In cooperation with the Texas Department of Transportation

Regression Equations for Estimation of Annual Peak-Streamflow Frequency for Undeveloped Watersheds in Texas Using an L-moment-Based, PRESS-Minimized, Residual-Adjusted Approach

Scientific Investigations Report 2009–5087
(Texas Department of Transportation Research Report 0–5521–1)

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

Front cover: "View looking upstream of the Colorado River below the mouth of the Concho River at the Miller Ranch, September [18,] 1936. Photograph by Miller at the peak stage. Discharge is 356,000 cubic feet per second by slope-area method and is greatest then known by local residents." Other note "near Leaday, Texas." Undated notes on back of photograph by Harold W. Albert.

Additional information: The photograph appears to have been taken about 1 mile south of Leaday, Texas, looking toward the north. The closest downstream U.S. Geological Survey (USGS) streamflow-gaging station is 08136700 Colorado River near Stacy, Texas, which had a peak streamflow on September 18, 1936, of 356,000 cubic feet per second at a gage height of 64.59 feet, although the period of record begins in 1969 to present (2009). A substantial part of the photograph appears to now be inundated by O.H. Ivie Reservoir monitored by USGS reservoir station 08136600 O.H. Ivie Reservoir near Voss, Texas. A brief history of O.H. Ivie is provided on inside of the back cover.

— Technical Report Documentation Page —

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA/TX–09/0–5521–1		

4. Title and Subtitle	5. Report Date
Regression Equations for Estimation of Annual Peak-Streamflow Frequency for Undeveloped Watersheds in Texas Using an L-moment-Based, PRESS-Minimized, Residual-Adjusted Approach	June 2009
7. Authors(s) William H. Asquith and Meghan C. Roussel	**8. Performing Organization Report No.** USGS SIR 2009–5087
9. Performing Organization Name and Address U.S. Geological Survey 8027 Exchange Drive Austin, Texas 78754	**10. Work Unit No. (TRAIS)** **11. Contract or Grant No.** Project 0–5521
12. Sponsoring Agency Name and Address Texas Department of Transportation Research and Technology Implementation Office P.O. Box 5080 Austin, Texas 78731	**13. Type of Report and Period Covered** Technical report on research from 2007 to 2009 **14. Sponsoring Agency Code**

15. Supplementary Notes

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

16. Abstract

Annual peak-streamflow frequency estimates are needed for flood-plain management; for objective assessment of flood risk; for cost-effective design of dams, levees, and other flood-control structures; and for design of roads, bridges, and culverts. Annual peak-streamflow frequency represents the peak streamflow for nine recurrence intervals of 2, 5, 10, 25, 50, 100, 200, 250, and 500 years. Common methods for estimation of peak-streamflow frequency for ungaged or unmonitored watersheds are regression equations for each recurrence interval developed for one or more regions; such regional equations are the subject of this report. The method is based on analysis of annual peak-streamflow data from U.S. Geological Survey streamflow-gaging stations (stations). Beginning in 2007, the U.S. Geological Survey, in cooperation with the Texas Department of Transportation and in partnership with Texas Tech University, began a 3-year investigation concerning the development of regional equations to estimate annual peak-streamflow frequency for undeveloped watersheds in Texas. The investigation focuses primarily on 638 stations with 8 or more years of data from undeveloped watersheds and other criteria. The general approach is explicitly limited to the use of L-moment statistics, which are used in conjunction with a technique of multi-linear regression referred to as PRESS minimization. The approach used to develop the regional equations, which was refined during the investigation, is referred to as the "L-moment-based, PRESS-minimized, residual-adjusted approach." For the approach, seven unique distributions are fit to the sample L-moments of the data for each of 638 stations and trimmed means of the seven results of the distributions for each recurrence interval are used to define the station-specific, peak-streamflow frequency. As a first iteration of regression, nine weighted-least-squares, PRESS-minimized, multi-linear regression equations are computed using the watershed characteristics of drainage area, dimensionless main-channel slope, and mean annual precipitation. The residuals of the nine equations are spatially mapped, and residuals for the 10-year recurrence interval are selected for generalization to 1-degree latitude and longitude quadrangles. The generalized residual is referred to as the OmegaEM parameter and represents a generalized terrain and climate index that expresses peak-streamflow potential not otherwise represented in the three watershed characteristics. The OmegaEM parameter was assigned to each station, and using OmegaEM, nine additional regression equations are computed. Because of favorable diagnostics, the OmegaEM equations are expected to be generally reliable estimators of peak-streamflow frequency for undeveloped and ungaged stream locations in Texas. The mean residual standard error, adjusted R-squared, and percentage reduction of PRESS by use of OmegaEM are $0.30\log_{10}$, 0.86, and -21 percent, respectively. Inclusion of the OmegaEM parameter provides a substantial reduction in the PRESS statistic of the regression equations and removes considerable spatial dependency in regression residuals. Although the OmegaEM parameter requires interpretation on the part of analysts and the potential exists that different analysts could estimate different values for a given watershed, the authors suggest that typical uncertainty in the OmegaEM estimate might be about $\pm 0.10\log_{10}$. Finally, given the two ensembles of equations reported herein and those in previous reports, hydrologic design engineers and other analysts have several different methods, which represent different analytical tracks, to make comparisons of peak-streamflow frequency estimates for ungaged watersheds in the study area.

17. Key Words	18. Distribution Statement
	No restrictions

19. Security Classif. (of report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	48	

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

Blank Page

Regression Equations for Estimation of Annual Peak-Streamflow Frequency for Undeveloped Watersheds in Texas Using an L-moment-Based, PRESS-Minimized, Residual-Adjusted Approach

By William H. Asquith and Meghan C. Roussel

In cooperation with the Texas Department of Transportation

Scientific Investigations Report 2009–5087
(Texas Department of Transportation Research Report 0–5521–1)

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

U.S. Department of the Interior
KEN SALAZAR, Secretary

U.S. Geological Survey
Suzette M. Kimball, Acting Director

U.S. Geological Survey, Reston, Virginia: 2009

For product and ordering information:
World Wide Web: http://www.usgs.gov/pubprod
Telephone: 1-888-ASK-USGS

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment:
World Wide Web: http://www.usgs.gov
Telephone: 1-888-ASK-USGS

Suggested citation:
Asquith, W.H., and Roussel, M.C., 2009, Regression equations for estimation of annual peak-streamflow frequency for undeveloped watersheds in Texas using an L-moment-based, PRESS-minimized, residual-adjusted approach: U.S. Geological Survey Scientific Investigations Report 2009–5087, 48 p. [http://pubs.usgs.gov/sir/2009/5087]

Typeset on June 16, 2009

Contents

Abstract 1
Introduction 2
    Purpose and Scope 2
    Identification of Annual Peak-Streamflow Data 2
    Acknowledgments 4
An L-moment-Based, PRESS-Minimized, Residual-Adjusted Approach 4
    L-moment-Based, Station-Specific Peak-Streamflow Frequency Estimates 4
        Sample L-moments 5
        Sampling Error 5
        Probability Distributions for Peak-Streamflow Frequency Curves 6
        Trimmed Mean Estimates of Peak-Streamflow Frequency 7
        Modeling Error 9
        Combining Sampling and Modeling Errors 9
    Weighted-Least-Squares, Multi-Linear Regression Analysis and PRESS Minimization 10
    Residual-Adjusted Regression and the OmegaEM Parameter 10
Regression Equations in Texas Using an L-moment-Based, PRESS-Minimized, Residual-Adjusted Approach 11
    Regression Equations without OmegaEM Parameter 11
    OmegaEM Parameter 12
    Regression Equations with OmegaEM Parameter 12
    Example Computations for a Single Station and Comparison to Estimates from Previously Published Equations 16
    Comparison of Regional Equations to Those of Previous Studies 18
        Comparability and Applicability of Peak-Streamflow Estimates from OmegaEM and Previous Equations 19
        Preference for OmegaEM Equations over Previous Equations 21
    Considerations for Application of Regression Equations to Gaged and Ungaged Watersheds 23
Summary 24
References 26

Appendix 1—Supplemental Information and External Data Files 29
Appendix 2—Results of Weighted-Least-Squares, Multi-Linear Regression Analysis Using PRESS Minimization of Peak-Streamflow Frequency to Watershed Characteristics of Drainage Area, Mean Annual Precipitation, and Dimensionless Main-Channel Slope 33
Appendix 3—Results of Weighted-Least-Squares, Multi-Linear Regression Analysis Using PRESS Minimization of Peak-Streamflow Frequency to Watershed Characteristics of Drainage Area, Mean Annual Precipitation, Dimensionless Main-Channel Slope, and OmegaEM 41

Figures

1. Graph showing example of peak-streamflow frequency curves with superimposed axis labels and explanation in real and logarithmic space rendered from the R environment (R Development Core Team, 2008) using functions of Asquith (2008) for seven probability distributions fit to sample L-moments of annual peak-streamflow data available in file `Appendix1_677annpks.txt` 8
2. Map showing hill-shade relief in Texas with superimposed values of OmegaEM parameter that represents a generalized terrain and climate index for regionalization of peak-streamflow frequency 13
3. Map showing rivers in Texas with superimposed values of OmegaEM parameter that represents a generalized terrain and climate index for regionalization of peak-streamflow frequency 14
4. Map showing level II and III ecoregions in Texas with superimposed values of OmegaEM parameter that represents a generalized terrain and climate index for regionalization of peak-streamflow frequency 15
5. Graph showing comparison by recurrence interval of peak-streamflow frequency estimates for the 638 stations using equations listed in table 2 (vertical axis) and using equations originally shown in Asquith and Thompson (2005, table 5) (horizontal axis, AT2005) that are based on station-specific, peak-streamflow frequency based from broad interpretation of Interagency Advisory Committee on Water Data (1982) (17B) 20
6. Graph showing comparison by recurrence interval of peak-streamflow frequency estimates for the 638 stations using equations listed in table 3 (vertical axis) and using equations originally shown in Asquith and Thompson (2005, table 5) (horizontal axis, AT2005) that are based on station-specific, peak-streamflow frequency based from broad interpretation of Interagency Advisory Committee on Water Data (1982) (17B) 22

Tables

1. Summary statistics of the watershed characteristics for 638 U.S. Geological Survey streamflow-gaging stations used to develop regional equations for estimation of peak-streamflow frequency 4
2. Summary of weighted-least-squares, PRESS-minimized, regional regression equations using drainage area, dimensionless main-channel slope, and mean annual precipitation 16
3. Summary of weighted-least-squares, PRESS-minimized, regional regression equations using drainage area, dimensionless main-channel slope, mean annual precipitation, and OmegaEM 17

Conversion Factors and Datum

Inch/Pound to SI

Multiply	By	To obtain
	Length	
inch (in.)	25.4	millimeter (mm)
mile (mi)	1.609	kilometer (km)
	Slope	
foot per mile (ft/mi)	.1894	meter per kilometer (m/km)
	Area	
square mile (mi^2)	2.590	square kilometer (km^2)
	Flow	
cubic foot per second (ft^3/s)	.02832	cubic meter per second (m^3/s)

SI to Inch/Pound

Multiply	By	To obtain
	Length	
meter (m)	3.281	foot (ft)
	Slope	
meter per kilometer (m/km)	5.28	foot per mile (ft/mi)
	Area	
square kilometer (km^2)	.3861	square mile (mi^2)
	Flow	
cubic meter per second (m^3/s)	35.31	cubic foot per second (ft^3/s)

Datum

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

Regression Equations for Estimation of Annual Peak-Streamflow Frequency for Undeveloped Watersheds in Texas Using an L-moment-Based, PRESS-Minimized, Residual-Adjusted Approach

By William H. Asquith and Meghan C. Roussel

Abstract

Annual peak-streamflow frequency estimates are needed for flood-plain management; for objective assessment of flood risk; for cost-effective design of dams, levees, and other flood-control structures; and for design of roads, bridges, and culverts. Annual peak-streamflow frequency represents the peak streamflow for nine recurrence intervals of 2, 5, 10, 25, 50, 100, 200, 250, and 500 years. Common methods for estimation of peak-streamflow frequency for ungaged or unmonitored watersheds are regression equations for each recurrence interval developed for one or more regions; such regional equations are the subject of this report. The method is based on analysis of annual peak-streamflow data from U.S. Geological Survey streamflow-gaging stations (stations). Beginning in 2007, the U.S. Geological Survey, in cooperation with the Texas Department of Transportation and in partnership with Texas Tech University, began a 3-year investigation concerning the development of regional equations to estimate annual peak-streamflow frequency for undeveloped watersheds in Texas. The investigation focuses primarily on 638 stations with 8 or more years of data from undeveloped watersheds and other criteria. The general approach is explicitly limited to the use of L-moment statistics, which are used in conjunction with a technique of multi-linear regression referred to as PRESS minimization. The approach used to develop the regional equations, which was refined during the investigation, is referred to as the "L-moment-based, PRESS-minimized, residual-adjusted approach." For the approach, seven unique distributions are fit to the sample L-moments of the data for each of 638 stations and trimmed means of the seven results of the distributions for each recurrence interval are used to define the station-specific, peak-streamflow frequency. As a first iteration of regression, nine weighted-least-squares, PRESS-minimized, multi-linear regression equations are computed using the watershed characteristics of drainage area, dimensionless main-channel slope, and mean annual precipitation. The residuals of the nine equations are spatially mapped, and residuals for the 10-year recurrence interval are selected for generalization to 1-degree latitude and longitude quadrangles. The generalized residual is referred to as the OmegaEM parameter and represents a generalized terrain and climate index that expresses peak-streamflow potential not otherwise represented in the three watershed characteristics. The OmegaEM parameter was assigned to each station, and using OmegaEM, nine additional regression equations are computed. Because of favorable diagnostics, the OmegaEM equations are expected to be generally reliable estimators of peak-streamflow frequency for undeveloped and ungaged stream locations in Texas. The mean residual standard error, adjusted R-squared, and percentage reduction of PRESS by use of OmegaEM are $0.30 \log_{10}$, 0.86, and -21 percent, respectively. Inclusion of the OmegaEM parameter provides a substantial reduction in the PRESS statistic of the regression equations and removes considerable spatial dependency in regression residuals. Although the OmegaEM parameter requires interpretation on the part of analysts and the potential exists that different analysts could estimate different values for a given watershed, the authors suggest that typical uncertainty in the OmegaEM estimate might be about $\pm 0.10 \log_{10}$. Finally, given the two ensembles of equations reported herein and those in previous reports, hydrologic design engineers and other analysts have several different methods, which represent different analytical tracks, to make comparisons of peak-streamflow frequency estimates for ungaged watersheds in the study area.

Introduction

Annual peak-streamflow frequency estimates are needed for flood-plain management; for objective assessment of flood risk; for cost-effective design of dams, levees, and other flood-control structures; and for design of roads, bridges, and culverts. Annual peak-streamflow frequency represents the peak streamflow for nine recurrence intervals of 2, 5, 10, 25, 50, 100, 200, 250, and 500 years. Common methods for estimation of peak-streamflow frequency for ungaged or unmonitored stream watersheds are regression equations for each recurrence interval developed for one or more regions (not strictly geographic); such regional equations are the subject of this report. The method is based on analysis of annual peak-streamflow data from U.S. Geological Survey (USGS) streamflow-gaging stations (stations).

Beginning in 2007, the USGS, in cooperation with the Texas Department of Transportation (Research Project 0–5521) and in partnership with Texas Tech University, began a 3-year investigation concerning the development of regional equations to estimate annual peak-streamflow frequency for undeveloped watersheds in Texas. The general approach was explicitly limited to use of L-moment statistics, which are used in conjunction with a technique of multi-linear regression referred to as PRESS minimization (Asquith and Thompson, 2008). The approach used to develop the regional equations, which was refined during the investigation, is referred to as the "L-moment-based, PRESS-minimized, residual-adjusted approach."

The study area for this investigation includes Texas and selected parts of neighboring states and essentially is the same as that considered by Asquith and Slade (1997, 1999). These two studies also represent previous research concerning estimation of peak-streamflow frequency for ungaged watersheds in Texas. Asquith (2001) provides analysis of peak-streamflow frequency in Texas using L-moment statistics and primarily reports the effects of regulation (flood control) on those statistics.

The background and discussion provided by Asquith and Thompson (2008), which is based heavily on Asquith and Thompson (2005), represents applicable discussion, analysis, and ideas that greatly influenced the line of research described in this report. To establish context, the major new contributions of this report relative to Asquith and Thompson (2005, 2008) are the additions of L-moment statistics, with subsequent use of numerous probability distributions, and the residual adjustment. These additions (topics or themes) are described in the "An L-moment-Based, PRESS-Minimized, Residual-Adjusted Approach" section of this report.

Purpose and Scope

The purpose of this report is to present regression equations from the L-moment-based, PRESS-minimized, residual-adjusted approach for estimation of annual peak-streamflow frequency for undeveloped watersheds in Texas and restricted mainly to ungaged watersheds. The approach, which required a complex computational framework, has three thematic components:

- L-moment statistics of the annual peak-streamflow data, by station, are used to fit seven probability distributions to the data. From these seven distributions, representative values (estimates) of station-specific, peak-streamflow frequency were extracted for each of the nine recurrence intervals;
- Weighted-least-squares, multi-linear regression analysis using the station-specific, peak-streamflow frequency and selected watershed characteristics is used to develop regression equations. The regression included a technique to minimize the PRediction Error Sum of Squares (PRESS) statistic by power transformation of drainage area; and
- An adjustment based on regression residuals is created that represents a generalized correction for climate, terrain, and other variables not otherwise expressed by the selected watershed characteristics.

The report is limited to the annual peak-streamflow data for 677 and 638 selected stations. Two distinct station counts are used in this report as explained in the next section. The data for these watersheds are provided in the two text files `Appendix1_677annpks.txt` and `Appendix1_638annpks.txt`, which are described in appendix 1. The report also is limited to three selected watershed characteristics (contributing drainage area, dimensionless main-channel slope, and mean annual precipitation), which previously have been shown to be important predictors of peak-streamflow frequency in Texas (Asquith and Slade, 1997; Asquith and Thompson, 2008) and are identified in the next section.

Identification of Annual Peak-Streamflow Data

From an initial candidate station count of 1,030 stations that had 1 or more years of annual peak-streamflow data in the study area, the number of stations used for this report were reduced by review of the data, watershed conditions, and a minimum record-length criterion.

Assessment of undeveloped conditions by peak-streamflow qualification codes in the authoritative USGS "peak value" database (U.S. Geological Survey, 2008) for the 1,030 stations was made along with historical and retrospective analysis concerning development in each watershed (such as the years of construction of substantial flood-control reservoirs or growth of municipalities). The assessment was further augmented by graphical depiction and review of a time series for each of the 1,030 stations. To complete the assessment, intervals of acceptable record from the beginning of record through a station-specific terminal year were identified.

Stations considered for this report include those with at least 8 years (through the 2006 water year[1]) of annual peak-streamflow data that are representative of undeveloped streamflow conditions in Texas (acceptable record in the previous paragraph). Undeveloped conditions, in regard to annual peak streamflow, are watershed conditions representative of ideals such as natural, rural, unregulated, and unurbanized. Furthermore, surface-water diversions or return flows are anticipated to be unsubstantial relative to the typical magnitude of individual annual peak-streamflow values.

After the assessment of undeveloped conditions, time series, and minimum record-length criterion, 677 stations remained. The data for these 677 stations were used for a specific component of the analysis described in the "Sampling Error" section of this report. The annual peak-streamflow data for these stations are available in text file `Appendix1_677annpks.txt`, which is described in appendix 1.

The selected watershed characteristics, which will be important in later regression analysis, are derived from Asquith and Slade (1997) and augmented with digital processing and verification (David B. Thompson and Lucia S. Barbato, Texas Tech University, written commun., 2007). The watershed characteristics of contributing drainage area (drainage area), dimensionless main-channel slope, and mean annual precipitation are used. Assessments of the reliability of these watershed characteristics, which were obtained in early stages of this investigation, were made.

