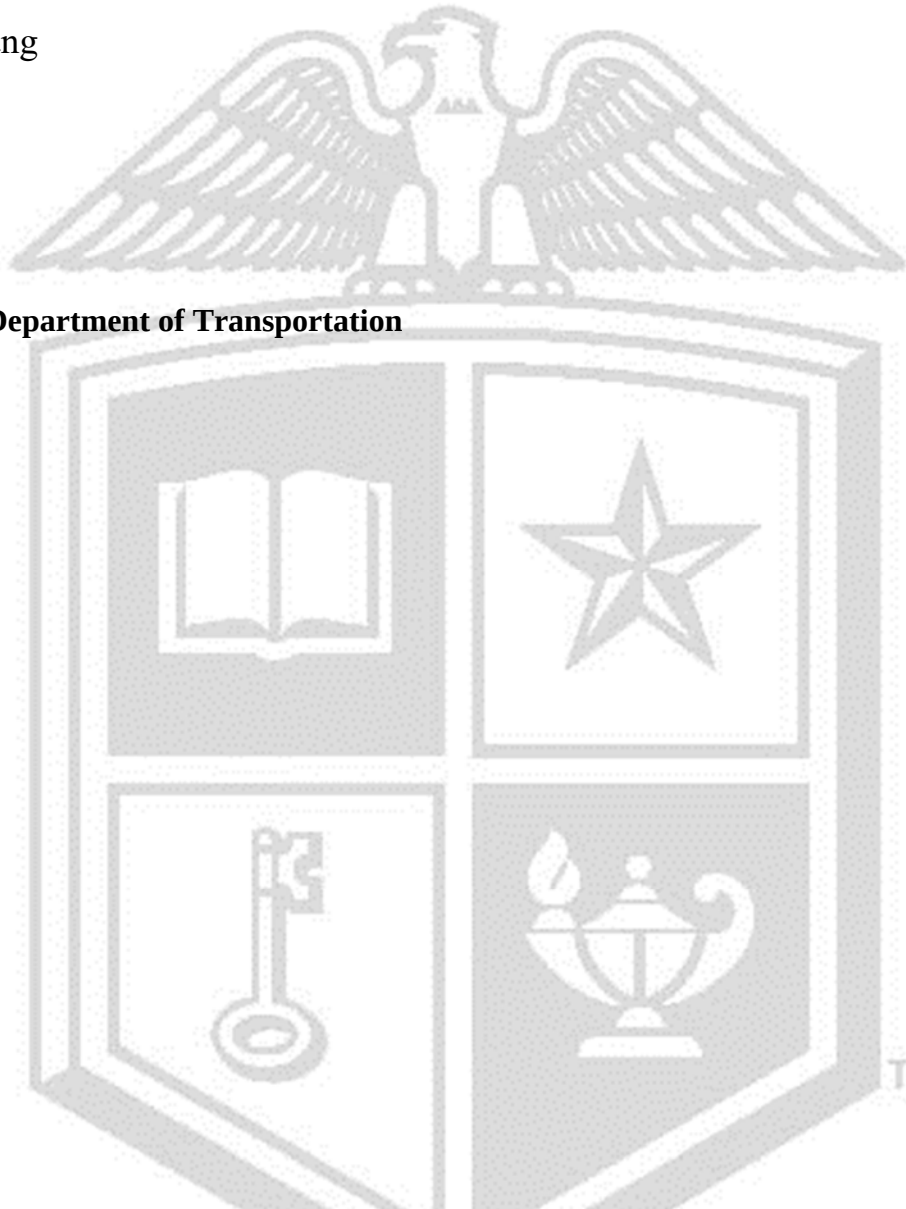
Texas Tech University

Multidisciplinary Research in Transportation

Generalized Skew Update and Regional Study of Distribution Shape for Texas Flood Frequency Analyses

Theodore G. Cleveland and Nick Z. Fang



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

Research Project 0-6977
Research Report 0-6977-1
https://www.depts.ttu.edu/techmrtweb/reports.php

August 2021
Published December 2021

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA/TX–21/0–6977–1		

4. Title and Subtitle	5. Report Date
Generalized Skew Update and Regional Study of Distribution Shape for Texas Flood Frequency Analyses	August 2021 Published December 2021
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Theodore G. Cleveland https://orcid.org/0000-0002-2232-2110 Nick Z. Fang https://orcid.org/0000-0001-9871-8405 with contributions from William H. Asquith, Monica V. Yesildirek, Jeremy S. McDowell, Jiaqi Zhang, Raven N. Landers, Lindsay D. Otto	0-6977-1

9. Performing Organization Name and Address	10. Work Unit No.
Center for Multidisciplinary Research in Transportation Texas Tech University, Box 41023 Lubbock, TX 79409-1023 University of Texas at Arlington Box 19308, Arlington, TX 76019-0308 U.S. Geological Survey Oklahoma-Texas Water Science Center 1505 Ferguson Lane, Austin, TX 78754	**11. Contract or Grant No.** 0-6977

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Texas Department of Transportation Research and Technology Implementation Division 125 E. 11th Street Austin, TX 78701	Technical Report (1 Nov 2018 – 31 Aug 2021)
	14. Sponsoring Agency Code

15. Supplementary Notes

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

Study title: Develop a Generalized Skew Update an Regional Study of Other Measures of Distribution Shape for Texas Flood Frequency Analyses

https://doi.org/10.18738/T8/SVLCOQ

16. Abstract

The research herein, performed under joint collaboration between researchers at Texas Tech University, the University of Texas at Arlington, and the U.S. Geological Survey (USGS) Oklahoma-Texas Water Science Center, aimed to update the Texas generalized skew map and its mean-square error for the Texas Department of Transportation (TxDOT) Hydraulic Design Manual. The research was conducted using at least 30 years of streamflows acquired from 444 USGS streamgages in Texas, Oklahoma, and eastern New Mexico. A novel application of a generalized additive model is used for 2-dimensional spatial regression on station-skew values from the streamgages to create predictions of generalized skew on the 1-kilometer USGS National Hydrogeologic Grid clipped to the study area.

Training examples on flood frequency analysis oriented around Bulletin 17C for Texas watersheds are included. Emergent studies of (1) climate sensitivity of peak streamflows, (2) evaluation of USGS multiorder hydrologic position metrics, and (3) experimental accommodation of effects of regulation are presented.

The report and data are archived at the Texas Digital Library at https://doi.org/10.18738/T8/SVLCOQ, and—along with the three USGS publications resulting from the research—establish a comprehensive update of generalized skew and set a foundation for future statistical research into Texas, Oklahoma, and eastern New Mexico flood hydrology.

17. Key Words	18. Distribution Statement
Texas Flood Frequency; Regional Skew Coefficient; Generalized Additive Model; Stream gages	No restrictions. This document is available to the public through the National Technical Information Service, Alexandria, Virginia 22312, https://www.ntis.gov/.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	151	

Generalized Skew Update and Regional Study of Distribution Shape for Texas Flood Frequency Analyses

by

Theodore G. Cleveland, Associate Professor,
Civil, Environmental, and Construction Engineering, Texas Tech University

and

Nick Z. Fang, Associate Professor,
Civil and Environmental Engineering, University of Texas at Arlington

with contributions from

William H. Asquith, Research Hydrologist (Project team member 2019–2021)
Monica V. Yesildirek, Hydrologist (Project team member 2019–2021)
Jeremy S. McDowell, Hydrologist (Project team member 2019–2021)
U.S. Geological Survey Oklahoma-Texas Water Science Center

and

Jiaqi Zhang, University of Texas at Arlington (Project team member 2019–2021)
Raven N. Landers, Texas Tech University (Project team member 2020–2021)
Lindsay D. Otto, University of Iowa (Project team member 2019–2020)

Research Project 0–6977
Research Report 0–6977–1

Project Title: *Develop a Generalized Skew Update an Regional Study of Other Measures of Distribution Shape for Texas Flood Frequency Analyses*

Sponsored by the Texas Department of Transportation

August 2021
Published December 2021

Center for Multidisciplinary Research in Transportation
Texas Tech University
Box 41023
Lubbock, Texas 79409-1023

Disclaimer

The contents of this report reflect the views of the authors (Cleveland, T.G. and Fang, Z.N.), who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official view or policies of the Federal Highway Administration (FHWA) or the Texas Department of Transportation (TxDOT). The chapter herein titled "Chapter 4: Technique to estimate generalized skew coefficients of annual peak streamflow for natural watershed conditions in Texas, Oklahoma, and eastern New Mexico" does represent the views of the U.S. Geological Survey. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. This report does not constitute a standard, specification, or regulation. This report is not intended for construction, bidding, or permit purposes. The State of Texas does not endorse products or manufacturers. Trade or manufacturers' names that appear herein are solely because they are considered essential to the object of this report.

The researcher in charge of this project was Dr. Theodore G. Cleveland, Ph.D., P.E., at Texas Tech University, Lubbock, Texas.

There was no invention or discovery conceived or first actually reduced to practice in the course of or under this contract (to date), including any art, method, process, machine, manufacture, design, or composition of matter, or any new useful improvement thereof, or any variety of plant, which is or may be patentable under the patent laws of the United States of America or any foreign country.

Abstract

Research Project 0–6977, performed in joint collaboration between researchers at Texas Tech University, the University of Texas at Arlington, and the U.S. Geological Survey (USGS) Oklahoma-Texas Water Science Center, aimed to update the Texas generalized skew map and its mean-square error (chap. 4) for the Texas Department of Transportation (TxDOT) Hydraulic Design Manual. The research was conducted using at least 30 years of streamflows acquired from 444 USGS streamgages in Texas, Oklahoma, and eastern New Mexico. A novel application of a generalized additive model is used for 2-dimensional spatial regression on so-called station-skew values from the streamgages to create predictions of generalized skew on the 1-kilometer USGS National Hydrogeologic Grid clipped to the study area.

Training materials (chap. 5) were produced from this research with examples on flood frequency analysis oriented around Bulletin 17C for Texas watersheds. The research also extended understanding of flood hydrology in the region by conducting (1) climate sensitivity of peak streamflows (chap. 6), (2) experimental evaluation of a USGS multiorder hydrologic position metrics (chap. 7), and (3) experimental accommodation of effects of regulation (chap. 8).

This report along with its applicable data, archived into the Texas Digital Library (`https://tdl.org`) Texas Data Repository (Cleveland and Fang, 2021, `https://doi.org/10.18738/T8/SVLCOQ`), and the three USGS publications resulting from Research Project 0–6977 establish a comprehensive and well-documented update of generalized skew and set a foundation for future statistical research into Texas, Oklahoma, and eastern New Mexico flood hydrology.

Contents

1. Introduction

The flood hydrology of Texas, Oklahoma, and eastern New Mexico (east of the Great Continental Divide) is complex because of a myriad of meteorological and physiographic factors, and flood hydrology is typically studied using the annual peak streamflow data collected by the U.S. Geological Survey (USGS) at streamgages. Hydraulic design engineers need standard of practice guidance for various tasks involving the analysis and application peak streamflow information. Analyses of this information materially influences bridge design, operational safety of drainage infrastructure, flood-plain management, and other decisions affecting society.

Research Project 0–6977, which was performed in joint collaboration between researchers at Texas Tech University, the University of Texas at Arlington, and the USGS Oklahoma-Texas Water Science Center, was tasked with a primary objective of updating the Texas generalized skew map and its mean-square error. Secondary objectives of the research were to provide, as shown in this report, training materials using Texas watershed examples on flood frequency analysis and example impacts of the updated generalized skew. Tertiary objectives of the research reported on herein further extend understanding of flood hydrology in the region inclusive of (1) climate sensitivity of peak streamflows, (2) experimental evaluation of a USGS multiorder hydrologic position (MOHP) metric that is gridded at continental scale and expresses stream reach position on the landscape, and (3) experimental accommodation of effects of regulation.

1.1. General Approach

The general approach was to aggregate foundational datasets of annual peak streamflows from the USGS National Water Information System (NWIS) (U.S. Geological Survey, 2018) for USGS streamgages. The research selected 1,703 streamgages in the study area as a basis for data retrieval. These streamgages have at least 6 years of annual peak streamflow data and constitute the vast majority of peak streamflow information for the study area housed in NWIS. Of the 1,703 streamgages, a subset of 444 streamgages was chosen for the primary objective of updating generalized skew. These 444 streamgages had at least 30 years of

data from unregulated and undeveloped watershed conditions. The term "natural watershed conditions" then is used elsewhere in this context.

Extensive watershed properties were computed for all 1,703 streamgages. NWIS traditionally only stores the official USGS drainage area and contributing drainage area of streamgages. Other watershed properties, such as mean annual precipitation or main-channel slope, though are useful for statistical study of flood hydrology. The USGS MOHP was determined for the streamgages. The MOHP is a recent advance by the USGS for depiction of position or drainage order of a given grid cell relative to the terrain and drainage divides. This MOHP has never been reviewed for its information content in statistical flood hydrology for the study area.

Peak streamflow regulation by dams (reservoirs) in the study area can substantially alter summary statistics and in turn complicate interpretation of standard of practice methods for hydraulic engineering decision and decisions. This project also aggregated the U.S. Army Corps of Engineers National Inventory of Dams (NID) through software to bind, that is, associate, cumulative year-by-year reservoir storage to year-by-year annual peak streamflows.

The academic investigators of the project team worked to assemble non-citable Technical Memoranda for TxDOT project managers during project tenure. These were intended to report on current progress towards contractual task obligations and elevate situational awareness of managers. Material from those technical memoranda has been abstracted and used as needed in this final report. This final report presents a comprehensive review of research conclusions for the project.

Accompanying this final report are citations to persistent, publicly-accessible publications by digital object identifier (`https://doi.org`). The project team has made the content supporting this final report transparently available in a novel combination of platforms including this final report (through the University of Texas Center for Transportation Research Library, `https://ctr.utexas.edu`), Texas Digital Library (`https://tdl.org`), and USGS publications (data releases at `https://www.sciencebase.gov` and software releases at `https://code.usgs.gov`). The research team considers this combination of information outlets maximally stewards for future reference the most critical content stemming from TxDOT research sponsorship.

1.2. Report Structure

The final report for Research Project 0–6977 is organized as a series of chapters. Each chapter contains its own bibliography and this design highlights the stand-alone features of these report components. Chapters 2 and 3 largely follow the workflow chronology and these chapters are abstracted synopses about the non-citable Technical Memoranda produced for

TxDOT product managers during the course of the research (2019–21). These documents are archived in a persistent, publicly-accessible database and maintained by the Texas Digital Library (`https://tdl.org`) Texas Data Repository (Cleveland and Fang, 2021).

The primary objective of the project was to update the generalized skew in the current (September 2019) Hydraulic Design Manual (Texas Department of Transportation, 2020). Chapter 4 is, therefore, the main product of this research and many readers could skip to this chapter to learn about the generalized updated skew values, how they were obtained, and where the results could appear in the Hydraulic Design Manual. Many readers will already know how to applied generalized skew in practice using Bulletin 17C (England et al., 2018) and USGS PeakFQ software (U.S. Geological Survey, 2020). However, for readers who have not previously used generalized skew, some training material specific to Texas examples of flood frequency analyses using the updated generalized skew (chap. 4) are presented in chapter 5. Finally, chapter 4 provides a suggested citation to that chapter ahead of its textual content; this feature is implemented because that particular chapter represents the views of the USGS, which means that the contents of the chapter have been peer reviewed and have received "bureau approval" by the Director of the USGS.

Chapters 6, 7, and 8 report on tertiary project objectives to extend understanding of flood hydrology in the study area. The chapters are to be read as stand-alone "white papers" as it were, and possible parlance would be to use the term "yellow paper" instead because "[a yellow paper] is a document containing research that has not yet been formally accepted or published" from `https://en.wikipedia.org/wiki/White_paper` (accessed June 27, 2021).

These three chapters present results of experimental or exploratory research conducted for this project thought to be of value related to flood hydrology in Texas, Oklahoma, and eastern New Mexico. These chapters have been titled as special studies; they are suitable for future investigation someday and represent some emergent thinking about flood hydrology stemming from the project. Finally, chapter 9 is the summary and conclusions for the entire research project.

1.3. Data Sources

The annual peak streamflow data used in this study reside in NWIS (U.S. Geological Survey, 2018). These data are readily retrieved using an Internet browser or from structured computer-based data retrievals. The authors regularly use the *dataRetrieval* package (DeCicco et al., 2020) in the R language (R Core Team, 2020) for these computer-based data retrievals. The *dataRetrieval* package is published by the USGS (canonical USGS home page is `https://code.usgs.gov/water/dataRetrieval` [accessed June 26, 2021]), is highly optimized for interfacing with `https://waterservices.usgs.gov/` (accessed June 26, 2021), and is the major automation portal to public NWIS data. The following two sections identify other

data sources, or more precisely, the archival of data and software sources associated with other aspects of the research project.

1.3.1. About Texas Data Repository Archival

Texas Digital Library (`https://tdl.org`) Texas Data Repository (Cleveland and Fang, 2021) houses the typesetting sources for this final report. More importantly, however, is that Cleveland and Fang (2021) contains in a directory-structured layout copies of USGS data as reported on in the next section but also the functional scripts in the R language (R Core Team, 2020) used for statistical computations. Within this report and at critical junctures, references to file names along directory paths of Cleveland and Fang (2021) are made using a `monospaced font` (as shown). This references are present to foster research transparency and science equity at the terminal end (August, 2021) of the research project.

1.3.2. About USGS Datasets and Software Publications

For study of flood hydrology in Texas, Oklahoma, and eastern New Mexico, the USGS as a participating project team member published a data release (Yesildirek et al., 2021) on watershed and ancillary properties for 1,703 USGS streamgages, specialized automation software (*scNIDaregis*) (Asquith et al., 2021) for working with the NID to preprocess such data into watershed- and streamgage-specific time series of cumulative reservoir storages for statistical hydrologic study, and software (Asquith, England, and Herrmann, 2020) for computation of the multiple Grubbs–Beck test following Bulletin 17C guidelines (England et al., 2018). The three USGS publications identified in this section represent the views of the USGS, which means, like chapter 4, that the contents have been peer reviewed and have received "bureau approval" by the Director of the USGS.

Chapter References

Asquith, W.H., Cleveland, T.G., Yesildirek, M.V., Zhang, J., Fang, Z.N., and Otto, L.D., 2021, *scNIDaregis*—Geospatial processing of dams in the United States from the National Inventory of Dams with a state-level aggregation scheme, demonstrated for selected dams in eight states in south-central region of the United States, and post-processing features for basin-specific tabulation: U.S. Geological Survey software release, Reston, Va., `https://doi.org/10.5066/P90NJVB9`.

Asquith, W.H., England, J.F., and Herrmann, G.R., 2020, *MGBT*—Multiple Grubbs-Beck low-outlier test: U.S. Geological Survey software release, R package, Reston, Va., accessed July 27, 2020, at `https://doi.org/10.5066/P9CW9EF0`.

Cleveland, T.G., and Fang, Z.N., 2021, Texas-Skew-Update-2021: Texas Data Repository, https://doi.org/10.18738/T8/SVLCOQ.

DeCicco, L., Hirsch, R., Lorenz, D., Read, J., Walker, J., Carr, L., and Watkins, D., 2020, dataRetrievalRetrieval functions for USGS and EPA hydrologic and water quality data: R package version 2.7.6, dated March 11, 2020, accessed October 3, 2020, at https://cran.r-project.org/package=dataRetrieval.

England, J.F., Cohn, T.A., Faber, B.A., Stedinger, J.R., Thomas Jr., W.O., Veilleux, A.G., Kiang, J.E., and Mason, R.R., 2018, Guidelines for determining flood flow frequency Bulletin 17C: U.S. Geological Survey Techniques and Methods, book 4, chap. 5.B, 148 p., https://doi.org/10.3133/tm4B5.

R Core Team, 2020, R—A language and environment for statistical computing: R Foundation for Statistical Computing, Vienna, Austria, version 4.0.2, accessed July 4, 2020, at https://www.r-project.org.

Texas Department of Transportation, 2020, Hydraulic design manual—September 2019: online resource accessed on July 26, 2020, at https://onlinemanuals.txdot.gov/txdotmanuals/hyd/index.htm and subsection https://onlinemanuals.txdot.gov/txdotmanuals/hyd/statistical_analysis_of_stream_gauge_data.htm.

U.S. Geological Survey, 2018, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed April 9, 2018, at https://doi.org/10.5066/F7P55KJN.

U.S. Geological Survey (USGS), 2020, PeakFQ—Flood frequency analysis based on Bulletin 17B and recommendations of the Advisory Committee on Water Information (ACWI) Subcommittee on Hydrology (SOH) Hydrologic Frequency Analysis Work Group (HFAWG), version 7.2, accessed January 29, 2019, at https://water.usgs.gov/software/PeakFQ/, version 7.3, accessed February 1, 2020, at https://water.usgs.gov/software/PeakFQ/.

Yesildirek, M.V., McDowell, J.S., Zhang, J., Asquith, W.H., 2021, Geospatial data of watershed characteristics for select U.S. Geological Survey streamgaging stations in New Mexico, Oklahoma, and Texas useful for statistical study of annual peak streamflows in and near Texas: U.S. Geological Survey data release, https://doi.org/10.5066/P9A91W4Z.

2. Literature Review

This chapter replicates much of TM2 which is archived at at the Texas Digital Library dataverse repository (Cleveland and Fang, 2021) in the directory `0-6177-dataverse-archive/technical_memos/`. The chapter is repeated here for continuity with the original research scope of work and to establish the knowledge the researchers had collectively accumulated circa 2019.

2.1. Analysis Methods

Tasker and Stedinger (1986) developed a weighted least squares (WLS) procedure for estimating regional skewness coefficients based on sample skewness coefficients for the logarithms of annual peak-streamflow data. Their method of regional analysis of skewness estimators accounts for the precision of the skewness estimator for each streamgage, which depends on the length of record for each streamgage and the accuracy of an ordinary least squares (OLS) regional mean skewness. These methods automated much of B17B process and were incorporated into software used for streamgage analysis.

More recently, Reis and others (2005), Gruber et al. (2007), and Gruber and Stedinger (2008) developed a Bayesian generalized least squares (B-GLS) regression model for regional skewness analyses. Use of a generalized least squares (GLS) model allows the incorporation of the cross correlation of skewness estimators. Cross correlation arises, as skewness estimators are dependent upon concurrent cross correlation flood records. The Bayesian method allows for the computation of a posterior distribution of both the regression parameters and the model error variance.

As shown in Reis and others (2005), for cases in which the model error variance is small compared to the sampling error of the at-site estimates, the Bayesian posterior distribution provides a more reasonable description of the model error variance than both the GLS method of moments and maximum likelihood point estimates (Veilleux, 2011).

Whereas weighted least squares (WLS) regression accounts for the precision of the regional model and the effect of the record length on the variance of skewness coefficient estimators, GLS regression also considers the cross correlations among the skewness coefficient estimators.

The B-GLS regression procedures extend the GLS regression framework by also providing a description of the precision of the estimated model error variance, a pseudo analysis of variance, and enhanced diagnostic statistics (Griffis and Stedinger, 2009).

