_p} error (time difference in peak streamflow) in hours that is defined as $\varepsilon^{T_{Q_p}} = T^{Q_p^{\text{obs}}} - T^{Q_p^{\text{model}}}$.

η — The second of two dimensionless shape parameters of the GLD (eq. 19).

$\eta^{[C_L]}$ — A dimensionless shape parameter of GLD (eq. 19) for C_L (eq. 21 and table 4).

$\eta^{[I_A]}$ — A dimensionless shape parameter of GLD (eq. 19) for I_A (eq. 20 and table 3).

$\Gamma(u)$ — The complete gamma function is expressed as an infinite integral: $\Gamma(u) = \int_0^\infty u^{k-1} e^{-u}\, du$. The $\Gamma(u)$ is in equation 4 and figure 2.

κ — The first of two dimensionless shape parameters of the GLD (eq. 19).

$\kappa^{[C_L]}$ — A dimensionless shape parameter of GLD (eq. 19) for C_L (eq. 21 and table 4).

$\kappa^{[I_A]}$ — A dimensionless shape parameter of GLD (eq. 19) for I_A (eq. 20 and table 3).

λ_1 — The first L-moment, which is the arithmetic mean and a measure of location on the real-number line of a distribution, in units of the data. The symbol is not actually represented in this report as the preference is to use "mean." Values are listed in table 2.

λ_2 — The second L-moment, which is a measure of the variability of a distribution, in units of the data. The symbol is not actually represented in this report as the preference is to use "L-scale." Values are listed in table 2.

φ — The power coefficient on L for regression equation 22.

$\varphi^{[C_L]}$ — The power coefficient on L for C_L equation (eq. 29).

$\varphi^{[I_A]}$ — The power coefficient on L for I_A equation (eq. 23).

ψ — The scale parameter of the GLD (eq. 19).

$\psi^{[C_L]}$ — Scale parameter of GLD (eq. 19) for C_L (eq. 21 and table 4). The parameter has units of watershed inches per hour.

$\psi^{[I_A]}$ — Scale parameter of GLD (eq. 19) for I_A (eq. 20 and table 3). The parameter has units of watershed inches.

σ — The standard error for an individual terminal node in the I_A tree (fig. 7) or C_L tree (fig. 8).

$\sigma^{[K]}$ — The standard error of equation 5.

$\sigma^{[T_p]}$ — The standard error of equation 11.

$\sigma^{[I_A]}$ — The standard error of equation 23.

$\sigma^{[C_L]}$ — The standard error of equation 29.

τ_3 — L-skew is a dimensionless L-moment statistic of a distribution and measures of distribution skewness (shape). Values are listed in table 2.

τ_4 — L-kurtosis is a dimensionless L-moment statistic of a distribution and can be thought of as a measure of distribution peakedness. Values are listed in table 2.

τ_5 — A dimensionless L-moment statistic of a distribution with no unique conceptualization of distribution shape. Values are listed in table 2.

ξ — The location parameter of the GLD (eq. 19).

$\xi^{[C_L]}$ — Location parameter of GLD (eq. 19) for C_L (eq. 21 and table 4) in watershed inches per hour.

$\xi^{[I_A]}$ — Location parameter of GLD (eq. 19) for I_A (eq. 20 and table 3) in watershed inches.

Latin Alphabet

A — Contributing drainage area in square miles. Values for the 92 watersheds are listed in appendix 1 (table 1.1).

A^{pred} — Contributing drainage area in square miles predicted by equation 1.1 in appendix 1.

$C_L(F)$ — The quantile function of the GLD (eq. 21) fit to the L-moments (table 2) of the 92 $\overline{C_L}$ values, which are listed in appendix 1 (table 1.1).

C_L — Constant loss of a watershed after I_A has been satisfied, in watershed inches per hour.

$C_{L\oplus}$ — C_L estimated by equation 29.

$^{\natural}C_L$ — Storm-specific C_L based on analysis described in this report.

$\overline{C_L}$ — Watershed-specific mean C_L based on analysis described in this report. This symbol also is used in the context of overall mean of the 92 $\overline{C_L}$ values listed in appendix 1 (table 1.1). Also see the glossary entry for $\overline{I_A{}^{\&}C_L}$ model.

$\check{C}_L$ — Overall median of the 92 watershed-specific mean constant-loss values based on analysis described in this report. Also see glossary entry for $\check{I}_A{}^{\&}\check{C}_L$ model.