The primary watershed characteristics used for this investigation are summarized as follows:

- DRAINAGE AREA is the horizontal projection of the area that directs water to the streamflow-gaging station and is measured in square miles;
- DIMENSIONLESS MAIN-CHANNEL SLOPE is defined as the magnitude of the change in elevation in feet between the two end points of the main channel divided by the MAIN-CHANNEL LENGTH in feet. Main-channel length is defined as the length in stream-course miles of the longest defined channel from the approximate watershed headwaters to the outlet. Although not used directly in the regression equations, these lengths are needed for computation of dimensionless main-channel slope. Because of the wide range in drainage area, various methods for length estimation have been used and are represented in the database. Asquith and Slade (1997) provide a specific definition based on 1:100,000-scale maps. For this investigation digital verification, through 30-meter or 90-meter digital elevation models and digital line graphs, was used to define the main channel. Manual techniques were occasionally used as station-specific circumstances required; and
- MEAN ANNUAL PRECIPITATION is the arithmetic mean of a suitably long period of time of total annual precipitation in inches. The mean annual precipitation was assigned based on the approximate center of the watersheds. Asquith and Slade (1997) generally provide the values used for this report. The period 1951–80 was used by Asquith and Slade because those authors felt that most of the data (measured in then active stations) in the database were within this time period. Given the many sources of uncertainty, the authors of the current (2009) report consider that any general and authoritative source of mean annual precipitation for any suitably long period (perhaps 30 years) is sufficient for substitution into the regional regression equations reported here.

Another assessment or restriction for the purpose of the regression analysis is that stations with drainage areas less than 10,000 square miles were used. This threshold was chosen from review of diagnostics of exploratory regression analysis—most watersheds having drainage areas this large are currently (modern times) considered regulated—and most design needs of the Texas Department of Transportation are for much smaller watersheds in Texas.

After the assessment of watershed characteristics and the less-than-10,000-square-mile criterion, 638 stations remained. The watershed characteristics for these stations are available in file `Appendix1_638wtrshdchr.txt`, which is described in appendix 1. The annual peak-streamflow data for these stations are available in file `Appendix1_638annpks.txt`, which also is described in

[1] A water year is the 12-month period between October 1 and September 30 and is designated by the calendar year in which it ends. Thus, the year ending September 30, 2006, is the "2006 water year."

Table 1. Summary statistics of the watershed characteristics for 638 U.S. Geological Survey streamflow-gaging stations used to develop regional equations for estimation of peak-streamflow frequency.

[*A*, drainage area in square miles; *S*, dimensionless main-channel slope; and *P*, mean annual precipitation in inches]

Characteristic	Minimum	1st quartile	Median	3rd quartile	Maximum
A	0.100	6.403	97.40	499	9,329
$100 \times S$	.023	.152	.269	.657	7.03
P	8	23	31	41	57

appendix 1. Summary statistics of the selected watershed characteristics for the 638 stations are listed in table 1.

Acknowledgments

The authors recognize the contributions of many Texas Department of Transportation engineers including Amy Ronnfeldt, Design Division, 0–5521 Project Director; David Zwernemann, Project Advisor; and George Herrmann, San Angelo District, former Project Director and Advisor. The authors also acknowledge the contributions of our three research supervisors for this project at Texas Tech University: Professors David B. Thompson (formerly Texas Tech University, 2007), Ken Rainwater (2008), and Theodore G. Cleveland (2009). The authors also acknowledge the geographic information system contributions of Lucia S. Barbato, Center for Geospatial Technology, Texas Tech University.

An L-moment-Based, PRESS-Minimized, Residual-Adjusted Approach

A generalized overview of the L-moment-based, PRESS-minimized, residual-adjusted approach (referred to as the "Approach") used for this investigation is described in this section. The Approach is complex, highly technical, and challenging to succinctly describe. Furthermore, the Approach relied on numerous single-purpose, highly specialized scripts, computer programs, and integration with features of the host-operating system as well as considerable judgement and iterative refinement on the part of the authors. The computational framework was ultimately implemented as a relatively turnkey process from start to finish, but nearly 6 months of effort were required to build, test, and refine the process to produce the equations reported in the "Regression Equations in Texas Using an L-moment-Based, PRESS-Minimized, Residual-Adjusted Approach" section of this report.

The generalized overview of the Approach consists of three thematic elements. Furthermore, several R code listings are presented to demonstrate critical computations. First, the method of computing station-specific, peak-streamflow frequency estimates using L-moment statistics is described as well as methods of obtaining estimates of uncertainty. Second, the use of weighted-least-squares, multi-linear regression in the context of PRESS minimization is described. Third and finally, the residual adjustment and subsequent recomputation of the regression equations incorporating a special watershed characteristic known as the $\mho$ parameter (also the term "OmegaEM"[2]) is described.

L-moment-Based, Station-Specific Peak-Streamflow Frequency Estimates

L-moments (Hosking, 1990) summarize samples and can be used to fit probability distributions. L-moments are based on linear combinations of differences of the expectations of order statistics as opposed to the more familiar product moments, which are based on powers (exponents) of differences from the mean. This distinction results in favorable sampling properties including unbiasness, robustness, and reliable measures of distribution shape. The mathematical theory and comparisons between L-moments and product moments are available in several sources and the references therein (Hosking, 1990; Stedinger and others, 1993; Hosking and Wallis, 1997; Asquith, 2001, 2006; Asquith and others, 2006). Furthermore, in a peak streamflow application, treatment for low outliers caused by typical $\log_{10}$ transformation (Stedinger and others, 1993, p. 18.5) of the data prior to statistical processing is not needed. Hosking (2006, p. 194) reports that L-moments "are now widely used in hydrology to summarize data and fit flood frequency [peak-streamflow frequency] distributions" and provides several citations of such practice.

In brief, L-moments were used for this investigation for station-specific computations of peak-streamflow frequency because the method removes or at least relaxes the need to consider concepts such as low-outlier thresholds, high-outlier thresholds, historical information, and

[2]A textual representation of the mathematical nomenclature also is needed (actually required) by the portable document format (PDF) standard for the headings, bookmarks, and other features of the PDF version of this report.

generalized skew as described in other methods for peak-streamflow frequency computation such as those in the Interagency Advisory Committee on Water Data (1982) guidelines.

For this report, the algorithms of Asquith (2008) were used within the R environment (R Development Core Team, 2008). These algorithms are based on many years of iterative development from various authentic and authoritative sources such as Hosking (1990, 1996) and Hosking and Wallis (1997). Specific functions from Asquith (2008) are identified in the following sections. The use of function names in the description of the L-moment-based portion of the Approach permits the exclusion of mathematics and succinct demonstration or examples of critical-to-document computations.

Sample L-moments

The sample L-moments represent specific metrics of "distribution geometry" for lack of a better term. The first L-moment is the arithmetic mean and the second L-moment is analogous, but not numerically equal, to the standard deviation. Higher-order L-moments are measures of distribution skewness (symmetry), kurtosis, and other concepts.

For the Approach, the first five sample L-moments were computed for each of the 677 stations using the `lmoms()` function of Asquith (2008). These L-moments, respectively, are the mean, L-scale, L-skew, L-kurtosis, and Tau5. The coefficient of L-variation or L-CV is defined as the ratio of L-scale to the mean. The sample L-moments by station were considered results of intermediate computations and are not reported here.

An example of the computation method of the sample L-moments for station 07153500 Dry Cimarron River near Guy, New Mex., is shown in the following R code listing. This station is the first listed in file `Appendix1_677annpks.txt` and has 33 years of annual peak-streamflow data identified for this investigation. This station will be used repeatedly in the example computations described herein.

```
> library(lmomco) # load in the lmoms() function
# annual peaks in cubic feet per second for
# station 07153500
> peaks <- c( 8200, 7120, 1440, 2960,  3000,
               980, 2140, 6880, 4350,  3350,
              1330, 3950, 2800, 8100,  8500,
              5200, 1820, 1610, 1190,  3270,
              4920,  475, 3930, 2070, 46100,
             22500, 4310, 3660,  987,   962,
              2890,  723,  510) # the peaks
> length(peaks) # returns 33 years as the total
# count of annual peak streamflow
[1] 33
> lmr <- lmoms(peaks) # compute sample L-moments
> str(lmr) # show the sample L-moments
List of 6
 $ lambdas : num [1:5] 5219 3011 1782 1463 1261
 $ ratios  : num [1:5]    NA 0.577 0.592 0.486 0.419
 $ trim    : num 0
 $ leftrim : NULL
 $ rightrim: NULL
 $ source  : chr "lmoms"
```

The listing shows that the mean and L-scale values are about 5,219 and 3,011 cubic feet per second, respectively. The listing also shows that the L-skew, L-kurtosis, and Tau5 values (dimensionless) are 0.592, 0.486, and 0.419, respectively. (The value 0.577 is L-CV.) These sample L-moments also will be used in the "Probability Distributions for Peak-Streamflow Frequency Curves" of this report.

Sampling Error

For the weighted-least-squares regression analysis, weight factors representing sampling error and modeling errors as distinct components of uncertainty are used. In this section, the concept and method of computation of a sampling error of peak streamflow for each recurrence interval by station is described. It is favorable to have stations with more record (smaller sampling error, more certainty) to have more weight in regression analysis.

For this report, sampling error is the concept that the peak-streamflow frequency for a given station is dependent on the number of years of available record. Stations having long periods of record (measured in many decades) have more defined and presumably more reliable peak-streamflow frequency values than the values for stations having short periods of record (measured in less than about 2 decades). Thus, sample size in years for a station acts to progressively decrease sampling error.

For this report, `weighted.mean()` values of the first five dimensionless sample L-moments of the 677 stations were computed by the so-named function in R. The years of record for each station in file `Appendix1_677annpks.txt` were used as weight factors.[3] The term dimensionless in this context implies a unit mean and the second L-moment equal to the L-CV. The weighted-mean, dimensionless L-moments are 1, 0.5055, 0.3939, 0.2496, and 0.1590 for the mean, L-scale, L-skew, L-kurtosis, and Tau5, respectively. These L-moments are referred to as the dimensionless regional L-moments, and these weighted-mean values provide an approximation of the geometry of the "parent" peak-streamflow frequency distribution for the study area.

[3] These weights are not to be confused with the weight factors used in the regression analysis reported herein.

The dimensionless regional L-moments were used to estimate a dimensionless regional Wakeby distribution (Landwehr and others, 1979) using the `parwak()` function of Asquith (2008). This fitted Wakeby distribution, because of its immense flexibility with five parameters, is assumed to provide a reasonable first-order approximation to the underlying (dimensionless) structure of peak-streamflow frequency for the study area. The use of the dimensionless regional Wakeby distribution is described in the remainder of this section.

Using statistical simulation and the `genci()` function of Asquith (2008) for each unique number of years of record for the 677 stations and the dimensionless regional Wakeby distribution, the L-CV for each of the nine recurrence intervals were estimated. An example of the computation method for station 07153500, which has 33 years of record, and the 2-year recurrence interval is shown in the following R code listing:

```
> library(lmomco) # load the vec2lmom(),
# parwak(), T2prob(), and genci() functions

# Manually set the dimensionless regional
# L-moments
> L <- vec2lmom(c(1,0.505,0.394,0.250,0.159))

# Fit a dimensionless Wakeby distribution to
# the regional L-moments in variable L
> W <- parwak(L)
# the algebraic form of distribution is shown
# in the text

# Compute nonexceedance probability (0.5)
# for 2-year event
> F <- T2prob(2)

> n <- 33; # sample size, years of record
> nsim <- 10000 # the num. of simulations and
# a large value to gain numerical accuracy
> genci(W, n, F=F, nsim=nsim)
  nonexceed_prob      lower       true      upper
       lscale       lcv
1          0.500  0.4725784  0.6667878  0.9042445
    0.07403137 0.1110269
```

The regional dimensionless Wakeby distribution fit to the L-moments shown in the previous R code listing is

$$Q(F) = -0.0266 + \frac{1.10}{6.10}(1-(1-F)^{6.10}) - \frac{0.692}{0.206}(1-(1-F)^{-0.206}), \qquad (1)$$

where $Q(F)$ is dimensionless peak streamflow (unit mean) for nonexceedance probability $F = 1 - 1/T$ for recurrence interval T.

In brief, the simulation process is as follows: For a simulation run of 10,000 iterations, a random sample of size n (33 years for the example) for each iteration was drawn from the regional dimensionless Wakeby distribution. Subsequently, for each iteration, “new” sample L-moments of the random sample were computed and a “new” Wakeby distribution fit to the new L-moments. If a Wakeby could not be fit, a generalized Pareto distribution was used instead; Hosking and Wallis (1997) provide details of the algorithm. The value of the distribution at a given T-year recurrence interval (2-year for the example) is stored. After the completion of the 10,000 iterations, the L-CV of the 10,000 stored estimates of the 2-year event was computed. This value represents the model error for a sample of size $n = 33$.

The previous R code listing shows that the number of interest is the value listed under the `lcv` heading. For the example, this value is about 0.111 (the last value in the output). Conceptually this value represents the relative (dimensionless) uncertainty in the estimation of the 2-year peak-streamflow value. Considerable computational effort was expended to estimate the L-CV values for each of nine selected recurrence intervals and 71 unique sample sizes (number of years of record) for the 677 stations. Thus, a total of 6.39 million simulations were conducted ($9 \times 71 \times 10{,}000$). The computation of modeling error was based on statistics of the 677 stations. From this point forward in the discussion of the Approach and results, the 638 stations become the subject.

In summary, these 639 L-CV values (9×71) represent the relative uncertainty in the peak-streamflow estimate at a given recurrence interval and for a given sample size. The uncertainty is relative to the “mean” estimate of the peak-streamflow magnitude. Discussion of this mean estimate is provided in the next section.

Probability Distributions for Peak-Streamflow Frequency Curves

For this report, as many as seven probability distributions are fit to the sample L-moments for each station and resultant frequency curves were graphically depicted and inspected by the authors in real and $\log_{10}$-transformed space.

The authors explicitly chose to avoid selection of a single form of a probability distribution to model the station-specific, peak-streamflow frequency through simultaneous use of a substantial number of three-parameter and more distributions. Further, by use of more than one probability distribution, estimates for modeling error (see next section) could be obtained.

The probability distributions considered and parameter and quantile functions of Asquith (2008) used for this report are as follows:

- Generalized extreme value distribution—supported by functions `pargev()` and `quagev()`;
- Generalized logistic distribution—supported by functions `parglo()` and `quaglo()`;
- Generalized normal distribution—supported by functions `pargno()` and `quagno()`;
- Generalized Pareto distribution—supported by functions `pargpa()` and `quagpa()`;
- Kappa distribution (if solution available), otherwise the generalized lambda distribution (if solution available)—supported by functions `parkap()` and `quakap()` for kappa and functions `pargld()` and `quagld()` for generalized lambda;
- Pearson Type III distribution—supported by functions `parpe3()` and `quape3()`; and
- log-Pearson Type III distribution, which is the same as Pearson Type III but the L-moments of the $\log_{10}$ values of data are used instead to fit the distribution.

The seven quantile functions for the itemized probability distributions are internally called from the `qlmomco()` function, which technically was used in algorithms by the authors.

In summary, for almost all stations, seven distributions were fit and seven estimates of peak streamflow for each of the nine recurrence intervals were stored for later use. An example of the computation method for only the generalized extreme value and generalized logistic distributions (two of the seven distributions), which are accessed using the `qlmomco()` function, for station 07153500 is shown in the following R code listing:

```
> library(lmomco) # load the lmoms(), qlmomco(),
                  # and lmom2par() functions
# The variable peaks was generated in the section
# ‘‘Sample L-moments’’
> lmr <- lmoms(peaks) # compute sample L-moments

# Compute the 2-year peak streamflow using two
# probability distributions: generalized extreme
# value and generalized logistic distributions
> Q2gev <- qlmomco(0.5, lmom2par(lmr, type="gev"))
> Q2glo <- qlmomco(0.5, lmom2par(lmr, type="glo"))

# Show the estimates in cubic feet per second
> print(Q2gev) # 2-year generalized extreme value
[1] 2724.746
> print(Q2glo) # 2-year generalized logistic
[1] 2754.674
```

The listing shows that the 2-year peak streamflow from the generalized extreme value is about 2,725 cubic feet per second, and the 2-year peak streamflow from the generalized logistic distribution is about 2,755 cubic feet per second. The difference is a natural and an expected result of the form of the fitted distributions. Generally, as recurrence interval increases, differences in estimates from different distributions tend to increase. These two values will be used with five others from five additional distributions in the next section.

For station 07153500, the seven probability distributions are shown in figure 1. Each graph depicts the peak-streamflow frequency curves in real and logarithmic units (A and B, respectively). The horizontal axis is rendered in standard normal deviates[4] and hence is a normal probability axis although not labeled in units of probability. Also on the figure, the individual plots are rendered with the respective station number to ensure consistency with the figure caption. The seven estimates, by distribution, by recurrence interval, and by station were considered results of intermediate computations and are not reported here. However, the more than 677-page, portable document format (PDF) file `Appendix1_677freqcurves.pdf` described in appendix 1 is available. This file provides graphical documentation of the computations described in this section and repeats figure 1 for completeness.

Trimmed Mean Estimates of Peak-Streamflow Frequency

For each of the nine recurrence intervals, a symmetrically trimmed L-moment (trimmed mean) was computed using the `TLmoms(QT,trim=1)` function of Asquith (2008) based on trimmed L-moments (Elamir and Seheult, 2003; Hosking, 2007), where `QT` is a vector that represents as many as (generally) seven unique values of peak streamflow for a given recurrence interval. The use of trimming, which computationally is more complex than simply dropping the largest and smallest values and computing an arithmetic mean, is anticipated to provide a more robust measure of central location for the formal estimation of station-specific, peak-streamflow frequency. It is important to note that $\log_{10}$ transformation of the peak-streamflow estimates was not made prior to computation of the trimmed mean.

The computation of the trimmed mean is demonstrated using results for station 07153500 for the 2-year peak streamflow. In the following R code listing, the estimates for each of the seven distributions are set into variables and

[4] Standard normal deviate—A normally distributed random variable (or nonexceedance probability for the plots) with an expected value of 0 and a variance of 1.

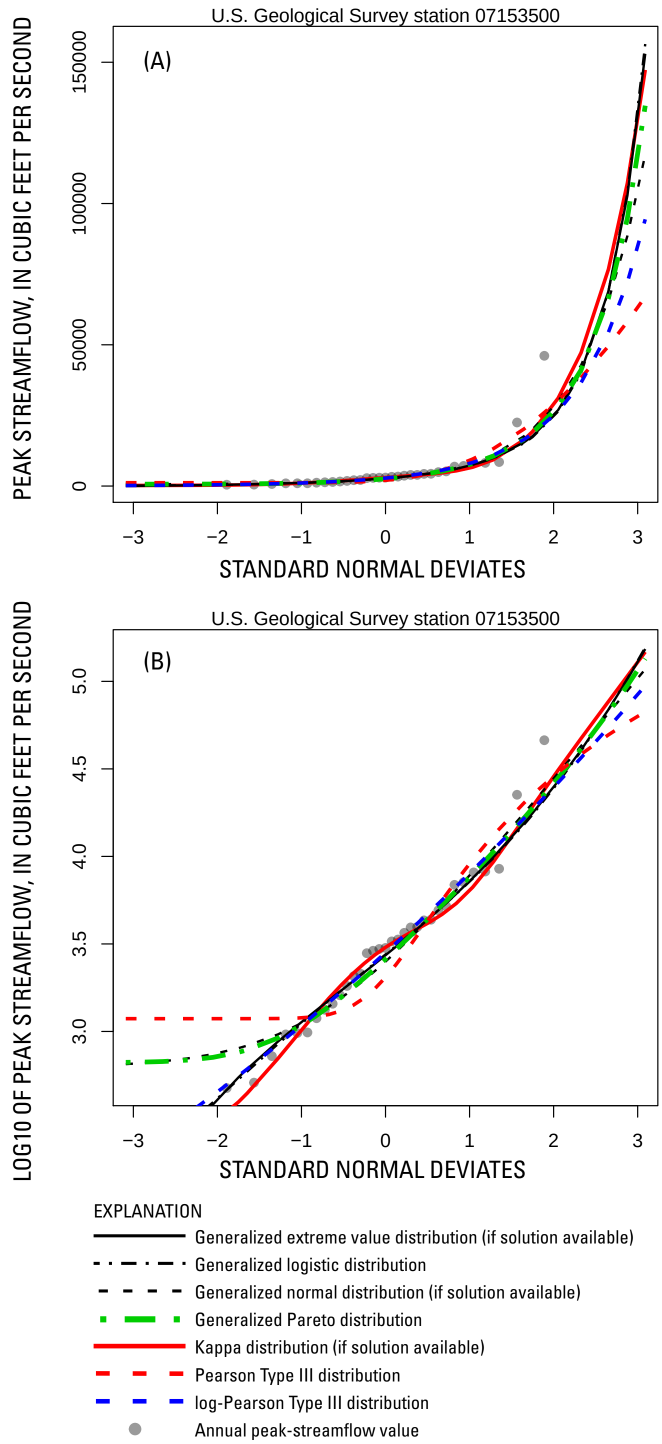


Figure 1. Example of peak-streamflow frequency curves with superimposed axis labels and explanation in real and logarithmic space rendered from the R environment (R Development Core Team, 2008) using functions of Asquith (2008) for seven probability distributions fit to sample L-moments of annual peak-streamflow data available in file `Appendix1_677annpks.txt` for station 07153500 Dry Cimarron River near Guy, New Mexico.

then concatenated into the variable `Q2`. The arithmetic and trimmed means are computed using the `TLmoms()` function of Asquith (2008). The values for `gev` and `glo` were computed in the "Probability Distributions for Peak-Streamflow Frequency Curves" section of this report.

```
> library(lmomco) # load the TLmoms() function
# Seven unique estimates for 2-year recurrence
# interval in cubic feet per second
> gev    <- 2725 # from generalized extreme value
> glo    <- 2755 # from generalized logistic
> gno    <- 2507 # from generalized normal
> gpa    <- 2590 # from generalized pareto
> kap    <- 3023 # from kappa
> pe3    <- 2037 # from pearson type III
> pe3log <- 2820 # from log-pearson type III
> Q2 <- c(gev,glo,gno,gpa,kap,pe3,pe3log)

# Now compute the usual mean (not used)
> TLmoms(Q2,nmom=1, trim=0)
$lambdas
[1] 2636.714

# Now compute the symmetrically trimmed mean
> TLmoms(Q2,nmom=1, trim=1)
$lambdas
[1] 2683.429
```

The listing shows that the usual arithmetic mean of the 2-year peak streamflow for the station is about 2,637 cubic feet per second. The trimmed mean is about 2,683 cubic feet per second and was used. The authors experimented with the arithmetic mean, median, and trimmed mean in the iterative development of the Approach; the authors judged that the trimmed mean provided an appropriate statistic to estimate station-specific, peak-streamflow frequency.