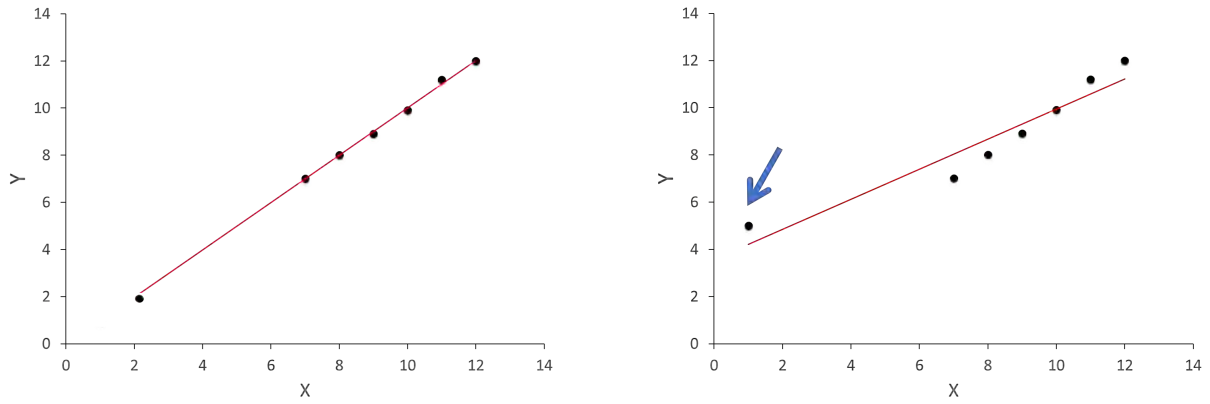
Because of complications introduced by the use of the Expected Moments Algorithm (EMA) (Cohn et al., 1997) and large cross-correlations between annual peak streamflows at pairs of streamgages sites, an alternate regression procedure was developed to provide both stable and defensible results for regional skewness coefficient models (Veilleux, 2011). This alternate procedure is referred to as the B-WLS/B-GLS regression framework (Veilleux, 2011; Veilleux et al., 2011). It uses an OLS analysis to fit an initial regional skewness model; that OLS model is then used to generate a stable regional skewness coefficient estimate for each site. That stable regional estimate is the basis for computing the variance of each at-site skewness coefficient estimator employed in the WLS analysis. Then, B-WLS is used to generate estimators of the regional skewness coefficient model parameters. Finally, B-GLS is used to estimate the precision of those WLS parameter estimators, to estimate the model error variance and the precision of that variance estimator, and to compute various diagnostic statistics, including Bayesian plausibility values, pseudo adjusted R squared, pseudo analysis of variance table, two diagnostic error variance ratios, as well as leverage and influence metrics. This method has been successfully used to generate regional skew estimates in various parts of the Nation but not the Nation as a whole.

2.1.1. Recent Techniques within/comparable to Bulletin 17C

In Bulletin 17C (England et al., 2018), the multiple Grubbs–Beck test (MGBT) are recommended for the detection of potentially influential low floods (PILFs) ("low outliers" remains the preferred term in Texas flood hydrology). The new MGBT was developed as an improvement to the Grubbs–Beck test (GBT) (Grubbs and Beck, 1972) used in Bulletin 17B. The MGBT is a statistically appropriate generalization of the GBT test, and is sensitive to the PILFs. The MGBT also correctly evaluates cases where one or more observations are zero, or are below a recording threshold (partial record sites). Thus, MGBT provides a consistent, objective, and statistically defensible algorithm that considers whether a range of the smallest observations should be classified as outliers (or PILFs) for a much wider range of situations (England et al., 2018).

2.1.2. Outlier Detection and Handling

Outliers are observations that depart from the overall pattern of a distribution. The presence of an outlier indicates some sort of problem (from a statistical perspective), such as a case that does not fit the model under study (an observation form a different process—hurricane as compared to a synaptic storm), or an error in measurement.



(a) Normalized observations with the far left plotting data point being influential that is not an outlier. The trend line fits all the data .

(b) Normalized observations with the far left plotting data point being influential and an outlier. The trend line is leveraged away with poor fit to the other data.

Figure 2.1. Example of normalized observations in the context of outlier detection.

Figures 2.1a and 2.1b illustrate the influence of outliers. The single value in figure 2.1b at the left end of the series changes the model result (the slope of the red curve is decreased). Outliers such as these can impact the overall analysis, and if the values are indeed outliers then the resulting extrapolations after distributional fitting will be suspect.

Low outliers (as depicted above) have an effect on computed skews (PILFs in the introductory sentence of this section). High outliers also exist. The skewness coefficient is sensitive to extreme values, and outliers left unconsidered will exert influence on the computed coefficient.

Figure 2.2 illustrates the analytical choices; the red curve passed through the data cloud is based only on the values lying above the "alternative low-outlier threshold." Were these values included in the analysis, the curve would be bent even further, and to preserve fit have an increased average slope that would result in the extrapolated estimates (right side of the curve) being larger than suggested by the outlier-free series.

Paretti et al. (2014) evaluated the EMA and a multiple low-outlier test on streamgage stations in Arizona (one such streamgage is depicted in the illustrative figure). The authors concluded that that EMA-MGB performed as well or better than B17B-GBT, especially when low outliers were present and (or) historical information was available. In addition, EMA-MGB properly addresses PILFs using MGBT to ensure that zero and low-flow peaks that depart from the trend of the data have remarkably little influence on the frequency fit. Finally, the findings support those of previous studies that indicate the efficacy of EMA-MGBT and that EMA-MGBT is a suitable successor to traditional B17B-GBT methods.

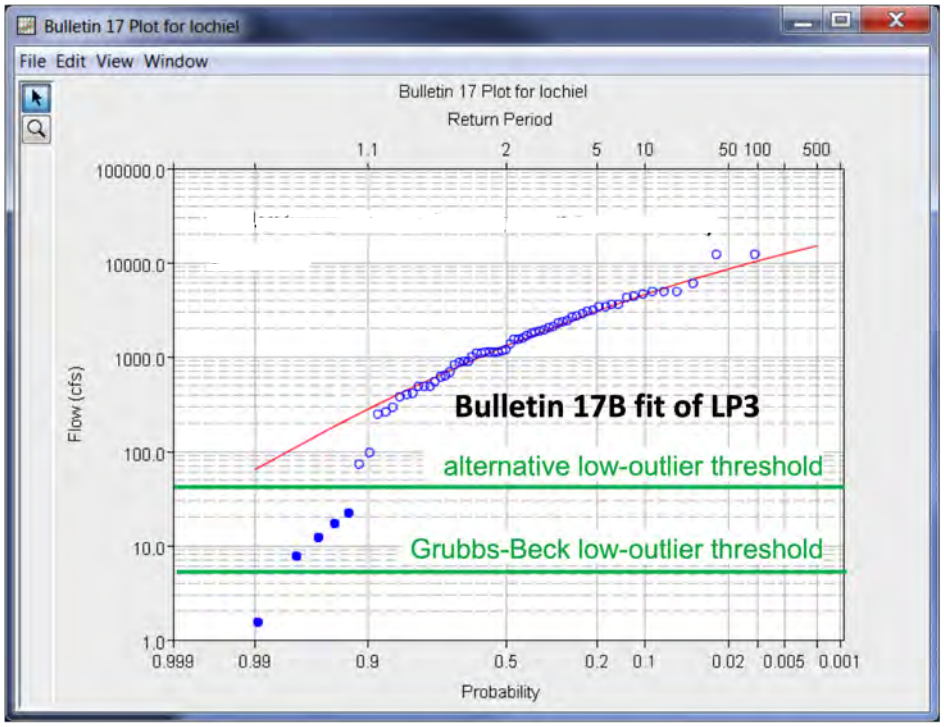


Figure 2.2. Results of low-outlier analysis for annual peak streamflows. Bulletin 17B (Grubbs–Beck) would only conditionally remove a single lowest value. An experienced practitioner would likely choose a threshold between 30 and 200 cubic feet per second. Bulletin 17C (multiple Grubbs–Beck test) automatically selects a threshold as alternatively shown that truncates the lowest five peak streamflows. Fit in the upper (right-hand) tail of the fitted Pearson type III distribution is enhanced and often the effect is that the skewness of the distribution is reduced as low-outliers are removed.

The substantial effort in Paretti et al. (2014) and other prior studies is embedded in the Bulletin 17C tools to identify both low- and high-outliers within a data series and provide mechanisms to censor or replace these values with values that honor the existence of the observation, but reduce its potential influence on the estimated distributional characteristics.

Because EMA allows for the censoring of low outliers, as well as the use of estimated interval peak streamflows for missing, censored, and historic data, it complicates the computations of effective record length (and effective concurrent record length) used to describe the precision of sample estimators because the peak streamflows are no longer solely represented by single values (Veilleux et al., 2012).

2.2. Texas Flood Distributions

The distribution shapes of Texas flood distributions might be expected to show complex interaction between space as in a conventional skew map (Judd et al., 1996) and ancillary watershed properties (Asquith, 2001). Regulation (discussed below) is one of these ancillary properties that are to be addressed.

2.2.1. Effects of Regulation by Dams and Reservoirs

Historically, a natural watershed in Texas has been defined by the USGS as a watershed with less than 10-percent impervious cover, with less than 10 percent of its drainage area controlled by reservoirs, and no other human-related factors that would affect peak streamflow. Benson (1962, 1964) determined that about 100 acre-feet per square mile (acre-ft/mi^2) of flood storage in the drainage area reduces the annual peak streamflow by about 10 percent in humid areas, and that about 50 acre-ft/mi^2 of flood storage reduces the annual peak streamflow by about 10 percent in arid areas. Such criteria serve as loose guidance for assessing whether a watershed is regulated or not.

The degree of actual control (regulation) in a given part of a watershed is more usefully assessed on a more qualitative than quantitative basis, because regulation in reality encompasses a wide spectrum of human influences on annual peak streamflow. For example, a specific reservoir or suite of reservoirs might dramatically alter low flows and midrange flows, but leaves higher flows, such as the annual peaks, largely unaffected. The exclusion of higher flows from records at streamgages in the watershed will bias frequency analysis for that watershed—in essence, categorization as regulated potentially excludes valuable information.

Peak streamflow in regulated watersheds is affected by runoff from the unregulated part of the watershed and by runoff from the regulated part of the watershed. The regulated runoff is affected by discharges from or retarded by dams or reservoirs and by the quantity, type, and spatial distribution of flood-detention or flood-retention structures, commonly termed "small floodwater-retarding structures."

Asquith (2001) quantified, and predicted effects of regulation on annual peak streamflow in Texas through changes in the L-moment statistics of the annual peak streamflow. The change in the L-moments of the streamflow was related to changes in reservoir characteristics (therefore, degree of regulation). Analysis of the relations between the L-moment and the variables representing regulation indicated that L-moments other than the mean are negligibly affected by regulation, and that as potential flood storage in a watershed increases, the mean annual peak streamflow decreases nonlinearly. The effect on the mean supports the conjecture above of a bias on the frequency analysis at a streamgage, and the non-liner behavior suggests a strong geographic influence is contained in the data.

The designation of a watershed as natural, urban (a type of regulation) or regulated (though impoundment) is not sufficient to accurately express the degree of regulation because of the presence of complex arrangements of flow altering structures in the majority of watersheds across the entire range of drainage areas in Texas. Such structures include contour-plowing of agricultural lands for erosion control and enhanced water retention, bulldozed earthen embankments on the smallest of usually dry tributaries forming "Texas stock tanks," low-head dams built nearly as high as the tops of stream banks, small floodwater-retarding structures, constant-level recreational and water-supply reservoirs, and colossal flood-control reservoirs on many of the rivers in Texas. Peak streamflows from reservoirs represent controlled discharges (releases) and uncontrolled discharges through pipes, spillways, or other structures. Controlled releases often are dictated by flood-management practices and water-supply concerns, which might have opposing objectives. There are more than 200 major reservoirs in Texas—those having maximum storage capacities in excess of 10,000 acre-feet (acre-ft). This total also includes a few off-channel and diversion-oriented reservoirs.

Until recently, the ability to use this information was not well articulated so the censoring in the streamflow records was a logical—however most tools and analyses ignore these records as part of routine analysis. Recent concepts and tools may allow the extension of records (for skew analysis) using geographic information systems to extract the portion of a watershed that is unregulated, as well as use the distance from the regulatory structure to the streamgage as a regionalization tool to extend the utility of existing databases by accounting for regulation, and its associated covariation.

2.2.2. Distal Tail Shape or Distribution Curvature

The general purpose for fitting a probability distribution is to represent the magnitude of floods across a wide spread of annual exceedance probabilities, and a reasonable probability distribution is especially important when extrapolations of the fitted frequency curve are to be made.

To illustrate, consider figure 2.3, which is a screen capture from a streamgage. The figure displays about 100 annual peaks. Of these 100 values, 86 values plot within the range of 10,000 to 30,000 cubic feet per second. Some 7 values plot above 30,000 cubic feet per second, and 7 values plot below 10,000 cubic feet per second.

The extrapolation of these observations for high flow tendencies relies on the 7 high values; the 7 low values also influence the analysis, whereas the 86 values between 10,000 and 30,000 cubic feet per second explain the mean behavior and variability with respect to ± 2 sample standard deviations.

Figure 2.4 is representative of the same type of information plotted on a cumulative relative frequency versus a normalized (about the mean value) magnitude axis. The information

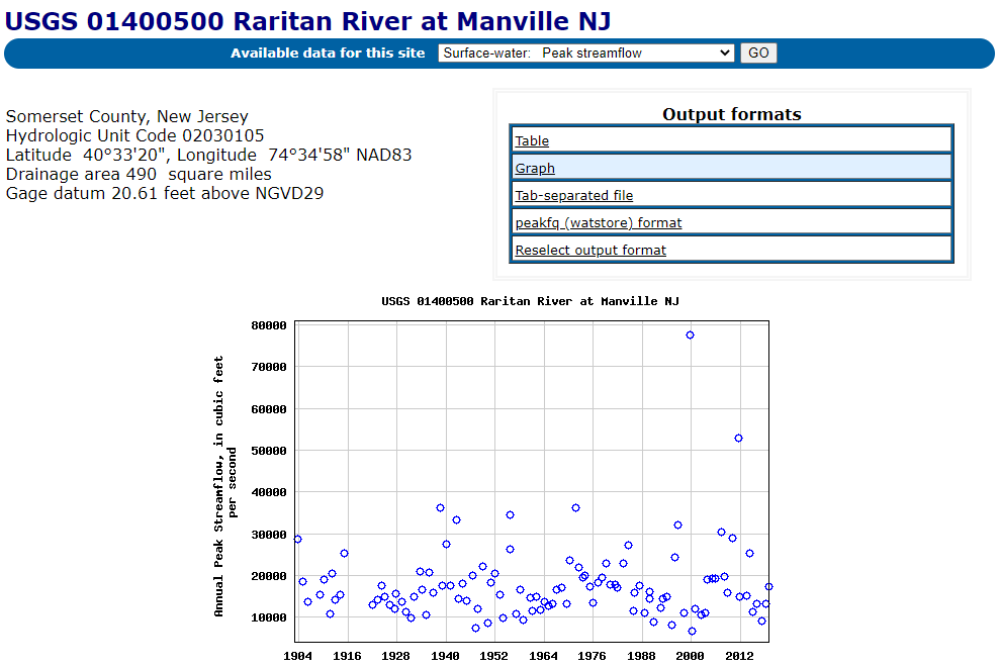


Figure 2.3. Annual peak streamflows at U.S. Geological Survey streamgage 01400500 Raritan River at Manville, New Jersey, used by Asquith et al. (2017) in study of several parameter estimation methods and distributions for modeling flood frequency within which much background of peak streamflow data in the USGS National Water Information System (NWIS) (U.S. Geological Survey, 2018) is also provided.

used to extrapolate and estimate streamflows for low annual exceedance probability (rare events, large magnitudes) is contained in the upper right-hand corner indicated by the circle on the figure. The lower left-hand corner of the figure contains low flow information. In the figure, both these regions of the cumulative distribution are in locations with substantial curvature, which itself presents much of the analytical challenge. The central portion contains information about the mean annual streamflow (central location on the real-number line).

Asquith et al. (2017) studied shape and distributions going deep into the distribution tail (the upper right hand corner of fig. 2.4). This low annual exceedance probability, high magnitude portion lies beyond the typical design requirements (in probability space), but is of immense policy importance for extremely rare events of large magnitude and destructive capability. Asquith et al. used data from two selected USGS streamgages, and invented a framework to quantify uncertainty in analyses of peak-streamflow frequency attributable to two sources: (1) the choice of the distribution to which the frequency curve is fit and (2) sampling error with respect to a chosen quantile and distribution. Emphasis was placed on the distribution choice uncertainty as a means to express or further explain however much uncertainty in extreme flood quantiles exists. The distribution choice is important because its

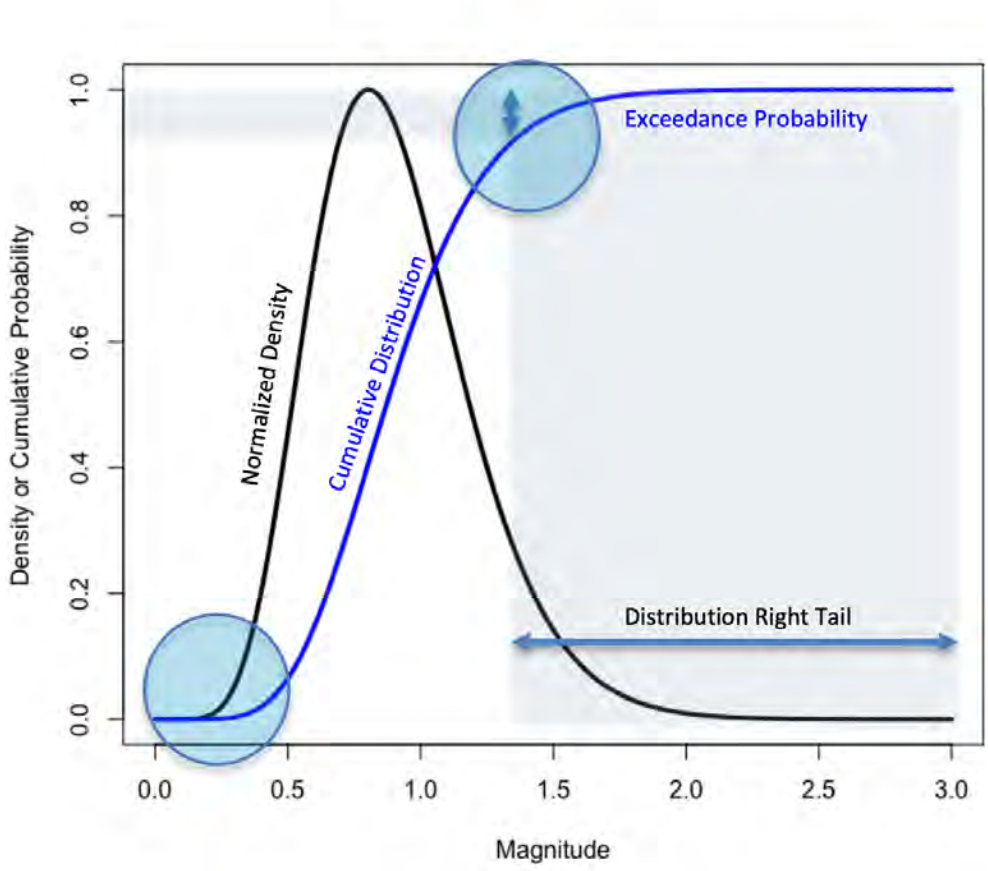


Figure 2.4. Representative probability density function and cumulative distribution function schematically derived from a time series of annual peak streamflows with a high light as to the relative position and mean of the rail tail (upper tail) of the distribution.

shape and behavior in the curved portions of the plot determines the extrapolated behavior value (yet one has only 7 percent of whatever sample population to assess the "fit" in this region).

Asquith et al. (2017) performed parameter estimation using product moments (EMA if historical information is present) by USGS PeakFQ software 7.2 (U.S. Geological Survey, 2020), and L-moments, resulting in frequency curves for the distributions. The effort considered alternative and complementary methods to fit probability distributions include product moments, maximum likelihood (MLE), expected moments algorithm (EMA), L-moments, and maximum product of spacings (MPS).

Asquith et al. (2017) concluded that estimation of at-site peak-streamflow frequency for very low annual exceedance probability is a difficult problem; in part because event rarity may be so extreme that the actual presence of such events in observational datasets will be extremely unusual. They further conclude that distribution choice uncertainty is considerably larger than sampling uncertainty for very low annual exceedance probabilities. However, the choices of the practically useful distributions within EMA, USGS PeakFQ software produce useful and comparable results at typical design probabilities of interest to Departments of Transportation.

2.3. Regionalization Methods

The third moment of the log-Pearson type III distribution (or other appropriate distribution) is the skewness coefficient, which is sensitive to extreme events, such as large floods. Thus, an accurate estimate of the skewness coefficient is important in flood frequency analysis because the majority of the interest is focused on the large flood events. However, short record lengths at gaged sites make a regional estimate of skew valuable in determining flood frequency estimates.

Historically, regional regression analysis used an OLS framework that considers the residual errors to be homoscedastic and independently distributed. However, the estimates of the variable of interest at different gaged sites have different precision because of differences in record length and possible differences in the precision of measurements and their variability Tasker and Stedinger (1989). Tasker and Stedinger (1986) and Tasker and Stedinger (1989) developed a Generalized Least Squares (GLS) framework, which considers both differences in record lengths and precisions, as well as cross-correlation among station estimators because station estimators are generally correlated.

Asquith, and Thompson (2008) and Asquith and Roussel (2009) developed regression equations for Texas borrowing these regionalization techniques and building an extension using a non-linear residuals minimization technique (PRESS) that incorporated locational, and geomorphic characteristics (expressed as a single geospatial residual correction parameter referred to as "OmegaEM"). Asquith et al. (2013) used regionalization techniques and generalized additive modeling (a type of regression technique) to develop from an extensive streamflow database to produce predictive models of important hydraulic variables for Texas

Veilleux et al. (2011) developed statistical techniques for estimating regional skewness coefficients for flood frequency analysis in the United States using a Bayesian generalized least squares (B-GLS) regression framework. In that framework, a normalized distance was used to determine the likelihood that two drainage watersheds are nested, whereas the drainage-area ratio method is used to determine if two nested watersheds are sufficiently similar in size that they are essentially or are at least in large part the same watershed for the purposes

of developing a regional hydrologic model. Asquith, Roussel, and Vrabel (2006) describe applications of the drainage-area ratio method based on Texas hydrology.

2.4. Chapter Conclusions

EMA-MGB is a clearly suitable successor to traditional B17B-GB methods and is to be employed (within B17C using PeakFQ) for outlier handling in the present research. An analytical framework (within B17C) allows for comparison and selection of distributional choice and should be leveraged for the present research to select an appropriate distributional model for Texas. Recent work using independent tools allows for a measure of quality control/quality assurances that did not exist two decades prior.

Recent regionalization tools will support the development of a new skewness coefficient for Texas, and longstanding concerns of "out of the box" B17B procedures (too few low outliers identified as a rule) for hydrology in the study area should be mitigated with presence of MGBT and the revised skewness coefficient. Bulletin 17C, an updated tool that incorporates methods (including those listed above) developed in the three decades since the Bulletin 17B and is the primary tool to be used in the present research. Data from otherwise regulated effected streamgages appears to be extractable and should further extend the capabilities of the B17C results and produce longer effective record lengths, that in-turn, lead to improved estimates of flood-frequency curves.

Chapter References

Aryal, Y.N., Villarini, G., Zhang, W., and Vecchi, G.A., 2018, Long term changes in flooding and heavy rainfall associated with North Atlantic tropical cyclones—Roles of the North Atlantic Oscillation and El Niño–Southern Oscillation: Journal of Hydrology, v. 559 pp. 698–710, https://doi.org/10.1016/j.jhydrol.2018.02.072.

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., https://doi.org/10.3133/wri014243.

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., https://doi.org/10.3133/wri964307.

Asquith, W.H., Herrmann, G.R., and Cleveland, T.G., 2013, Generalized additive regression models of discharge and mean velocity generally associated with direct-runoff conditions in Texas—The utility of the U.S. Geological Survey discharge measurement database: American

Society of Civil Engineers, Journal of Hydrologic Engineering, v. 18, no. 10, pp. 1331–1348, https://doi.org/10.1061/(ASCE)HE.1943-5584.0000635.

Asquith, W.H., Kiang, J.E., and Cohn, T.A., 2017, Application of at-site peak-streamflow frequency analyses for very low annual exceedance probabilities: U.S. Geological Survey Scientific Investigation Report 2017–5038, 93 p., https://doi.org/10.3133/sir20175038.

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. https://doi.org/10.3133/sir20095087.