${}^{\downarrow}C_{L\oplus}^{\alpha/2}$ and ${}^{\uparrow}C_{L\oplus}^{\alpha/2}$ — The lower (↓) and upper (↑) $100 \times (1-\alpha)$ prediction interval limits of C_L (eq. 29).

CN — The CN is a parameter used in the Curve Number (CN) method to estimate the maximum potential retention of rainfall in a watershed (Natural Resources Conservation Service, 2006) and reportedly accounts for differences between soil types, land-cover classifications, and other hydrologic conditions of the land surface that affect watershed storage of rainfall. Values for the 92 watersheds are listed in appendix 1 (table 1.1).

$CN_\oplus$ — A watershed-specific CN value.

df — Degrees of freedom of a regression equation, which in the context of this report are used to determine the value of the t-distribution (see table 2.1 in appendix 2) for computations of prediction limits.

$\mathrm{df}^{[C_L]}$ — Degrees of freedom of the C_L equation (eq. 29).

$\mathrm{df}^{[I_A]}$ — Degrees of freedom of the I_A equation (eq. 23).

$\mathrm{df}^{[K]}$ — Degrees of freedom of the K equation (eq. 5).

$\mathrm{df}^{[T_p]}$ — Degrees of freedom of the T_p equation (eq. 11).

D — The binary watershed development classification: $D = 0$ for undeveloped watersheds and $D = 1$ for developed watersheds. Values of D for the 92 watersheds are listed in appendix 1 (table 1.1).

$D_\oplus$ — A watershed-specific D value.

F — Nonexceedance probability (eqs. 19, 20, and 21).

$h_\oplus$ — The leverage statistic for a specific prediction, which is made from the $\oplus$-vector of the watershed characteristics or regressor variables.

$h_\oplus^{[C_L]}$ — The leverage for a $C_{L\oplus}$ prediction for the watershed from equation 29.

$h_{\max}^{[C_L]}$ — The maximum leverage value of data on which equation 29 is based.

$h_\oplus^{[I_A]}$ — The leverage for a $I_{A\oplus}$ prediction for the watershed from equation 23.

$h_{\max}^{[I_A]}$ — The maximum leverage value of data on which equation 23 is based.

$h_\oplus^{[K]}$ — The leverage for a $K_\oplus$ prediction for the watershed from equation 5.

$h_{\max}^{[K]}$ — The maximum leverage value of data on which equation 5 is based.

$h_\oplus^{[T_p]}$ — The leverage for a $T_{p\oplus}$ prediction for the watershed from equation 11.

$h_{\max}^{[T_p]}$ — The maximum leverage value of data on which equation 11 is based.

$I_A(F)$ — The quantile function of the GLD (eq. 20) fit to the L-moments (table 2) of the 92 $\overline{I_A}$ values listed in appendix 1 (table 1.1).

I_A — Initial abstraction of a watershed and is measured in watershed inches.

$I_{A\oplus}$ — I_A estimated by equation 23.

${}^{\lightning}I_A$ — Storm-specific I_A based on analysis described in this report.

$\overline{I_A}$ — Watershed-specific mean I_A based on analysis described in this report. This symbol also is used in the context of overall mean of the 92 $\overline{I_A}$ values listed in appendix 1 (table 1.1). Also see glossary entry for $\overline{I_A{}^{\&}C_L}$ model.

$\check{I}_A$ — Overall median of the 92 watershed-specific mean initial-abstraction values based on analysis described in this report. Also see glossary entry for $\check{I}_A{}^{\&}\check{C}_L$ model.

$I_A^{[CN]}$ — Initial abstraction estimated by the *CN* method (eq. 35).

$I_A{}^{\&}C_L$ — The collective reference to the initial-abstraction, constant-loss model with the specific context of the model as a "representation" of losses of rainfall.

$\overline{I_A{}^{\&}C_L}$ model — The watershed-loss model using the mean values of $\overline{I_A}$ and $\overline{C_L}$ with consideration of watershed development (table 1):
$\overline{I_A} = 1.106 \mid 0.690$
$\overline{C_L} = 0.617 \mid 0.512$

$\check{I}_A{}^{\&}\check{C}_L$ model — The watershed-loss model using the median values of $\overline{I_A}$ and $\overline{C_L}$ with consideration of watershed development (table 1):
$\check{I}_A = 1.111 \mid 0.564$
$\check{C}_L = 0.481 \mid 0.520$

$^{\natural}I_A{}^{\&}{}^{\natural}C_L$ — The collective reference to the initial-abstraction, constant-loss model when set to storm-specific ($^{\natural}$) values of I_A and C_L. These values were computed by specialized software (see the next entry).