In summary, the trimmed mean was used to estimate the station-specific peak streamflow for each of the nine recurrence intervals. The trimmed means are used in the regression analysis described in the "Weighted-Least-Squares, Multi-Linear Regression Analysis and PRESS Minimization" section of this report. The trimmed means by recurrence interval and by station were considered intermediate computations, but these results are available in file `Appendix1_677trimmedQTs.txt`, which is described in appendix 1.

Modeling Error

Because seven distributions (models) were fit for each station, seven unique estimates of peak streamflow were obtained for each recurrence interval. For a given recurrence interval, these estimates have a distribution; the relative variability of this distribution is treated as a measure of modeling error. Stations with more variability in peak-streamflow estimates, which is attributable to distribution choice, have more uncertainty than those stations with less variability. Therefore, it is favorable for stations with more uncertainty to have less weight in regression analysis.

The computation of the modeling error is demonstrated using results for station 07153500 for the 2-year peak streamflow. In the following R code listing, the estimates for each of the seven distributions are available in the variable `Q2` set in the previous code listing. The L-moments (not trimmed) are computed using the `lmoms()` function of Asquith (2008).

```
> library(lmomco) # to access lmoms() function
# Compute sample L-moments of seven distributions
# for the 2-year recurrence interval
> lmoms(Q2,nmom=2) # no trimming
$lambdas
[1] 2636.7143  178.5238
$ratios
[1]         NA  0.06770692
# Thus, after rounding off the decimal
# Lambda_1 (mean) = 2637; Lambda_2 (L-scale) = 179
```

The listing shows that the L-scale of the seven fits is about 179 cubic feet per second. An approximate variance in R code for the model fit, and representative of model error, is thus `(179*sqrt(pi))^2` or 100,660. (The `sqrt()` function is the square root.)

Combining Sampling and Modeling Errors

Combination of the modeling error with sampling error is straightforward. The standard deviation of a peak-streamflow frequency estimate can be estimated by the Pythagorean distance of the variances of the sampling and modeling errors.

In the "Sampling Error" section of this report, station 07153500 has a sampling error measured in L-CV of about 0.111. If the trimmed mean estimate for the station is 2,683 cubic feet per second for the 2-year event (see section "Trimmed Mean Estimates of Peak-Streamflow Frequency"), then the sampling error has an approximate variance in R code of `(0.111*(2683)*sqrt(pi))^2` or 278,636. The previous section shows that the modeling error variance is about 100,660.

These two variances can be aggregated in R code by `sqrt(278636 + 100660)` for a combined error of about 616 cubic feet per second. Hence, the standard deviation of the 2,683 cubic feet per second estimate of the 2-year peak streamflow for station 07153500 is 616 cubic feet per second, which can be expressed as a relative variability as the ratio `616/2683` or about 0.230. This relative error, actually its inverse, was used as one of the 5,742 weight factors in weighted-least-squares regression analysis. The weight factor count is the product of 638 stations and nine recurrence intervals (638 × 9).

Weighted-Least-Squares, Multi-Linear Regression Analysis and PRESS Minimization

For this report, weighted-least-squares, multi-linear regression analysis is used to develop the statistical relation between the station-specific, trimmed-mean, T-year peak-streamflow values and the three watershed characteristics (explanatory variables previously described). The sampling and modeling errors for each station were combined as described in the "Combining Sampling and Model Errors" section of this report. The combined error is used to derive weight factors for the weighted-least-squares regression between peak-streamflow frequency and the watershed characteristics of drainage area, dimensionless main-channel slope, mean annual precipitation, and, when used, the Ω parameter described in the next section.

PRESS minimization is formally described by Asquith and Thompson (2008) and is only summarized in this report. The PRESS computation used for this investigation was not weighted by the regression weight factors as used by Asquith and Thompson (2008). The PRESS statistic generally is regarded as a measure of regression performance when the model is used to predict new data (Montgomery and others, 2001, p. 153). Prediction of new data is what analysts and hydrologic engineers do when they estimate peak streamflow from a regression equation. Regression equations with small PRESS values are desirable; thus, PRESS minimization is an appropriate goal. Helsel and Hirsch (2002, p. 247) state that, "Minimizing PRESS means that the equation produces the least error when making new predictions." Conceptually, PRESS minimization identifies the appropriate transformation of drainage area to "press" the bias (residual curvature) out of the equations.

Other variables were not "pressed" because curvature in the residuals is sufficiently removed by use of drainage area. Further, considerable computational complexities are introduced in investigation of simultaneous multi-transform optimization.

Because the PRESS statistic is an overall measure of regression fit (like residual standard error) and is a validation statistic (unlike residual standard error), minimization of PRESS is desirable. The most "valid" regression is produced when the PRESS statistic is minimized. The following transformation on drainage area was used:

$$A' = A^{\lambda}, \tag{2}$$

where A' is the transformed value of drainage area for the regression, A is drainage area, and λ is a real number. For this report specific values of λ were determined by exhaustive search to three significant figures for each regression equation. Thus, values for λ can be thought of as regression-specific parameters, but are not formally counted as such in diagnostic statistics of the regression. The λ values computed by the PRESS minimization are reported in the "Regression Equations in Texas Using an L-moment-Based, PRESS-Minimized, Residual-Adjusted Approach" section of this report.

Residual-Adjusted Regression and the OmegaEM Parameter

The terminal steps towards computation of the final (preferred by authors) regression equations reported here involved an iterative reprocessing through the development of a generalization of regression residuals. The generalized regression residual has been previously identified herein as the Ω parameter, and Ω can be thought of as a special watershed characteristic, which is analogous to some other unknown, but spatially varying variable that expresses peak-streamflow potential in the study area.

The residual-adjusted, regression process comprises the following steps:

1. For each of the nine recurrence intervals, weighted-least-squares regression analysis using PRESS minimization was made using only the watershed characteristics of drainage area, dimensionless main-channel slope, and mean annual precipitation for the 638 stations. The raw outputs from scripts operating in the R environment for each recurrence interval are shown in appendix 2. The information contained therein is interpreted and formally presented in the "Regression Equations in Texas Using an L-moment-Based, PRESS-Minimized, Residual-Adjusted Approach" section of this report;

2. The residuals for each of the nine regression equations were computed. Maps depicting the spatial variation of the residuals with symbols determined by the magnitude and direction of the residual were created and are formally presented in appendix 1 of this report. Spatial dependency is evident for all

recurrence intervals. For each recurrence interval, the median residual was computed for each square of 1-degree quadrangle of latitude and longitude in the study area containing at least one station;

3. The authors, through exploratory analysis and judgement, selected the median values of the residuals to pool or combine into 1-degree quadrangles for the 10-year recurrence interval. The authors, through manual smoothing, consultation of various geologic and ecological region maps, interpretation of regional topographic maps, and resident familiarity with the study area, estimated the magnitude and sign of the 10-year recurrence interval residuals for each 1-degree quadrangle in the study area. These estimates are referred to as the Ω parameter, are in units of $\log_{10}$(streamflow), and are formally in the "OmegaEM Parameter" section of this report;
4. For each station, the Ω parameter was assigned based on the 1-degree quadrangle containing the station. In practice, the Ω parameter could be assigned similarly or by judgement considering the location and spatial extent of the watershed in question;
5. For each of the nine recurrence intervals, weighted-least-squares regression using PRESS minimization is made using the watershed characteristics of drainage area, dimensionless main-channel slope, mean annual precipitation, and Ω parameter for the 638 stations. The raw output from scripts operating in the R environment for each recurrence interval are shown in appendix 3. The information contained therein is interpreted and formally presented in the "Regression Equations in Texas Using an L-moment-Based, PRESS-Minimized, Residual-Adjusted Approach" of this report. New power transformations by recurrence interval on drainage area are computed by the PRESS minimization. Further, for each recurrence interval a separate regression coefficient on the Ω parameter is computed; and
6. As in the second step, the residuals of the regressions using the Ω parameter were computed and mapped. Considerable reduction in spatial dependency is evident. The authors interpret that the residuals from the regressions using the Ω parameter are reasonably spatially invariant.

Introduction of a final regression diagnostic is needed. Akaike Information Criterion (AIC) statistic is reported here and is a measure of information content of a regression model. The statistic accounts for a tradeoff between number of parameters and the fit of the model; small values are sought.

Regression Equations in Texas Using an L-moment-Based, PRESS-Minimized, Residual-Adjusted Approach

The final results of the L-moment-based, PRESS-minimized, residual-adjusted approach are described in this section. This section is organized as follows. First, the regional equations using drainage area, dimensionless main-channel slope, and mean annual precipitation are formally presented and discussed. Second, maps that depict the Ω parameter are presented and discussed. Third, the regional equations using drainage area, dimensionless main-channel slope, mean annual precipitation, and the Ω parameter also are formally presented and discussed. Fourth, example computations and comparison to estimates from the regression equations to those of previous studies is provided and interpretations are made. Finally, this section ends with example computations and discussion about considerations for application of the regional equations to gaged and ungaged watersheds.

Regression Equations without OmegaEM Parameter

Nine weighted-least-squares, PRESS-minimized regression equations were computed using the watershed characteristics of drainage area, dimensionless main-channel slope, and mean annual precipitation and each of the 638 trimmed-mean, peak-streamflow estimates for each of the nine recurrence intervals. The multi-linear regression was made using the `lm()` function of R Development Core Team (2008). The weights for the regression equations were derived from the sampling and modeling errors described in previous sections. The PRESS-minimization was made using a user-guided search algorithm. Separate PRESS-minimizations are made and a unique exponent on drainage area is determined for each recurrence interval.

The final regression equations are listed in table 2 with important diagnostic statistics. Additional computational details for each equation are listed in appendix 2, in which potentially informative statistics that should be archived include summary statistics of the residuals, the p-values for the coefficients, maximum leverage statistics, and the inverted-covariance matrix. Finally, the residuals for each station and for each 1-degree quadrangle by recurrence interval are available in file `Appendix1_residualmaps.pdf` described in appendix 1.

Several general observations about the equations listed in table 2 can be made. First, there are 634 degrees of freedom in each equation—this is a substantially large number relative to many other hydrologic equations such as those by Asquith and Slade (1997), Asquith (2001), or Asquith and Roussel (2007). The mean of the nine residual standard errors is $0.34 \log_{10}$. The residual standard errors are generally larger than about $1/3 \log_{10}$, which by the authors' experience is difficult to eclipse in many hydrostatological models. The adjusted R-squared values are substantially large, which indicate that a reasonably large amount of collective variation in peak-streamflow frequency for the 638 stations is explained by the equations. The exponent on drainage area increases with increasing recurrence interval, which suggests that increasing curvature in the relation between T-year peak streamflow and drainage area exists (Asquith and Thompson, 2005, 2008).

OmegaEM Parameter

The Ω parameter represents a generalized terrain and climate index that expresses relative differences in peak-streamflow potential across the study area. The Ω parameter is interpreted as an expression of peak-streamflow potential not represented in the watershed characteristics of drainage area, dimensionless main-channel slope, and mean annual precipitation. Maps depicting, by 1-degree quadrangle and plotted at the quadrangle center, the Ω parameter superimposed on maps of hill-shade relief, Texas rivers (U.S. Geological Survey, 2003), and ecoregions (Commission for Environmental Cooperation, 1997), respectively, are shown in figures 2–4.

The Ω parameter was derived from interpretive analysis of the spatial distributions of residuals from the first regression analysis (table 2). The spatial interpretation of the residuals, in particular, was focused on those for the 10-year recurrence interval. The 10-year recurrence interval was chosen because either the PRESS statistic is at a minimum or the residual standard error is at a minimum (see table 2). This is not a new finding because the authors observe similar patterns in other studies (Asquith and Slade, 1997).

The authors interpret the minimum (PRESS $= 297$) at the 10-year recurrence interval as representing maximum predictability of peak-streamflow potential in relation to the three watershed characteristics. In particular, for recurrence intervals greater than about 10 years the authors envision that such peak streamflows are produced by storms of sufficient size (depth, volume, duration) that the watershed characteristics used in this study (dominated by drainage area) explain a maximum fraction of the variability of peak streamflow.

As recurrence interval increases beyond 10 years, the three watershed characteristics still play a primary role in production of peak streamflow; however the increasing magnitude of the PRESS statistics (decreasing adjusted R-squared, increasing residual standard error) is explained by increasing error in the peak-streamflow frequency estimates because quantile estimations are being made in increasingly more distal, right-hand regions of the probability distributions. In other words, regional equations for the 10-year recurrence interval represent a situation of some sort of maximum of information content.

The median residuals by 1-degree quadrangle are shown in file `Appendix1_residualmaps.pdf` described in appendix 1. These medians subsequently were spatially interpreted and further generalized through "moving" weighted-mean values—that is, overlapping averaging—of two or more neighboring 1-degree quadrangles. When weighted means were computed, the number of stations for each quadrangle were used as weights. These averaged or generalized residuals are plotted in figures 2–4 and labeled as the Ω parameter. The $\{30–31°, 104–105°\}$ quadrangle shows two Ω values. Because of the mountainous region in the approximately northern one-half of the quad, separate values were selected as better representation of the parameter than a single value.

The authors explicitly acknowledge that the development of the Ω parameter represents an ad hoc process and that alternative and potentially more optimal means of residual generalization could exist. As shown in the next section, the use of the Ω parameter as another variable in regional regression development provides a marked reduction in the PRESS statistic with associated increases (decreases) in adjusted R-squared and residual standard error.

Regression Equations with OmegaEM Parameter

Nine weighted-least-squares, PRESS-minimized regression equations were computed using the watershed char-

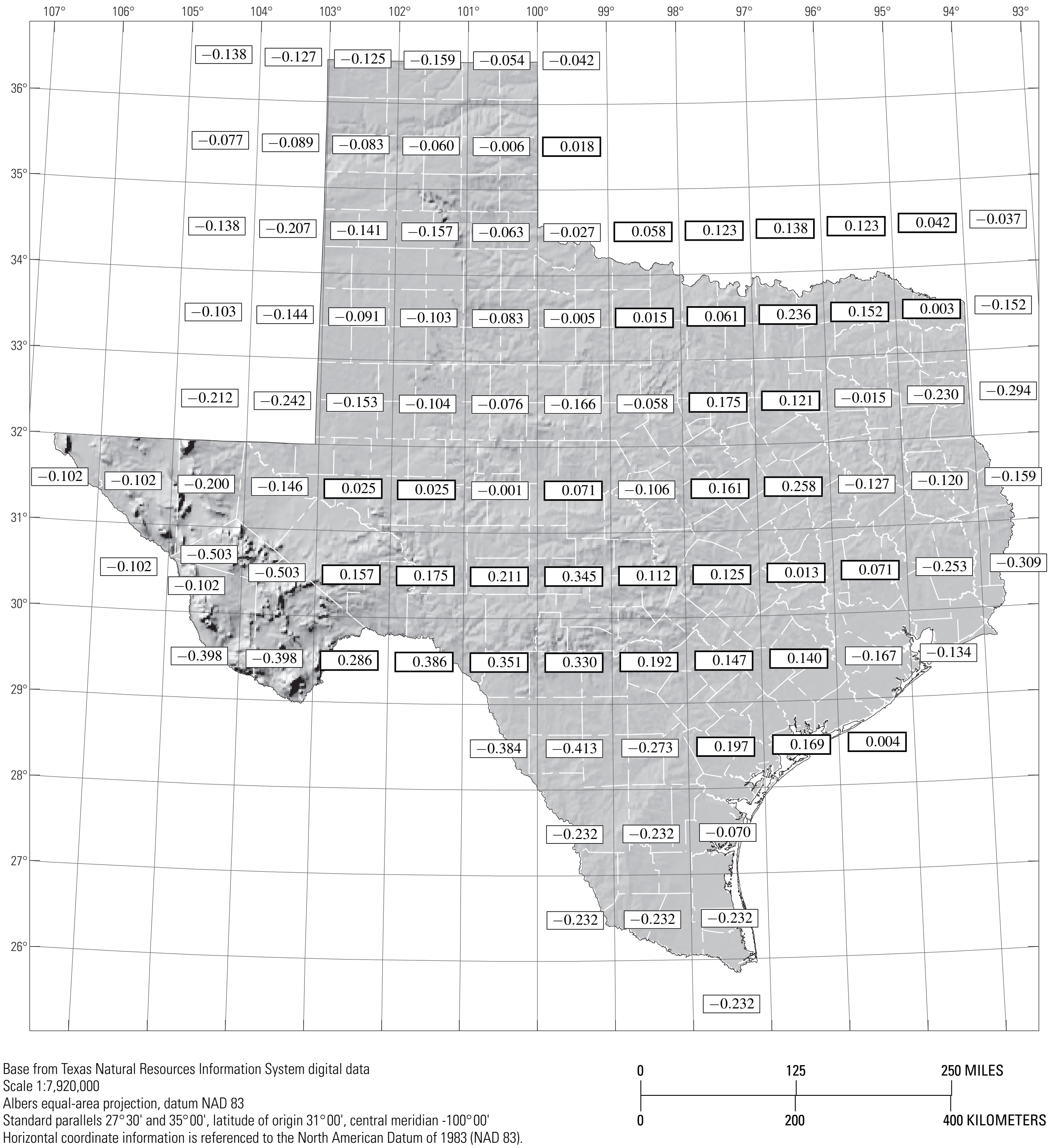


Figure 2. Hill-shade relief in Texas with superimposed values of OmegaEM parameter that represents a generalized terrain and climate index for regionalization of peak-streamflow frequency.