Asquith, W.H., Roussel, M.C., and Vrabel, J., 2006, Statewide analysis of the drainage-area ratio method for 34 streamflow percentile ranges in Texas: U.S. Geological Survey Scientific Investigations Report 2006–5286, 34 p., https://doi.org/10.3133/sir20065286.

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., https://doi.org/10.3133/sir20085084.

Benson, M.A., 1962, Factors influencing the occurrence of floods in a humid region of diverse terrain: U.S. Geological Survey Water-Supply Paper 1580–B, 64 p., https://doi.org/10.3133/wsp1580B.

Benson, M.A., 1964, Factors influencing the occurrence of floods in the southwest: U.S. Geological Survey Water-Supply Paper 1580–D, 72 p., https://doi.org/10.3133/wsp1580D.

Cleveland, T.G., and Fang, Z.N., 2021, Texas-Skew-Update-2021: Texas Data Repository, https://doi.org/10.18738/T8/SVLCOQ.

Cohn, T.A., England, J.F., Berenbrock, C.E., Mason, R.R., Stedinger, J.R., and Lamontagne, J.R., 2013, A generalized Grubbs–Beck test statistic for detecting multiple potentially influential low outliers in flood series: Water Resources Research, v. 49, no. 8, pp. 5047–5058, https://doi.org/10.1002/wrcr.20392.

Cohn, T.A., Lane, W.L., and Baier, W.G., 1997, An algorithm for computing moments based flood quantile estimates when historical flood information is available: Water Resources Research, v. 33, no. 9, pp. 2089–2096, https://doi.org/10.1029/97WR01640.

England, J.F., Cohn, T.A., Faber, B.A., Stedinger, J.R., Thomas Jr., W.O., Veilleux, A.G., Kiang, J.E., and Mason, R.R., 2018, Guidelines for determining flood flow frequency Bulletin 17C: U.S. Geological Survey Techniques and Methods, book 4, chap. 5B, 148 p., https://doi.org/10.3133/tm4B5.

Griffis, V.W., and Stedinger, J.R., 2009, Log-Pearson type 3 distribution and its application in flood frequency analysis, III—Sample skew and weighted skew estimators: Journal of Hydrologic Engineering, v. 14, no. 2, pp. 121–120, https://doi.org/10.1061/(ASCE)1084-0699(2009)14:2(121).

Grubbs, F.E., and Beck, G., 1972, Extension of sample sizes and percentage points for significance tests of outlying observations: Technometrics, v. 14, no. 4, pp. 847–854, https://doi.org/10.2307/1267134.

Gruber, A.M., Dirceu, S.R., Jr., and Stedinger, J.R., 2007, Models of regional skew based on Bayesian GLS regression: Paper 40927-3285, World Environmental and Water Resources Conference—Restoring our Natural Habitat, K.C. Kabbes ed., Tampa, Florida, May 15–18.

Gruber, A.M., and Stedinger, J.R., 2008, Models of LP3 regional skew, data selection and Bayesian GLS regression: Paper 596, World Environmental and Water Resources Congress—Ahupua'a, Babcock, R.W. and R. Watson eds., Honolulu, Hawaii, May 12–16.

Interagency Advisory Committee on Water Data (IACWD), 1982, Guidelines for determining flood flow frequency: Bulletin 17B, 28 p., Hydrology Subcommittee, Washington, D.C.

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., https://doi.org/10.3133/wri964117.

Paretti, N.V., Kennedy, J.R., and Cohn, T.A., 2014, Evaluation of the expected moments algorithm and a multiple low-outlier test for flood frequency analysis at streamgaging stations in Arizona: U.S. Geological Survey Scientific Investigations Report 2014–5026, 61 p., https://doi.org/10.3133/sir2014502.

Pettitt, A.N., 1978, A non-parametric approach to the change-point problem: Applied Statistics, v. 28, no. 2, pp. 126–135, https://doi.org/10.2307/2346729.

Reis, D.S., Stedinger, J.R., and Martins, E.S., 2005, Bayesian generalized least squares regression with application to log Pearson type 3 regional skew estimation: Water Resources Research, v. 41, W10419, https://doi.org/10.1029/2004WR003445.

St. George, S., and Mudelsee M., 2019, The weight of the flood-of-record in flood frequency analysis. Journal of Flood Risk Management, v. 12, suppl. 1:e12512, 8 p., https://doi.org/10.1111/jfr3.12512.

Stedinger, J.R. and Griffis, V.W., 2008, Flood frequency analysis in the United States—Time to Update (editorial): Journal of Hydrologic Engineering, April, pp. 199–204, https://doi.org/10.1061/(ASCE)1084-0699(2008)13:4(199).

Tasker, G.D., and Stedinger, J.R., 1986, Regional skew with weighted LS regression: Journal of Water Resources Planning and Management, v. 112, no. 2, pp. 225–237, https://doi.org/10.1061/(ASCE)0733-9496(1986)112:2(225).

Tasker, G.D., and Stedinger, J.R., 1989, An operational GLS model for hydrologic regression: Journal of Hydrology, v. 111, no. 1–4, pp. 361–375, https://doi.org/10.1016/0022-1694(89)90268-0.

U.S. Geological Survey, 2018, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed April 9, 2018, at https://doi.org/10.5066/F7P55KJN.

U.S. Geological Survey (USGS), 2020, PeakFQ—Flood frequency analysis based on Bulletin 17B and recommendations of the Advisory Committee on Water Information (ACWI) Subcommittee on Hydrology (SOH) Hydrologic Frequency Analysis Work Group (HFAWG), version 7.2, accessed January 29, 2019, at https://water.usgs.gov/software/PeakFQ/, version 7.3, accessed February 1, 2020, at https://water.usgs.gov/software/PeakFQ/.

Veilleux, A.G., 2011, Bayesian GLS regression for regionalization of hydrologic statistics, floods, and Bulletin 17 skew: Ph.D. dissertation, Cornell University.

Veilleux, A.G., Cohn, T.A., Flynn, K.M., Mason, R.R., Jr., and Hummel, P.R., 2014, Estimating magnitude and frequency of floods using the PeakFQ 7.0 program: U.S. Geological Survey Fact Sheet 2013–3108, 2 p., https://doi.org/10.3133/fs20133108.

Veilleux, A.G., Stedinger, J.R., and Eash, D.A., 2012, Bayesian WLS/GLS regression for regional skewness analysis for regions with large crest stage gage networks, *in* World Environmental and Water Resources Congress 2012: Crossing Boundaries, pp. 2253–2263.

Veilleux, A.G., Stedinger, J.R. and Lamontagne, J.R., 2011, Bayesian WLS/GLS regression for regional skewness analysis for regions with large cross-correlations among flood flows: Paper 3103, World Environmental and Water Resources Congress 2011 *in* Bearing Knowledge for Sustainability, American Society of Civil Engineers, Palm Springs California, May 22–26.

3. Foundational Data Assembly

This chapter summarizes contents of interim reports TM3A, TM3B, TM4A, and TM4B. The interim reports are archived in Cleveland and Fang (2021) located in the directory `0-6177-dataverse-archive/technical_memos/`. File paths in those documents, then contemporaneously sent to TxDOT, refer to contents of a repository within the physical Texas Tech University data center, which was decommissioned by the lead author on August 31, 2021.

TM3 reports details the foundational data assembly for streamflow and watershed properties and TM4 reports the preparatory data processing applied to streamflow records, leveraged with associated watershed properties, to produce the derivative databases used for skew regionalization. The results of these two activities comprise the foundational data set and subsequent analysis for method development.

3.1. Synopsis of TM3A

TM3A presents the process used to identify suitable U.S. Geological Survey (USGS) streamgages in Texas and neighboring states and prepare datasets for further analysis (as well as archival tables). Research into extending the utility of streamgages, or more specifically, their annual peak streamflows, that are coded as regulated is included. The scope included:

1. Research into suitability of streamgages in the research area, Texas and proximal to borders, Oklahoma and New Mexico east of the Great Continental Divide. Streamgages in association with special circumstances, reservoir side weirs, canals, and springs are identified and removed from consideration. Remaining streamgages with at least 6 years of record are retained for further consideration. A list of USGS streamgage identification numbers and summaries of periods of record was produced; and

   • The relevant data files from this step are at the Texas Digital Library dataverse repository Cleveland and Fang (2021) at: `0-6177-dataverse-archive/final_report/src/GenSkew/sitetable/GenSkewMasterSiteList.txt` and discussed later in chapter 4.

2. The annual peak streamflow data for the streamgages are segregated into undeveloped (urban watersheds are given a code "C" in USGS databases) and unregulated (regulated watersheds are given a code "6" in USGS databases). Wagner et al. (2017) provide background on the USGS peak streamflow database. This segregation produced two collections of data: one in the "WATSTORE" format for USGS PeakFQ software 7.3 (U.S. Geological Survey, 2020), another in ASCII format and simpler tab-delimited peak streamflows. These formats are both ASCII (plain-text) but organized differently. Special manual adjustments were performed to discharge qualification codes to ensure that each streamgage has at least two years of non-code 6 or C (Wagner et al., 2017) regardless of the original data because the USGS PeakFQ software 7.3 (U.S. Geological Survey, 2020) requires at least two data points to generate an output screen, otherwise the program exits with an error code.

   • The relevant data files from this step are also in Cleveland and Fang (2021) located in the directory: `0-6177-dataverse-archive/final_report/src/GenSkew/data/pkfq/...` and therein are individual directories titled in the pattern `.../07148350d/` for the first streamgage with the identification number 07148350 and so on for the remaining 444 steamgages ultimately retained for constructing the regional skew estimates discussed in chapter 4.

3.2. Synopsis of TM3B

TM3B presents the watershed properties assembled to support subsequent regionalization for the Texas generalized skew update. The scope included:

1. Determine relevant watershed metrics such as contributing drainage area, main-channel slope, 10-85 percent (between 10 to 85 percent of the channel length) slope, shape (as in the definition of main-channel length squared divided by area), mean annual precipitation, centroid (of watershed) location, and solar radiation as explanatory variables for regionalization of skew and other distribution shapes for Texas. Most of these properties already exist in some content in the USGS National Water Information System (NWIS) (U.S. Geological Survey, 2018) of the related streamgages and the effort was to extract these values and organize them is a fashion specifically useful to the present project;

2. Aggregate non-sensitive dam (reservoir) information related to peak streamflow from the United States Army Corps of Engineers (USACE) National Inventory of Dams (NID) (U.S. Army Corps of Engineers, 2020). The relations are based on hydraulic and hydrologic principles and specifically deal with storage information, pool elevations, and release behavior. Stemming from Research Project 0–6977, the USGS has

published software to process the NID for purposes of statistical analyses of USGS annual peak streamflow (Asquith et al., 2021);

3. Derived properties including the functional drainage area (FA) and functional distance (FD) for streamgages identified as containing regulatory structure (for example, the intersection of streamgage drainage area polygons and reservoir locations);

4. Relevant watershed hydrologic properties classified as morphometric, hydrologic (other than the common metrics), pedologic (soils)/geologic/land use, and climatic.

   • The relevant data files for steps 1, 3, and 4 above are containted in Cleveland and Fang (2021) located at: `0-6177-dataverse-archive/data/1703gages/`; for step 2 the relevant data are located in the directory: `0-6177-dataverse-archive/data/scNIDaregis_1703/rawdata/`.[1]

3.3. Synopsis of TM4A

TM4A presents results of trend analyses performed on the streamflow databases built during the data assembly steps of the previous chapter. The research scope included:

1. Perform a simple trend analysis to produce a table of streamgage identification number, total years of record, record range, Kendall's Tau and attained significance (p-value), and potentially other salient trend testing results, such as change point analyses (PELT or Pettit test), a temporal integration of the construction of dams in the NID (Asquith et al., 2021): streamgage, water year, and cumulative reservoir storage details;

   • The results for this step were reported in TM4A and incorporated into other databases used herein. A copy is archived in Cleveland and Fang (2021) located at `0-6177-dataverse-archive/data/various_interim_results/gagesIIIrecordSummary_alltail.txt`;[2]

2. Build and deploy a multiple Grubbs–Beck Test (MGBT) software for outlier identification software tool for independent, meaning separate from Bulletin 17C, code base as implemented in USGS PeakFQ software 7.3 (U.S. Geological Survey, 2020), for

[1] The enclosing directory `scNIDaregis_1703/` contains the scripts to process the data files and is a copy of the v1.0.0 release of Asquith et al. (2021), which is the canonical homepage of that USGS released software.

[2] File is included in Cleveland and Fang (2021) for completeness, its contents are not used as-is, but instead were incorporated into other database files used in subsequent processing; the `README.md` file in the enclosing directory has an explanation of the purpose of file retention.

comparative estimation of low-outlier thresholds. A software package (Asquith, England, and Herrmann, 2020) was developed and approved by the USGS in September 2019 and was used in evaluating streamgage data as a prelude to skew regionalization to form a generalized skew estimate; and

3. Process annual peak streamflow of site-specific skew for the 444 streamgages using B17C-EMA+MGBT methodology in the PeakFQ software 7.3 (U.S. Geological Survey, 2020) to produce the primary generalized skew estimates from data in the NWIS at the time of processing (peak streamflows up to water year 2017).[3] Results germane to this step are major components of chapters 4 and 5, and discussion is deferred to those chapters. Listing this step here is to set some context to workflow chronology.

3.4. Synopsis of TM4B

TM4B discussed the application of Asquith, England, and Herrmann (2020) to the 444 long-record streamgages that formed the basis for the regional skew study as well as exploratory investigation into climate state classification (wet or dry) starting from the 1,703 master streamgage list for this study. The research scope included:

1. Coordinate with contemporaneous studies by USGS New Mexico and Oklahoma for those respective state-based Departments of Transportation that are either ongoing or contemporaneous; the coordination provided a quality assessment relative to adjacent states and these are reported in chapter 4;

2. Explore use of a wet or dry climate classification state with such a binary classification based on monthly climate indices aligning with the months of the annual peak streamflows that approximately bifurcate each streamgage record into halves. This effort is reported herein as chapter 6; and

3. Use MGBT estimates from Asquith, England, and Herrmann (2020) to process annual peak streamflow of site-specific skew for the systematic record of the 1,703 streamgages to generate skew and higher moment (shape) estimates using various methods and skew estimation procedures. The results of this step help inform the generalized additive modeling effort reported in chapter 4.

[3] There is a couple of year gap from 2017 to about the time PeakFQ analyses were started for this project because peak streamflows for 2018 and 2019 where not accessible from public USGS data interfaces because the USGS was is transition from one enterprise software system supporting the National Water Information System to another. Internal to the USGS, the so-called `PKENTRY` program (entry of peak streamflows into the database) was disabled.

3.5. Persistent Archives of Applicable Data

The watershed and ancillary property databases compiled for TM3A and TM3B were synthesized into the USGS data release by Yesildirek et al. (2021). Figure 3.1 is a screen capture of the public facing web-landing page (`https://doi.org/10.5066/P9A91W4Z`) for the data release providing a persistent (long-term support) location for the data that was used in the research describe herein.

The reservoir storage data compiled for TM3A and TM3B were synthesized into the USGS software release by Asquith et al. (2021) as shown in figure 3.2. This software was used to support experimental research reported in chapter 8. As always and forever, the peak streamflow data are acquirable from NWIS (or future derivatives) (U.S. Geological Survey, 2018).

For purposes of archival data immediately associated with the 0–6977 Research Project, the aforementioned data are placed into the project archive in Cleveland and Fang (2021) along the file path: `0-6177-dataverse-archive/data/peaks_props_NID_1703.feather.zip`. The file contains records for 1,703 USGS streamgages with 185 columns of information and the number of peak streamflows represented equals 59,663 through about the 2020 water year. This datafile was used in experimental research described in chapters 7 and 8 in this report.

It is useful to conclude this chapter through a depiction of the relative frequency of peak streamflows by water year, which is shown in figure 3.3. The figure caption credits TxDOT for responsibility for part of the streamgage increase beginning about the 2006 water year with a small-watershed streamgaging program with the USGS (Asquith and others, 2018; Asquith and Harwell, 2018; Harwell and Asquith, 2011).

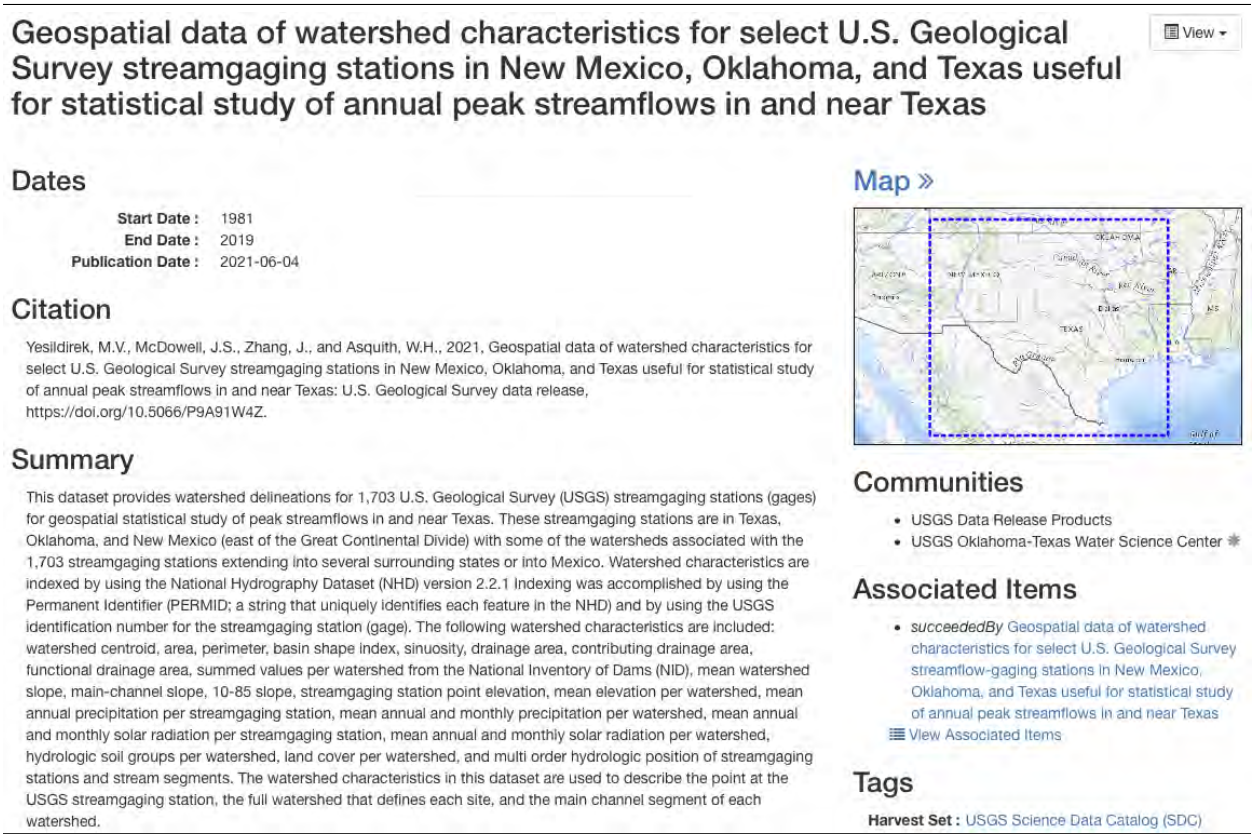


Figure 3.1. Web-landing page screenshot for U.S. Geological Survey (USGS) data release by Yesildirek et al. (2021) comprehensively listing watershed properties and ancillary data for the 1,703 USGS streamgages of this study. There exists possibility to augment this master list of streamgages for study of Texas flood hydrology with more USGS streamgages as future USGS streamgages come online or other agency gages are incorporated and stakeholder interest exists. The title of the data release was chosen to not state a date range, which opens the possibility for expansion as TxDOT is currently (2021) sponsoring unrelated research activities, to the project that this report represents, involving additional and now (summer 2021) operational USGS streamgages specific to TxDOT needs that are not in the 1,703 streamgage count.

Figure 3.2. Web-landing page screenshot for U.S. Geological Survey (USGS) software release by Asquith et al. (2021) that facilitates temporal integration (accumulation) of the U.S. Army Corps of Engineers National Inventory of Dams (NID) for arbitrary watershed polygons and optional binding of year-by-year cumulative reservoir storages to USGS annual peak streamflow data from the USGS National Water Information System (U.S. Geological Survey, 2018). The software contains state-based, text-file copies of the NID with minor preparation to be used by the software. Updating the NID components therein or software revision as needed is possible.

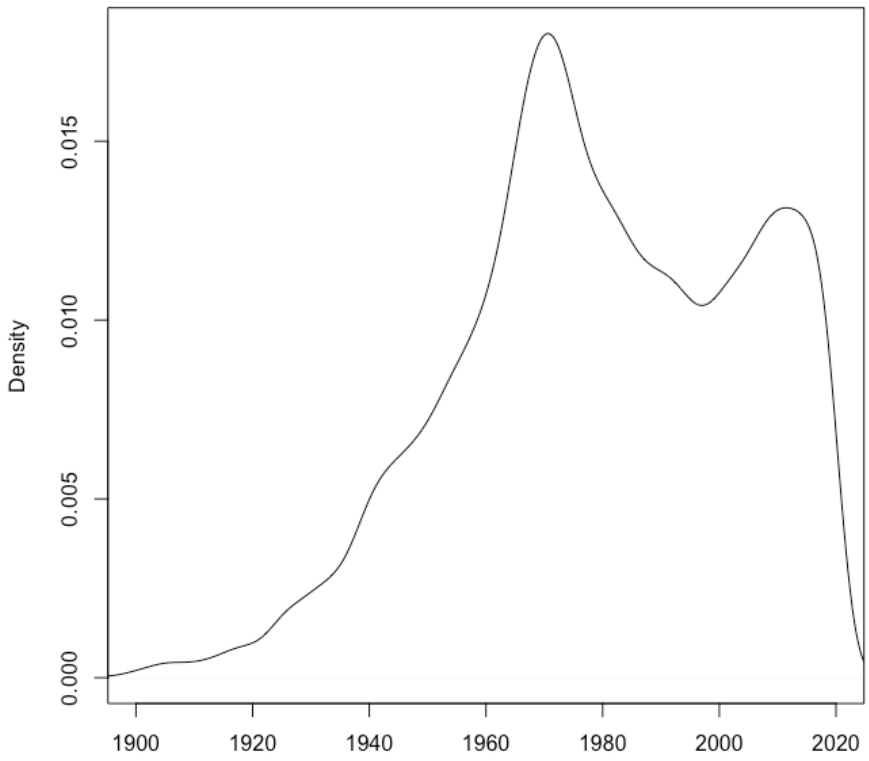


Figure 3.3. Density estimation showing relative frequency of annual peak streamflows for U.S. Geological Survey streamgages by water year represented in data archived in Cleveland and Fang (2021). This includes Texas, Oklahoma, and eastern New Mexico. The long-term persistent decline in number of peak streamflows from about water year 1970 to about water year 2000 represents a decline in the number of operational USGS streamgages. The decline was reversed at the beginning of the 21st century with increased stakeholder sponsorship of streamgages along with about 50 streamgages sponsored by the Texas Department of Transportation starting in about 2006 (Asquith and others, 2018; Asquith and Harwell, 2018; Harwell and Asquith, 2011) and that are still operational as of 2021. The density is only an approximation and the rapid fall off about 2020 is synthetic (artificial) by the mathematics of the curve, which means it is not presentative of the exact count of USGS streamgages in operation.

Chapter References

Asquith, W.H., England, J.F., and Herrmann, G.R., 2020, *MGBT*—Multiple Grubbs-Beck low-outlier test: U.S. Geological Survey software release, R package, Reston, Va., accessed July 27, 2020, at https://doi.org/10.5066/P9CW9EF0.