$^{\natural}I_A{}^{\&}{}^{\natural}C_L$ software — Custom-built, analyst-directed software for estimation of $^{\natural}I_A$ and $^{\natural}C_L$ values from observed rainfall and runoff data and $K^{\&}T_p$-GUH.

$^{\downarrow}I_{A\oplus}^{\alpha/2}$ and $^{\uparrow}I_{A\oplus}^{\alpha/2}$ — The lower (↓) and upper (↑) $100 \times (1-\alpha)$ prediction interval limits of I_A (eq. 23).

K — The shape parameter of GUH defined by equation 2.

$\boxed{K}$ — The K of a GUH in hours that is defined in item 5 on page 27 for the algorithm to compensate for the influence of modeling time step that begins on page 27.

$K_\oplus$ — The shape parameter of the GUH for a specific prediction, which is made from the ⊕-vector of the watershed characteristics or regressor variables.

$K^{\&}T_p$ — The collective reference to the K and T_p equations (eqs. 5 and 11 of this report) by Asquith and others (2005).

$K^{\&}T_p$-GUH — Reference to a GUH set by the K and T_p equations from Asquith and others (2005).

$^{\downarrow}K_\oplus^{\alpha/2}$ and $^{\uparrow}K_\oplus^{\alpha/2}$ — The lower (↓) and upper (↑) $100 \times (1-\alpha)$ prediction interval limits of K (eq. 5).

L — Main-channel length: The L is defined as the length in stream-course miles of the longest defined channel shown in a 30-meter digital elevation model from the approximate watershed headwaters to the outlet. Values for the 92 watersheds are listed in appendix 1 (table 1.1).

$L_\oplus$ — A watershed-specific L value.

n — The number of samples (watersheds) for an individual terminal node in the I_A tree (fig. 7) or C_L tree (fig. 8).

P_m — The mth value of excess rainfall in watershed inches for convolution with the unit hydrograph U_{n-m+1} (eq. 1).

q — Watershed-depth streamflow in watershed inches per hour.

$q(t)$ — The watershed-depth streamflow in watershed inches per hour at time t of the GUH defined by equation 2.

q_p — The watershed-depth peak streamflow in watershed inches per hour of the GUH defined by equation 2.

$\boxed{q_p}$ — The q_p of a GUH in hours that is defined in item 2 on page 27 for the algorithm to compensate for the influence of modeling time step that begins on page 27.

Q — Streamflow in cubic feet per second.

Q_n — The nth value of streamflow in cubic feet per second from convolution of the unit hydrograph with excess rainfall (eq. 1).

Q_p — The peak or maximum instantaneous streamflow in cubic feet per second.

$\mathring{Q}_p$ — The Q_p from the combined $I_A{}^{\&}C_L$ model (eq. 38).

Q_p^{model} — The Q_p of a modeled runoff hydrograph. As context requires, an equivalent symbol is $Q_p^{\text{a model}}$.

Q_p^{obs} — The Q_p of an observed runoff hydrograph.

$Q_p^{\overline{I_A{}^{\&}C_L}\text{ model}}$ — The Q_p from the $\overline{I_A{}^{\&}C_L}$ model (eq. 38).

$Q_p^{\check{I}_A{}^{\&}\check{C}_L\text{ model}}$ — The Q_p from the $\check{I}_A{}^{\&}\check{C}_L$ model (eq. 38).

$Q_p^{I_A{}^{\&}C_L\text{ equations}}$ — The Q_p from the $I_A{}^{\&}C_L$ equations (eq. 38).

$Q_p^{I_A{}^{\&}C_L\text{ trees}}$ — The Q_p from the $I_A{}^{\&}C_L$ trees (eq. 38).

R — The binary classification of watersheds that are rock-dominated with thin soils, rock channels, and karstic features: $R = 0$ for non-rocky watersheds and $R = 1$ for rocky watersheds. Values for the 92 watersheds are listed in appendix 1 (table 1.1).

$R_{\oplus}$ — A watershed-specific R value.

S — Dimensionless main-channel slope: The S is defined as the change in elevation in feet (ΔE) between the two end points of L divided by L in feet: $S = \Delta E/(5{,}280 \times L)$. A 30-meter digital elevation model was used to compute S for this report. Values for the 92 watersheds are listed in appendix 1 (table 1.1).

$S_{\oplus}$ — A watershed-specific S value.

$S_{\max}$ — Maximum potential retention for the CN method (eq. 36).

t — The time in hours from inception of streamflow of the GUH defined by equation 2.