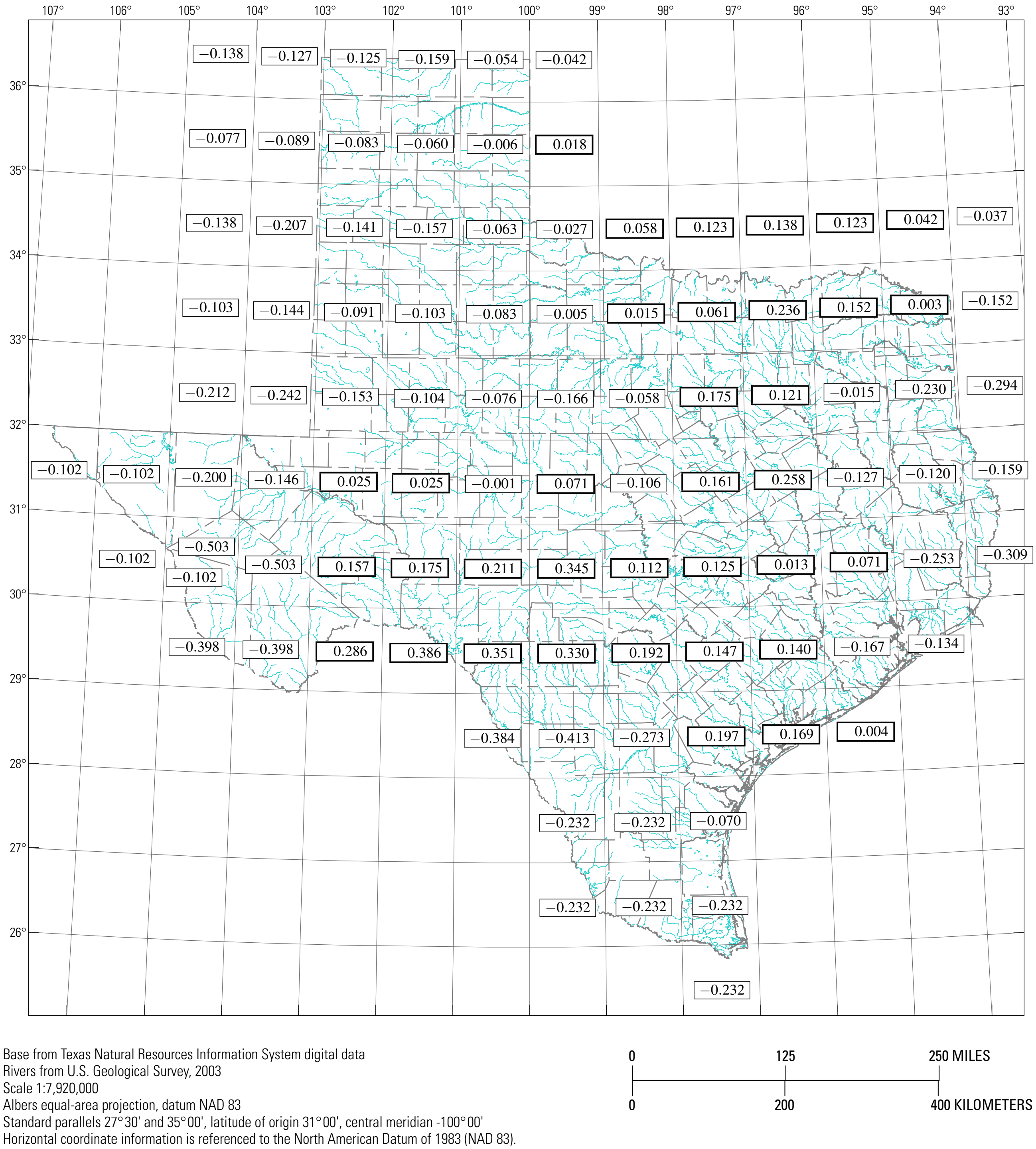
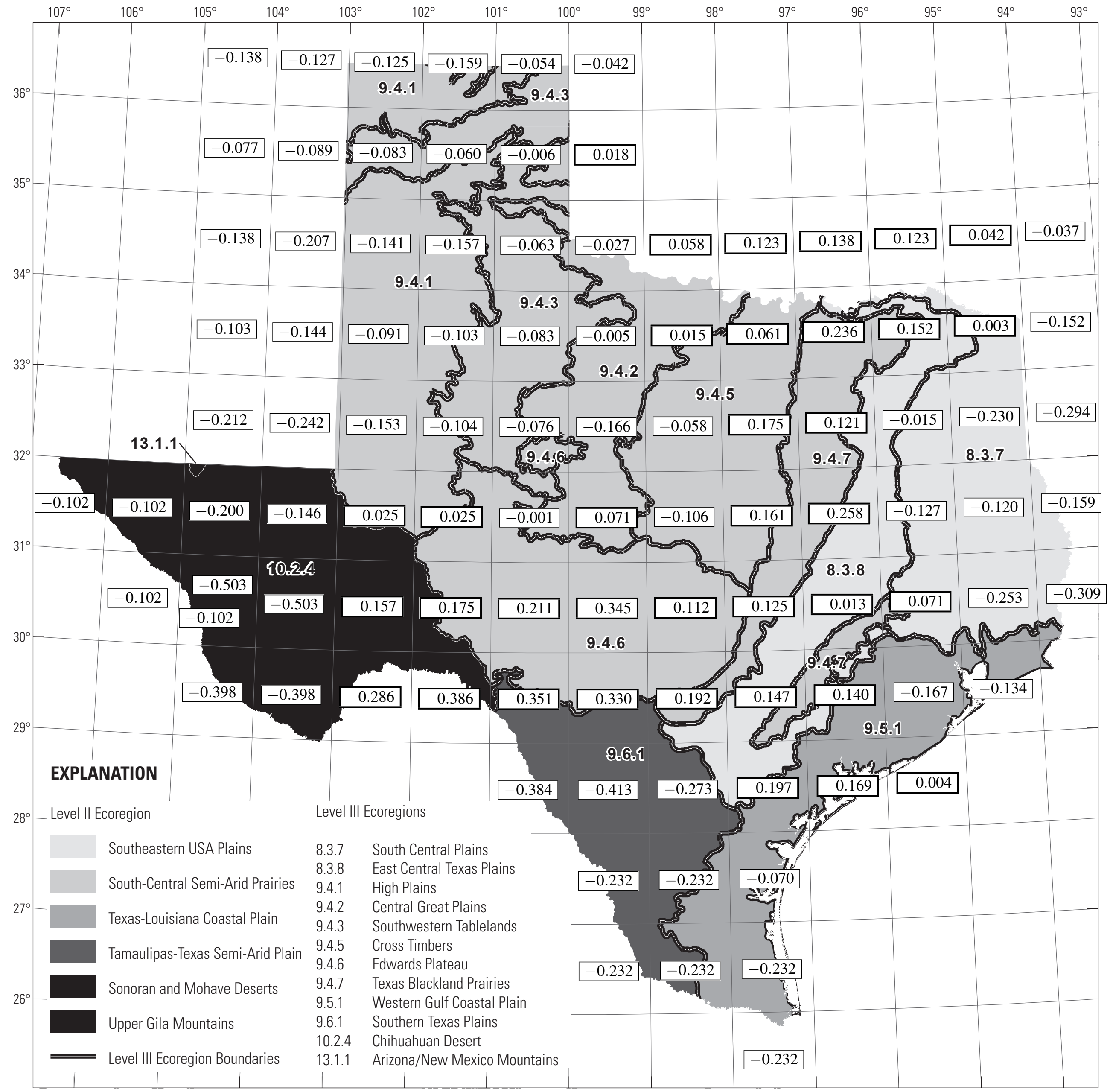


Figure 3. Rivers in Texas with superimposed values of OmegaEM parameter that represents a generalized terrain and climate index for regionalization of peak-streamflow frequency.



Base from Texas Natural Resources Information System digital data
Ecoregions from Commission for Environmental Cooperation, 1997
Scale 1:7,920,000
Albers equal-area projection, datum NAD 83
Standard parallels 27°30' and 35°00', latitude of origin 31°00', central meridian -100°00'
Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83).

0 125 250 MILES
0 200 400 KILOMETERS

Figure 4. Level II and III ecoregions in Texas with superimposed values of OmegaEM parameter that represents a generalized terrain and climate index for regionalization of peak-streamflow frequency.

Table 2. Summary of weighted-least-squares, PRESS-minimized, regional regression equations using drainage area, dimensionless main-channel slope, and mean annual precipitation.

[RSE, residual standard error in $\log_{10}$-units of cubic feet per second; Adj., adjusted; AIC, Akaike Information Criterion; PRESS, PRediction Error Sum of Squares; Q_T, peak streamflow for T-year recurrence interval in cubic feet per second; P, mean annual precipitation in inches; S, dimensionless main-channel slope; and A, drainage area in square miles]

Regression equation	RSE	Adj. R-squared	AIC statistic	PRESS statistic
$Q_2 = P^{1.562}\,S^{0.385} \times 10^{[38.64-37.94A^{-0.0080}]}$	0.32	0.81	396	77.4
$Q_5 = P^{1.491}\,S^{0.500} \times 10^{[16.35-15.04A^{-0.0228}]}$	.30	.84	314	65.2
$Q_{10} = P^{1.377}\,S^{0.530} \times 10^{[13.67-11.99A^{-0.0299}]}$	.30	.85	297	63.7 ← minimum
$Q_{25} = P^{1.316}\,S^{0.574} \times 10^{[12.00-9.992A^{-0.0380}]}$	.31	.84	348	67.1
$Q_{50} = P^{1.285}\,S^{0.608} \times 10^{[11.40-9.195A^{-0.0429}]}$	.33	.83	414	73.5
$Q_{100} = P^{1.257}\,S^{0.640} \times 10^{[11.07-8.673A^{-0.0470}]}$	.35	.82	493	82.8
$Q_{200} = P^{1.221}\,S^{0.665} \times 10^{[10.86-8.282A^{-0.0506}]}$	.37	.80	586	95.3
$Q_{250} = P^{1.209}\,S^{0.676} \times 10^{[10.83-8.181A^{-0.0517}]}$	.38	.79	617	100
$Q_{500} = P^{1.181}\,S^{0.705} \times 10^{[10.67-7.846A^{-0.0554}]}$	.41	.77	713	117

acteristics of drainage area, dimensionless main-channel slope, mean annual precipitation, and Ω (figs. 2–4) as well as each of the 638 trimmed-mean, peak-streamflow estimates for each of the nine recurrence intervals. The weights for the regression equations were the same as those used for the regression equations listed in table 2. Separate PRESS-minimizations were made for each recurrence interval and a unique exponent on drainage area was determined. These exponents differ from those listed in table 2.

The final regression equations are listed in table 3 with important diagnostic statistics. Additional computational details for each equation are listed in appendix 3, in which potentially informative statistics that should be archived include the summary statistics of the residuals, p-values for the coefficients, maximum leverage statistics, and inverted-covariance matrix. The residuals for each station and for each 1-degree quadrangle for each recurrence interval are available in file `Appendix1_residualmaps.pdf`.

Several general observations about the equations listed in table 2 can be made and are similar to those made for the equations listed in table 3. The mean of the nine residual standard errors is $0.30 \log_{10}$. The mean adjusted R-squared is about 0.86, and the mean percentage reduction in PRESS is about −21 percent. The residual standard errors are arguably smaller than $1/3 \log_{10}$, which was a professional goal of the authors for the Approach. The adjusted R-squared values are substantially large, which indicate a reasonably large amount of variation in peak-streamflow frequency for the 638 stations is explained. The exponent on drainage area increases with increasing recurrence interval, which suggests that increasing curvature in the relation between T-year peak streamflow and drainage area exists (Asquith and Thompson, 2005, 2008).

Of the two suites of regional equations shown in this report, the authors prefer the use of those based on Ω. Further, by noting the spatial dependence of residuals for the equations in table 2 (equations not based on Ω) and, by generality, such dependence for the equations in Asquith and Thompson (2005, 2008) must exist, the authors conclude that equations in table 3 are preferred over those in the two reports (Asquith and Thompson) because the equations in table 3 generally lack substantial spatial dependency in their residuals.

Example Computations for a Single Station and Comparison to Estimates from Previously Published Equations

This section provides example computations that demonstrate how the regional regression equations (tables 2 and 3) could be used in practice, and a comparison to estimates from previously published equations. The focus of the computations is on the 100-year peak streamflow, but application to other recurrence intervals is straightforward. For the computations, the watershed represented by station 08190000 Nueces River at Laguna, Tex., was arbitrarily chosen.

The suggested equation for estimation of the 100-year peak streamflow (see table 3) has the form

$$Q_{100} = P^c\,S^d \times 10^{[e\Omega+a+bA^{\lambda}]}, \qquad (3)$$

Table 3. Summary of weighted-least-squares, PRESS-minimized, regional regression equations using drainage area, dimensionless main-channel slope, mean annual precipitation, and OmegaEM.

[RSE, residual standard error in $\log_{10}$-units of cubic feet per second; Adj., adjusted; AIC, Akaike Information Criterion; PRESS, PRediction Error Sum of Squares; Percent change, percent change from PRESS in table 2 to PRESS listed to the left; Q_T, peak streamflow for T-year recurrence interval in cubic feet per second; P, mean annual precipitation in inches; S, dimensionless main-channel slope; Ω, OmegaEM parameter in figures 2–4; and A, drainage area in square miles]

Regression equation	RSE	Adj. R-squared	AIC statistic	PRESS statistic	Percent change
$Q_2 = P^{1.398}\, S^{0.270} \times 10^{[0.776\,\Omega + 50.98 - 50.30A^{-0.0058}]}$	0.29	0.84	273	64.6	−16.5
$Q_5 = P^{1.308}\, S^{0.372} \times 10^{[0.885\,\Omega + 16.62 - 15.32A^{-0.0215}]}$	.26	.88	122	49.1	−24.7
$Q_{10} = P^{1.203}\, S^{0.403} \times 10^{[0.918\,\Omega + 13.62 - 11.97A^{-0.0289}]}$	.25	.89	86.5	46.6	−26.8
$Q_{25} = P^{1.140}\, S^{0.446} \times 10^{[0.945\,\Omega + 11.79 - 9.819A^{-0.0374}]}$	.26	.89	140	49.5	−26.2
$Q_{50} = P^{1.105}\, S^{0.476} \times 10^{[0.961\,\Omega + 11.17 - 8.997A^{-0.0424}]}$	.28	.87	220	55.6	−24.4
$Q_{100} = P^{1.071}\, S^{0.507} \times 10^{[0.969\,\Omega + 10.82 - 8.448A^{-0.0467}]}$	.30	.86	320	64.8	−21.7
$Q_{200} = P^{1.034}\, S^{0.531} \times 10^{[0.975\,\Omega + 10.61 - 8.058A^{-0.0504}]}$	.33	.84	436	77.2	−19.0
$Q_{250} = P^{1.021}\, S^{0.541} \times 10^{[0.977\,\Omega + 10.56 - 7.943A^{-0.0516}]}$	.34	.83	474	81.9	−18.1
$Q_{500} = P^{0.988}\, S^{0.569} \times 10^{[0.976\,\Omega + 10.40 - 7.605A^{-0.0554}]}$	.37	.81	591	98.7	−15.6

where a, b, c, d, and e are regression coefficients (listed in this sentence in the order as produced by the R software) specific for the 100-year recurrence interval, λ is a power determined by iterative PRESS-minimization for the 100-year recurrence interval, P is mean annual precipitation in inches, S is dimensionless main-channel slope, A is drainage area in square miles, and the Ω parameter is a generalized terrain and climate index (figs. 2–4), which was derived from the residuals of the 10-year equation in table 2.

For the example computations, the 100-year regression equation (table 3) is

$$Q_{100} = P^{1.071}\, S^{0.507}\, 10^{[0.969\,\Omega + 10.82 - 8.448A^{-0.0467}]}, \quad (4)$$

for which the residual standard error is $0.30 \log_{10}$ and adjusted R-squared is 0.86.

Station 08190000 has a latitude and longitude of 29°25'42" and 99°59'49", respectively, with the following estimates from Asquith and Slade (1997, table 1) of the three watershed characteristics: $P = 24.5$ inches, $S = 0.00326$ (or about 17.21 feet per mile), $A = 737$ square miles. The value of $\Omega = 0.33$ is derived from figures 2–4 given the latitude and longitude of the station.

Solving the equation for Q_{100} produces an estimate of about 145,000 cubic feet per second (Q_{100}^{Ω}). Using the Q_{100} equation without the Ω parameter in table 2, the Q_{100} is about 73,400 cubic feet per second (Q_{100}^{AR}). For reference, the station-specific, trimmed-mean, Q_{100} streamflow for this station is about 302,000 cubic feet per second (Q_{100}^{LM}).

Using the Asquith and Thompson (2005, table 5; 2008, table 7) equation for Q_{100}, which is

$$Q_{100} = P^{0.9883}\, S^{0.6295} \times 10^{7.307 - 6.714A^{-0.0601}},$$

and solved for the watershed characteristics

$$Q_{100} = 24.5^{0.9883} \times 17.21^{0.6295} \times 10^{7.307 - 6.714 \times 737^{-0.0601}}, \quad (5)$$

or about 139,000 cubic feet per second (Q_{100}^{AT}).

Asquith and Slade (1997, table 1) report for station 08190000 that the Q_{100} streamflow was then (circa 1997) estimated as 336,000 cubic feet per second (Q_{100}^{17B}). This station resides in region 5 of Asquith and Slade (1997) and the regional equation provided by those authors (Asquith and Slade, 1997, table 2) is $Q_{100} = 9180A^{0.594}H^{-0.420}$, which when solved for the watershed characteristics[5] is $Q_{100} = 9180 \times 737^{0.594} \times 5.8^{-0.420}$ or about 223,000 cubic feet per second (Q_{100}^{AS}).

A discussion of the example computations for the arbitrarily selected watershed follows; unfortunately, the scope of the investigation is far too large for an effective discussion to be provided for other watersheds. However, the discussion is intended to help guide analysts in making comparisons between various methods of peak-streamflow frequency estimation.

[5]The basin-shape factor is $H = 5.80$ as reported by Asquith and Slade (1997, table 1).

The example computations can be summarized according to two different analytical tracks. Each track is summarized in separate itemized lists:

ANALYTICAL TRACK 1—For the results using station-specific, peak-streamflow frequency based on broad interpretation of Interagency Advisory Committee on Water Data (1982) made by Asquith and Slade (1997), the three Q_{100} estimates are:

- The station-specific Q_{100}^{17B} estimate from Asquith and Slade (1997, table 1) using data through about 1994 for the station is about 336,000 cubic feet per second;
- The regional-equation Q_{100}^{AS} estimate from Asquith and Slade (1997, table 2) is about 223,000 cubic feet per second; and
- The regional-equation Q_{100}^{AT} estimate from alternative interpretation of the regional regression method based on PRESS minimization, drainage area, dimensionless main-channel slope, and mean annual precipitation by Asquith and Thompson (2005, 2008) is about 139,000 cubic feet per second.

ANALYTICAL TRACK 2—For the results using station-specific, peak-streamflow frequency based on the L-moment approach described in this report, the Q_{100} estimates are:

- The station-specific Q_{100}^{LM} estimate from the current investigation by the trimmed mean of the seven distributions using data through about 2006 for the station is about 302,000 cubic feet per second;
- The regional-equation Q_{100}^{AR} estimate (table 2) from the current investigation using the equation lacking the $\mho$ parameter is about 73,400 cubic feet per second; and
- The regional-equation $Q_{100}^{\mho}$ estimate (table 3) from the current investigation using the equation with the $\mho$ parameter is about 145,000 cubic feet per second.

The computations of the first analytical track can be compared or interpreted in several ways. First, Q_{100}^{AS} and Q_{100}^{AT} are each smaller than the station-specific estimate of Q_{100}^{17B}. Hence, the watershed represented by station 08190000 has larger peak-streamflow frequency relative to many other stations in either region 5 (Q_{100}^{AS}) or statewide (Q_{100}^{AT}). The Q_{100}^{AS} estimate is larger than Q_{100}^{AT}. The Q_{100}^{AS} is closer to the Q_{100}^{17B} than Q_{100}^{AT} is. These differences can be attributed in part to the many fewer degrees of freedom for the regressions used in region 5 of Asquith and Slade (1997) than in Asquith and Thompson (2008).

The computations of the second analytical track also can be interpreted in several ways. First, Q_{100}^{AR} and $Q_{100}^{\mho}$ are each less than Q_{100}^{LM}. Hence, again, it can be concluded that the watershed represented by station 08190000 has greater peak-streamflow frequency relative to many other stations. The differences between the estimates appear substantial but can be generally said to be within about $0.40 \log_{10}$.

The conclusion that the station has greater peak-streamflow frequency than estimated by the regional models considered can be stated in another way—the peak-streamflow data for this station appear to represent comparatively greater peak values for the watershed characteristics relative to similarly characterized watersheds in the study area. In fact, many of the watersheds in the general region of Texas containing the station are well known to produce some of the largest peak-streamflow values for their respective drainage areas in Texas as well as the nation (Burnett, 2008). The $\mho$ maps in figures 2–4 clearly support this conclusion by the positive values for the $\mho$ parameter.

Continuing, the values for Q_{100}^{17B} and Q_{100}^{LM} are within $0.05 \log_{10}$ of each other—some validation of the basic computational steps used in this report thus is provided. This is an important observation, which is generalized in the next section.