Asquith, W.H., Cleveland, T.G., Yesildirek, M.V., Zhang, J., Fang, Z.N., and Otto, L.D., 2021, *scNIDaregis*—Geospatial processing of dams in the United States from the National Inventory of Dams with a state-level aggregation scheme, demonstrated for selected dams in eight states in south-central region of the United States, and post-processing features for basin-specific tabulation: U.S. Geological Survey software release, Reston, Va., https://doi.org/10.5066/P90NJVB9.

Asquith, W.H., and Harwell, G.R., 2018, Database of peak streamflow derived from interpretations of indirect measurements for a crest-stage gage network in Texas through water year 2015: U.S. Geological Survey data release, https://doi.org/10.5066/F7057D39.

Asquith, W.H., Harwell, G.R., and Winters, K.E., 2018, Annual and approximately quarterly series peak streamflow derived from interpretations of indirect measurements for a crest-stage gage network in Texas through water year 2015: U.S. Geological Survey Scientific Investigations Report 2018–5107, 24 p., https://doi.org/10.3133/sir20185107.

Cleveland, T.G., and Fang, Z.N., 2021, Texas-Skew-Update-2021: Texas Data Repository, https://doi.org/10.18738/T8/SVLCOQ.

Harwell, G.R., and Asquith, W.H., 2011, Annual peak streamflow and ancillary data for small watersheds in central and western Texas: U.S. Geological Survey Fact Sheet 2011–3082, 4 p., https://pubs.usgs.gov/fs/2011/3082/.

U.S. Army Corps of Engineers, 2020, CorpsMap—National Inventory of Dams, accessed on July 7, 2020 at https://nid.sec.usace.army.mil/ords/f?p=105:1::::::.

U.S. Geological Survey, 2018, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed April 9, 2018, at https://doi.org/10.5066/F7P55KJN.

U.S. Geological Survey (USGS), 2020, PeakFQ—Flood frequency analysis based on Bulletin 17B and recommendations of the Advisory Committee on Water Information (ACWI) Subcommittee on Hydrology (SOH) Hydrologic Frequency Analysis Work Group (HFAWG), version 7.2, accessed January 29, 2019, at https://water.usgs.gov/software/PeakFQ/, version 7.3, accessed February 1, 2020, at https://water.usgs.gov/software/PeakFQ/.

Wagner, D.M., Kiang, J.E., and Asquith, W.H., 2017, U.S. Geological Survey streamgaging methods and annual peak streamflow, appendix 1 *of* Asquith, W.H., Kiang, J.E., and Cohn, T.A., 2017, Application of at-site peak-streamflow frequency analyses for very low annual exceedance probabilities: U.S. Geological Survey Scientific Investigations Report 2017–5038, 93 p., https://doi.org/10.3133/sir20175038.

Yesildirek, M.V., McDowell, J.S., Zhang, J., and Asquith, W.H., 2021, Geospatial data of watershed characteristics for select U.S. Geological Survey streamgaging stations in New Mexico, Oklahoma, and Texas useful for statistical study of annual peak streamflows in and near Texas: U.S. Geological Survey data release, https://doi.org/10.5066/P9A91W4Z.

4. Technique to Estimate Generalized Skew Coefficients of Annual Peak Streamflow for Natural Watershed Conditions in Texas, Oklahoma, and Eastern New Mexico

Suggested Citation:

Asquith, W.H., Yesildirek, M.V., Landers, R.N., Cleveland, T.G., Fang, Z.N., Zhang, J., 2021, Technique to estimate generalized skew coefficients of annual peak streamflow for natural watershed conditions in Texas, Oklahoma, and eastern New Mexico, *in* Cleveland, T.G., and Fang, Z.N., 2021, Generalized skew update and regional study of distribution shape for Texas flood frequency analyses: Texas Department of Transportation Research Report 0–6977–1, chap. 4, pp. 31–58.

Abstract

Reliable information about the frequency of annual peak streamflow is needed for floodplain management, objective assessment of flood risk, and cost-effective design of dams, levees, other flood-control structures, and roads, bridges, and culverts. Generalized skew coefficients are among the data needed for log-Pearson type III peak-streamflow frequency analyses of annual peak streamflows. A technique is presented to estimate generalized skew coefficients used for log-Pearson type III peak-streamflow frequency analyses of annual peak streamflow from natural watersheds (minimal regulation and minimal impervious cover). The estimation of generalized skew coefficients was based on annual and historical peak streamflow data from an initial set of 444 selected USGS streamgaging stations (streamgages) with at least 30 years of recorded annual peak streamflows from natural watersheds in Texas, Oklahoma, and the part of New Mexico east of the Great Continental Divide. The primary focus was to obtain information that could be used to update previously published generalized skew coefficients in Texas.

Of the 444 candidate streamgages, 341 were used in the final construction of statistical models. Two generalized additive models (GAMs) were used to predict generalized skew based on a 2-dimensional smooth on projected Albers equal area coordinates of either (1) the locations of the centroids of the gaged watersheds or (2) the streamgage locations. To create maps of generalized skew coefficients, predictions were made on a 1-kilometer grid and contour lines were superimposed. The centroid-location map, with a mean-squared error (MSE) of 0.216, is preferred. Generalized skew coefficients from the centroid-location map, along with the MSE, are useful for computing weighted-skew values when conducting frequency analyses of annual peak streamflow following the guidelines set forth in Bulletin 17C. Based on the results of the study, text revision of the TxDOT Hydraulic Design Manual could be made.

4.1. Introduction

Reliable information about the frequency of annual peak streamflow is needed for flood-plain management, objective assessment of flood risk, and cost-effective design of dams, levees, other flood-control structures, and roads, bridges, and culverts. In 2019, the U.S. Geological Survey (USGS), in cooperation with the Texas Department of Transportation (TxDOT) and research colleagues at Texas Tech University and the University of Texas at Arlington, began a 3-year investigation into the statistical properties of floods in Texas; Oklahoma and the part of New Mexico east of the Great Continental Divide were included in the analysis because watersheds important for the analysis of generalized skew coefficients in Texas cross state boundaries. This chapter presents the specific results of the study, the goals of which were to update the 1996 generalized skew coefficients (generalized skews) in Texas (Judd et al., 1996) and the current (September 2019) TxDOT Hydraulic Design Manual (Texas Department of Transportation, 2020). An update is deemed useful because 25 years of additional data collection has occurred, and Federal guidance and governing mathematical steps on flood-frequency analyses have been updated (England et al., 2018). Also, the inclusion of annual peak-flow records from eastern New Mexico and all of Oklahoma yielded a more rectangular-like study area relative to that used in the 1996 analysis (Judd et al., 1996). All germane data and data-processing workflows are documented in Cleveland and Fang (2021) and Yesildirek et al. (2021). A top-level directory `0-6177-dataverse-archive/final_report/src/GenSkew/` in Cleveland and Fang (2021) contains the archival material supporting this chapter.

4.1.1. Physical Setting

The climate and physiography of Texas vary considerably across the State (Carr, 1967). Accordingly, climatic and physiographic factors typically cause the annual peak streamflows at individual streamgages to be generally non-lognormally distributed and range by as much as five orders of magnitude. The non-lognormality and extreme range in peak streamflow

values make it difficult to estimate the frequency of annual peak streamflow (Asquith et al., 2017). Climatic variability in Texas contributes substantially to the non-lognormality and extreme range of annual peak streamflow at streamgages in the State. Many near world-record precipitation events have occurred in Texas (Asquith, 1998; Watson et al., 2018). Long-term droughts that result in small annual peak streamflows referred to as low floods (low outliers) (England et al., 2018) at streamgages also can occur statewide (Winters, 2013).

Much of the western one-half of the State contains alluvial basins, where high rates of evapotranspiration can cause substantial reduction of the smaller, more frequent annual peak streamflows than substantial floods; thus, the observed range in annual peak streamflow can be large at many streamgages. In contrast, many streams in the eastern half of the State gain streamflow from shallow groundwater. Additionally, in south-central Texas, the range in observed annual peak streamflow for streamgages can be large because of the loss of streamflow into fractured limestone bedrock during drought and because of extraordinarily large streamflows resulting from periods of abundant precipitation in conjunction with runoff from thin soils and steep slopes of the surrounding terrain (Asquith et al., 1995; O'Connor and Costa, 2018).

4.1.2. Importance of Generalized Skew Coefficients

Generalized skew coefficients are an important component of peak-streamflow frequency analyses of observed annual peak streamflows. The skew coefficient of the observed annual peak streamflows from a streamgage (station skew) is weighted with generalized skew to yield a more accurate estimate of skew (Interagency Committee on Water Data, 1982; England et al., 2018) and attendant enhancement of peak streamflow reliability.

Bulletin 17C (England et al., 2018) is the most recent update to Federal guidance for frequency analysis of annual peak streamflow and supersedes Bulletin 17B (Interagency Committee on Water Data, 1982). Bulletin 17C recommends frequency analysis of annual peak streamflow using a Pearson type III (PE3) distribution fit to the logarithms of annual peak streamflow at a particular streamgage using parameter estimation. A skew value near zero for this distribution means that generally a lognormal-like distribution form predominates.

The USGS computer program PeakFQ (U.S. Geological Survey, 2020a) often is used to compute frequency analysis of annual peak streamflow and compute station skew for each of the 444 streamgages included in this study. The canonical list of streamgages and basic identification information is presented by Cleveland and Fang (2021) along the file path: `0-6177-dataverse-archive/final_report/src/GenSkew/sitetable/GenSkewMasterSiteList.txt`. The PeakFQ input and output information also is in Cleveland and Fang (2021) along the file path: `0-6177-dataverse-archive/final_report/`

`src/GenSkew/data/` and therein are individual directories titled in the pattern `pkfq/07148350d/` for the first streamgage with the identification number 07148350.

PeakFQ follows the Bulletin 17C guidelines for frequency analysis of annual peak streamflows, including incorporation of the expected moments algorithm (EMA) to compute the PE3 distribution and the multiple Grubbs–Beck test (MGBT) for identifying low outliers referred to as "potentially influential low floods" (PILFs). Both annual and historical peak-streamflow records for each streamgage are used; a historical peak streamflow represents a major flood that occurred before, after, or during a gap in the gaged period of record that can be used to define an extended period during which the largest floods, either recorded or historical, are known.

The station skew influences the shape or curvature of the final PE3 distribution for a frequency analysis. A skew value of zero for a fitted PE3 distribution results in a symmetrical distribution or linear plot on log-probability graphing scales. All other statistical parameters being equal, estimates of annual peak streamflow corresponding to small annual exceedance probabilities (less than 0.10) will be larger for positively skewed distributions and smaller for negatively skewed distributions (Asquith et al., 2017).

Station skew computed for a streamgage with a period of record less than about 30 years tends to be less reliable for use in frequency analysis of annual peak streamflow than station skew computed for a streamgage having a longer period of record (greater than about 30 years). Therefore, Bulletin 17C recommends using a weighted skew in frequency analysis of annual peak streamflow (England et al., 2018, eq. 7.20). The weighted skew is computed by weighting the station skew with a generalized skew representative of the surrounding region, resulting in the weighted skew. The weights are based on the inverse of the respective mean-squared errors (MSEs) of the station skew and generalized skew. More thorough discussion of the weighted skew and its effects on the computations is available in Cohn et al. (2019) and England et al. (2018).

4.1.3. Peak Streamflow Data

This study estimates generalized skews based on station skews computed using annual and historical peak streamflow data from 444 selected USGS streamgages with at least 30 years of gaged records of annual peak streamflow that are also unaffected by regulation or urbanization in Texas, Oklahoma, and the part of New Mexico east of the Great Continental Divide. Streamgages in Oklahoma and part of New Mexico were deemed to provide information useful for determining generalized skews for Texas. It is a critical part of introductory information related to original data sources that this geographic area be mentioned because it pertains to how certain metadata are treated in historically separate databases.

A "natural" watershed is defined as having less than 10 percent of the drainage area controlled by reservoirs (a measure of regulation) (U.S. Army Corps of Engineers, 2020; Asquith et al., 2021) and less than 10-percent impervious cover (a measure of urbanization). This definition is a legacy germane to streamgages in Texas. Conversely, for Oklahoma, Lewis et al. (2019, p. 4) states "substantial regulation is defined as a contributing drainage area where 20 percent or more of the drainage area is upstream from dams and floodwater-retarding structures." Wagner et al. (2017) reviewed "discharge qualification codes" assigned to annual peak streamflows by the USGS in the National Water Information System (NWIS) database (U.S. Geological Survey, 2018). The qualification codes are not always clear on a peak-by-peak basis regarding the usefulness of the respective peak for a particular application.

Regulation or urbanization qualification codes applied to peak streamflow values stored in NWIS require consideration in the determination of station skews and are discussed in brief. The USGS does not provide authoritative metrics for regulation or urbanization in the qualification codes. From information contained within the peak-flow files for streamgages in NWIS, it is not possible to definitively assess whether individual gaged peak streamflows are, in fact, even semi-quantitatively affected by regulation or urbanization. However, a general metric to identify periods of record unaffected by urbanization or regulation is the absence of qualification codes 6 (regulation) and C (urbanization) assigned to annual peak streamflows (Wagner et al., 2017). Qualification code 5 (affected to unknown degree by regulation or diversion) was ignored; peak streamflows were included because the default setting of the USGS PeakFQ software is to treat code 5 as "unregulated."

After retrieval of the streamflow data from NWIS (U.S. Geological Survey, 2018), software (Asquith, England, and Herrmann, 2020) designed to identify systematic records of peak streamflow or gaged annual peak streamflows was used. Systematic record was considered any sequence of water-year records where two or more peaks form a continuous one-year incrementing set. As a result, the entirety of systematic record for a streamgage is composed of one or more of such sets. These data were then reviewed for the presence or absence of qualification codes 6 and C, and some streamgage-by-streamgage judgment was required to identify streamgage-specific systematic record used in PeakFQ analysis for the purposes of this study. Detailed investigation of the effects of regulation on annual peak streamflows is complex, although some insight was gained by the use of large-scale statistical binding of annual peak streamflows to temporal cumulative storage (regulation) (Asquith et al., 2021). The watersheds for some streamgages became regulated and (or) urbanized during their period of record (according to discharge qualification codes 6 and C), and the data for such periods were systematically excluded from this study with few exceptions.

4.1.4. Purpose and Scope

The purpose of this chapter is to present a technique for estimating generalized skews of annual peak streamflow values for Texas as part of a larger study area that also includes Oklahoma and part of New Mexico. Although Oklahoma and the part of New Mexico east of the Great Continental Divide were included in the analysis, the primary focus was to obtain information that could be used to update previously published generalized skews in Texas. Generalized additive modeling was used for the analysis of station skew from long-term USGS streamgages where the streamflow is derived from natural watersheds. Long-term streamgages are defined as those having at least 30 years of annual peak streamflows from natural watersheds. Annual peak data through the 2016 water year were used, if available.

4.2. Methods

The steps used in the computation of station skew and related statistics for use in the determination of generalized skews for Texas were as follows:

1. Retrieve annual peak streamflow data for 444 long-term streamgages in Texas, Oklahoma, and part of New Mexico from NWIS (444 streamgages in the study area that met the 30-year criterion and formed the core dataset for the initial analysis) (U.S. Geological Survey, 2018). Of the 444 streamgages initially considered for use in the study, 341 streamgages were used in the final construction of statistical models. To clarify and be consistent with later statements herein, this 341 streamgage count represents the number of streamgages used in statistical modeling but not necessarily the computation of a specific statistical error that would have implications for the end user of generalized skew values documented herein. The USGS station number, name, and other pertinent information for each streamgage can be found in Cleveland and Fang (2021) and Yesildirek et al. (2021).

2. Plot the annual time series of peak streamflow for each streamgage to find unusual observations that require further investigation. In addition to identifying the PILFs by statistical means, the `MGBT::plotPeaks()` function (Asquith, England, and Herrmann, 2020) enables the visualization of annual peak streamflow and graphical depiction of USGS discharge qualification codes (Wagner et al., 2017).

3. Evaluate if there are statistically significant trends in annual peak streamflows using the Kendall's tau statistic (Helsel et al., 2020), visualization, and regional context. It is sometimes possible to isolate a suitable subperiod within the greater period of record available for purposes of seeking at least 30 years of annual peak streamflow unaffected by trends considered attributable to nonnatural watershed conditions.

4. Set lower and upper bounds of streamflows intervals assigned to missing years of record (data gaps) when possible and set streamflow perception thresholds for historical and gaged period(s) of streamgage operation. Perception thresholds are defined and discussed in detail by England et al. (2018), but in short, perception thresholds define the range of streamflow for which a flood event could have been observed. The inherent assumption and consequence is that any year for which an event was not observed and recorded must have had a peak streamflow outside of (usually below) the perception threshold.

5. Run the EMA/MGBT analyses using version 7.3 of PeakFQ software with the station skew option set. Verify that here have not been major land-use changes (urbanization) altering the watershed from natural conditions. This verification is qualitative and was done for only a few locations by inspection of aerial imagery collected over time because urban watersheds with streamgages are generally self evident; they are within the major metropolitan areas of the study area. The study area is predominantly rural and at the time of data preparation (U.S. Geological Survey, 2018), the 30-year criterion automatically removed the many streamgages used in urban studies of hydrology. The trend detection in step 3 assists in whole-streamgage rejection or isolation of a particular record of 30 or more years in early time. These decisions are augmented by the PeakFQ interface simultaneously plotting the peak streamflow time series. Consider the following example. There are 60 years of record at a given streamgage but only the first 30 years appear representative of natural conditions. Urban expansion into the watershed of the streamgage has not occurred in those first 30 years, and trend analyses of the streamflow record or visual cues in the time series of streamflow indicate the streamflow record in most recent 30 years is different from the first 30 years of record. In this example only the first 30 years of record would be used in this study. Finally, obtain the mean, standard deviation, skew, and MSE of the skew from PeakFQ output. For technical completeness, the MSE is the "EMA ESTIMATE OF MSE OF SKEW WITHOUT REG SKEW" result lines in the PeakFQ output file.

6. Review the peak-streamflow frequency curve to consider if the curve adequately fits the annual peak streamflow data and evaluate low outliers (low peak streamflows associated with drought conditions) using the MGBT function followed by incorporation of a user-specified low-outlier threshold as deemed necessary.

7. Keep a record of streamgages for which valid PE3 analyses could not be computed because of fatal software errors caused by too many censored values or for which the PE3 analyses were otherwise deemed unreliable because of trends in annual peak streamflows or major land-use changes.

8. Assess redundancy (streamgage pairs not collecting independent information from each other) and keep a record of these streamgages for potential removal from the final analysis to facilitate the "best estimate" of uncertainty in generalized skew.

9. Compute generalized skew and its MSE as part of the regional statistical analysis. Steps 1–8 resulted in the removal of certain streamgages from the final statistical analysis. Of the 444 candidate streamgages, 341 were retained.

4.2.1. Individual Streamgage Analyses

The current (2021) Federal guidance for the estimation of peak-streamflow frequency statistics is described in Bulletin 17C (England et al., 2018). Version 7.3 of USGS PeakFQ software was used to perform these computations (U.S. Geological Survey, 2020a).

For purposes of computing station skew, the analyst is required to select the "station skew" option within the PeakFQ software in order to not weight the estimate of station skew with an estimate of generalized skew. The EMA is a method for fitting the PE3 distribution that has been shown effective at incorporating information about historical annual peak streamflows into the frequency analysis. EMA is an improvement over the methodology recommended by Bulletin 17B (Interagency Committee on Water Data, 1982) because it correctly accounts for uncertainty in estimates of the frequency of annual peak streamflow related to uncertainty in the skew as computed during software operation and the shape parameter (the skew coefficient) of the fitted PE3 distribution (Asquith et al., 2017). The uncertainty in the skew of the fitted PE3 distribution is represented by the MSE of that skew, which is related to the variance of the skew estimate and reported in the output files of the PeakFQ software.

The EMA can accommodate interval data, which simplifies analysis of datasets containing nongaged historical data, and PILFs (low outliers) (Asquith et al., 1995; Asquith, England, and Herrmann, 2020; Cohn et al., 2013) and uncertain data points while simultaneously providing enhanced confidence intervals on estimated peak streamflows. The PeakFQ software (Veilleux et al., 2014; U.S. Geological Survey, 2020a) version 7.2 (for this study, spring 2019) and subsequent audit and secondary check using version 7.3 (for this study, spring 2020) was used to compute peak-streamflow frequency for the 444 long-term streamgages previously defined in step 1 of the Methods section (sec. 4.2).

PeakFQ automates many of the procedures for peak-streamflow frequency analysis, including identifying and adjusting for low outliers and historical periods and fitting the PE3 distribution to the logarithms of annual peak streamflow data. The program includes the EMA procedure for frequency analysis and MGBT screening for PILFs. The Bulletin 17C method uses gaged peaks (observed or estimated annual peak streamflows that occurred during the gaged period at a streamgage) and historical peaks (annual peak streamflows

observed outside the gaged period). Flow intervals are used to describe the knowledge of the peak flow in each year, and perception thresholds are used to describe the range of measurable streamflow in each year. When possible, historical peaks are used to define the upper threshold of annual peak streamflow for missing years (years without peak streamflow data) by using perception thresholds to accommodate missing data in either the historical record (for example, miscellaneous records of large floods) or from the gaged record. Adjusting the settings in PeakFQ to skip years of missing record becomes necessary when missing years cannot be adequately canvased with perception thresholds. PeakFQ also executes the Mann–Kendall test for Kendall's Tau and reports the Kendall's tau, p-value, and Sen Slope for the detection of monotonic trends in peak streamflows found in the gaged period of record.

4.2.2. Regional Statistical Methods and Provisional Study

Three methods for the development of generalized skews by regional analysis are suggested by the authors of Bulletin 17C: (1) plot station skews on a map and construct skew isolines, (2) use multiple linear regression techniques to develop a skew-prediction equation relating station skews to selected watershed characteristics, or (3) use the arithmetic mean of station skews from long-term streamgages in the region. Two generalized additive models (GAMs) were used (a centroid-location GAM and a streamgage-location GAM) to explore all three methods within one unified statistical framework. A GAM is a type of regression model that also has features for 2-dimensional smoothing to construct isolines (fig. 4.1); the use of a GAM in a 2-dimensional context is described by Asquith (2020).

Method 1 (maps of skew isolines), method 2 (regression modeling), a hybrid of methods 1 and 2, and method 3 (arithmetic mean) were assessed through GAMs by using the dataset of 341 streamgages. The modeling weights for the response variable (station skew) were the inverses of the MSEs of the station skews, and the explanatory variables for method 2 were watershed characteristics, including drainage area, main-channel slope, watershed-averaged mean annual precipitation, mean land-surface elevation of the watershed, land-surface elevation of the location of the streamgage, mean annual solar radiation at the location of the streamgage, the location of the streamgage, and the location of the watershed centroid (Yesildirek et al., 2021). The GAM framework (fig. 4.1) provides the analyst the ability to assess a hybrid of method 1 (spatial mapping) and method 2 (regression modeling) using the additive 2-dimensional smooth on easting and northing coordinates of the streamgage and watershed centroids. The availability of analyst-led choice of incorporating the locations of the streamgages or the centroids of their watersheds permitted testing different easting and northing coordinates using the smoothing function.

A generalized additive model (GAM) uses relations between a response variable and an additive combination of various parametric terms and smooth terms (smooth functions) (Wood, 2017). A GAM is a type of regression, but the inclusion of smooth functions can be an advantage to GAMs over multi-linear regression because such functions provide linearly additive terms of nonlinear relations in the data. The algorithms of the `mgcv` package (Wood, 2020) were used by default parameters (arguments) of the `gam()` function in the R language (R Core Team, 2020). Furthermore, the additive components can be analyzed simultaneously in two or more dimensions, which is not really a feature of multi-linear regression.