$\mathrm{t}_{[\alpha/2,\mathrm{df}]}$ — The t-distribution for α and degrees of freedom df. The t-distribution is used to compute prediction limits of a regression equation.

T_b — The time base of a runoff hydrograph or the overall width in time of a runoff hydrograph.

T_p — The time to q_p of a GUH in hours, which is defined by equation 2.

$\boxed{T_p}$ — The T_p of a GUH in hours that is defined in item 4 on page 27 for the algorithm to compensate for the influence of modeling time step that begins on page 27.

$T_{p\oplus}$ — The T_p of a GUH in hours, for a specific prediction, which is made from the $\oplus$-vector of the watershed characteristics or regressor variables.

${}^{\downarrow}T_{p\oplus}^{\alpha/2}$ and ${}^{\uparrow}T_{p\oplus}^{\alpha/2}$ — The lower ($\downarrow$) and upper ($\uparrow$) $100 \times (1-\alpha)$ prediction interval limits of T_p (eq. 11).

T^{Q_p} — The time of Q_p occurrence.

$\mathring{T}^{Q_p}$ — The T^{Q_p} from the combined $I_A{}^{\&}C_L$ model (eq. 40).

$T^{Q_p^{\text{model}}}$ — The T^{Q_p} for a modeled runoff hydrograph; represents the actual date and time that Q_p of a modeled runoff hydrograph occurs.

$T^{Q_p^{\text{obs}}}$ — The T^{Q_p} for an observed runoff hydrograph; represents the actual date and time that Q_p of an observed runoff hydrograph occurs.

$T_p^{\overline{I_A{}^{\&}C_L}\text{ model}}$ — The T^{Q_p} from the $\overline{I_A{}^{\&}C_L}$ model (eq. 40).

$T_p^{\check{I}_A{}^{\&}\check{C}_L\text{ model}}$ — The T^{Q_p} from the $\check{I}_A{}^{\&}\check{C}_L$ model (eq. 40).

$T_p^{I_A{}^{\&}C_L\text{ equations}}$ — The T^{Q_p} from the $I_A{}^{\&}C_L$ equations (eq. 40).

$T_p^{I_A{}^{\&}C_L\text{ trees}}$ — The T^{Q_p} from the $I_A{}^{\&}C_L$ trees (eq. 40).

U_{n-m+1} — The $n-m+1$th value of the unit hydrograph in cubic feet per second for convolution with excess rainfall P_m (eq. 1).

V — The volume in watershed depth of a runoff hydrograph and for a unit hydrograph $V = 1$.

$\mathring{V}$ — The V from the combined $I_A{}^{\&}C_L$ model (eq. 39)

V^{model} — The V of a modeled runoff hydrograph in watershed inches.

V^{obs} — The V of an observed runoff hydrograph in watershed inches.

$V_p^{\overline{I_A{}^{\&}C_L}\,\text{model}}$ — The V from the $\overline{I_A{}^{\&}C_L}$ model (eq. 39).

$V_p^{\check{I}_A{}^{\&}\check{C}_L\,\text{model}}$ — The V from the $\check{I}_A{}^{\&}\check{C}_L$ model (eq. 39).

$V_p^{I_A{}^{\&}C_L\,\text{equations}}$ — The V from the $I_A{}^{\&}C_L$ equations (eq. 39).

$V_p^{I_A{}^{\&}C_L\,\text{trees}}$ — The V from the $I_A{}^{\&}C_L$ trees (eq. 39).

w_i — The number of suitable storms processed for a given watershed (i). The values are listed in appendix 1 (table 1.1).

x — A quantile value of the GLD (eq. 19).

$x(F)$ — The quantile x as a function of F.

Technical Notes

[a] A hydrograph is the temporal distribution of streamflow at a given stream location following substantial rainfall on a watershed; specifically, rainfall that produces a comparatively distinct rise and recession of streamflow over durations of hours to a couple of days. A runoff hydrograph also is referred to as stormflow, storm runoff, and other similar terms.

[b] Design in this context emphasizes the connection between the model results reported and the criteria used to guide and constrain the *design* of drainage infrastructure. A design storm is a hypothetical storm having a specified duration and depth. The depth of a design storm is specified by risk criteria (design criteria) such as an annual nonexceedance probability. Annual nonexceedance probability commonly is expressed as annual recurrence interval such as the "25-year event," which has a 0.96 annual nonexceedance probability. The depth and duration of design storms in Texas is considered by Asquith and Roussel (2004).