Comparison of Regional Equations to Those of Previous Studies

A comparison of the regional equations reported in tables 2 and 3 to those of previous studies is made in this section. Two substantially important questions must be asked to further frame the comparisons and are answered in succession in the following two sections:

1. How are the station-specific, peak-streamflow estimates based on the trimmed mean of the seven distributions fit to the sample L-moments comparable and applicable to more established estimates based on broad interpretation of Interagency Advisory Committee on Water Data (1982) and documented by Asquith and Slade (1997)?
2. Should the $\mho$ equations in table 3 (authors prefer those in table 3 over those lacking $\mho$ in table 2) be preferred over equations in Asquith and Slade (1997)? Asquith and Thompson (2005, 2008) did

not directly ask this question of their equations in comparison to those of Asquith and Slade (1997), but only identify the Asquith and Thompson equations as "alternative."

Comparability and Applicability of Peak-Streamflow Estimates from OmegaEM and Previous Equations

To facilitate the comparisons, the watershed characteristics of the 638 stations will be used to make computations. To explore the answer to the first question, the relations between estimates of peak streamflow for the 2-, 5-, 10-, 25-, 50-, and 100-year recurrence intervals using the equations in table 2 and the equations in Asquith and Thompson (2005, table 5) are shown in figure 5. The vertical axis in the plots represents the regional model-derived features of the Approach including L-moments, weights based on sampling and modeling errors, and PRESS minimization. The horizontal axis in the plots represents a similarly parameterized regional model by Asquith and Thompson (2005) using equivalent years of record as weights (Asquith and Slade, 1997) based on broad interpretation of the Interagency Advisory Committee on Water Data (1982) guidelines. An equal value line is superimposed on the plots.

By inspection of figure 5, the estimates are close to parallel—that is, show remarkable similarity—to the equal value line and generally show increasing divergence (offset to the right) as recurrence interval increases.

An encompassing conclusion can be drawn, but unfortunately it is dependent on technical familiarity with peak-streamflow frequency analysis. The conclusion is that the use of L-moments, multiple distributions, and trimmed mean produces a regional model (table 2) of peak-streamflow frequency that is congruent[6] with the regional model (Asquith and Thompson, 2005, 2008) having a foundation on broad interpretation of the Interagency Advisory Committee on Water Data (1982) guidelines. Several natural extensions to this conclusion can be drawn and address long-held research questions of the senior author (Asquith) that date back to the senior author's lead in application of the Interagency Advisory Committee on Water Data (1982) guidelines for the Asquith and Slade (1997) report. These extensions are:

1. LOW-OUTLIER THRESHOLDS—The authors of Asquith and Slade (1997) had a first priority to use a custom equation[7] to estimate low-outlier thresholds, which have the effect of adjusting the fit of the log-Pearson Type III distribution. For many fits of the distribution, the authors then abandoned their own equation and selected low-outlier thresholds in an ad hoc fashion to improve the visual fit of the distribution. The use of low-outlier thresholds to make the procedures of Interagency Advisory Committee on Water Data (1982) work for Texas data was judged critical for Asquith and Slade (1997) to produce reliable estimates.

   The effect of the low-outlier threshold often is to twist or rotate the fitted distribution to acquire generally reduced values of peak streamflow in the right tail of the distribution. The senior author (Asquith) and associates have long held concerns about the arbitrary and capricious application of a low-outlier threshold. Yet, Texas peak-streamflow analysis generally requires low-outlier treatment when computations occur within the Interagency Advisory Committee on Water Data (1982) framework. The use of the L-moments, which avoid $\log_{10}$ transformation of the data and the resultant magnification of low outliers on peak-streamflow frequency relations can be thought of as a technique to avoid low-outlier treatment altogether;

2. HIGH OUTLIERS AND HISTORICAL RECORD—The Interagency Advisory Committee on Water Data (1982) guidelines provide considerable emphasis on refinement of statistical computations using historical information. For example, a historical peak streamflow is a peak that has a historically documented nonexceedance interval greater than the period of record represented by the station. Such historical information can be difficult to interpret but can also enhance peak-streamflow estimation at particular stations. For the use of the L-moments as implemented in the Approach, such adjustment was not performed.

   The effect of historical adjustment, when available, is to dilate or stretch the far right-hand side (large recurrence interval) of the peak-streamflow relation and thus generally reduce values of peak streamflow. A conclusion, based on the fact that the two regional models in figure 5 are so similar, is that historical peak-streamflow information as represented for the study area through Asquith and Slade (1997) and

[6]Congruent—in agreement or harmony.

[7]The equation is $\log_{10}(L) = 1.09\mu - 0.584\sigma + 0.14\gamma - 0.799$, where L is the low-outlier threshold in cubic feet per second; and μ, σ, and γ are the respective product moment mean, standard deviation, and skew of the annual peak streamflow for a given station in $\log_{10}$(cubic feet per second) with exception of γ (dimensionless).

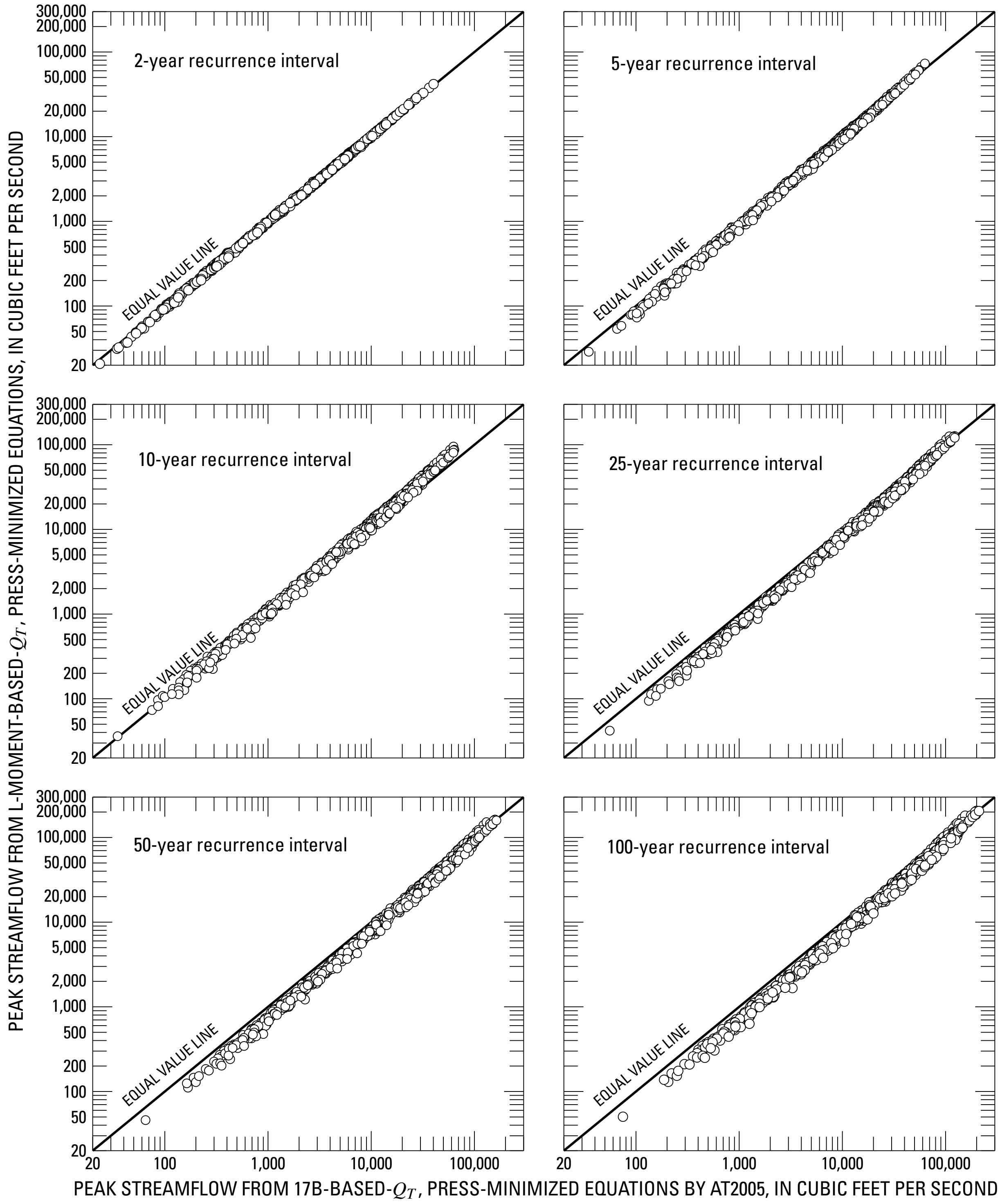


Figure 5. Comparison by recurrence interval of peak-streamflow frequency estimates for the 638 stations using equations listed in table 2 (vertical axis) and using equations originally shown in Asquith and Thompson (2005, table 5) (horizontal axis, AT2005) that are based on station-specific, peak-streamflow frequency based from broad interpretation of Interagency Advisory Committee on Water Data (1982) (17B).

references therein provides little overall information to the regional model of peak-streamflow frequency for the study area. The authors consider this conclusion to be consistent with the opinion of Hosking and Wallis (1997, p. 161), who, although in the context of a structurally different regional model (index-flood procedure), are "skeptical about the practical utility of historical information;" and

3. GENERALIZED SKEW—For the use of the L-moments, the skewness (and kurtosis) of the distribution were estimated solely on the data on a per station basis. There was no treatment for a spatially varying parameter (generalized skew) supposedly representing distribution symmetry (Judd and others, 1996). The authors interpret, should generalized skew exist, that the influence of this skew as well as other characteristics influencing peak-streamflow potential of the watershed have been lumped into the Ω parameter.

On further interpretation of figure 5, it is seen that the divergence from the equal value line is offset to the right as opposed to the left. This suggests that the Asquith and Thompson (2005) equations (those based on Asquith and Slade [1997] results) produce larger estimates of peak streamflow, up to about $0.10\log_{10}$, more than those derived from table 2. Thus, the equations in table 2 produce slightly smaller estimates. The authors ask further questions: Are the Asquith and Slade (1997) estimates of station-specific, peak-streamflow frequency then too large? Are the L-moment estimates too low? Are the differences related to sample size? The apparently larger values occur even with the considerable treatment of low outliers and high outliers in the Asquith and Slade (1997) analysis. However, the differences could be thought of as slight. Specifically, what differences in station-specific, peak-streamflow frequency can be attributed to the addition of more record between about 1994 and 2006 for some stations? The two suites of regional equations represented in figure 5 differ slightly by degrees of freedom—that is, stations included in the regression analysis—does this contribute to the divergence (or accidental convergence) in the figure? The answers to these questions are not known and exploration is beyond the scope of this report.

Preference for OmegaEM Equations over Previous Equations

The discussion of the first question provides an authoritative account that the trimmed mean L-moment-based estimates of station-specific, peak-streamflow frequency are reliable.

The regional regression equations listed in this report that are preferred by the authors are the Ω equations listed in table 3 and thus use of the Ω parameter is suggested. These equations should be preferred over the equations listed in table 2 as justified in part by the large percentage reduction in the PRESS statistics.

To have some parallelism with the discussion in the previous section, a comparison of the 638 estimates from the Ω equations to those in Asquith and Thompson (2005, table 5) is made in figure 6. The data points scatter around the equal value line. In part, this shows that the inclusion of the Ω parameter does not have a deleterious[8] effect on the regression analysis. The scatter does not represent and should not be interpreted as a lack-of-fit, but rather demonstrates that the Ω parameter has the effect of pulling the data points in the vertical direction according to the sign of the parameter.

Perhaps the most important question, which will undoubtedly be asked of the authors and many of the acknowledged colleagues, deserves repeating: Should the Ω equations in table 3 be preferred over equations in Asquith and Slade (1997)?

The authors suggest that the "OmegaEM equations" listed in table 3 should be preferred over the equations and procedures of Asquith and Slade (1997) for the following reasons and supporting ancillary discussion:

1. The Ω equations have solutions for 200-, 250-, and 500-year recurrence intervals, which are not available in Asquith and Slade (1997);

2. The Ω equations have nearly 18 times more degrees of freedom than those in Asquith and Slade (1997). The equations would be expected to have broader applicability across a wider range of watersheds;

3. There are six Ω equations compared to 96 in Asquith and Slade (1997)—the three greater than 100-year recurrence intervals in table 3 cannot be included in this particular comparision;

4. The use of L-moment statistics for the Ω equations has removed:

   (a) the largely ad hoc process of low-outlier adjustment to station-specific, peak-streamflow frequency estimates;

[8]Deleterious—causing harm or damage.

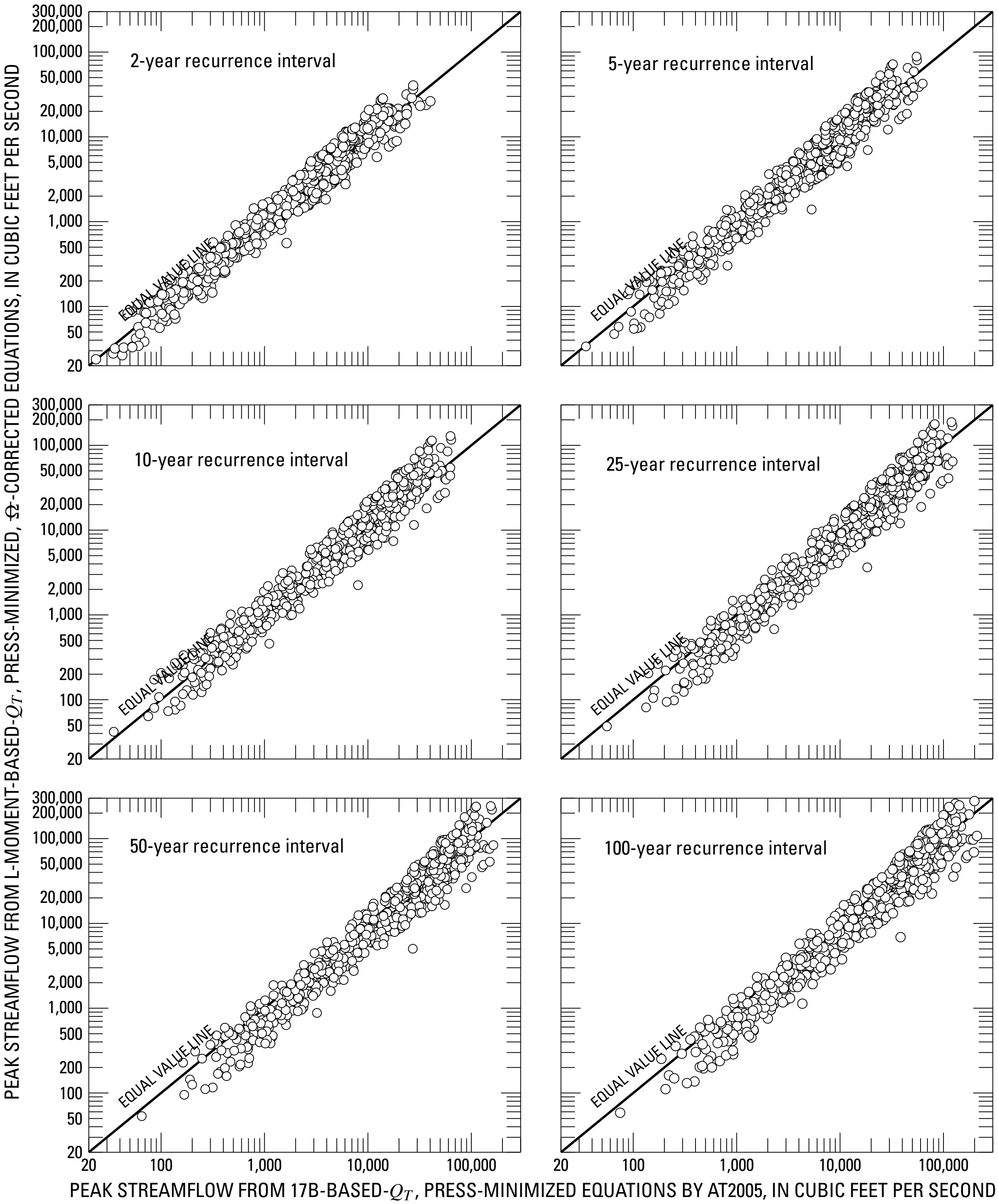


Figure 6. Comparison by recurrence interval of peak-streamflow frequency estimates for the 638 stations using equations listed in table 3 (vertical axis) and using equations originally shown in Asquith and Thompson (2005, table 5) (horizontal axis, AT2005) that are based on station-specific, peak-streamflow frequency based from broad interpretation of Interagency Advisory Committee on Water Data (1982) (17B).

(b) the need to use the sometimes difficult-to-interpret or inconsistent station-to-station historical record; and

(c) the concept of generalized skew, at least as implemented in the Approach.

5. The Ω equations generally have similar residual standard errors and adjusted R-squared values as those in Asquith and Slade (1997), but more importantly, the Ω equations have minimal PRESS statistics;

6. The minimization of PRESS for the Ω equations should be expected to have enhanced extrapolation characteristics to small watersheds (Asquith and Thompson, 2005, 2008);

7. The maps in figures 2–4 of the Ω parameter supplant the need for separate and sometimes arbitrary regions of Asquith and Slade (1997). Although it certainly can be argued that the 1-degree quadrangles of the Ω parameter have themselves become "microregions of arbitrary size and extent," the authors fully acknowledge that complications and ambiguities to the selection of the Ω will arise. The authors conclude that the smoother regional structure of the map relative to the regions of Asquith and Slade (1997) elucidates[9] previously suspected, but inadequately documented, relative peak-streamflow potential in Texas. All other variables being equal (drainage area, dimensionless main-channel slope, mean annual precipitation), peak-streamflow potential is relatively less in far east Texas and far west Texas, greater in a swath through much of the central part of the state (the region demarked by positive Ω values), and highest along the Balcones escarpment in south-central Texas. O'Connor and Costa (2003, p. 9) identify this region (Balcones escarpment) of the nation as having "concentrations of large floods."

In conclusion, the Ω equations are expected to be reliable estimators of peak-streamflow frequency for undeveloped and ungaged stream locations in Texas with watershed characteristics within the distribution of those characteristics used to develop the equations. Although the Ω parameter requires interpretation on the part of analysts and the potential exists that different analysts could estimate different values for a given watershed, the authors suggest that typical uncertainty in the Ω estimate might be about $\pm 0.10 \log_{10}$, which is small relative to other uncertainties, such as those measured by residual standard error, in hydrostatological models. Finally, given the two ensembles of equations reported herein and those in previous reports, hydrologic design engineers and other analysts in Texas now have several different methods, which were derived from differing analytical tracks, to make comparisons of peak-streamflow frequency estimates for ungaged stream locations in Texas.

[9] Elucidates—makes (something) clear; explains.

Considerations for Application of Regression Equations to Gaged and Ungaged Watersheds

There are three remaining aspects of regional equation use in practice that require discussion: (1) application to "gaged" watersheds, (2) practice of watershed subdivision, and (3) estimation of prediction limits.

The first aspect is that the authors observe that the equations reported herein will occasionally be applied at stream locations coincident with an operational streamflow-gaging station or locations with historical annual peak-streamflow data. Further these locations could also have peak-streamflow frequency estimates from combinations of synthetic or calibrated rainfall and runoff models or other hydrologic methods. Procedures or suggestions for combining the various estimates of peak-streamflow frequency from two or more methods explicitly are not provided in this report.

The second aspect has not been previously discussed in either Asquith and Slade (1997), Asquith (2001), Asquith and Thompson (2005), or Asquith and Thompson (2008); although the authors have had considerable communication with practitioners on the following topic. Occasionally, some users, who often are concerned with equation applicability for "unusual" watersheds, (1) subdivide watersheds into one or more subbasins of more "typical" characteristics relative to those used to develop the equations, (2) compute peak-streamflow frequency from regional regression equations for each ungaged subbasin, and (3) then combine the estimated values by summation. The authors suggest that users are seeking to enhance prediction accuracy by reducing extrapolation to watershed characteristics not well represented in the equations. (This is a particular issue with the limited degree-of-freedom equations in Asquith and Slade[1997].) Such a practice is not optimal.

The authors observe that annual peak-streamflow data from stations represent aggregate or integrated peak-streamflow potential (information) of *whole* watersheds. The regression process seeks to explain the variability of peak-streamflow frequency for whole watersheds and not the distributed processes within subbasins. The authors suggest that the summation explicitly assumes that the joint probability of contemporaneous peak-streamflow occurrence within each subbasin is 100 percent. Logical extension suggests that the practice of subdivision could yield peak-streamflow frequency curves that are generally too large and hence poorly reflect the risk level as specified by the T-year recurrence interval.