The general form of the GAMs considered for this study is

$$y_i = \mathbf{X}_i\Theta + f(x_i;\Psi) + \cdots + s(E_i, N_i;\eta) + \epsilon_i,$$

where

- y_i is the response variable (station skew) for the ith streamgage record;
- $\mathbf{X}_i$ is a model vector including an optional intercept for strictly parametric and suitably transformed predictor variables,
- Θ is a parameter matrix;
- $f(x_i;\Psi)$ is a smooth function of the jth predictor variable $x_{i,j}$ controlled by smoothing settings Ψ_i;
- $\cdots$ represent additional smooth terms as needed;
- $s(E_i, N_i;\eta)$ is the smooth on the easting (E_i) and northing (N_i) Albers-Equal Area projected coordinates of the respective longitude and latitude and optional smoothing parameters η; and
- ϵ_i are random errors following a declared error distribution (for example, Gaussian).

The $\mathbf{X}_i\Theta$ term is the familiar multi-linear (parametric) regression component of a GAM, and the Θ are regression coefficients (conventional slope terms). Finally, weights for each record can optionally be include, and for this study, the inverses of MSEs of the individual station skews were used.

Figure 4.1. Description of a generalized additive model (GAM) that encompasses schemes evaluated in this study and specific workflow details for the purposes of this study are available in Cleveland and Fang (2021). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

4.3. Generalized Skew Coefficients for Texas, Oklahoma, and eastern New Mexico

In the study area, method 1, a hybrid of method 1 and method 2, and method 2 yielded superior statistical performance to method 3; therefore, method 3 (arithmetic mean) was not further considered. Iterative analysis of generalized skews was accomplished using various combinations of the aforementioned explanatory variables. In general, prediction performance approaching that achieved by the 2-dimensional smooth on spatial positions exists with the supposedly non-spatial variables, such as main-channel slope. It is hypothesized that sufficient cross correlation exists between variables (for example, elevations increasing with distance from the coast and main-channel slopes increasing with mean land-surface elevation) such that it is more straightforward to use the spatial model generated using method 1 than method 2. The predictive performance of the GAM was not sufficiently enhanced by adding explanatory variables, such as drainage area or mean land-surface elevation, to the 2-dimensional smooth on the projected Albers equal area coordinates of either the centroids of the watersheds or streamgages.

The station skews derived from the 341 streamgages that were retained were used to develop the centroid-location GAM and the streamgage-location GAM. The results of these two GAMs are shown in figures 4.2 and 4.3. The GAMs discussed here are deemed more applicable than methods and results available in Judd et al. (1996) because additional annual peak streamflows, updated statistical methods for estimation of peak-streamflow frequency, and GAMs were used. The GAMs were weighted using the inverses of MSEs of the station skews. For each GAM, streamgages that were removed for reasons including insufficient (processable) record or redundancy, which are also within the depicted counties, are shown on the maps.

Rejected watersheds include those with fatal errors in peak-flow frequency analysis occurring within the PeakFQ software (6 streamgages), those with less than 30 years of unregulated or nonurbanized record lacking monotonic trends in annual peak streamflows (24 streamgages), those with possible mislabeling of discharge qualification codes (3 streamgages), and those having a high degree of redundancy (56 streamgages). After removal of these streamgages from the analysis, those having drainage areas less than 1 square mile (7 streamgages) or greater than 35,000 square miles (7 streamgages) were eliminated prior to computing the GAM. This minor trimming of the input data was deemed useful to remove a few small watersheds in New Mexico with a preponderance of censored peak streamflows. A few of the largest riverine systems were removed because they represent major transport corridors through the region and are quite regulated in modern times for which the concept of generalized skew is implicitly less applicable. As a result, the records of annual peak streamflow for $444 - 6 - 24 - 3 - 56 - 7 - 7 = 341$ streamgages were used in the final GAMs and are

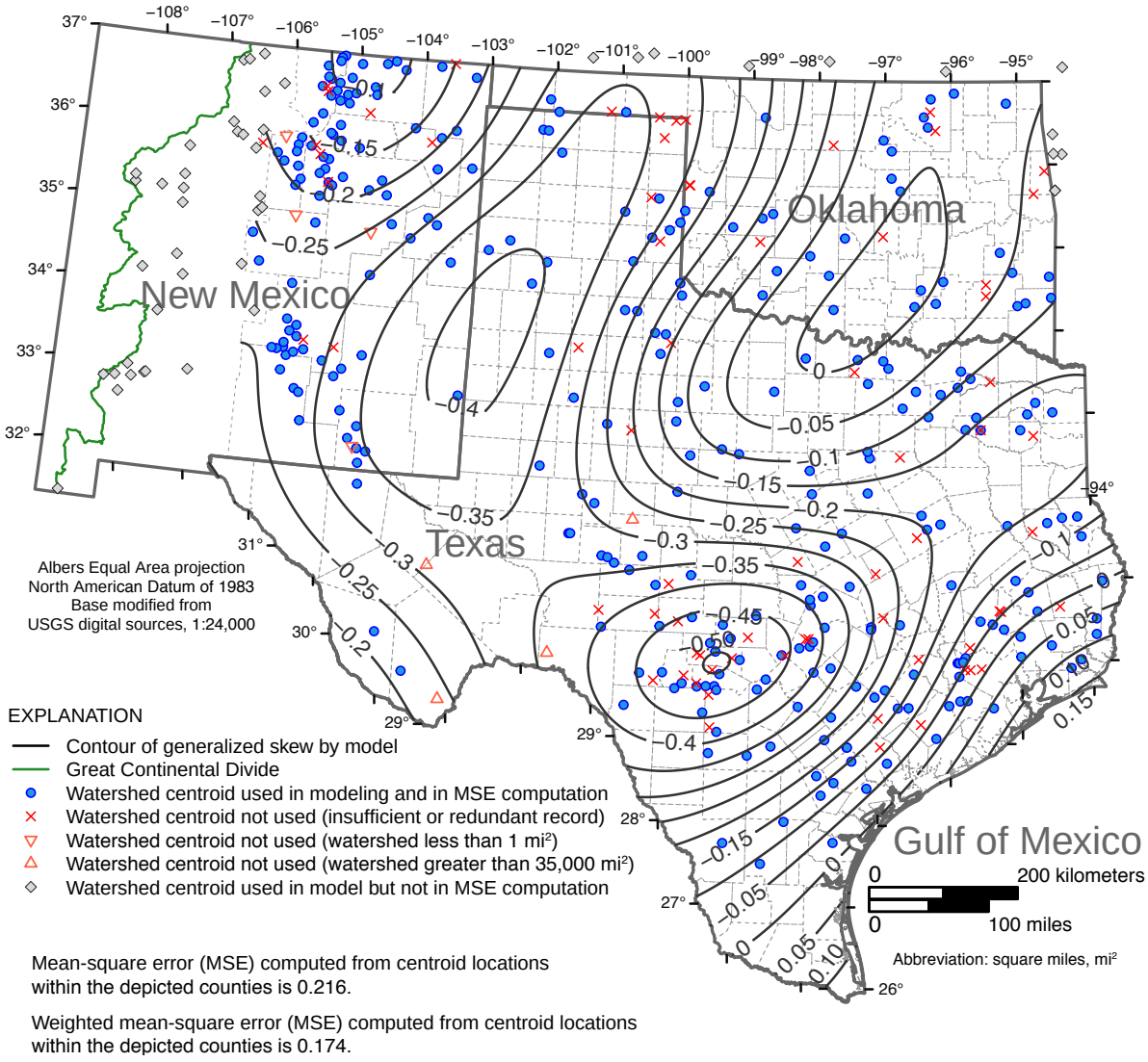


Figure 4.2. Map of generalized skew coefficients determined by using centroid location generalized additive modeling techniques (centroid-location GAM) with 2-dimensional smooth function on projected Albers equal area coordinates, Texas, Oklahoma, and eastern New Mexico.

expected to yield updated predictions of generalized skews in the State of Texas compared to those available in Judd et al. (1996).

The generalized skew map created using the GAM based on the locations of the centroids of the watersheds (fig. 4.2) is presented first. The streamgage-location skew map is also included because the locations of the streamgages were previously used by Judd et al. (1996). A quick comparison between figures 4.2 and 4.3 shows that much similarity exists and the

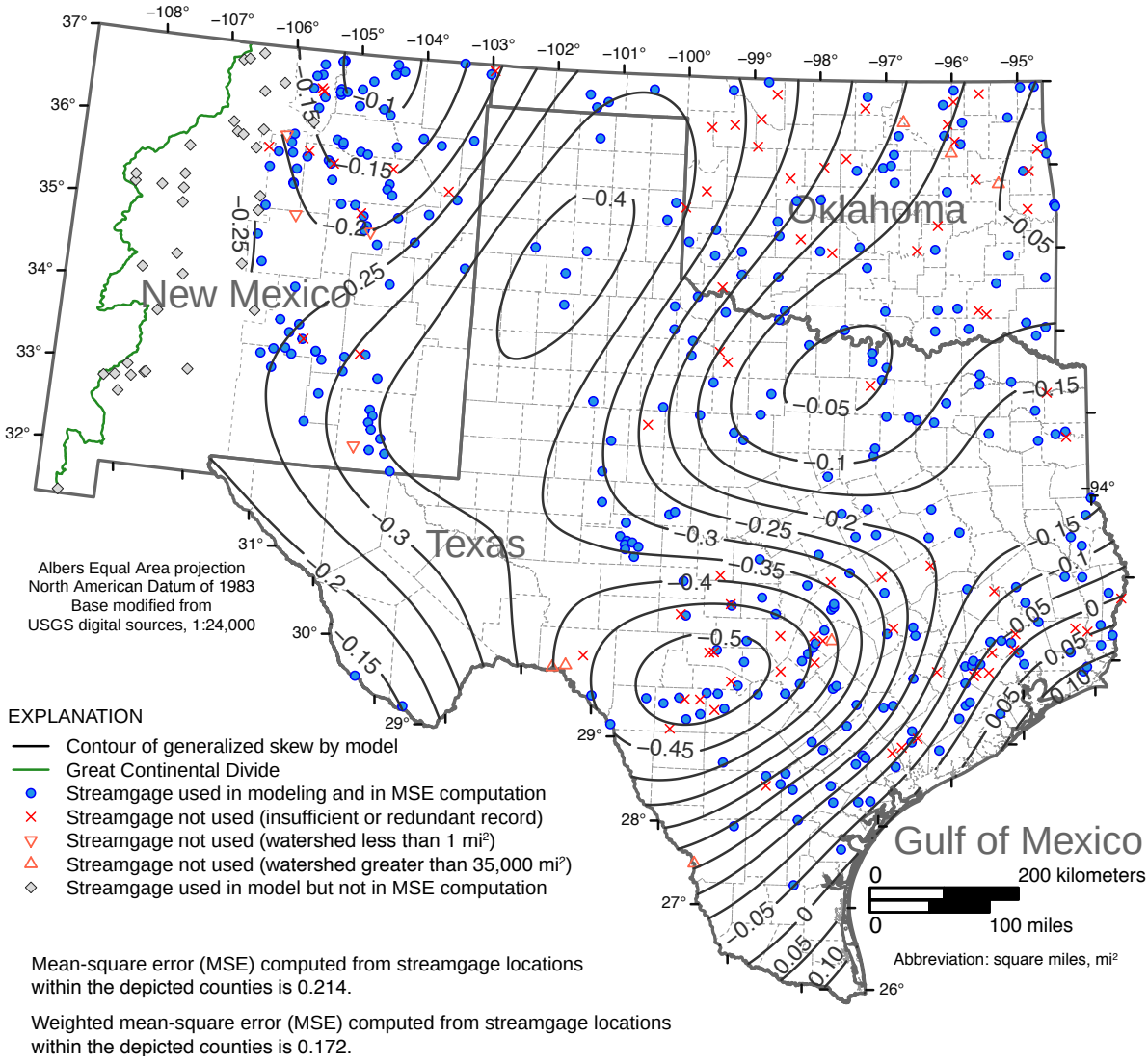


Figure 4.3. Map of generalized skew coefficients determined by using streamgage location generalized additive modeling techniques (the streamgage-location GAM) with 2-dimensional smooth function on projected Albers equal area coordinates, Texas, Oklahoma, and eastern New Mexico.

choice to reference the location of the centroid of the watershed or the streamgage makes little difference.

In order to develop a grid on which the contour lines could be drawn, both maps used a 1-kilometer grid spacing and an Albers equal area projection (Clark et al., 2018) clipped to the boundary of the counties depicted on the two maps. The depicted counties do not include the entire State of New Mexico or even the entire extent of New Mexico east of the Great Continental Divide; this mapping style is deliberate to emphasize that, by design of the study, the generalized skews are intended primarily for use in Texas. (The generalized skews also have some applicability in Oklahoma and eastern New Mexico). The counties in New Mexico shown correspond to those used in Asquith et al. (2006) in a study of storm statistics focused on Texas.

The unweighted MSEs corresponding to each GAM are indicated on the maps, and it is noted that the MSEs are computed using only centroid and streamgage locations within the extent of the depicted counties. This means that the sample sizes for MSE computations are not going to equal 341 but will be smaller and are reported as follows. There are 293 centroid locations (fig. 4.2) and 306 streamgage locations (fig. 4.3) within the depicted counties in Texas, Oklahoma, and the part of New Mexico east of the Great Continental Divide (figs. 4.2 and 4.3). These two counts are the respective sample sizes used to compute the reported MSEs of 0.216 (centroid locations) and 0.214 (streamgage locations). The respective MSEs are published on the maps because they are required in the weighted-skew computation in PeakFQ and thus are important to document. Conversely, the weighted MSE of the centroid-location map is 0.174, and the weighted MSE of the streamgage-location map is 0.172. These values are based on the MSEs of the station skews themselves. The weighted MSEs are not appropriate for use in weighted-skew computations and are not further mentioned in this chapter.

There are other informative metrics to report. The intercept of the centroid-location GAM is -0.204, and the intercept of the streamgage-location GAM intercept is -0.207; these are conceptually similar to an arithmetic mean. Therefore, a "rule-of-thumb" for the greater study area (including the locations shown in New Mexico east of the Great Continental Divide) is that the generalized skew value is -0.206 before spatial adjustment in the additive structure of the GAMs. The adjusted R-squared (coefficient of determination) (Helsel et al., 2020) is 0.11 for both GAMs. This demonstrates that, although the maps presented herein are deemed reliable, there is inherently much unexplained (and perhaps unexplainable) variance in skews across the region, which is the core justification for recommending the use of a weighted skew computation in peak-flow frequency analysis (England et al., 2018).

4.3.1. Comparison of Texas Results to 1996 Texas Results

Qualitative comparison of the skew maps presented in this study to those from a previous 1996 study (Judd et al., 1996) can be made. Therefore, it is useful to reproduce the map from the previous 1996 study in this chapter as figure 4.4. In the previous 1996 study, the estimation of generalized skews were based on station skews from 255 streamgages having at least 20 years of annual peak streamflow data from natural watersheds exclusively in Texas. A form of kriging (Papritz and Stein, 1999) was used as the statistical basis for producing the 1996 skew map that has an MSE of 0.35.

The contours are similar between the studies (compare contours of figs. 4.2 and 4.3 to those in fig. 4.4) with the exception of the more negative skew in northwestern Texas in this study relative to the previous 1996 study (Judd et al., 1996). A region of large negative skew in south-central Texas is thus common between the studies. Both studies indicate a spatial trend towards positive skew in extreme southeastern Texas. Finally, both studies show similar magnitudes and sign of skew in north-central and northeastern Texas.

The boundaries of the 1996 study by Judd et al. (1996) were exclusively those of Texas, whereas the boundaries for this study were expanded beyond the state. This was done to avoid "faults" in generalized skews at the state boundaries and alleviate confusion for end-users working in watersheds that are near to or extend outside of the state boundaries. Also because of the irregular shape of Texas and the inclusion of streamgages in Oklahoma and eastern New Mexico, the more rectangular shape of this study area results in more authoritative statistical mapping. Streamgages in eastern New Mexico and western Oklahoma help the GAM span the more sparsely gaged part of the region in western Texas. Because of differences in geographic areas and streamgages used in each investigation, the MSEs from this study and those from Judd et al. (1996) are not directly comparable and cannot be used to judge one study as "better" than the other. The greater number of streamgages and longer annual peak-flow records available for use in this study are expected to provide more accurate estimates of generalized skew for Texas than those presented by Judd et al. (1996).

4.3.2. Comparison of Generalized Skew Coefficients in Texas to Generalized Skew Coefficients Published for Other States

A comparison of generalized skews in Texas to generalized skews in Arkansas and Louisiana (not geographically depicted herein, but immediately east of Oklahoma and Texas) can be made. Using Bayesian weighted least squares/Bayesian generalized least-squares (B-WLS/B-GLS) regression, Wagner et al. (2016) concluded that neither a map nor regression equations were appropriate to estimate generalized skews. Rather, the generalized skews for Arkansas and Louisiana were estimated to be the constant value of -0.17 with an MSE of 0.12. Inspection of the skew maps for this study indicates, from latitudes of about 32° to 33°

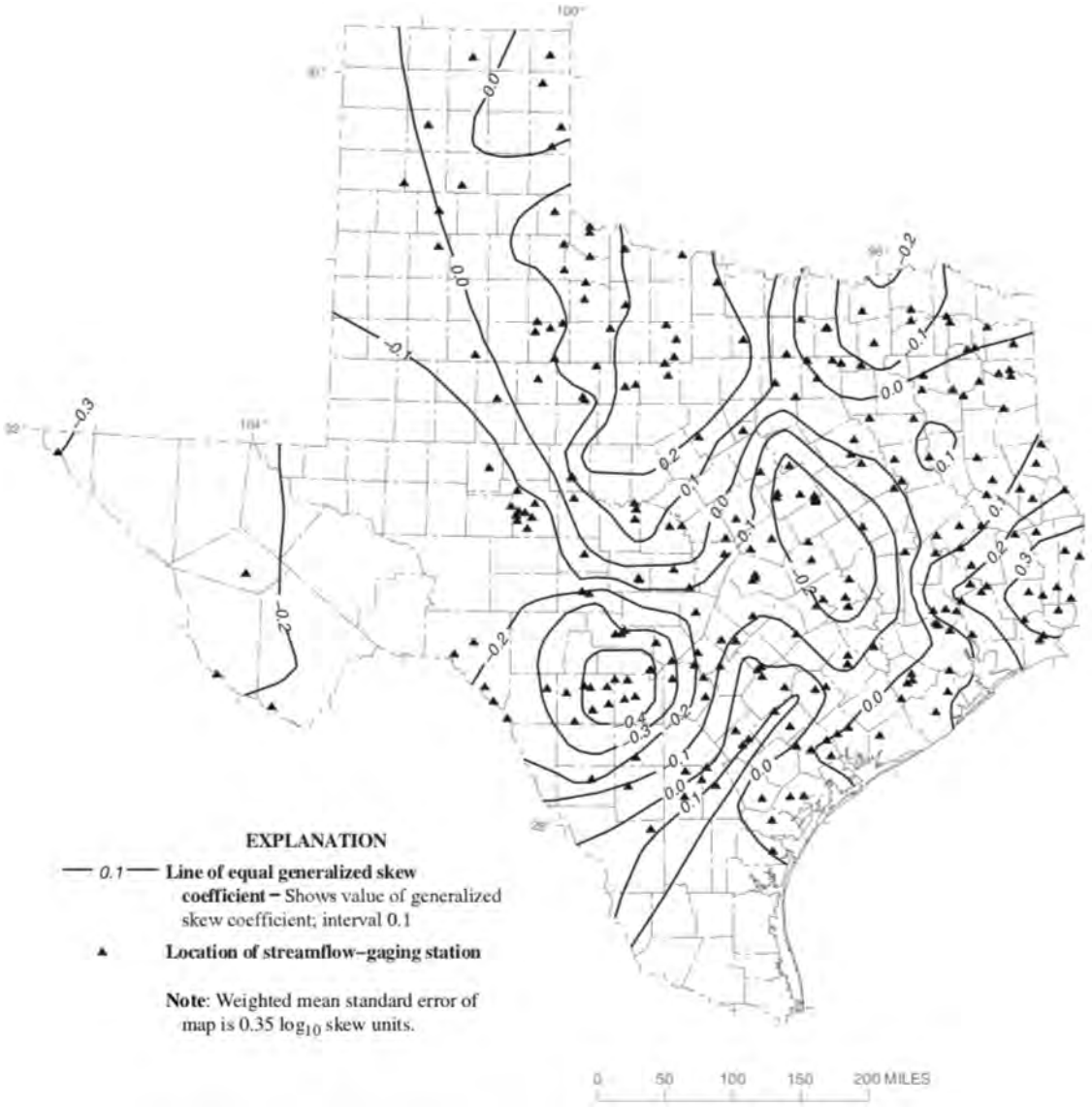


Figure 4.4. Map of generalized skew coefficients for Texas. Reproduced from figure 3 in Judd et al. (1996). Note: This figure is reproduced from a U.S. Geological Survey (USGS) series report. By convention in USGS series reports, longitude values are reported as positive values.

North, that the mapped skew value of approximately −0.15 (figs. 4.2 and 4.3) is similar to the skew value of −0.17 developed by Wagner et al. (2016). This provides evidence of the general compatibility of the GAM models of generalized skew in Texas with the B-WLS/B-GLS model of constant skew for the neighboring states to the east.

Comparisons of generalized skews in Texas to generalized skew in eastern New Mexico can also be made. The median of station skew values computed in prior skew investigations in New Mexico ranged from −0.220 to +0.297 for eight different flood regions in the state (Waltemeyer, 1986). It is noted from Waltemeyer (1986, table 2) that the northwest plains (median skew value of −0.220) and southwest plains (median skew value of −0.166) regions in that study align well with the similar magnitude negative skews in western Texas shown in figures 4.2 and 4.3. More than 20 years later, Waltemeyer (2008) used a generalized skew value of zero for the entire state of New Mexico with an MSE of 0.31. It is difficult to make direct comparisons of the results from these previous studies in New Mexico to the results of this study because the computation methods available in "PeakFQ equivalent software" (then a program on USGS mainframe computers) in the mid-1980s (Interagency Committee on Water Data, 1982) are different from those available in PeakFQ versions 7.1 and later (England et al., 2018).

Generalized skews in Texas and Oklahoma can be compared because the study area fully encompasses Oklahoma. The generalized skew map for Oklahoma (Lewis et al., 2019, fig. 2), which is reproduced herein as figure 4.5, was created using station skew computed following methods detailed in Bulletin 17C (England et al., 2018) and by "iterative interpolation" of the values across Oklahoma using geographic information system tools for interpolating and smoothing isolines. The interpolation and smoothing process was iterated four times, progressively refining the skew map by eliminating outlying skews and adjusting skew contours. The Oklahoma generalized skew has an MSE of about 0.148. The Oklahoma skew map is similar to this study's skew maps (figs. 4.2 and 4.3) in the magnitude of skew west of about −99° longitude.

The central part of Oklahoma is a region of very positive skew value (approximately +0.6) and much of Oklahoma shows positive skew (fig. 4.5). The results reported by Lewis et al. (2019) contrast with the results reported in this study. The two studies were conducted independently, with differences in streamgage selection methods, data retrieval, setup of peak-streamflow frequency analyses, and terminal statistical processing.

The results reported here are potentially affected by the approach that was used; streamgages with trends in annual peak streamflows were removed, the requirement for at least 30 years of annual peak streamflow record (the Oklahoma study by [Lewis et al., 2019] used a 20-year minimum period of record), mitigation for high cross-correlation of annual peak streamflows (redundancy), and the general avoidance of using annual peak streamflows assigned code 6 (regulated peaks) or C (urban peaks). It is also possible that differences in

the suite of streamgages used for the maps presented herein contribute to the differences between skews from the Oklahoma study and this study.

4.3.3. Generalized Skew Map

The maps presented in figures 4.2 and 4.3 depict generalized skews for the study area along with a summary of the supporting data locations. To reveal subtle differences in region skew values, it is useful to show an alternative representation of figure 4.2 with the locations of the watershed centroids removed and major physiographic provinces (Fenneman and Johnson, 1946; U.S. Geological Survey, 2020) shown instead (fig. 4.6).