[c] The authors express specific semantic intent with the term "applicable watersheds." The analyses and subsequent results reported here are founded on only 92 watersheds located in a central north-to-south swath through Texas. Whereas these watersheds represent what are believed to be a generally diverse set of watersheds, and they nearly represent the collective extent of USGS data in Texas suitable for the purposes of this report, the 92 watersheds represent a minute fraction of potential watershed types throughout Texas. The results reported here are most applicable to watersheds similar to the 92 watersheds; however, determination of the applicability of the results reported here is a site-specific, hydrologic-engineering decision, which is best left to analysts. This report provides comprehensive discussion to help guide decisions regarding applicability.

[d] Excess rainfall is rainfall that is not retained or stored and represents the volume of runoff per unit area of the watershed. An excess rainfall hyetograph is the temporal distribution of the rainfall that will become a runoff hydrograph after being "processed" by the watershed. Excess rainfall also is referred to as effective rainfall in some hydrologic literature.

[e] A design storm hyetograph is the hypothetical time distribution of rainfall intensity (a hyetograph) for the design storm. Rainfall hyetographs in the context of Texas hydrology are considered by Asquith, Roussel, and others (2004) and Williams-Sether and others (2004).

[f] A suitable storm is a storm (pair of rainfall and runoff data files) that could actually be processed by the ${}^{\natural}I_A$ & ${}^{\natural}C_L$ software. An example of an unsuitable storm is one that has more observed (recorded) runoff than observed rainfall. Other, less-easily described situations occurred for a comparatively small number of storms compared to the more than 1,600 storms available for processing.

[g] Although it is physically meaningless to have either ${}^{\natural}I_A$ or ${}^{\natural}C_L$ less than zero, statistically optimal solutions, as measured by minimization of squared residuals, can exist at ${}^{\natural}I_A < 0$ or ${}^{\natural}C_L < 0$ with the volume of the observed and modeled runoff hydrographs being equal. Conceptual causes for such a situation might be poor estimation of the observed runoff hydrograph, inadequate representation of the rainfall hyetograph for the watershed, or violations of fundamental assumptions of unit hydrograph theory. For ideal processing, the rainfall hyetograph in terms of depth simultaneously at every time step must be "uniformly distributed" and the percentage of the watershed actually contributing runoff is 100 percent. (The unit hydrograph method assumes that 100 percent of the watershed contributes to runoff.) Spatial and temporal rainfall estimation likely is the greatest source of error because generally few (and commonly just one) rain gages are operated in a watershed. Thus, the negative values for ${}^{\natural}I_A$ and ${}^{\natural}C_L$ have the effect of generating "rainfall" to compensate for misspecification of the excess rainfall hyetograph.

[h] Percentage impervious cover was calculated by using delineated drainage areas and intersecting them with the 1992 version of the National Land Cover Dataset (NLCD) (U.S. Geological Survey, 2006). After the two datasets were combined, the NLCD classification system was used to determine which codes were developed, areas characterized by a high percentage of constructed materials including asphalt, concrete, and buildings. The area associated with each developed classification type within each hand-delineated drainage

area was then used to calculate the percentage impervious cover within each classification. For example, classification-type 21 includes areas with a mixture of constructed materials and vegetation. Constructed materials account for 30–80 percent of the cover. Therefore, for all areas designated with a classification code 21, 50 percent (approximate average) of the area was designated as impervious cover. The three codes designated as developed were 21 (low intensity residential), 22 (high intensity residential), and 23 (commercial/industrial/transportation). For each drainage area, and each code designated as developed, impervious cover was calculated and summed to obtain the total approximate impervious cover per hand-delineated drainage area. The percentage impervious cover analyzed were general averages of what the impervious cover was for each drainage area in 1992.

[i] Karst is defined (Bates, 1987, p. 356) as a type of terrain associated with limestone and other readily dissolved rock, which is characterized by sinkholes, caves, and considerable underground drainage. The limestone and karst-like watersheds in and around the Austin and San Antonio areas also have an abundance of fractures and faults. These $R = 1$ watersheds therefore are expected to have considerable capacity to abstract an absolute value of rainfall and substantially affect losses after initial abstraction is satisfied.

[j] The watershed scale dependency of this conclusion is not explored in this report. Whereas, about 15 minutes is the mean $\varepsilon^{T_{Q_p}}$, as a percentage of T^{Q_p} for a given watershed, 15 minutes might represent a substantial timing error (for small watersheds or those with small L) or an unsubstantial timing error (for large watersheds or those with large L).

Prepared by the USGS Lafayette Publishing center.

Information regarding water resources in Texas is available at
http: //tx.usgs.gov/