The third aspect is that the prediction limits can be useful for expressing uncertainty when a equation is used. A $100 \times (1-\alpha)$ prediction interval for Q_T is given by the probability

$$\Pr\{{}^{\downarrow}Q_T \leq Q_T \leq {}^{\uparrow}Q_T\} \geq 1-\alpha, \tag{6}$$

where the down (up) arrow signifies the lower (upper) prediction limit for Q_T at the α significance level. The lower and upper limits are computed by

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

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

where Q_T is the prediction from a T-year equation, $\mathrm{t}_{[\alpha/2,\mathrm{df}]}$ is the upper tail of the t-distribution for $\mathrm{df}^{[Q_T]}$ degrees of freedom at the α significance level, $\sigma^{[Q_T]}$ is the residual standard error of the T-year equation, and $h_o^{[Q_T]}$ is the leverage of the prediction for the watershed. The $h_o^{[Q_T]}$ requires matrix multiplication—the requisite matrixes are listed in appendixes 2 and 3 under the `INVERTED-COVARIANCE MATRIX` headings.

The authors observe that a rapid method for estimation of prediction limits can be conceived because of the large degrees of freedom for the regressions. The rapid method is more straightforward than the method involving matrix computation described in Asquith and Slade (1997) and mathematically outlined in Asquith and Roussel (2007). Other observations follow. First, the number of stations in the equations is $n = 638$ and the number of parameters (variables plus 1 for the regression intercept) are either $p = 4$ (table 2) or $p = 5$ (table 3). Second, the average leverage for a regression is about p/n so for this investigation, this ratio is less than 0.008. The authors suggest that $h_o^{[Q_T]} \approx 0.008$. Thus, the term $\sqrt{1+0.008} \approx 1$. Third, again because the degrees of freedom ($\mathrm{df} = n - p$) is large, $\mathrm{t}_{[\alpha/2,\mathrm{df}]}$ can be well approximated by the quantile function[10] of the standard normal distribution $\mathrm{N}(F)$ for nonexceedance probability F.

$${}^{\downarrow}Q_T = 10^{\log_{10}(Q_T) - \mathrm{N}(1-\alpha/2)\,\sigma^{[Q_T]}}, \text{ and} \tag{9}$$

$${}^{\uparrow}Q_T = 10^{\log_{10}(Q_T) + \mathrm{N}(1-\alpha/2)\,\sigma^{[Q_T]}}. \tag{10}$$

The rapid method is demonstrated in the following example. Suppose that the 90th-percentile prediction limits need rapid estimation for the 100-year peak streamflow estimate by the respective equation in table 3. The estimate is $Q_T = 9{,}000$ cubic feet per second and $\sigma^{[Q_{100}]} = 0.30$. The $\mathrm{N}(0.95) = 1.6$ where $0.95 = (1-\alpha/2) = (1-(1-0.90)/2)$ for the 90th-percentile (two-tailed) limits. (In R code, the quantity 1.6 is `qnorm(0.95)`.) As a result, the 90th-percentile prediction limits are estimated as

$${}^{\downarrow}Q_T = 10^{\log_{10}(9{,}000) - 1.6\times 0.30} = 2{,}980, \text{ and} \tag{11}$$

$${}^{\uparrow}Q_T = 10^{\log_{10}(9{,}000) + 1.6\times 0.30} = 27{,}200. \tag{12}$$

Thus, the prediction limits for the 100-year peak streamflow estimate of $Q_{100} = 9{,}000$ cubic feet per second can be written as $2{,}980 \leq Q_{100}^{(90\mathrm{th})} \leq 27{,}200$ cubic feet per second. The authors, however, suggest the more compact notation of $Q_{100}^{(90\mathrm{th})} = 9{,}000 \pm 0.48 \log_{10}$ cubic feet per second be used instead where the quantity 0.48 is 1.6×0.30. Finally, the authors observe that prediction limits constructed as shown will be generally too small (narrow) because of inherent uncertainties of the Ω parameter, which are difficult to propagate through the preceding computations.

Summary

Annual peak-streamflow frequency estimates are needed for flood-plain management; for objective assessment of flood risk; for cost-effective design of dams, levees, and other flood-control structures; and for design of roads, bridges, and culverts. Annual peak-streamflow frequency represents the peak streamflow for nine recurrence intervals of 2, 5, 10, 25, 50, 100, 200, 250, and 500 years. Common methods for estimation of peak-streamflow frequency for ungaged or unmonitored stream watersheds are regression equations for each recurrence interval developed for one or more regions (not strictly geographic); such regional equations are the subject of this report. The method is based on statistical analysis of annual peak-streamflow data from U.S. Geological Survey (USGS) streamflow-gaging stations (stations).

Beginning in 2007, the USGS, in cooperation with the Texas Department of Transportation and in partnership with Texas Tech University, began a 3-year investigation

[10]The quantile function of the standard normal distribution for nonexceedance probability F could be approximated for the purposes of this report by $Q(F) = 5.063[F^{0.135} - (1-F)^{0.135}]$.

concerning the development of regional equations to estimate annual peak-streamflow frequency for undeveloped watersheds in Texas. The general approach was explicitly limited to the use of L-moment statistics, which are used in conjunction with a technique of multi-linear regression referred to as PRESS minimization. The approach used to develop the regional equations, which was refined during the investigation, is referred to as the "L-moment-based, PRESS-minimized, residual-adjusted approach." The study area for this investigation includes Texas and selected parts of neighboring states and essentially is the same as that considered in previous studies.

The primary purpose of this report is to present regression equations from the L-moment-based, PRESS-minimized, residual-adjusted approach for estimation of annual peak-streamflow frequency for undeveloped watersheds in Texas and primarily restricted to ungaged watersheds.

The primary scope of the report is limited to the annual peak-streamflow data for 677 and 638 selected stations. The scope of the report also is limited to three selected watershed characteristics, which have previously been shown as important predictors of peak-streamflow frequency in Texas. The watershed characteristics are drainage area, dimensionless main-channel slope, and mean annual precipitation.

From an initial candidate station count of 1,030 stations that had 1 or more years of annual peak-streamflow data in the study area, the number of stations used for this report were reduced by review of the data, watershed conditions, reliability of watershed characteristics, and a minimum record-length criteria (8 years), leaving 677 stations. The data for these 677 stations are used for a specific component of the analysis. Another assessment or restriction for the purpose of the regression analysis is that stations with drainage areas less than 10,000 square miles are used. After further assessment of watershed characteristics and the less-than-10,000-square-mile criterion, 638 stations remained.

The L-moment-based, PRESS-minimized, residual-adjusted approach (Approach) used for this investigation is complex and highly technical. Furthermore, the Approach relied on numerous single-purpose, highly specialized scripts, computer programs, and integration with features of the host-operating system as well as considerable judgement and iterative refinement on the part of the authors.

L-moments are used for station-specific computations of peak-streamflow frequency with no consideration for concepts such as low-outliers, high-outliers, historical information, and generalized skew as described in other methods for peak-streamflow frequency computation. For the Approach, the first five sample L-moments were computed for each of the 677 stations. These L-moments, respectively, are the mean, L-scale, L-skew, L-kurtosis, and Tau5.

For the weighted-least-squares regression analysis, weight factors representing sampling error and modeling errors as distinct components of uncertainty are used. It is favorable to have stations with more record (smaller sampling error, more certainty) to have more weight in regression analysis.

Sampling error estimates were based on dimensionless regional L-moments, which are weighted-mean values for the study area, and these weighted-mean values provide an approximation of the geometry of the "parent" peak-streamflow frequency distribution for the study area. The dimensionless regional L-moments were used to estimate a dimensionless regional Wakeby distribution. This fitted Wakeby distribution is assumed to provide a reasonable first-order approximation to the underlying (dimensionless) structure of peak-streamflow frequency for the study area.

For this report, as many as seven probability distributions are fit to the sample L-moments for each station. The authors explicitly chose to avoid selection of a single form of a probability distribution to model the station-specific, peak-streamflow frequency through use of a substantial number of three-parameter and more distributions. For almost all stations, seven distributions were fit and seven estimates of peak streamflow for each of the nine recurrence intervals were computed.

Because seven distributions were fit for each station, seven unique values of peak streamflow were obtained for each recurrence interval. At a given recurrence interval, these values have a distribution; the relative variability of this distribution is treated as a measure of modeling error. It is favorable for stations with more variability in peak-streamflow frequency estimates by distribution to have less weight in regression analysis because more uncertainty exists at such stations.

For each of the nine recurrence intervals, a symmetrically trimmed L-moment (trimmed mean) was computed from each of seven unique values of peak streamflow for a given recurrence interval. The use of trimming, which computationally is more complex than simply dropping the largest and smallest values and computing an arithmetic mean, is anticipated to provide a more robust measure of central location for the formal estimation of station-specific, peak-streamflow frequency. The trimmed mean was used to estimate the station-specific, peak streamflow for each of the nine recurrence intervals.

For this report, weighted-least-squares, multi-linear regression analysis is used to develop the statistical relation between the station-specific, trimmed-mean, T-year peak-streamflow values and the three watershed characteristics (explanatory variables). The sampling and modeling errors are combined and are used to derive weight factors for the weighted-least-squares regression between peak-streamflow frequency and the watershed characteristics of drainage area, dimensionless main-channel slope, mean annual precipitation, and, when used, the Ω parameter.

Nine weighted-least-squares, PRESS-minimized regression equations are computed using the watershed characteristics of drainage area, dimensionless main-channel slope, and mean annual precipitation and each of the 638 trimmed-mean, peak-streamflow estimates for each of the nine recurrence intervals. The PRESS-minimization was made using a user-guided search algorithm. Separate PRESS-minimizations were made and a unique exponent on drainage area is determined for each recurrence interval.

The residuals for each of the nine regression equations were computed. Maps depicting the spatial distribution of the residuals with symbols determined by the magnitude and direction of the residual are shown. Spatial dependency is evident for all recurrence intervals. For each recurrence interval, the median residual for each square of 1-degree quadrangle of latitude and longitude in the study area was computed.

The median values of the residuals were computed for 1-degree quadrangles for the 10-year recurrence interval for generalization. The authors, through manual smoothing, consultation of various geologic and ecological region maps, interpretation of regional topographic maps, and resident familiarity with the study area, estimated the magnitude and sign of the 10-year recurrence interval residuals for each 1-degree quadrangle in the study area. These estimates, referred to as the Ω parameter, are in units of $\log_{10}$(streamflow); maps of the study area depicting, by 1-degree quadrangle, the Ω parameter superimposed on hill-shade relief, Texas rivers, and ecoregions are provided.

For each station, the Ω parameter was assigned based on the 1-degree quadrangles containing the station. The Ω parameter represents a generalized terrain and climate index that expresses relative differences in peak-streamflow potential across the study area. The Ω parameter is interpreted as an expression of peak-streamflow potential not represented in the watershed characteristics of drainage area, main-channel slope, and mean annual precipitation.

Nine weighted-least-squares, PRESS-minimized regression equations were computed using the watershed characteristics of drainage area, dimensionless main-channel slope, mean annual precipitation, and Ω as well as each of the 638 trimmed-mean, peak-streamflow estimates for each of the nine recurrence intervals. The weights for the regression equations were the same as those used for the nine regression equations not including Ω. Separate PRESS-minimizations were made for each recurrence interval and a unique exponent on drainage area was determined. The mean residual, standard error, adjusted R-squared, and mean percentage reduction in PRESS by use of Ω are $0.30\log_{10}$, 0.86, and about -21 percent, respectively.

Example computations are provided to demonstrate how the regional regression equations could be used in practice. A comparison of the regression equations to results from a previous study is made, and the results are discussed. In brief, a similar regional model or structure of the relation between peak-streamflow frequency and the three watershed characteristics in Texas results from the Approach.

The Ω equations are expected to be reliable estimators of peak-streamflow frequency for undeveloped and ungaged stream locations in Texas. Although the Ω parameter requires interpretation on the part of analysts and the potential exists that different analysts could estimate different values for a given watershed, the authors suggest that typical uncertainty in the Ω estimate might be about $\pm 0.10\log_{10}$, which is small relative to other uncertainties, such as those measured by residual standard error, in hydrostatological models. Finally, given the two ensembles of equations reported herein and those in previous reports, hydrologic design engineers and other analysts in Texas now have several different methods, which represent different analytical tracks, to make comparisons of peak-streamflow frequency estimates for ungaged stream locations in the study area.

References

Asquith, W.H., 2001, Effects of regulation on L-moments of annual peak streamflow in Texas: U.S. Geological Survey Water-Resources Investigations Report 01–4243, 66 p. [http://pubs.usgs.gov/wri/wri014243/]

Asquith, W.H., 2006, L-moments and TL-moments of the generalized lambda distribution: Computational Statistics and Data Analysis, v. 51, p. 4,484–4,496.

Asquith, W.H., 2008, lmomco—L-moments, Trimmed L-moments, L-comoments, and many distributions: R package version 0.93.3, February 2, 2008. [http://www.cran.r-project.org/package=lmomco]

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]

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, ISBN 1–411–31041–1, 299 p. [http://pubs.usgs.gov/pp/pp1725/]

Asquith, W.H., and Slade, R.M., 1997, Regional equations for estimation of peak-streamflow frequency for natural basins in Texas: U.S. Geological Survey Water-Resources Investigations Report 96–4307, 68 p. [http://pubs.usgs.gov/wri/wri964307/]

Asquith, W.H., and Slade, R.M., 1999, Site-specific estimation of peak-streamflow frequency using generalized least-squares regression for natural basins in Texas: U.S. Geological Survey Water-Resources Investigations Report 99–4172, 19 p. [http://pubs.usgs.gov/wri/wri994172/]

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. [http://library.ctr.utexas.edu/pdf/4405-2.pdf]

Asquith, W.H., and Thompson, D.B., 2008, Alternative regression equations for estimation of annual peak-streamflow frequency for undeveloped watersheds in Texas using PRESS minimization: U.S. Geological Survey Scientific Investigations Report 2008–5084, 40 p. [http://pubs.usgs.gov/sir/2008/5084]

Burnett, Jonathan, 2008, Flash floods in Texas: College Station, Texas A&M Univesity Press, 330 p.

Commission for Environmental Cooperation, 1997, Ecological regions of North America—Toward a common perspective: Montreal, Quebec, Canada, 71 p.

Elamir, E.A., and Seheult, A.H., 2003, Trimmed L-moments: Computational Statistics and Data Analysis, v. 43, p. 299–314.

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.

Hosking, J.R.M., 1996, FORTRAN routines for use with the method of L-moments, version 3: Yorktown Heights, N.Y., IBM Research Division, T.J. Watson Research Center, Research Report RC20525.

Hosking, J.R.M., 2006, On the characterization of distributions by their L-moments: Journal of Statistical Planning and Inference, v. 136, p. 193–198.

Hosking, J.R.M., 2007, Some theory and practical uses of trimmed L-moments: Journal of Statistical Planning and Inference, v. 137, p. 3,024–3,039.

Hosking, J.R.M., and Wallis, J.R., 1997, Regional frequency analysis—An approach based on L-moments: Cambridge, Cambridge University Press, ISBN 0–521–43045–3, 224 p.

Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood flow frequency: Reston, Va., U.S. Geological Survey, Office of Water Data Coordination, Hydrology Subcommittee Bulletin 17B [variously paged].

Judd, L.J., Asquith, W.H., and Slade, R.M., 1996, Techniques to estimate generalized skew coefficients of annual peak streamflow for natural basins in Texas: U.S. Geological Survey Water-Resources Investigations Report 96–4117, 28 p.

Landwehr, J.M., Matalas, N.C., and Wallis, J.R., 1979, Estimation of parameters and quantiles of Wakeby distributions: Water Resources Research, v. 15, no. 5, p. 1,362–1,379.

Montgomery, D.C., Peck, E.A., and Vining, G.G., 2001, Introduction to linear regression analysis (3rd ed.): New York, Wiley, 641 p.

O'Connor, J.E., and Costa, J.E., 2003, Large floods in the United States—Where they happen and why: U.S. Geological Survey Circular 1245, ISBN 0–607–89380–X, 13 p.

R Development Core Team, 2008, R—A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing, ISBN 3–900051–07–0, version 2.7.0, http://www.R-project.org

Stedinger, J.R., Vogel, R.M., and Foufoula-Georgiou, E., 1993, Frequency analysis of extreme events, chapter 18, *in* Maidment, D.A., ed., Handbook of hydrology: New York, McGraw-Hill, p. 18.1–18.66.

U.S. Geological Survey, 2003, The national atlas of the United States of America—Streams and waterbodies of the United States: Available for download at http://nationalatlas.gov/atlasftp.html?openChapters=chpwater#chpwater

U.S. Geological Survey, 2008, National Water Information System: accessed on various dates in 2007 at http://nwis.waterdata.usgs.gov/usa/nwis/peak

Appendix 1—Supplemental Information and External Data Files

Blank Page

The purpose of this appendix is to provide a central reference to several text or portable document format (PDF) files that accompany this report and are referenced. Each of these are listed below and available from the report Web site:

1. `Appendix1_638annpks.txt`—This text file is a comma-delimited file, which lists by station, the annual peak-streamflow values used for statistical computation as described in the text. The number of stations is 638 and the criteria leading to this count are described in the report in the "Identification of Annual Peak Streamflow Data" section of this report;
2. File `Appendix1_638wtrshdchr.txt`—This text file is a comma-delimited file, which lists by station, the watershed characteristics of drainage area, dimensionless main-channel slope, and mean annual precipitation that are used for regression analyses as described in the text. The number of stations is 638 and the criteria leading to this count are described in the report in the "Identification of Annual Peak-Streamflow Data" section of this report. A statistical summary of the selected watershed characteristics in this file is provided in table 1 on page 4;
3. File `Appendix1_677annpks.txt`—This text file is a comma-delimited file, which lists by station, the annual peak-streamflow values used for statistical computation as described in the text. The number of stations is 677 and the criteria leading to this count are described in the report in the "Identification of Annual Peak-Streamflow Data" section of this report;
4. File `Appendix1_677trimmedQTs.txt`—This text file is a comma-delimited file, which lists by station, the annual peak-streamflow values used for statistical computation as described in the text. The number of stations is 677 and the criteria leading to this count are described in the report in the "Trimmed Mean Estimates of Peak-Streamflow Frequency" section of this report;
5. File `Appendix1_677freqcurves.pdf`—This PDF file provides a graphical archive of the L-moments fits of as many as seven probability distributions to the annual peak-streamflow data listed in file `Appendix1_677annpks.txt`. The results reported in this file were used to compute the trimmed mean estimates of station-specific, peak-streamflow frequency reported in file `Appendix1_677trimmedQTs.txt`; and
6. File `Appendix1_residualmaps.pdf`—This PDF file provides a graphical archive of the spatial distribution of the residuals from the regression equations reported in tables 2 and 3.

Each file also contains either a heading (text files) or, for the PDF files, a file-specific "Introduction" and "References" section. These headings or sections provide further documentation of the respective contents of each file. The file `Appendix1_677freqcurves.pdf` is listed before the other PDF file; therefore, the figures within that file are numbered with a "1A" prepended to the sequential figure number. Similarly, figures in file `Appendix1_residualmaps.pdf` are numbered with a "1B" prepended to the sequential figure number.