For this analysis, the western Basin and Range physiographic province was combined with the Colorado Plateaus physiographic province (Fenneman and Johnson, 1946) in the western part of the study area. The Ouachita and Ozark Plateaus physiographic provinces were combined in the northeastern part of the study area. The Southern Rocky Mountains physiographic province extends into north-central New Mexico (fig. 4.6) (Fenneman and Johnson, 1946) and serves as a reminder of the common effects of major mountain ranges on streamflow (Fonstad, 2003) near the distal northeastern part of the contoured area and of the extreme precipitation events found in mountain ranges that can contribute to large floods (such as rain on snow events) (Li et al., 2019).

Insights into generalized skews for different parts of Texas were gained by analyzing skew patterns. An overview of the implications of different skew values, including the effect of the extremes, will help frame the discussion (England et al., 2018; Asquith et al., 2017). First, large negative skews result in PE3 distributions with upper bounds or finite upper limits; this implies that the peak streamflow producible by a watershed with negative skew has an asymptotic upper limit. Second, zero to positive skews result in PE3 distributions with infinite upper bounds meaning that the peak streamflow producible from a watershed with a positive skew has no apparent upper limit.

Consider the subregion of negative skew values of <-0.30 in the central part of the study area, which is in an area that generally aligns with the Great Plains physiographic province with <-0.40 skew values in northwestern Texas and of <-0.50 skew values in central Texas. The authors suggest that the nearly flat main-channel slopes of streams in the high plains of northwestern Texas, along with scant regional precipitation (interrupted only rarely by large storms), result in watersheds having finite upper bounds on their annual peak streamflows. In contrast, the upland, steep, incised-limestone bedrock watersheds of central Texas produce some of the largest peak streamflows per unit area in the United States (Asquith and Slade, 1995; O'Connor and Costa, 2018), yet skew values are quite negative (the largest negative values in Texas) (fig. 4.6) and substantiated by a large amount of data from many streamgages.

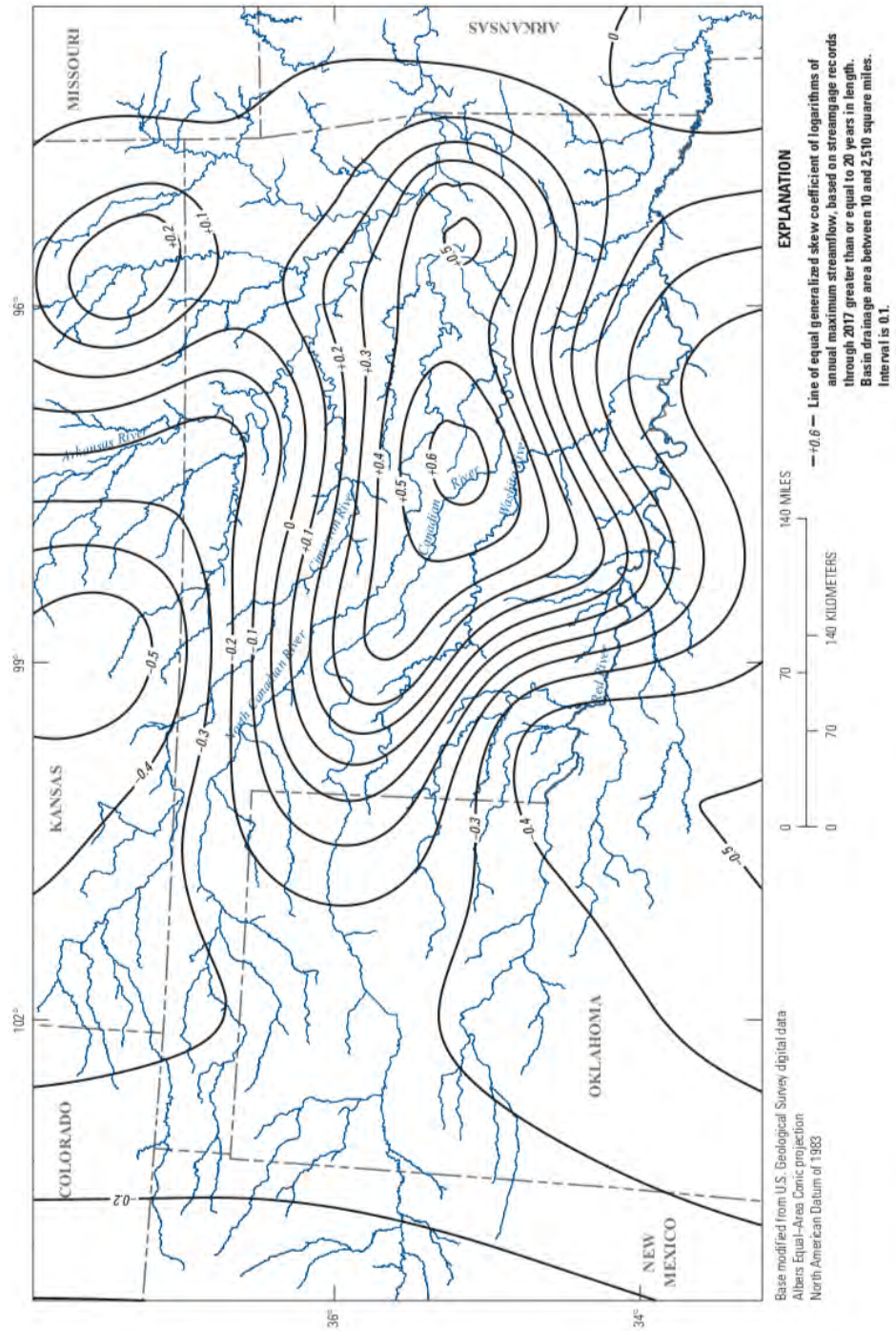


Figure 4.5. Map of generalized skew coefficients for Oklahoma. Reproduced from figure 2 in Lewis et al. (2019). Note: This figure is reproduced from a U.S. Geological Survey (USGS) series report. By convention in USGS series reports, longitude values are reported as positive values.

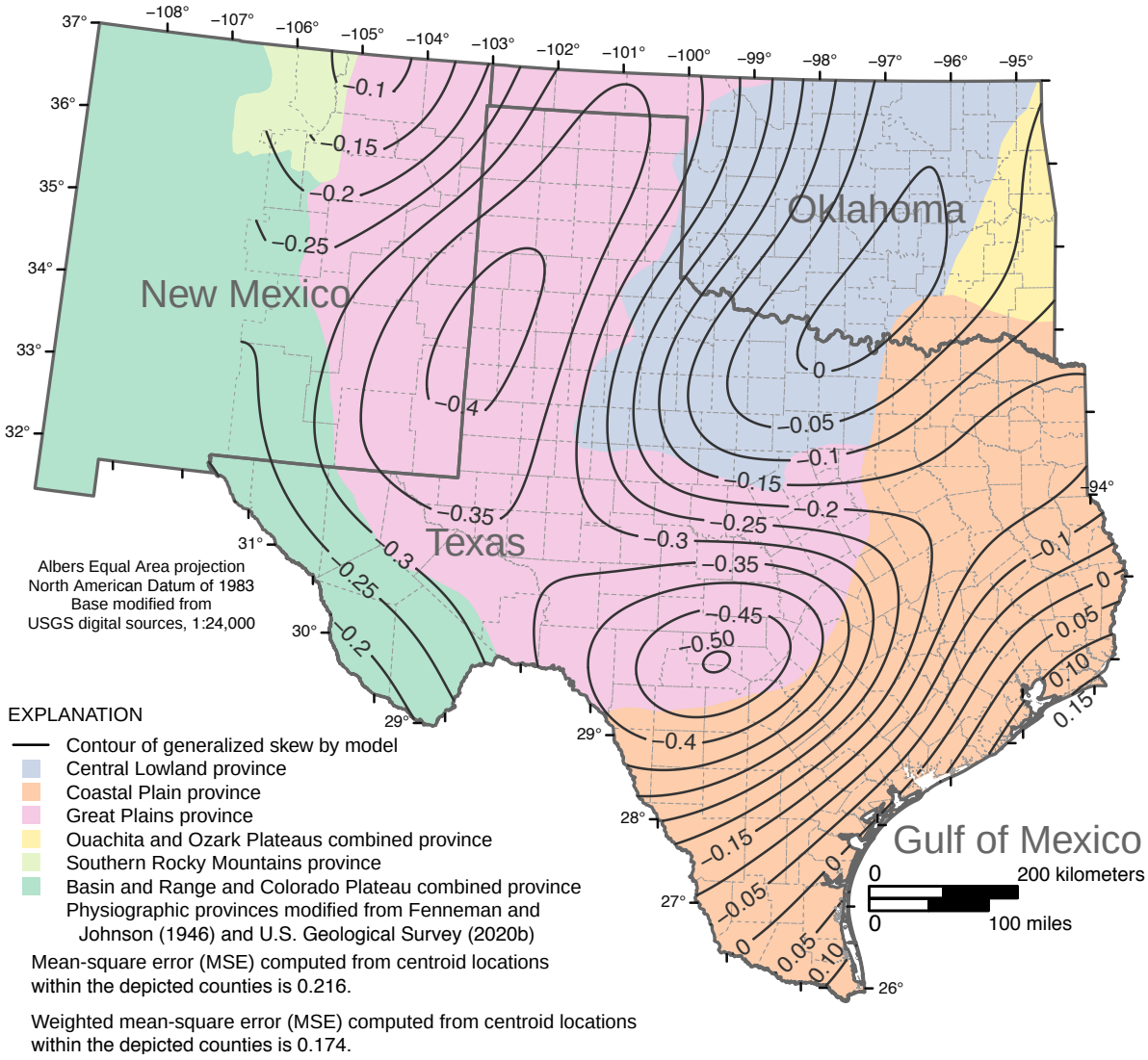


Figure 4.6. Map of generalized skew coefficients determined by using centroid location generalized additive modeling techniques (centroid-location GAM) with 2-dimensional smooth function on projected Albers equal area coordinates with annotation of major physiographic provinces modified from Fenneman and Johnson (1946) and U.S. Geological Survey (2020), Texas, Oklahoma, and eastern New Mexico.

In south-central Texas, sufficient peak-streamflow records exist and perhaps did at least as early as 1996 (Judd et al., 1996) to indicate that upper limits to streamflow have in large part have been defined. The distal upper tail of empirical distributions of annual peak streamflow for the longest-record streamgages in central Texas often indicate the rate of increase in peak streamflows becomes small as the annual exceedance probability becomes small as seen in Asquith et al. (1996, table 1). Phenomenologically, this relation between the rate of increase in peak streamflows and the annual exceedance probability implies that annual peak streamflows reach an upper limit as the annual exceedance probability nears zero. Perhaps in south-central Texas the watersheds are so efficient in converting precipitation to peak streamflow that additional precipitation inputs relative to the timing of peak streamflow in the watersheds cannot increase streamflow because a portion of the flood water is always rapidly exiting the watershed in this part of Texas where channel slopes can exceed 100 feet per mile (Asquith et al., 1996; Asquith and Slade, 1997).

The -0.30 skew contour in far western Texas generally aligns with the boundary between the Basin and Range and Great Plains physiographic provinces. It is not that the magnitude -0.30 is expected to align with the boundary between these physiographic provinces but that coincidentally the contour takes on similar curvature for almost 400 miles from the southern Texas border near the -102.5° longitude. Skew becomes less negative in far southwestern Texas, similar to the better monitored and climatically similar montane regions of eastern New Mexico. Skew appears to become increasingly less negative in the northwestern Great Plains and Southern Rocky Mountains physiographic provinces. The general increase in skew values from negative values to about zero from north-central Texas northeast into Oklahoma appears coincident and aligned with the transition into the Central Lowlands physiographic province.

Skew contours along the southeastern coast of Texas are subparallel to the coastline. Phenomenologically, the authors note that, for many coastal streamgages, one or two high outliers (large historical floods) are present in the streamgage records, and these appear to be the result of major tropical cyclones. In the low-slope region near the Gulf Coast, major tropical cyclones could represent either a unique population of annual peak streamflows resulting from the abnormally large amounts of precipitation produced by such storms or that watershed boundaries in the region become irrelevant during such extreme region-wide precipitation events. The Coastal Plain province (Fenneman and Johnson, 1946; U.S. Geological Survey, 2020) seems a bit too large (extends too far inland to the north) for purposes of interpreting skew near the coast of Texas. The authors suggest that skew values >-0.15 might be more suitable for this part of Texas where tropical cyclones affect the distribution of annual peak streamflows.

A skew contour of 0.0 extends into southeastern Texas and appears congruent in shape to a region of rapidly enlarging multi-day precipitation frequency values, especially 100-year, 10-day storm depths (National Oceanic and Atmospheric Administration, 2020a) or other

extreme precipitation contours (Asquith and Roussel, 2004). The paucity of streamgages in the southern part of Texas (south of about 28° North latitude where the Texas coast bends sharply) contributes to extrapolation concerns of statistical methods. The contoured results for the extreme southern tip of Texas were manually decreased to skew values of about +0.10 in an effort provide consistency across the remainder of coastal Texas. South of about 28° North latitude, the southern part of Texas is where the skew values are about zero; this area is also a part of the Texas coast that appears to have a lower probability of tropical cyclone landfall relative to either extreme south Texas or the Texas coast east of about −97° longitude (National Oceanic and Atmospheric Administration, 2020b). Tropical cyclones also make landfall less frequently near the sharp bend in the Texas coast (approximate latitude 28° North) relative to the rest of the Texas Gulf Coast, and that difference in tropical-cyclone landfall contributes to less positive skews near this sharp bend in the coast.

The prior discussion on skew patterns in Texas has implications for peak-streamflow distribution behavior as measured by the PE3 distribution fit by EMA and MGBT-based removal of low outliers (small annual peak streamflows). For the generalized skews to be most applicable and produce reliable peak-streamflow frequency computations, future analysis of annual peak streamflow in Texas would need to generally adhere to MGBT or otherwise aggressively truncate parts of the left-hand (non-flood) tail of the distribution of the annual peak streamflows. Generalized skews are inherently coupled to left-tail truncation of low outliers using MGBT.

The MSE of the centroid-location map is 0.216. This value corresponds to $\sqrt{0.216} = 0.465$ so it can be said that generalized skews have an uncertainty in their native units (though dimensionless) of about ±0.465. There is conceptually much uncertainty in skew that remains purely statistical, which means that the uncertainty is a function of sample size (number of annual peak streamflows at any given streamgage). For example, the square root of the mean of the MSEs of station skews for a subgroup 293 streamgages (corresponding to the 293 centroid locations) used in the study is about 0.371; this supports that uncertainty in skew is fundamental to peak-streamflow frequency analysis. The mean systematic record length is about 57 years for the streamgages within the depicted counties on the centroid-location map (fig. 4.2).

4.4. Chapter Conclusions

Reliable information about the frequency of annual peak streamflow is needed for floodplain management, objective assessment of flood risk, and cost-effective design of dams, levees, other flood-control structures, and roads, bridges, and culverts. This chapter presents the results of a study to update the 1996 generalized skew coefficients (generalized skews) in Texas (Judd et al., 1996) and the current (September 2019) TxDOT Hydraulic Design

Manual (Texas Department of Transportation, 2020). Generalized skews are a vital element of frequency analysis of observed annual peak streamflows, and are used, along with station skews computed for USGS streamgages to provide weighted (and hence more accurate) estimates of individual skew values (England et al., 2018).

A technique is presented for estimating generalized skews of annual peak streamflow for streamgages in Texas. The technique is based on analysis of station skews from long-term USGS streamgages (minimum 30-year period of record from natural watersheds with minimal regulation and minimal impervious cover). Annual peak streamflow data through the 2016 water year were used, if available. A total of 444 streamgages in Texas, Oklahoma, and eastern New Mexico study area, were identified that met the regulation and impervious cover criteria and formed the core dataset for initial analysis. Of the 444 streamgages, 341 were used in the final skew model. Generalized skews were modeled using the centroids of the watersheds of the streamgages as the positions in generalized additive modeling (centroid-location GAM) using a 2-dimensional smooth on projected coordinates (Albers Equal Area projection).

The geographic boundaries of this study area are more extensive than in prior work in an effort to avoid state-line faults in the skew and confusion for end-users working in watersheds that extend outside of Texas. The inclusion of annual peak-flow records from eastern New Mexico and all of Oklahoma yielded a more rectangular-like study area relative to that used in the previous 1996 study (Judd et al., 1996). A more rectangular-like boundary facilitated more authoritative statistical mapping than in the 1996 study. Streamgages in both Oklahoma and eastern New Mexico help the GAM span the more sparsely gaged watersheds in west and northwestern Texas. The larger number of streamgages and longer periods of record used in the mapping are expected to provide more reliable estimates of generalized skews for Texas than those presented by Judd et al. (1996).

Figure 4.6 in this chapter could be considered as a replacement for figure 4.4 within the current (September 2019) TxDOT Hydraulic Design Manual (Texas Department of Transportation, 2020). The following text accompanying figure 4.4 in the Hydraulic Design Manual reads:

> "$\text{MSE}_{\overline{G}}$ = mean square error of $\overline{G}$ [generalized skew] for Texas is = 0.123 (RMSE = 0.35) [root-mean-square error] (Judd et al., 1996), which replaces the value of 0.302 (RMSE = 0.55) presented in Bulletin 17B [(Interagency Committee on Water Data, 1982)]."

That verbatim text could be replaced with the following text to align to this study and figure 4.6:

> "$\text{MSE}_{\overline{G}}$ = mean square error of $\overline{G}$ [generalized skew] for Texas is = 0.216 (RMSE = 0.465) [root-mean-square error], which replaces the value of 0.123 (RMSE = 0.35) presented in Judd et al. (1996)."

Acknowledgments

This work was made possible by the collaboration with the Texas Department of Transportation and research colleagues at Texas Tech University and the University of Texas at Arlington. We sincerely thank the hydrologic technicians at the U.S. Geological Survey in the Oklahoma-Texas and New Mexico Water Science Centers that collecting the streamflow data used in this analysis. We also thank reviewers John Gordon, George R. Herrmann, Anne C. Tillery, and Daniel M. Wagner for assistance with improving this manuscript. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Chapter References

Asquith, W.H., 1998, Depth-duration frequency of precipitation for Texas: U.S. Geological Survey Water-Resources Investigations Report 98–4044, 107 p., https://doi.org/10.3133/wri984044.

Asquith, W.H., 2020, The use of support vectors from support vector machines for hydrometeorologic monitoring network analyses: Journal of Hydrology, v. 583, April (2020) 124522, 10 p., https://doi.org/10.1016/j.jhydrol.2019.124522.

Asquith, W.H., Cleveland, T.G., Yesildirek, M.V., Zhang, J., Fang, Z.N., and Otto, L.D., 2021, *scNIDaregis*—Geospatial processing of dams in the United States from the National Inventory of Dams with a state-level aggregation scheme, demonstrated for selected dams in eight states in south-central region of the United States, and post-processing features for basin-specific tabulation: U.S. Geological Survey software release, Reston, Va., https://doi.org/10.5066/P90NJVB9.

Asquith, W.H., England, J.F., and Herrmann, G.R., 2020, *MGBT*—Multiple Grubbs-Beck low-outlier test: U.S. Geological Survey software release, R package, Reston, Va., accessed July 27, 2020, at https://doi.org/10.5066/P9CW9EF0.

Asquith, W.H., Kiang, J.E., and Cohn, T.A., 2017, Application of at-site peak-streamflow frequency analyses for very low annual exceedance probabilities: U.S. Geological Survey Scientific Investigation Report 2017–5038, 93 p., https://doi.org/10.3133/sir20175038.

Asquith, W.H., and Roussel, M.C., 2004, Atlas of depth-duration frequency of precipitation annual maxima for Texas: U.S. Geological Survey Scientific Investigations Report 2004–5041, 106 p., https://doi.org/10.3133/sir20045041.

Asquith, W.H., Roussel, M.C., Cleveland, T.G., Fang, X., and Thompson, D.B., 2006, Statistical characteristics of storm interevent time, depth, and duration for eastern New Mexico, Oklahoma, and Texas: U.S. Geological Survey Professional Paper 1725, 299 p., https://doi.org/10.3133/pp1725.

Asquith, W.H., and Slade, R.M., 1995, Documented and potential extreme peak discharges and relation between potential extreme peak discharges and probable maximum flood peak discharges in Texas: U.S. Geological Survey Water-Resources Investigations Report 95–4249, 58 p., 1 pl., https://doi.org/10.3133/wri954249.

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., https://doi.org/10.3133/wri964307.

Asquith, W.H., Slade, R.M., and Judd, Linda, 1995, Analysis of low-outlier thresholds for log-Pearson type III peak-streamflow frequency analysis in Texas, in Texas Water *95, American Society of Civil Engineers First International Conference, San Antonio, Texas, 1995, Proceedings: San Antonio, Texas, American Society of Civil Engineers, pp. 379–384.

Asquith, W.H., Slade, R.M., and Lanning-Rush, J., 1996, Peak-flow frequency and extreme flood potential for streams in the vicinity of the Highland Lakes, central Texas: U.S. Geological Survey Water Resources Investigations Report 96–4072, 1 sheet, https://doi.org/10.3133/wri964072.

Carr, J.T., 1967, The climate and physiography of Texas: Austin, Texas Water Development Board Report 53, 27 p., accessed July 27, 2020, at https://www.twdb.texas.gov/publications/reports/numbered_reports/doc/R53/R53.pdf.

Clark, B.R., Barlow, P.M., Peterson, S.M., Hughes, J.D., Reeves, H.W., and Viger, R.J., 2018, National-scale grid to support regional groundwater availability studies and a national hydrogeologic database: U.S. Geological Survey data release, https://doi.org/10.5066/F7P84B24.

Cleveland, T.G., and Fang, Z.N., 2021, Texas-Skew-Update-2021: Texas Data Repository, https://doi.org/10.18738/T8/SVLCOQ.

Cohn, T.A., Barth, N.A., England, J.F., Jr., Faber, B.A., Mason, R.R., Jr., and Stedinger, J.R., 2019, Evaluation of recommended revisions to Bulletin 17B: U.S. Geological Survey Open-File Report 2017–1064, 141 p., https://doi.org/10.3133/ofr20171064.

Cohn, T.A., England, J.F., Berenbrock, C.E., Mason, R.R., Stedinger, J.R., and Lamontagne, J.R., 2013, A generalized Grubbs–Beck test statistic for detecting multiple potentially influential low outliers in flood series: Water Resources Research, v. 49, no. 8, pp. 5047–5058, https://doi.org/10.1002/wrcr.20392.

England, J.F., Cohn, T.A., Faber, B.A., Stedinger, J.R., Thomas Jr., W.O., Veilleux, A.G., Kiang, J.E., and Mason, R.R., 2018, Guidelines for determining flood flow frequency Bulletin 17C: U.S. Geological Survey Techniques and Methods, book 4, chap. 5.B, 148 p., https://doi.org/10.3133/tm4B5.

Fenneman, N.M., and Johnson, D.W., 1946, Physical divisions of the United States: U.S. Geological Survey, 1 sheet, scale 1:7,000,000, https://doi.org/10.3133/70207506.

Fonstad, M.A., 2003, Spatial variation in the power of mountain streams in the Sangre de Cristo Mountains, New Mexico: Geomorphology, v. 55, pp. 75–96, https://doi.org/10.1016/S0169-555X(03)00133-8.

Helsel, D.R., Hirsch, R.M., Ryberg, K.R., Archfield, S.A., and Gilroy, E.J., 2020, Statistical methods in water resources: U.S. Geological Survey Techniques and Methods, book 4, chapter A3, 458 p., https://doi.org/10.3133/tm4a3. [Supersedes USGS Techniques of Water-Resources Investigations, book 4, chapter A3, version 1.1.]

Interagency Committee on Water Data, 1982, Guidelines for determining flood flow frequency, Bulletin 17B: Interagency Committee on Water Data, Hydrology Subcommittee, Technical Report.

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., https://doi.org/10.3133/wri964117.