Appendix 2—Results of Weighted-Least-Squares, Multi-Linear Regression Analysis Using PRESS Minimization of Peak-Streamflow Frequency to Watershed Characteristics of Drainage Area, Mean Annual Precipitation, and Dimensionless Main-Channel Slope

Blank Page

```
# START NEW EQUATION
[1] "Doing␣2␣year"
The power is -0.008

Call:
lm(formula = log10(Q) ~ CDA1 + log10(P) + log10(S), weights = mywgts)

Residuals:
      Min        1Q    Median        3Q       Max
-1.230070 -0.198423 -0.000294  0.202798  1.059189

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  38.63528    1.01911  37.911  < 2e-16 ***
CDA1        -37.94274    1.03466 -36.672  < 2e-16 ***
log10(P)      1.56157    0.10683  14.617  < 2e-16 ***
log10(S)      0.38458    0.05262   7.308 8.17e-13 ***
---
Signif. codes:  0:"***", 0.001:"**", 0.01:"*", 0.05:".", 0.1:"␣", 1

Residual standard error: 0.3191 on 634 degrees of freedom
Multiple R-Squared: 0.8099,     Adjusted R-squared: 0.809
F-statistic: 900.6 on 3 and 634 DF,  p-value: < 2.2e-16

PRESS= 77.3954345522906
[1] "LATEX:$Q_{2}␣=␣P^{1.562},S^{0.385}\times␣10^{[38.635-37.943,A^{-0.008]}},$␣&␣0.32␣&␣0.81␣&␣77.4␣&␣395.8\\"
[1] "INVERTED-COVARIANCE␣MATRIX"
           [,1]        [,2]        [,3]        [,4]
[1,]  10.2015098 -10.2857298  0.55103002  0.42527734
[2,] -10.2857298  10.5152921 -0.64343158 -0.42570485
[3,]   0.5510300  -0.6434316  0.11210399  0.03740539
[4,]   0.4252773  -0.4257049  0.03740539  0.02719925
[1] "MAXIMUM␣LEVERAGE"
[1] 0.05328113


# START NEW EQUATION
[1] "Doing␣5␣year"
The power is -0.0228

Call:
lm(formula = log10(Q) ~ CDA1 + log10(P) + log10(S), weights = mywgts)

Residuals:
     Min       1Q   Median       3Q      Max
-1.04803 -0.18715  0.02627  0.19603  0.96727

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  16.34951    0.35560   45.98   <2e-16 ***
CDA1        -15.03710    0.35712  -42.11   <2e-16 ***
log10(P)      1.49059    0.09973   14.95   <2e-16 ***
log10(S)      0.49965    0.04934   10.13   <2e-16 ***
---
Signif. codes:  0:"***", 0.001:"**", 0.01:"*", 0.05:".", 0.1:"␣", 1

Residual standard error: 0.3029 on 634 degrees of freedom
Multiple R-Squared: 0.8373,     Adjusted R-squared: 0.8365
F-statistic:  1087 on 3 and 634 DF,  p-value: < 2.2e-16

PRESS= 65.1550224898523
[1] "LATEX:$Q_{5}␣=␣P^{1.491},S^{0.5}\times␣10^{[16.35-15.037,A^{-0.0228]}},$␣&␣0.3␣&␣0.84␣&␣65.2␣&␣313.5\\"
[1] "INVERTED-COVARIANCE␣MATRIX"
          [,1]       [,2]        [,3]        [,4]
[1,]  1.3781673 -1.3125511  0.13748791  0.15309923
```

```
[2,] -1.3125511  1.3899001 -0.22737095 -0.15251351
[3,]  0.1374879 -0.2273710  0.10839049  0.03579911
[4,]  0.1530992 -0.1525135  0.03579911  0.02653628
[1] "MAXIMUM_LEVERAGE"
[1] 0.05091717


# START NEW EQUATION
[1] "Doing_10_year"
The power is -0.0299

Call:
lm(formula = log10(Q) ~ CDA1 + log10(P) + log10(S), weights = mywgts)

Residuals:
     Min       1Q   Median       3Q      Max
-1.05995 -0.17900  0.04071  0.19795  0.95176

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.67349    0.27797   49.19   <2e-16 ***
CDA1        -11.98655    0.27300  -43.91   <2e-16 ***
log10(P)      1.37713    0.09827   14.01   <2e-16 ***
log10(S)      0.52954    0.04838   10.95   <2e-16 ***
---
Signif. codes:  0:"***", 0.001:"**", 0.01:"*", 0.05:".", 0.1:"_", 1

Residual standard error: 0.2996 on 634 degrees of freedom
Multiple R-Squared: 0.8465,     Adjusted R-squared: 0.8458
F-statistic:  1165 on 3 and 634 DF,  p-value: < 2.2e-16

PRESS= 63.6909387784546
[1] "LATEX:$Q_{10}_=_P^{1.377},S^{0.53}\times_10^{[13.673-11.987,A^{-0.0299]}},$_&_0.3_&_0.85_&_63.7_&_297.4\\"
[1] "INVERTED-COVARIANCE_MATRIX"
           [,1]       [,2]        [,3]        [,4]
[1,]  0.86102385 -0.7734697  0.08289624  0.11796275
[2,] -0.77346969  0.8305043 -0.17342680 -0.11704920
[3,]  0.08289624 -0.1734268  0.10760071  0.03496131
[4,]  0.11796275 -0.1170492  0.03496131  0.02607787
[1] "MAXIMUM_LEVERAGE"
[1] 0.04964972


# START NEW EQUATION
[1] "Doing_25_year"
The power is -0.038

Call:
lm(formula = log10(Q) ~ CDA1 + log10(P) + log10(S), weights = mywgts)

Residuals:
     Min       1Q   Median       3Q      Max
-1.10198 -0.18766  0.03567  0.19749  0.92946

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  11.9959     0.2389   50.22   <2e-16 ***
CDA1         -9.9918     0.2286  -43.72   <2e-16 ***
log10(P)      1.3156     0.1021   12.89   <2e-16 ***
log10(S)      0.5740     0.0500   11.48   <2e-16 ***
---
Signif. codes:  0:"***", 0.001:"**", 0.01:"*", 0.05:".", 0.1:"_", 1

Residual standard error: 0.3107 on 634 degrees of freedom
Multiple R-Squared: 0.8415,     Adjusted R-squared: 0.8408
```

```
F-statistic:  1122 on 3 and 634 DF,  p-value: < 2.2e-16

PRESS= 67.1310969481496
[1] "LATEX:$Q_{25}␣=␣P^{1.316},S^{0.574}\times␣10^{[11.996-9.992,A^{-0.038]}},$␣&␣0.31␣&␣0.84␣&␣67.1␣&␣347.8\\"
[1] "INVERTED-COVARIANCE␣MATRIX"
            [,1]        [,2]        [,3]        [,4]
[1,]  0.59101965 -0.49267483  0.04866039  0.09480674
[2,] -0.49267483  0.54106276 -0.14043067 -0.09393455
[3,]  0.04866039 -0.14043067  0.10786137  0.03469492
[4,]  0.09480674 -0.09393455  0.03469492  0.02589693
[1] "MAXIMUM␣LEVERAGE"
[1] 0.04922661


# START NEW EQUATION
[1] "Doing␣50␣year"
The power is -0.0429

Call:
lm(formula = log10(Q) ~ CDA1 + log10(P) + log10(S), weights = mywgts)

Residuals:
     Min       1Q   Median       3Q      Max
-1.13394 -0.20740  0.02724  0.20993  0.92946

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.40360    0.23024   49.53   <2e-16 ***
CDA1        -9.19510    0.21648  -42.48   <2e-16 ***
log10(P)     1.28501    0.10730   11.98   <2e-16 ***
log10(S)     0.60823    0.05255   11.57   <2e-16 ***
---
Signif. codes:  0:"***", 0.001:"**", 0.01:"*", 0.05:".", 0.1:"␣", 1

Residual standard error: 0.3265 on 634 degrees of freedom
Multiple R-Squared: 0.8309,     Adjusted R-squared: 0.8301
F-statistic:  1039 on 3 and 634 DF,  p-value: < 2.2e-16

PRESS= 73.514894167704
[1] "LATEX:$Q_{50}␣=␣P^{1.285},S^{0.608}\times␣10^{[11.404-9.195,A^{-0.0429]}},$␣&␣0.33␣&␣0.83␣&␣73.5␣&␣414.2\\"
[1] "INVERTED-COVARIANCE␣MATRIX"
            [,1]        [,2]        [,3]        [,4]
[1,]  0.49734992 -0.39486813  0.03482157  0.08539406
[2,] -0.39486813  0.43964842 -0.12689784 -0.08455224
[3,]  0.03482157 -0.12689784  0.10801117  0.03469890
[4,]  0.08539406 -0.08455224  0.03469890  0.02590499
[1] "MAXIMUM␣LEVERAGE"
[1] 0.04914469


# START NEW EQUATION
[1] "Doing␣100␣year"
The power is -0.047

Call:
lm(formula = log10(Q) ~ CDA1 + log10(P) + log10(S), weights = mywgts)

Residuals:
     Min       1Q   Median       3Q      Max
-1.18887 -0.21895  0.02182  0.23338  0.95608

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.07336    0.23019   48.10   <2e-16 ***
CDA1        -8.67252    0.21349  -40.62   <2e-16 ***
```

```
log10(P)      1.25687    0.11400   11.03   <2e-16 ***
log10(S)      0.64039    0.05591   11.45   <2e-16 ***
---
Signif. codes:  0:"***", 0.001:"**", 0.01:"*", 0.05:".", 0.1:"␣", 1

Residual standard error: 0.3462 on 634 degrees of freedom
Multiple R-Squared: 0.8159,     Adjusted R-squared: 0.815
F-statistic: 936.7 on 3 and 634 DF,  p-value: < 2.2e-16

PRESS= 82.8434210751132
[1] "LATEX:$Q_{100}␣=␣P^{1.257},S^{0.64}\times␣10^{[11.073-8.673,A^{-0.047]}},$␣&␣0.35␣&␣0.82␣&␣82.8␣&␣493.4\\"
[1] "INVERTED-COVARIANCE_MATRIX"
            [,1]        [,2]        [,3]        [,4]
[1,]  0.44212745 -0.33758804  0.02658168  0.07958441
[2,] -0.33758804  0.38029216 -0.11876646 -0.07889804
[3,]  0.02658168 -0.11876646  0.10844215  0.03500918
[4,]  0.07958441 -0.07889804  0.03500918  0.02608425
[1] "MAXIMUM_LEVERAGE"
[1] 0.04942995


# START NEW EQUATION
[1] "Doing␣200␣year"
The power is -0.0506

Call:
lm(formula = log10(Q) ~ CDA1 + log10(P) + log10(S), weights = mywgts)

Residuals:
     Min       1Q   Median       3Q      Max
-1.25427 -0.23860  0.01366  0.24947  1.01623

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.8633     0.2359  46.059   <2e-16 ***
CDA1         -8.2818     0.2154 -38.444   <2e-16 ***
log10(P)      1.2209     0.1221   9.996   <2e-16 ***
log10(S)      0.6647     0.0600  11.078   <2e-16 ***
---
Signif. codes:  0:"***", 0.001:"**", 0.01:"*", 0.05:".", 0.1:"␣", 1

Residual standard error: 0.3712 on 634 degrees of freedom
Multiple R-Squared: 0.7975,     Adjusted R-squared: 0.7965
F-statistic: 832.3 on 3 and 634 DF,  p-value: < 2.2e-16

PRESS= 95.3359086701712
[1] "LATEX:$Q_{200}␣=␣P^{1.221},S^{0.665}\times␣10^{[10.863-8.282,A^{-0.0506]}},$␣&␣0.37␣&␣0.8␣&␣95.3␣&␣586.4\\"
[1] "INVERTED-COVARIANCE_MATRIX"
            [,1]        [,2]        [,3]        [,4]
[1,]  0.40368172 -0.29668113  0.01974134  0.07508040
[2,] -0.29668113  0.33676662 -0.11169444 -0.07431564
[3,]  0.01974134 -0.11169444  0.10824824  0.03499739
[4,]  0.07508040 -0.07431564  0.03499739  0.02612667
[1] "MAXIMUM_LEVERAGE"
[1] 0.04934428


# START NEW EQUATION
[1] "Doing␣250␣year"
The power is -0.0517

Call:
lm(formula = log10(Q) ~ CDA1 + log10(P) + log10(S), weights = mywgts)

Residuals:
```

```
     Min       1Q   Median       3Q      Max
-1.25699 -0.23952  0.01588  0.25038  1.05275

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.82773    0.23832  45.433   <2e-16 ***
CDA1        -8.18108    0.21707 -37.689   <2e-16 ***
log10(P)     1.20879    0.12510   9.662   <2e-16 ***
log10(S)     0.67555    0.06145  10.993   <2e-16 ***
---
Signif. codes:  0:"***", 0.001:"**", 0.01:"*", 0.05:".", 0.1:"␣", 1

Residual standard error: 0.3797 on 634 degrees of freedom
Multiple R-Squared: 0.7903,     Adjusted R-squared: 0.7893
F-statistic: 796.5 on 3 and 634 DF,  p-value: < 2.2e-16

PRESS= 100.050979279010
[1] "LATEX:$Q_{250}␣=␣P^{1.209},S^{0.676}\times␣10^{[10.828-8.181,A^{-0.0517]}},$␣&␣0.38␣&␣0.79␣&␣100.1␣&␣
     617.2\\"
[1] "INVERTED-COVARIANCE␣MATRIX"
            [,1]        [,2]        [,3]        [,4]
[1,]  0.39397766 -0.28676632  0.01822596  0.07395883
[2,] -0.28676632  0.32683298 -0.11041327 -0.07332429
[3,]  0.01822596 -0.11041327  0.10856462  0.03515585
[4,]  0.07395883 -0.07332429  0.03515585  0.02619725
[1] "MAXIMUM␣LEVERAGE"
[1] 0.04953543


# START NEW EQUATION
[1] "Doing␣500␣year"
The power is -0.0554

Call:
lm(formula = log10(Q) ~ CDA1 + log10(P) + log10(S), weights = mywgts)

Residuals:
     Min       1Q   Median       3Q      Max
-1.29275 -0.26326  0.01617  0.26951  1.11027

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.67132    0.24615   43.35   <2e-16 ***
CDA1        -7.84563    0.22150  -35.42   <2e-16 ***
log10(P)     1.18100    0.13451    8.78   <2e-16 ***
log10(S)     0.70541    0.06626   10.65   <2e-16 ***
---
Signif. codes:  0:"***", 0.001:"**", 0.01:"*", 0.05:".", 0.1:"␣", 1

Residual standard error: 0.4073 on 634 degrees of freedom
Multiple R-Squared: 0.7669,     Adjusted R-squared: 0.7658
F-statistic: 695.2 on 3 and 634 DF,  p-value: < 2.2e-16

PRESS= 116.921491737753
[1] "LATEX:$Q_{500}␣=␣P^{1.181},S^{0.705}\times␣10^{[10.671-7.846,A^{-0.0554]}},$␣&␣0.41␣&␣0.77␣&␣116.9␣&␣
     712.7\\"
[1] "INVERTED-COVARIANCE␣MATRIX"
            [,1]        [,2]        [,3]        [,4]
[1,]  0.36529409 -0.25706556  0.01343282  0.07057228
[2,] -0.25706556  0.29580421 -0.10558462 -0.07012783
[3,]  0.01343282 -0.10558462  0.10909026  0.03560090
[4,]  0.07057228 -0.07012783  0.03560090  0.02646788
[1] "MAXIMUM␣LEVERAGE"
[1] 0.05004109
```

Blank Page

Appendix 3—Results of Weighted-Least-Squares, Multi-Linear Regression Analysis Using PRESS Minimization of Peak-Streamflow Frequency to Watershed Characteristics of Drainage Area, Mean Annual Precipitation, Dimensionless Main-Channel Slope, and OmegaEM