Lewis, J.M., Hunter, S.L., and Labriola, L.G., 2019, Methods for estimating the magnitude and frequency of peak streamflows for unregulated streams in Oklahoma developed by using streamflow data through 2017: U.S. Geological Survey Scientific Investigations Report 2019–5143, 39 p., https://doi.org/10.3133/sir20195143.

Li, D., Lettenmaier, D.P., Margulis, S.A., and Andreadis, K., 2019, The role of rain-on-snow in flooding over the conterminous United States: Water Resources Research, v. 55, pp. 8429–8513, https://doi.org/10.1029/2019WR024950.

National Oceanic and Atmospheric Administration, 2020a, NOAA Atlas 14 cartographic maps of precipitation frequency estimates for selected frequencies and durations: National Weather Service, Hydrometeorological Design Studies Center, Precipitation Frequency Data Server (PFDS), accessed on July 27, 2020, at https://hdsc.nws.noaa.gov/hdsc/pfds/pfds_maps.html, accessed Texas 100-year, 10-day rainfall on July 27, 2020, at https://hdsc.nws.noaa.gov/pub/hdsc/data/tx/tx100y10d.pdf.

National Oceanic and Atmospheric Administration, 2020b, Tropical cyclone climatology: National Hurricane Center and Central Pacific Hurricane Center, accessed July 27, 2020, at https://www.nhc.noaa.gov/climo/.

O'Connor, J.E., and Costa, J.E., 2018, Large floods in the United States—Where they happen and why: U.S. Geological Survey Circular 1245, 13 p., https://doi.org/10.3133/cir1245.

Papritz A., and Stein A., 1999, Spatial prediction by linear kriging *in* Stein A., Van der Meer F., Gorte B. (eds) Spatial Statistics for Remote Sensing: Remote Sensing and Digital Image Processing, v. 1. Springer, Dordrecht, https://doi.org/10.1007/0-306-47647-9_6.

R Core Team, 2020, R—A language and environment for statistical computing: R Foundation for Statistical Computing, Vienna, Austria, version 4.0.2, accessed July 4, 2020, at https://www.r-project.org.

Texas Department of Transportation, 2020, Hydraulic design manual—September 2019: online resource accessed on July 26, 2020, at https://onlinemanuals.txdot.gov/txdotmanuals/hyd/index.htm and subsection https://onlinemanuals.txdot.gov/txdotmanuals/hyd/statistical_analysis_of_stream_gauge_data.htm.

U.S. Army Corps of Engineers, 2020, CorpsMap—National Inventory of Dams, accessed on July 7, 2020 at https://nid.sec.usace.army.mil/ords/f?p=105:1:::::.

U.S. Geological Survey, 2018, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed April 9, 2018, at https://doi.org/10.5066/F7P55KJN.

U.S. Geological Survey (USGS), 2020a, PeakFQ—Flood frequency analysis based on Bulletin 17B and recommendations of the Advisory Committee on Water Information (ACWI) Subcommittee on Hydrology (SOH) Hydrologic Frequency Analysis Work Group (HFAWG), version 7.2, accessed January 29, 2019, at https://water.usgs.gov/software/PeakFQ/, version 7.3, accessed February 1, 2020, at https://water.usgs.gov/software/PeakFQ/.

U.S. Geological Survey, 2020b, Physiographic divisions of the conterminous U.S., accessed July 24, 2020, at https://water.usgs.gov/GIS/metadata/usgswrd/XML/physio.xml.

Veilleux, A.G., Cohn, T.A., Flynn, K.M., Mason, R.R., Jr., and Hummel, P.R., 2014, Estimating magnitude and frequency of floods using the PeakFQ 7.0 program: U.S. Geological Survey Fact Sheet 2013–3108, 2 p., https://doi.org/10.3133/fs20133108.

Wagner, D.M., Krieger, J.D., and Veilleux, A.G., 2016, Methods for estimating annual exceedance probability discharges for streams in Arkansas, based on data through water year 2013: U.S. Geological Survey Scientific Investigations Report 2016–5081, 136 p., https://doi.org/10.3133/sir20165081.

Wagner, D.M., Kiang, J.E., and Asquith, W.H., 2017, U.S. Geological Survey streamgaging methods and annual peak streamflow, appendix 1 *of* Asquith, W.H., Kiang, J.E., and Cohn, T.A., 2017, Application of at-site peak-streamflow frequency analyses for very low annual exceedance probabilities: U.S. Geological Survey Scientific Investigations Report 2017–5038, 93 p., https://doi.org/10.3133/sir20175038.

Waltemeyer, S.D., 1986, Techniques for estimating flood-flow frequency for unregulated streams in New Mexico: U.S. Geological Survey Water-Resources Investigations Report 86–4104, 56 p., https://doi.org/10.3133/wri864104.

Waltemeyer, S.D., 2008, Analysis of the magnitude and frequency of peak discharge and maximum observed peak discharge in New Mexico and surrounding areas: U.S. Geological Survey Scientific Investigations Report 2008–5119, 105 p., https://doi.org/10.3133/sir20085119.

Watson, K.M., Harwell, G.R., Wallace, D.S., Welborn, T.L., Stengel, V.G., and McDowell, J.S., 2018, Characterization of peak streamflows and flood inundation of selected areas in southeastern Texas and southwestern Louisiana from the August and September 2017 flood resulting

from Hurricane Harvey: U.S. Geological Survey Scientific Investigations Report 2018–5070, 44 p., https://doi.org/10.3133/sir20185070.

Winters, K.E., 2013, A historical perspective on precipitation, drought severity, and streamflow in Texas during 1951–56 and 2011: U.S. Geological Survey Scientific Investigations Report 2013–5113, 24 p., https://doi.org/10.3133/sir20135113.

Wood, S.N., 2017, Generalized additive models—An introduction with R (2d ed.): Chapman and Hall / CRC, Boca Raton, Florida, ISBN 978–1–498–72833–1.

Wood, S.N., 2020, *mgcv*—Mixed GAM computation vehicle with automatic smoothness estimation: R package version 1.8–31, accessed July 4, 2020, at https://CRAN.R-project.org/package=mgcv.

Yesildirek, M.V., McDowell, J.S., Zhang, J., Asquith, W.H., 2021, Geospatial data of watershed characteristics for select U.S. Geological Survey streamgaging stations in New Mexico, Oklahoma, and Texas useful for statistical study of annual peak streamflows in and near Texas: U.S. Geological Survey data release, https://doi.org/10.5066/P9A91W4Z.

5. Using the Generalized Skew Coefficients of Annual Peak Streamflow for Natural Basins in Texas—A Training Framework

5.1. Introduction

Reliable peak-streamflow frequency information is needed for floodplain management, objective assessment of flood risk, and cost-effective design of dams, levees, other flood-control structures, and roads, bridges, and culverts. The generalized skew map shown in chapter 4 is proposed to update the current (September 2019) TxDOT Hydraulic Design Manual (Texas Department of Transportation, 2020). Should that map be adopted, it is informative to demonstrate the use of this "new" generalized skew with some Texas-based streamgages. The purpose of this chapter is to demonstrate the use of generalized skew coefficients in Texas. The chapter is semi-tutorial in nature, and is intended as a framework for stand-alone training materials.

5.1.1. When to Apply the Updated Generalized Skew

The use of generalized skew coefficients and the USGS PeakFQ software 7.3 (U.S. Geological Survey, 2020) arises when the designer intends to estimate peak discharges for TxDOT design and evaluation, and the facility site is near the streamgage on the same stream and watershed (direct application of flood frequency results); or if the facility site is on the same stream, but not nearby the streamgage, transposition of the streamgage analysis results may be used (Texas Department of Transportation, 2020, section 9). The generalized skew is based on analysis of natural watershed conditions. Complete suitability or extension of this skew to watersheds otherwise remains unknown though application of generalized skew for streamgages with short (fewer than about 15 years) to modest records (less than about 20–30 years) likely has some benefit towards achieving some stability in the distal upper tail.

The discussion hereinafter largely follows guidelines for statistical analyses of streamgage data in England et al. (2018). The Hydraulic Design Manual (Texas Department of Transportation, 2020)) stipulates that application of B17C in Texas should:

1. Obtaining a sufficiently large sample of streamflow data for statistical analysis. Chapter 3 largely satisfies this requirement and much of the work in these chapters reflect this goal. The databases produced and archived for this research provide an extensive collection of gages already examined and meeting minimum record length criteria. Facility sites located on these streams and watersheds could use the databases immediately;

2. Using the Pearson type III distribution fitting procedure, which is implicit in the use of PeakFQ 7.3, and chapter 4 adheres to this requirement;

3. Using a weighted skew value—The weighted skew value is computed using Equation 5.1, which is structurally identical to eq. 4-6 in the Hydraulic Design Manual (Texas Department of Transportation, 2020), with the term G_{GAM} replacing $\bar{G}$ and the values sourced from figure 4.6;

$$G_W = \frac{(MSE_{G_{GAM}})(G_{STA}) + (MSE_{G_{STA}})(G_{GAM})}{MSE_{G_{GAM}} + MSE_{G_{STA}}} \tag{5.1}$$

   PeakFQ 7.3 makes the requisite computation in eq. 5.1 if the designer supplies $MSE_{G_{GAM}}$ and G_{GAM} both of which are obtained from figure 4.6;

4. Accommodating outliers—Outliers are identified directly in the PeakFQ 7.3 software and the designer is relieved of manual computations to identify and censor low- and high-outlier thresholds;

5. Transposing streamgage analysis results, if necessary and appropriate—Transposition is unchanged using the tools herein; if streamgage data are not available at the design location, discharge values can be estimated by transposition if a flood-frequency curve is available at a nearby gaged location (from PeakFQ 7.3). Transposition is appropriate for hydrologically similar watersheds that differ in area by less than 50 percent with outlet locations less than 100 miles apart (Asquith, and Thompson, 2008). eq. 5.2 relates the gaged result to the design estimate. The equation is structurally identical to eqs. 4-10 and 4-11 in the Hydraulic Design Manual (Texas Department of Transportation, 2020); and

$$Q_{estimate} = Q_{gage} \left(\frac{A_{estimate}}{A_{gage}} \right)^e \tag{5.2}$$

where $Q_{estimate}$ is the estimated annual exceedance probability streamflow at the ungaged watershed, Q_{gage} is the known annual exceedance probability streamflow at the gaged watershed, $A_{estimate}$ is the drainage area associated with the assessment point of the ungaged watershed, A_{gage} is the drainage area associated with the assessment point of the gaged watershed, e is an exponent ranging from 1/2 (eq. 4-10) up to 9/10 (eq. 4-11). Values for e are sourced from Asquith, Roussel, and Vrabel (2006) and Asquith, and Thompson (2008).

5.1.2. How to Apply These Tools—Illustrative Example 1

Suppose there is intent to evaluate the transportation infrastructure facility relatively near USGS streamgage 08167000 Guadalupe River at Comfort, Texas, as shown in figure 5.1, hence a need to estimate discharge at a specific annual recurrence interval. The pin in the image is approximately at the stream thalweg, which is located at about latitude $29^o57'54.86"$ and longitude $-98^o53'49.80"$.

Figure 5.1. Google Earth image near USGS streamgage 08167000 Guadalupe River at Comfort, Texas.

This location is ideal for direct application of flood-frequency tools herein because the facility is proximal to the streamgage. The designer can access the USGS National Water Information System (NWIS) (U.S. Geological Survey, 2018) for the streamgage as shown in figure 5.2.

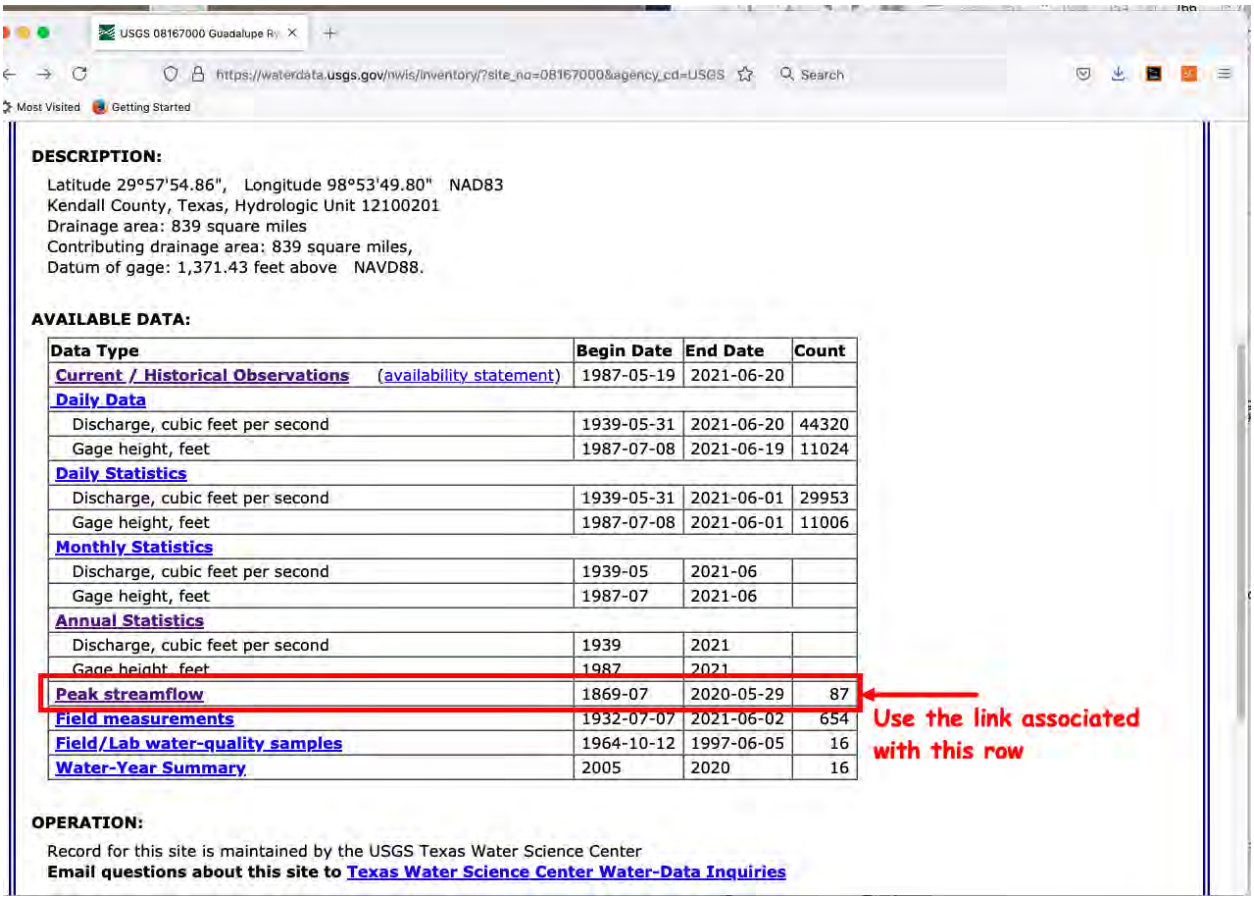


Figure 5.2. USGS National Water Information System landing page for USGS streamgage 08167000 Guadalupe River at Comfort, Texas.

The designer would then select the indicated link and choose to download the annual peak streamflows from figure 5.3. After the download is complete the designer would then start PeakFQ 7.3 and proceed as illustrated in the following figures:

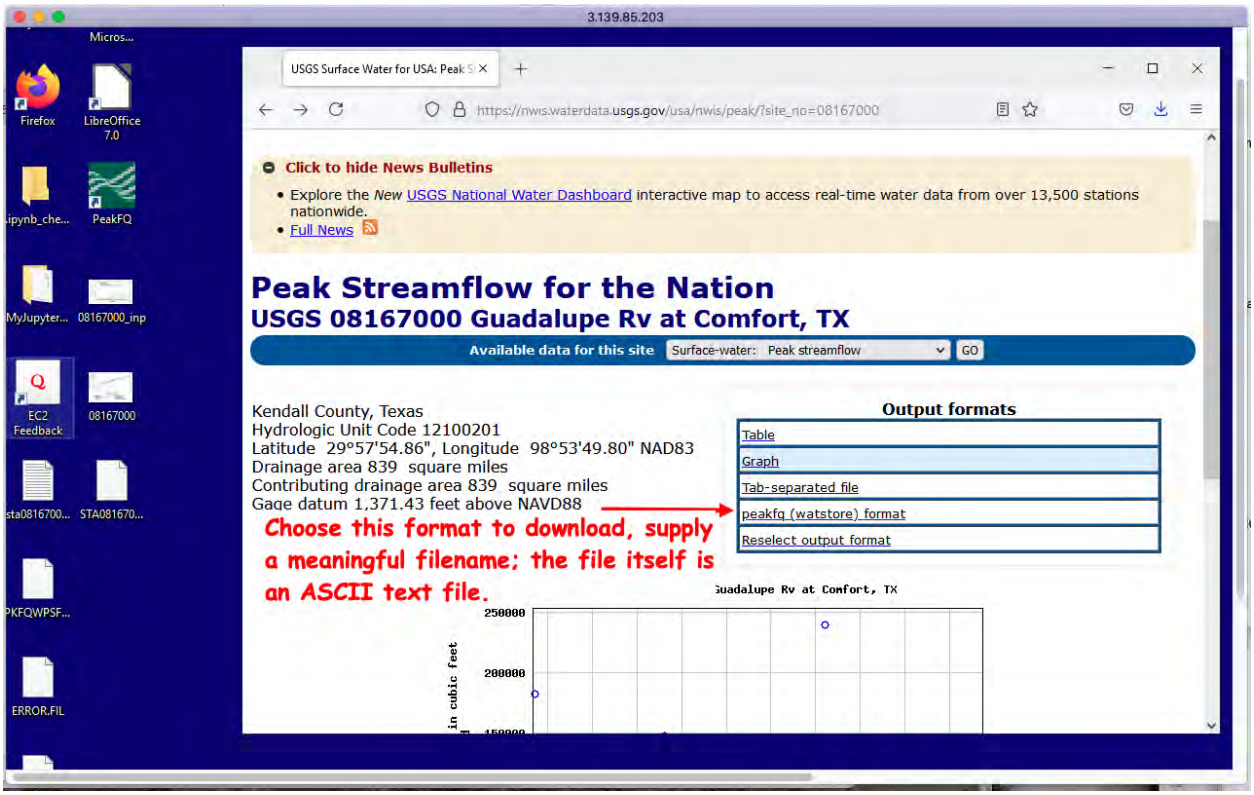


Figure 5.3. USGS National Water Information System annual peak streamflow data selection for USGS streamgage 08167000 Guadalupe River at Comfort, Texas.

Figure 5.4 depicts selecting the file (either just downloaded or from the previously downloaded research database).

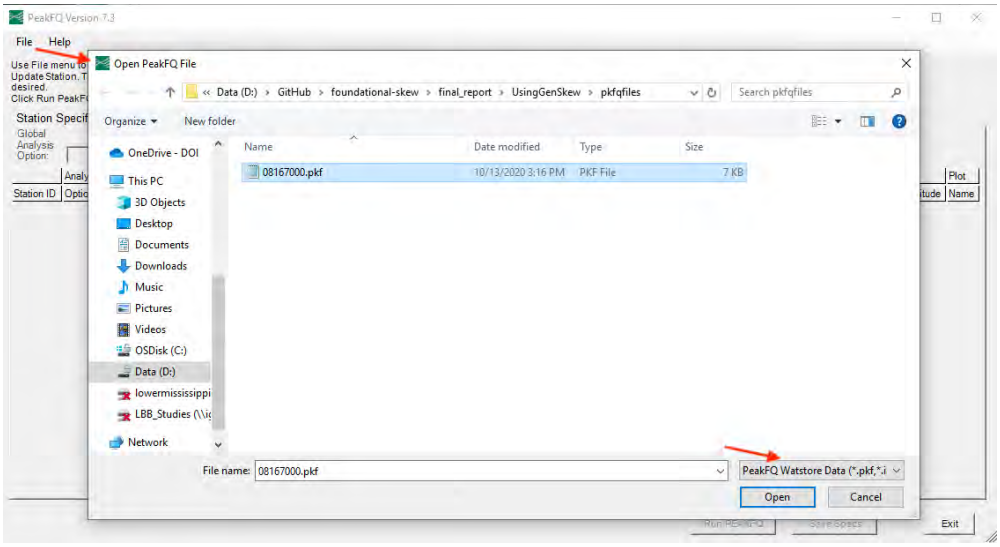


Figure 5.4. Loading of USGS streamgage 08167000 Guadalupe River at Comfort, Texas, annual peak streamflows in the so-called "WATSTORE" format to USGS PeakFQ software 7.3.

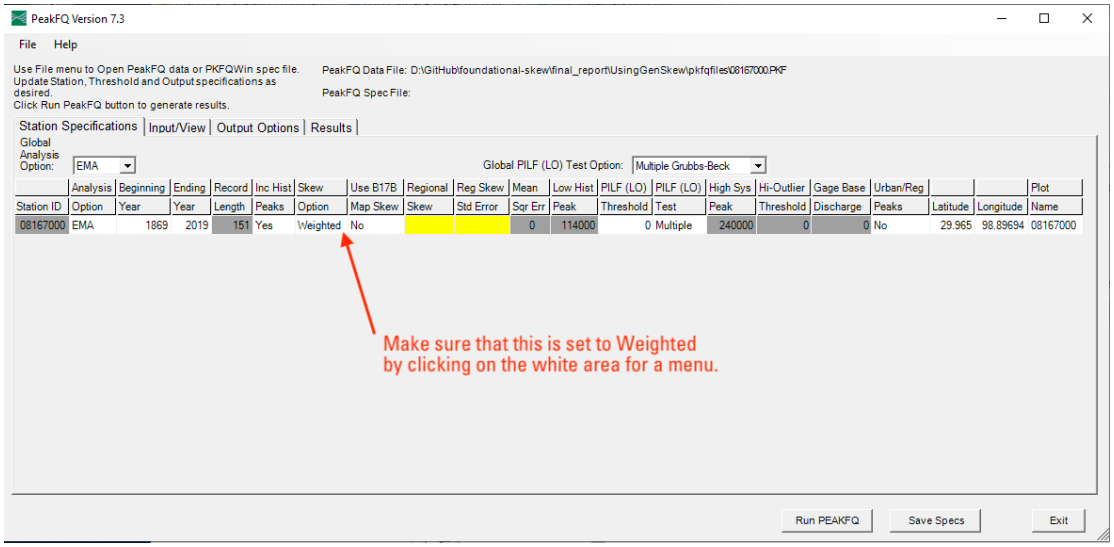


Figure 5.5. Station specification screen of USGS PeakFQ software 7.3 showing setting (default) to weighted skew computation along with the yellow highlighting that the generalized skew and its standard error needs input by the user.