Blank Page

```
# START NEW EQUATION
[1] "Doing␣2␣year"
The power is -0.0058

Call:
lm(formula = log10(Q) ~ CDA1 + log10(P) + log10(S) + OMEGAEM,
    weights = mywgts)

Residuals:
      Min        1Q    Median        3Q       Max
-1.117311 -0.164565 -0.003824  0.156962  1.112398

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  50.97508    1.27640  39.936  < 2e-16 ***
CDA1        -50.29910    1.29205 -38.930  < 2e-16 ***
log10(P)      1.39817    0.09800  14.266  < 2e-16 ***
log10(S)      0.27018    0.04874   5.543 4.36e-08 ***
OMEGAEM       0.77587    0.06629  11.705  < 2e-16 ***
---
Signif. codes:  0:"***", 0.001:"**", 0.01:"*", 0.05:".", 0.1:"␣", 1

Residual standard error: 0.2896 on 633 degrees of freedom
Multiple R-Squared: 0.8437,     Adjusted R-squared: 0.8427
F-statistic: 854.1 on 4 and 633 DF,  p-value: < 2.2e-16

PRESS= 64.601770644145
[1] "LATEX:$Q_{2}␣=␣P^{1.398},S^{0.27}\times␣10^{[50.975-50.299,A^{-0.0058}␣+␣0.776OmegaEMs]␣}$␣&␣0.29␣&␣0.84␣&␣
     64.6␣&␣273.2\\"
[1] "INVERTED-COVARIANCE␣MATRIX"
            [,1]        [,2]        [,3]         [,4]         [,5]
[1,]  19.4253954 -19.5922208  0.81164439  0.597916940 -0.109309193
[2,] -19.5922208  19.9044211 -0.90384486 -0.598196650  0.108244879
[3,]   0.8116444  -0.9038449  0.11451891  0.039052288 -0.011122022
[4,]   0.5979169  -0.5981967  0.03905229  0.028323546 -0.007694694
[5,]  -0.1093092   0.1082449 -0.01112202 -0.007694694  0.052392588
[1] "MAXIMUM␣LEVERAGE"
[1] 0.05358737


# START NEW EQUATION
[1] "Doing␣5␣year"
The power is -0.0215

Call:
lm(formula = log10(Q) ~ CDA1 + log10(P) + log10(S) + OMEGAEM,
    weights = mywgts)

Residuals:
     Min       1Q   Median       3Q      Max
-0.99501 -0.13423  0.01847  0.14796  0.83026

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  16.61519    0.32417  51.255   <2e-16 ***
CDA1        -15.31993    0.32617 -46.968   <2e-16 ***
log10(P)      1.30764    0.08662  15.096   <2e-16 ***
log10(S)      0.37159    0.04327   8.588   <2e-16 ***
OMEGAEM       0.88461    0.05897  15.000   <2e-16 ***
---
Signif. codes:  0:"***", 0.001:"**", 0.01:"*", 0.05:".", 0.1:"␣", 1

Residual standard error: 0.2604 on 633 degrees of freedom
Multiple R-Squared: 0.8799,     Adjusted R-squared: 0.8791
F-statistic:  1159 on 4 and 633 DF,  p-value: < 2.2e-16
```

```
PRESS= 49.1241230911915
[1] "LATEX:$Q_{5}␣=␣P^{1.308},S^{0.372}\times␣10^{[16.615-15.32,A^{-0.0215}␣+␣0.885OmegaEMs]␣}$␣&␣0.26␣&␣0.88␣&␣
     49.1␣&␣121.6\\"
[1] "INVERTED-COVARIANCE␣MATRIX"
            [,1]        [,2]        [,3]        [,4]        [,5]
[1,]  1.54962628 -1.48774667  0.15709427  0.16635099 -0.03302225
[2,] -1.48774667  1.56885660 -0.24679239 -0.16563232  0.03206169
[3,]  0.15709427 -0.24679239  0.11064184  0.03735295 -0.01066574
[4,]  0.16635099 -0.16563232  0.03735295  0.02760925 -0.00742009
[5,] -0.03302225  0.03206169 -0.01066574 -0.00742009  0.05128716
[1] "MAXIMUM␣LEVERAGE"
[1] 0.05118202


# START NEW EQUATION
[1] "Doing␣10␣year"
The power is -0.0289

Call:
lm(formula = log10(Q) ~ CDA1 + log10(P) + log10(S) + OMEGAEM,
    weights = mywgts)

Residuals:
     Min       1Q   Median       3Q      Max
-1.05107 -0.12910  0.01404  0.13987  0.81665

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.62460    0.24348  55.957   <2e-16 ***
CDA1        -11.96593    0.23971 -49.917   <2e-16 ***
log10(P)      1.20257    0.08398  14.320   <2e-16 ***
log10(S)      0.40314    0.04175   9.657   <2e-16 ***
OMEGAEM       0.91773    0.05794  15.838   <2e-16 ***
---
Signif. codes:  0:"***", 0.001:"**", 0.01:"*", 0.05:".", 0.1:"␣", 1

Residual standard error: 0.2537 on 633 degrees of freedom
Multiple R-Squared:  0.89,      Adjusted R-squared: 0.8893
F-statistic:  1281 on 4 and 633 DF,  p-value: < 2.2e-16

PRESS= 46.6043351340521
[1] "LATEX:$Q_{10}␣=␣P^{1.203},S^{0.403}\times␣10^{[13.625-11.966,A^{-0.0289}␣+␣0.918OmegaEMs]␣}$␣&␣0.25␣&␣0.89␣
     &␣46.6␣&␣86.5\\"
[1] "INVERTED-COVARIANCE␣MATRIX"
            [,1]        [,2]         [,3]         [,4]         [,5]
[1,]  0.92075433 -0.83429385  0.092911995  0.124886828 -0.023884308
[2,] -0.83429385  0.89247573 -0.183156553 -0.123763614  0.022313589
[3,]  0.09291200 -0.18315655  0.109537106  0.036345323 -0.009971505
[4,]  0.12488683 -0.12376361  0.036345323  0.027067908 -0.007181418
[5,] -0.02388431  0.02231359 -0.009971505 -0.007181418  0.052145254
[1] "MAXIMUM␣LEVERAGE"
[1] 0.04984506


# START NEW EQUATION
[1] "Doing␣25␣year"
The power is -0.0374

Call:
lm(formula = log10(Q) ~ CDA1 + log10(P) + log10(S) + OMEGAEM,
    weights = mywgts)

Residuals:
     Min       1Q   Median       3Q      Max
```

```
-1.09229 -0.14103  0.02173  0.15698  0.78448

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.79063    0.20629   57.16   <2e-16 ***
CDA1        -9.81896    0.19772  -49.66   <2e-16 ***
log10(P)     1.13976    0.08735   13.05   <2e-16 ***
log10(S)     0.44556    0.04322   10.31   <2e-16 ***
OMEGAEM      0.94541    0.06013   15.72   <2e-16 ***
---
Signif. codes:  0:"***", 0.001:"**", 0.01:"*", 0.05:".", 0.1:"␣", 1

Residual standard error: 0.2637 on 633 degrees of freedom
Multiple R-Squared: 0.886,      Adjusted R-squared: 0.8853
F-statistic:  1230 on 4 and 633 DF,  p-value: < 2.2e-16

PRESS= 49.4591620341938
[1] "LATEX:$Q_{25}␣=␣P^{1.14},S^{0.446}\times␣10^{[11.791-9.819,A^{-0.0374}␣+␣0.945OmegaEMs]␣}$␣&␣0.26␣&␣0.89␣&␣
     49.5␣&␣139.5\\"
[1] "INVERTED-COVARIANCE␣MATRIX"
            [,1]        [,2]         [,3]         [,4]         [,5]
[1,]  0.61180545 -0.51351782  0.054197182  0.098699053 -0.018955130
[2,] -0.51351782  0.56202692 -0.145648642 -0.097592316  0.017195210
[3,]  0.05419718 -0.14564864  0.109693565  0.036021717 -0.009706565
[4,]  0.09869905 -0.09759232  0.036021717  0.026858442 -0.007064779
[5,] -0.01895513  0.01719521 -0.009706565 -0.007064779  0.051983466
[1] "MAXIMUM␣LEVERAGE"
[1] 0.04938781


# START NEW EQUATION
[1] "Doing␣50␣year"
The power is -0.0424

Call:
lm(formula = log10(Q) ~ CDA1 + log10(P) + log10(S) + OMEGAEM,
    weights = mywgts)

Residuals:
     Min       1Q   Median       3Q      Max
-1.11420 -0.16605  0.02835  0.17509  0.87985

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.17250    0.20031   55.78   <2e-16 ***
CDA1        -8.99737    0.18862  -47.70   <2e-16 ***
log10(P)     1.10458    0.09286   11.90   <2e-16 ***
log10(S)     0.47647    0.04593   10.37   <2e-16 ***
OMEGAEM      0.96111    0.06364   15.10   <2e-16 ***
---
Signif. codes:  0:"***", 0.001:"**", 0.01:"*", 0.05:".", 0.1:"␣", 1

Residual standard error: 0.2801 on 633 degrees of freedom
Multiple R-Squared: 0.8757,     Adjusted R-squared: 0.8749
F-statistic:  1115 on 4 and 633 DF,  p-value: < 2.2e-16

PRESS= 55.6054009326601
[1] "LATEX:$Q_{50}␣=␣P^{1.105},S^{0.476}\times␣10^{[11.172-8.997,A^{-0.0424}␣+␣0.961OmegaEMs]␣}$␣&␣0.28␣&␣0.87␣&
     ␣55.6␣&␣219.9\\"
[1] "INVERTED-COVARIANCE␣MATRIX"
            [,1]        [,2]         [,3]        [,4]         [,5]
[1,]  0.51122238 -0.40861290  0.039403326  0.08863528 -0.017386707
[2,] -0.40861290  0.45333014 -0.131154457 -0.08755453  0.015615535
[3,]  0.03940333 -0.13115446  0.109858491  0.03603858 -0.009718802
[4,]  0.08863528 -0.08755453  0.036038576  0.02687713 -0.007077780
```

```
[5,] -0.01738671  0.01561553 -0.009718802 -0.00707778  0.051603573
[1] "MAXIMUM␣LEVERAGE"
[1] 0.04930603

# START NEW EQUATION
[1] "Doing␣100␣year"
The power is -0.0467

Call:
lm(formula = log10(Q) ~ CDA1 + log10(P) + log10(S) + OMEGAEM,
    weights = mywgts)

Residuals:
     Min       1Q   Median       3Q      Max
-1.15429 -0.19251  0.03230  0.18716  0.98946

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.81851    0.20276   53.36   <2e-16 ***
CDA1        -8.44803    0.18823  -44.88   <2e-16 ***
log10(P)     1.07107    0.10032   10.68   <2e-16 ***
log10(S)     0.50665    0.04968   10.20   <2e-16 ***
OMEGAEM      0.96895    0.06848   14.15   <2e-16 ***
---
Signif. codes:  0:"***", 0.001:"**", 0.01:"*", 0.05:".", 0.1:"␣", 1

Residual standard error: 0.302 on 633 degrees of freedom
Multiple R-Squared: 0.8601,     Adjusted R-squared: 0.8593
F-statistic: 973.2 on 4 and 633 DF,  p-value: < 2.2e-16

PRESS= 64.7959688613551
[1] "LATEX:$Q_{100}␣=␣P^{1.071},S^{0.507}\times␣10^{[10.819-8.448,A^{-0.0467}␣+␣0.969OmegaEMs]␣}$␣&␣0.3␣&␣0.86␣&
     ␣64.8␣&␣320\\"
[1] "INVERTED-COVARIANCE␣MATRIX"
            [,1]        [,2]         [,3]         [,4]         [,5]
[1,]  0.45078440 -0.34598773  0.030317559  0.082225298 -0.016017361
[2,] -0.34598773  0.38848667 -0.122200320 -0.081320632  0.014418496
[3,]  0.03031756 -0.12220032  0.110351187  0.036377618 -0.009880846
[4,]  0.08222530 -0.08132063  0.036377618  0.027065505 -0.007100056
[5,] -0.01601736  0.01441850 -0.009880846 -0.007100056  0.051424464
[1] "MAXIMUM␣LEVERAGE"
[1] 0.04958827


# START NEW EQUATION
[1] "Doing␣200␣year"
The power is -0.0504

Call:
lm(formula = log10(Q) ~ CDA1 + log10(P) + log10(S) + OMEGAEM,
    weights = mywgts)

Residuals:
     Min       1Q   Median       3Q      Max
-1.20847 -0.21382  0.02703  0.20050  1.08574

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.60993    0.21114  50.251   <2e-16 ***
CDA1        -8.05788    0.19298 -41.756   <2e-16 ***
log10(P)     1.03433    0.10945   9.451   <2e-16 ***
log10(S)     0.53118    0.05428   9.786   <2e-16 ***
OMEGAEM      0.97538    0.07474  13.051   <2e-16 ***
---
Signif. codes:  0:"***", 0.001:"**", 0.01:"*", 0.05:".", 0.1:"␣", 1
```

```
Residual standard error: 0.3298 on 633 degrees of freedom
Multiple R-Squared: 0.8404,     Adjusted R-squared: 0.8394
F-statistic: 833.5 on 4 and 633 DF,  p-value: < 2.2e-16

PRESS= 77.1727531680508
[1] "LATEX:$Q_{200}␣=␣P^{1.034},S^{0.531}\times␣10^{[10.61-8.058,A^{-0.0504}␣+␣0.975OmegaEMs]␣}$␣&␣0.33␣&␣0.84␣&
    ␣77.2␣&␣436.3\\"
[1] "INVERTED-COVARIANCE␣MATRIX"
           [,1]        [,2]         [,3]         [,4]         [,5]
[1,]  0.40989959 -0.30259286  0.022951245  0.077351356 -0.014780835
[2,] -0.30259286  0.34242026 -0.114612001 -0.076376797  0.013236336
[3,]  0.02295124 -0.11461200  0.110139892  0.036347426 -0.009837713
[4,]  0.07735136 -0.07637680  0.036347426  0.027090365 -0.007032369
[5,] -0.01478084  0.01323634 -0.009837713 -0.007032369  0.051356645
[1] "MAXIMUM␣LEVERAGE"
[1] 0.04948925


# START NEW EQUATION
[1] "Doing␣250␣year"
The power is -0.0516

Call:
lm(formula = log10(Q) ~ CDA1 + log10(P) + log10(S) + OMEGAEM,
    weights = mywgts)

Residuals:
     Min       1Q   Median       3Q      Max
-1.22113 -0.22338  0.02842  0.20574  1.13024

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.55893    0.21427  49.279   <2e-16 ***
CDA1        -7.94317    0.19520 -40.693   <2e-16 ***
log10(P)     1.02134    0.11271   9.061   <2e-16 ***
log10(S)     0.54069    0.05591   9.671   <2e-16 ***
OMEGAEM      0.97717    0.07685  12.716   <2e-16 ***
---
Signif. codes:  0:"***", 0.001:"**", 0.01:"*", 0.05:".", 0.1:"␣", 1

Residual standard error: 0.3391 on 633 degrees of freedom
Multiple R-Squared: 0.833,      Adjusted R-squared: 0.8319
F-statistic: 789.2 on 4 and 633 DF,  p-value: < 2.2e-16

PRESS= 81.8535513717814
[1] "LATEX:$Q_{250}␣=␣P^{1.021},S^{0.541}\times␣10^{[10.559-7.943,A^{-0.0516}␣+␣0.977OmegaEMs]␣}$␣&␣0.34␣&␣0.83␣
    &␣81.9␣&␣474.1\\"
[1] "INVERTED-COVARIANCE␣MATRIX"
           [,1]        [,2]        [,3]         [,4]         [,5]
[1,]  0.39917437 -0.29156312  0.02125209  0.076122309 -0.014815036
[2,] -0.29156312  0.33128176 -0.11313034 -0.075265114  0.013195683
[3,]  0.02125209 -0.11313034  0.11046049  0.036517920 -0.009856940
[4,]  0.07612231 -0.07526511  0.03651792  0.027175928 -0.007087477
[5,] -0.01481504  0.01319568 -0.00985694 -0.007087477  0.051347383
[1] "MAXIMUM␣LEVERAGE"
[1] 0.04968222


# START NEW EQUATION
[1] "Doing␣500␣year"
The power is -0.0554

Call:
lm(formula = log10(Q) ~ CDA1 + log10(P) + log10(S) + OMEGAEM,
```

```
    weights = mywgts)

Residuals:
     Min       1Q   Median       3Q      Max
-1.27882 -0.23558  0.02474  0.22946  1.23151

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.40251    0.22475  46.284  < 2e-16 ***
CDA1        -7.60456    0.20224 -37.601  < 2e-16 ***
log10(P)     0.98824    0.12329   8.016  5.3e-15 ***
log10(S)     0.56851    0.06132   9.272  < 2e-16 ***
OMEGAEM      0.97583    0.08382  11.643  < 2e-16 ***
---
Signif. codes:  0:"***", 0.001:"**", 0.01:"*", 0.05:".", 0.1:"␣", 1

Residual standard error: 0.3699 on 633 degrees of freedom
Multiple R-Squared: 0.808,      Adjusted R-squared: 0.8068
F-statistic:   666 on 4 and 633 DF,  p-value: < 2.2e-16

PRESS= 98.6828845882508
[1] "LATEX:$Q_{500}␣=␣P^{0.988},S^{0.569}\times␣10^{[10.403-7.605,A^{-0.0554}␣+␣0.976OmegaEMs]␣}$␣&␣0.37␣&␣0.81␣
     &␣98.7␣&␣590.9\\"
[1] "INVERTED-COVARIANCE␣MATRIX"
            [,1]        [,2]        [,3]         [,4]         [,5]
[1,]  0.36919005 -0.26055941  0.01622655  0.072556520 -0.014143283
[2,] -0.26055941  0.29893745 -0.10809000 -0.071907268  0.012683503
[3,]  0.01622655 -0.10809000  0.11109360  0.037023770 -0.010141912
[4,]  0.07255652 -0.07190727  0.03702377  0.027478471 -0.007203266
[5,] -0.01414328  0.01268350 -0.01014191 -0.007203266  0.051343459
[1] "MAXIMUM␣LEVERAGE"
[1] 0.05019887
```

Prepared by the USGS Lafayette Publishing Service Center.

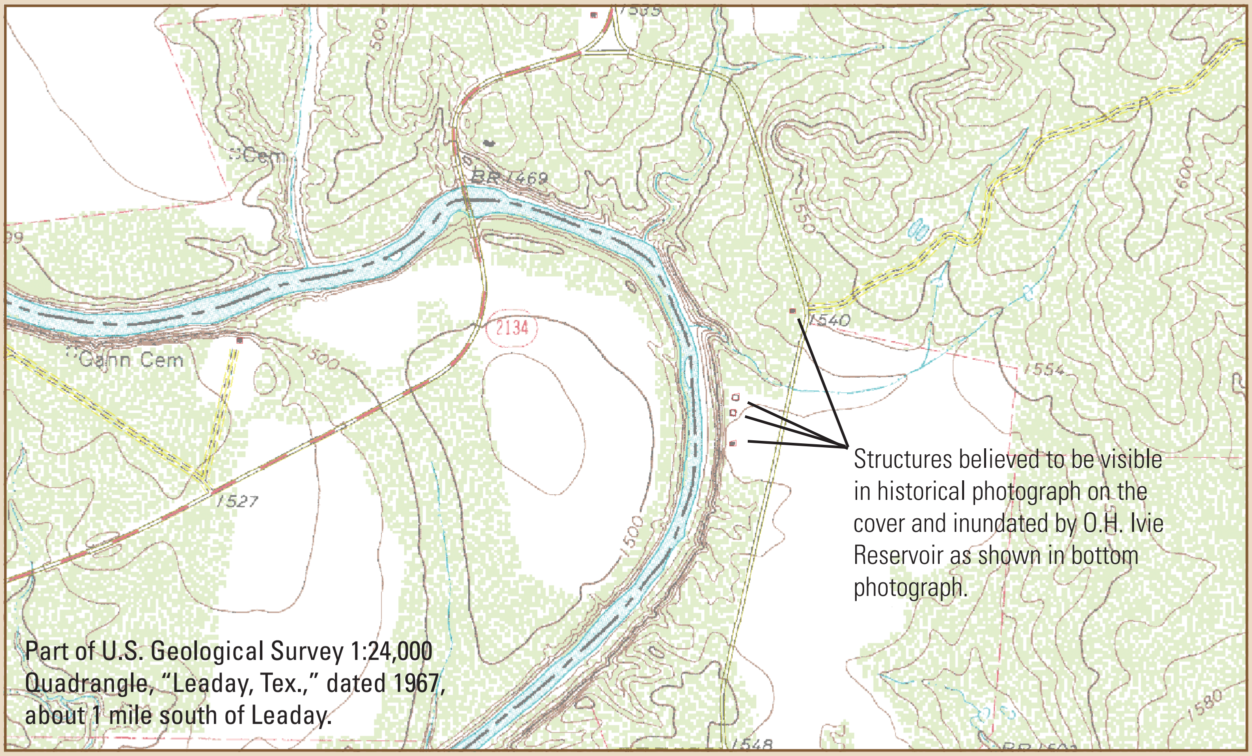
Information regarding water resources in Texas is available at
http: //tx.usgs.gov/

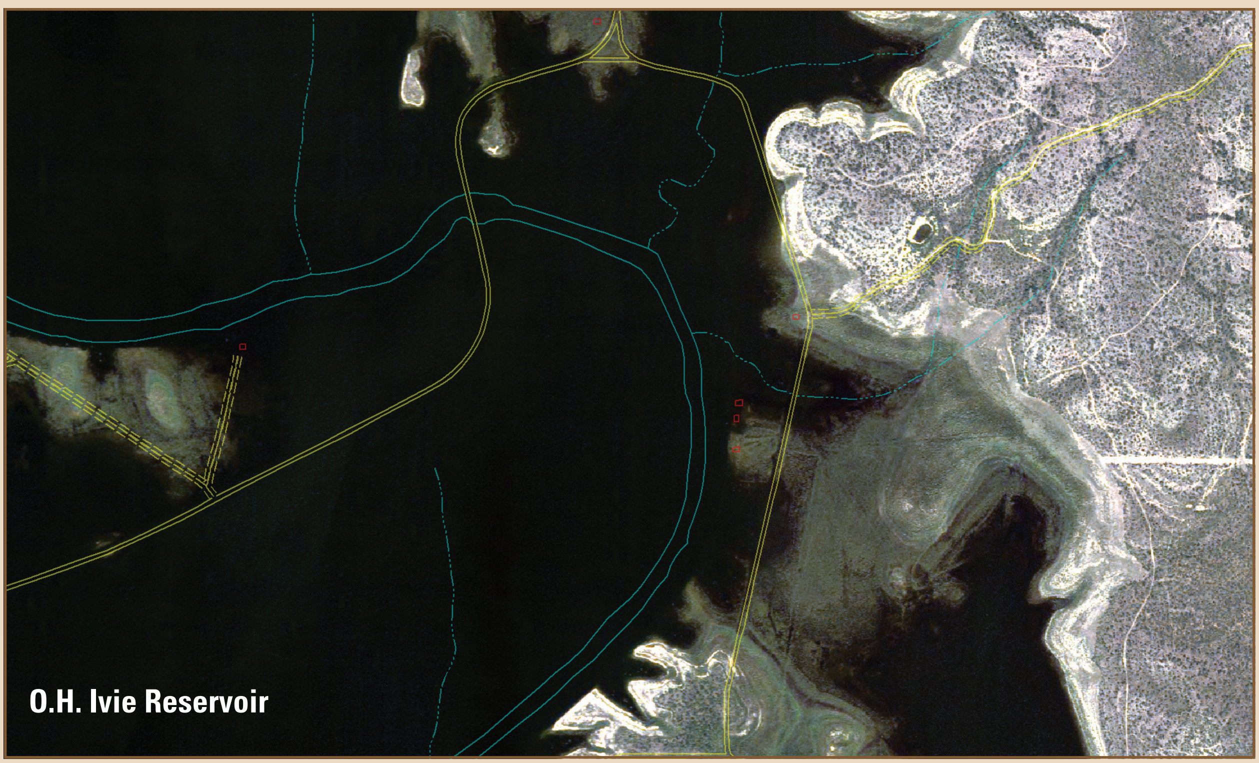
Back cover: O.H. IVIE RESERVOIR: The O.H. Ivie Reservoir, once called Stacy Reservoir, is impounded by the S.W. Freese Dam at the Concho-Coleman county line. It is located in Concho, Coleman, and Runnels Counties. In 1938 the U.S. Army Corps of Engineers expressed a desire for a reservoir site near the confluence of the Concho and Colorado Rivers. An agreement was finally reached in 1985, when the Texas Water Commission granted permission to impound 554,000 acre-feet of water on the Colorado River at Stacy, 16 miles below the confluence. The project was delayed by negotiations to preserve the endangered Concho water snakes, and to relocate several local family cemeteries. The reservoir was to be named for the Stacy settlement, but it was later decided instead to honor the water district's general manager, O.H. Ivie, and to name the dam for Simon W. Freese, a Fort Worth engineer whose firm had worked on major reservoir projects since 1949. The lake waters are used for domestic and municipal water supply for a number of West Texas cities and towns. The conservation surface area of the lake is 20,000 surface acres. The reservoir and its 2-mile rolled earthfill dam, constructed by Brown and Root USA, were dedicated in 1990 and are owned and operated by the Colorado River Municipal Water District. The reservoir is surrounded by a recreation area.

Bibliography: Dallas Morning News, April 10, 1989. John Peterson, "Trouble Rising behind Stacy Dam," Texas Observer, December 10, 1982. Robert Thomas, "Stacy Dam Gets Approval," Ranch Magazine, July 1985. Ed Todd, Cultural Resource Inventory and Assessment of the Proposed Stacy Reservoir, Concho, Coleman, and Runnels Counties, Texas (3 volumes, Austin, 1980).

Handbook of Texas Online, *http://www.tshaonline.org/handbook/online/articles/OO/roogh.html* (accessed April 15, 2009).

The Handbook of Texas Online is a project of the Texas State Historical Association, *http://www.tshaonline.org*





Printed on recycled paper