Figure 5.5 depicts PeakFQ 7.3 after loading the file into the program and presenting the default tab sheet **Station Specifications**. The designer will select the **Weighted Skew** option (column 7) and then will enter values in the **Regional Skew** (column 9) and **Regional Skew Standard Error** (column 10) entry cells.[1]

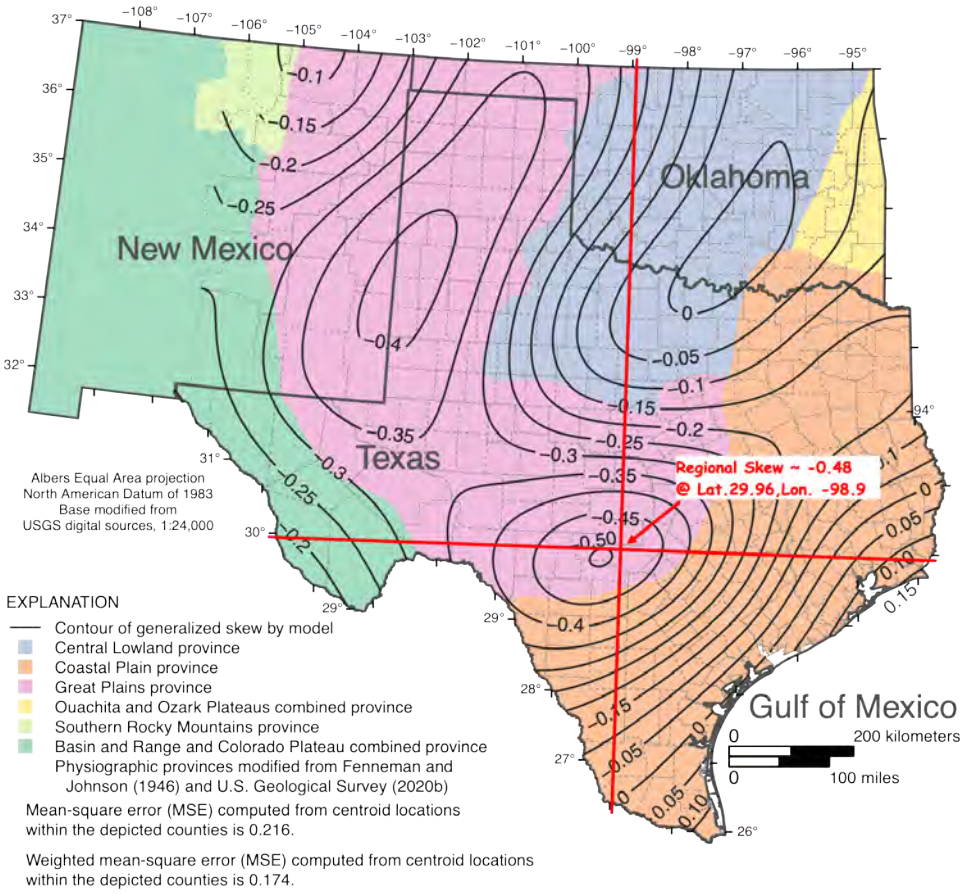


Figure 5.6. Estimating regional skew from map (fig. 4.6 of chapter 4) for USGS streamgage 08167000 Guadalupe River at Comfort, Texas.

Figure 5.6 depicts reading the skew value from the map for the particular streamgage location, which in this case is a value of -0.48 to be entered into the **Regional Skew** (column 9) entry cell. The mean square error is also reported in the map legend as 0.216; the square root of this value ($\sqrt{0.216} = 0.465$) is supplied to the **Regional Skew Standard Error** (column 10) entry cell.

[1] The values for these entires are sourced from figure 4.6.

Figure 5.7 depicts the entry of the regional skew and its standard error; upon completion, the designer would check the dates, choose the **EMA** option, and address any warnings or error messages then proceed to the **Input/View** tab.

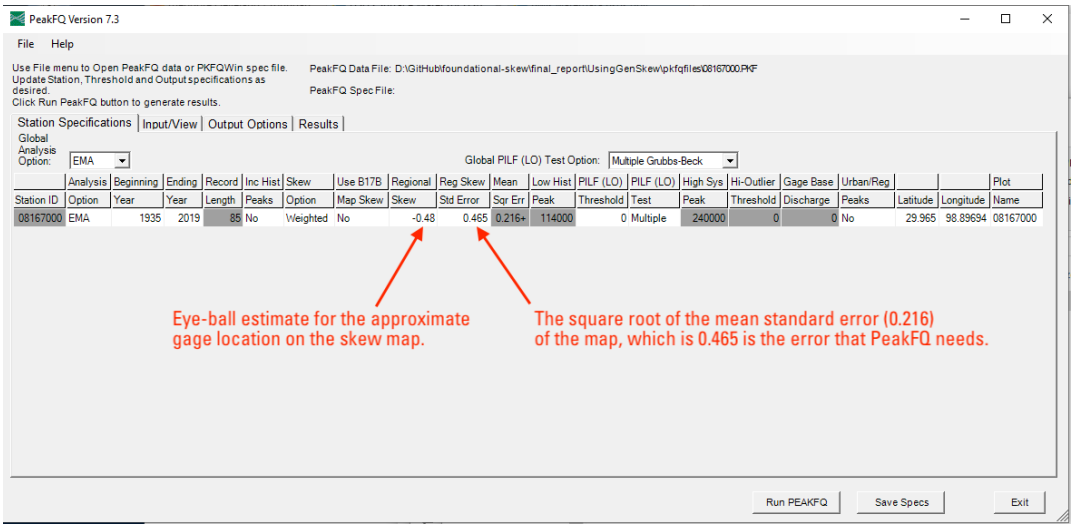


Figure 5.7. Station specification screen of USGS PeakFQ software 7.3 showing setting of the generalized skew and its standard error by the user for USGS streamgage 08167000 Guadalupe River at Comfort, Texas.

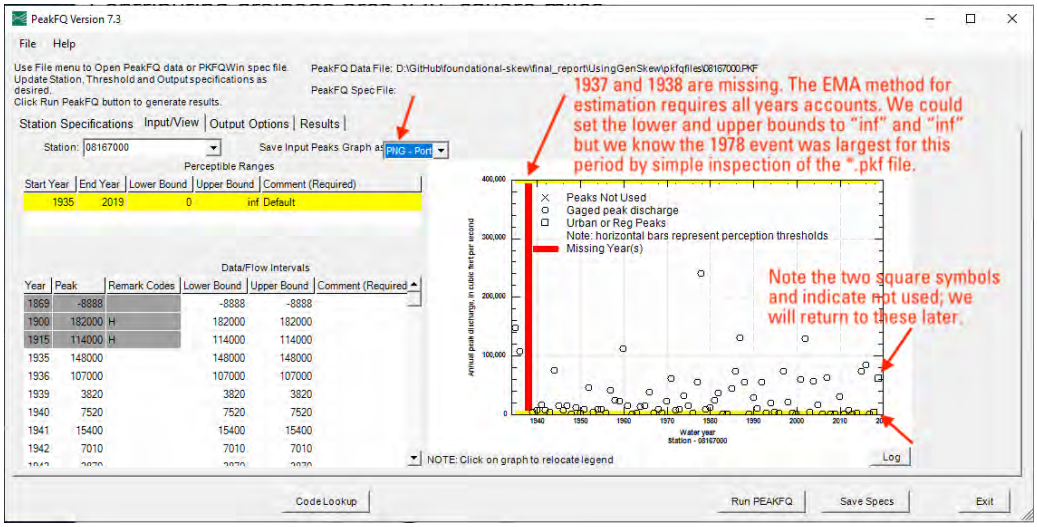


Figure 5.8. Input/View screen of USGS PeakFQ software 7.3 showing that a 1937–38 gap in record has been detected and accommodation by the user is needed and the setting of the output graphic format for USGS streamgage 08167000 Guadalupe River at Comfort, Texas.

Figure 5.8 depicts the **Input/View** tab, an important input control dialog that presents a graphical representation of peak streamflows loaded from the file (annotation in the figure), and explicitly identifies missing dates. The expected moments algorithm (**EMA**) requires temporally contiguous data, hence missing dates must be explicitly addressed.

In this example, a choice was made to substitute the 1978 annual peak and declare it as a lower bound for any missing dates, essentially distributing the empirical probability of the observed maximum among those missing years.

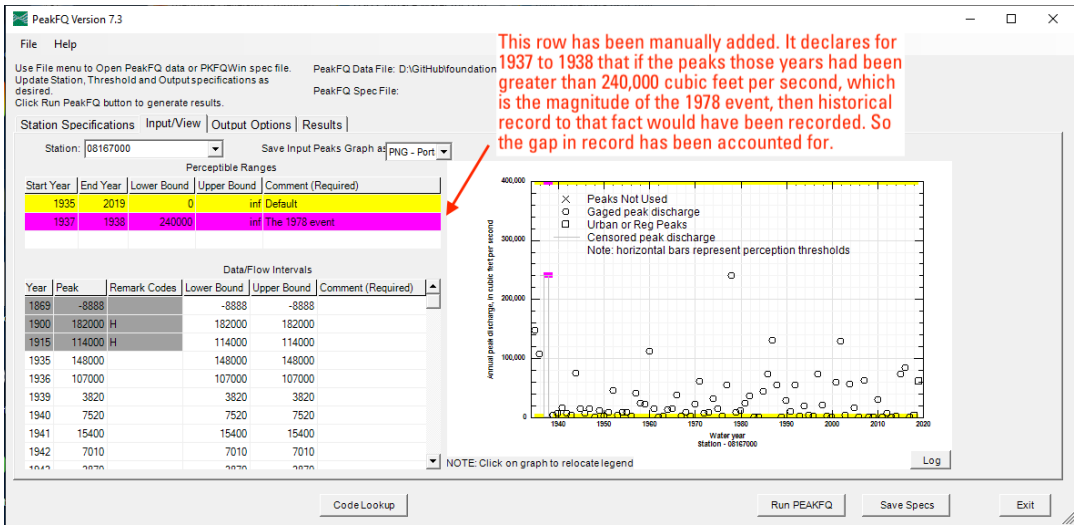


Figure 5.9. Input/View screen of USGS PeakFQ software 7.3 showing the accommodation for the gap in record by treating the 1978 peak as the perception threshold for the 1937–38 gap for USGS streamgage 08167000 Guadalupe River at Comfort, Texas.

Figure 5.9 depicts such an entry. Other choices could have been implemented. The decision to use the largest observed accounts for the likelihood (colloquial, not statistical) that had those years been historical maxima, there would have been some notation to that effect, hence these years probably did not exceed that value—and this is the source of the term "perception threshold" in the the expected moments algorithm.

On this input control dialog box, one chooses graphical output file type of `.png` portable network graphics, a common image format that most operating systems can render, and more importantly most browsers can render.

Figure 5.10 is the **Output Options** tab. Here we also exercise control of output conditions, after these various inputs are set, the designer runs the analysis from here. Upon completion of the analysis (less than a few seconds) the designer can examine results using the **Results** tab.

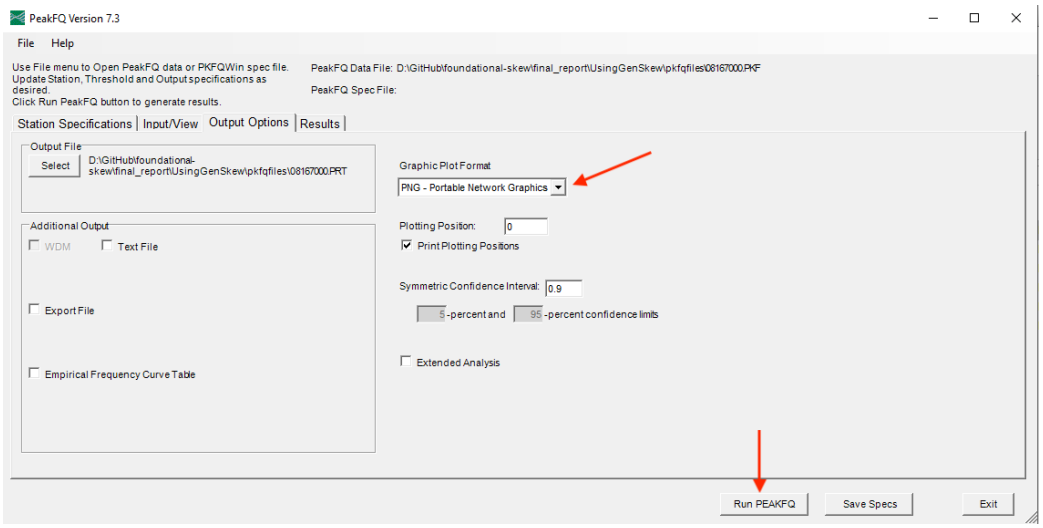


Figure 5.10. Output options screen of USGS PeakFQ software 7.3 showing a setting of the output graphic format and the button to perform the analysis for USGS streamgage 08167000 Guadalupe River at Comfort, Texas.

Figure 5.11 is a screen capture of the results tab; it is informative to examine the software generated flood-frequency curve, by selecting the appropriate curve (in this case there is only one, but multiple stations can be analyzed in a single instantiation of PeakFQ)

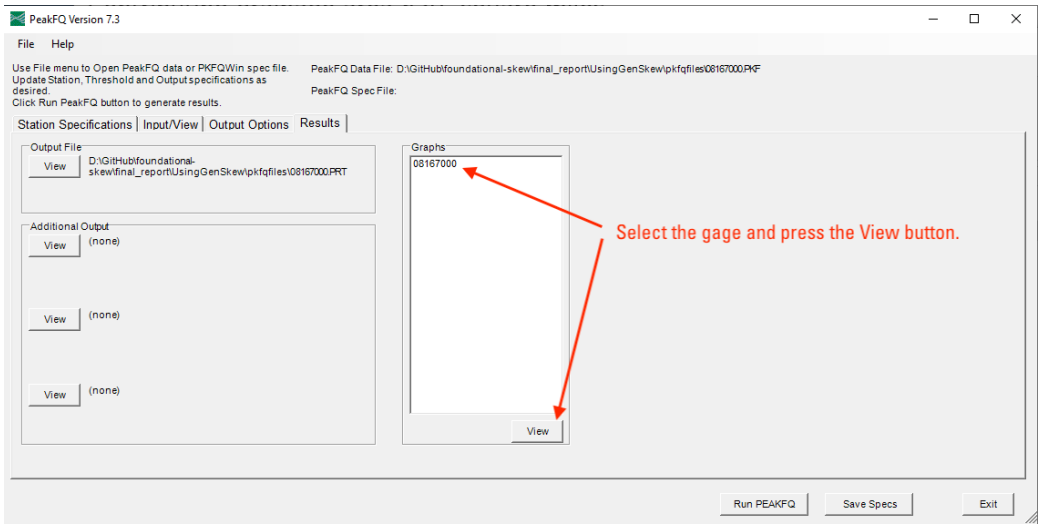


Figure 5.11. Results screen of USGS PeakFQ software 7.3 showing the selection of the site to show using the view button (see fig. 5.12) for USGS streamgage 08167000 Guadalupe River at Comfort, Texas.

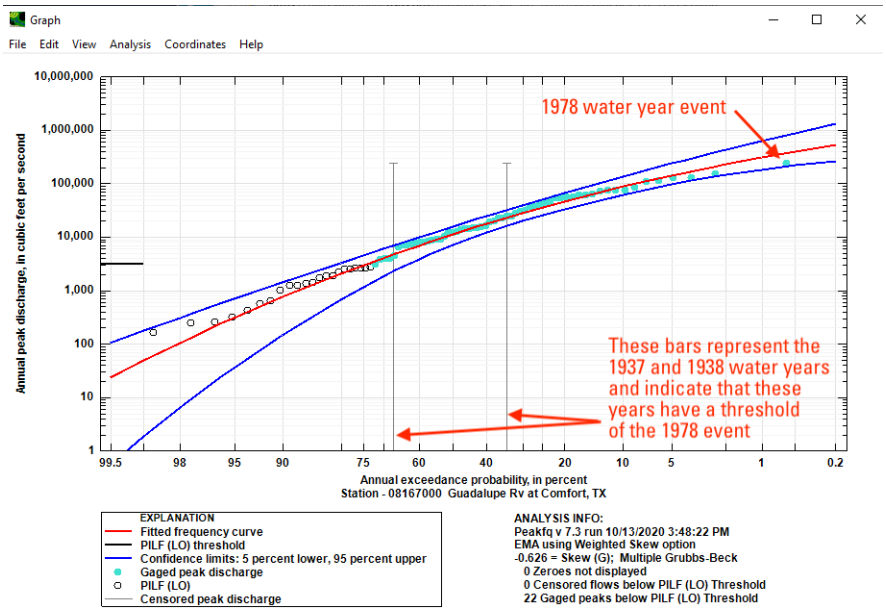


Figure 5.12. Results of flood-frequency analysis for USGS streamgage 08167000 using 81 peak streamflows where the 2018 and 2019 water year peak streamflows were "accidentally" not used for USGS streamgage 08167000 Guadalupe River at Comfort, Texas.

Figure 5.12 is the software generated flood-frequency curve. The red curve is the fitted curve (the Pearson type III [log] model); the markers are the annual peak magnitudes and their plotting position using the Weibull plotting formula (handled in the software) (technically Hirsch–Stedinger plotting positions are used when historical information is included). Open markers are low-outliers identified by the software, closed markers are peak streamflows that are used in the EMA analysis. The blue curves are the 90-percent confidence interval estimates.

The two missing years perception thresholds are shown as grey vertical bars that partition the probability axis into three equal parts—the perception thresholds themselves are stationed at $\approx$ 34-percent and $\approx$ 68-percent. If there were more missing years, the partition count increases proportional to the number of perception thresholds applied.

Additional output information is supplied in an ASCII file (extension is `.prt`) as shown in figure 5.13. The beginning of the file repeats or echos the input settings used.

Further into the file, a tabular representation of the flood-frequency curve is presented as shown in Figure 5.14. The portion of the file shown lists the station skew and the weighted

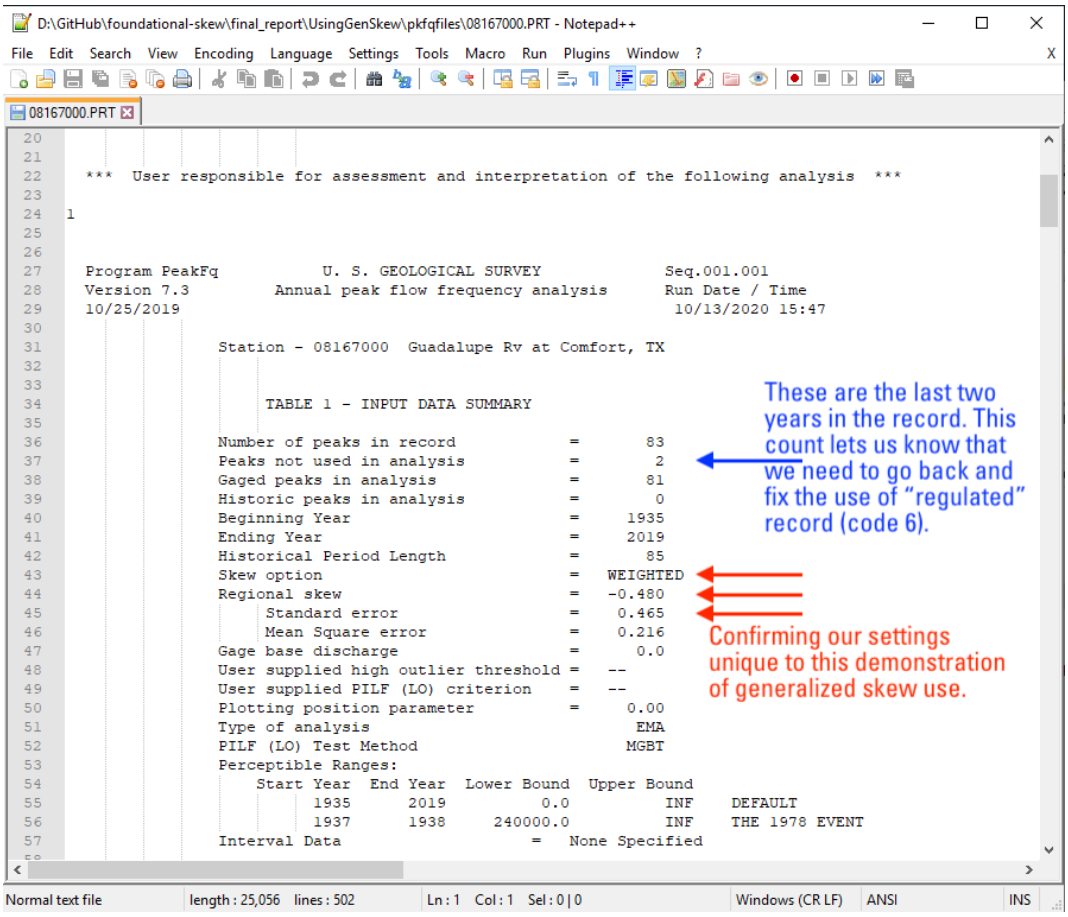


Figure 5.13. Inspection of the PRT output (text) file of USGS PeakFQ software 7.3 showing 81 peak streamflows used and confirmation that generalized skew settings were correct for USGS streamgage 08167000 Guadalupe River at Comfort, Texas.

skew based on the input values sourced from figure 4.6. The annotations in the figure highlight and show the result of using regional skew in contrast to individual station skew.

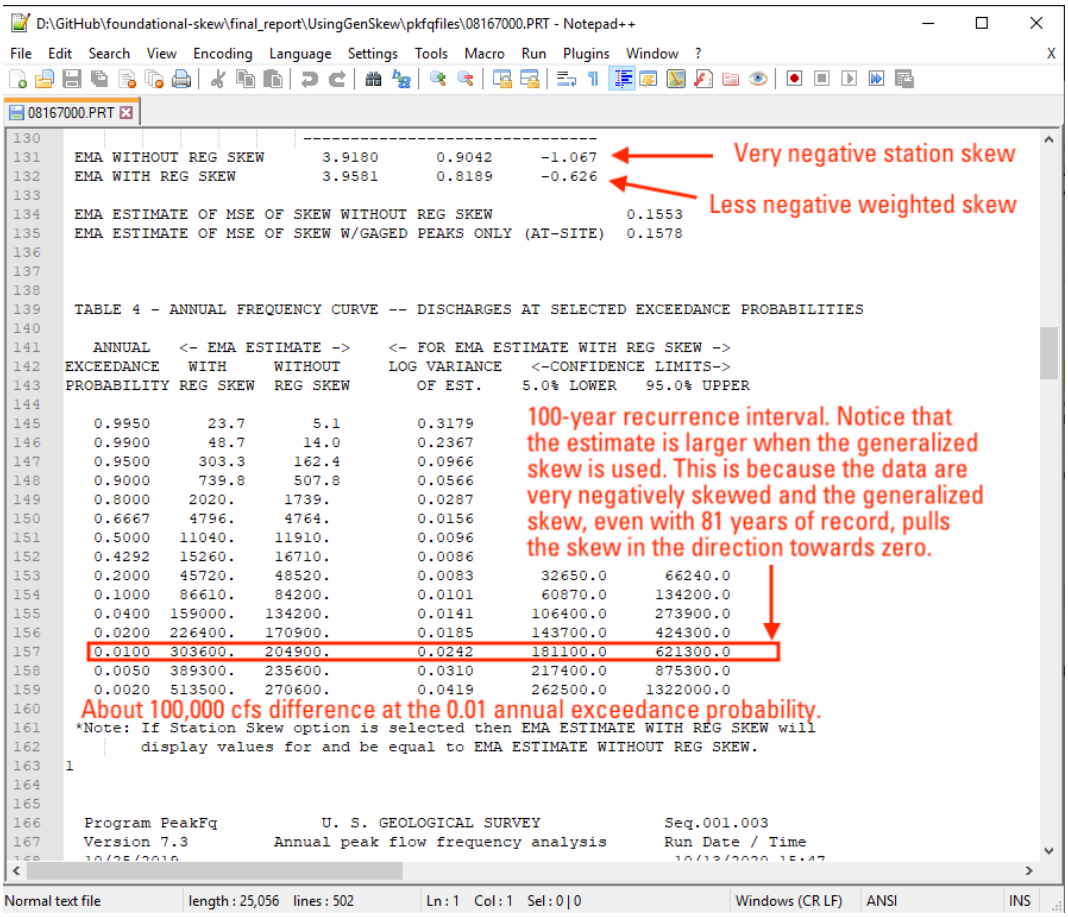


Figure 5.14. Inspection of the PRT output (text) file of USGS PeakFQ software 7.3 showing the flood-frequency values (see fig. 5.12) with emphasis on the 100-year recurrence interval with water years 2018 and 2019 omitted by accident for USGS streamgage 08167000 Guadalupe River at Comfort, Texas.

The MGBT code (Asquith, England, and Herrmann, 2020) developed as part of this research can read and parse NWIS files, and produces an informative graphic to guide designer use of PeakFQ. Figure 5.15 demonstrates and example for the example streamgage. The script produces a plot of recorded peak streamflows versus water year; missing years are displayed as a vertical dashed line segment, and specific NWIS discharge qualification codes are displayed with the respective water year where included in the data file. However, PeakFQ has its own overprint of discharge code nomenclature and documentation provides more detail.

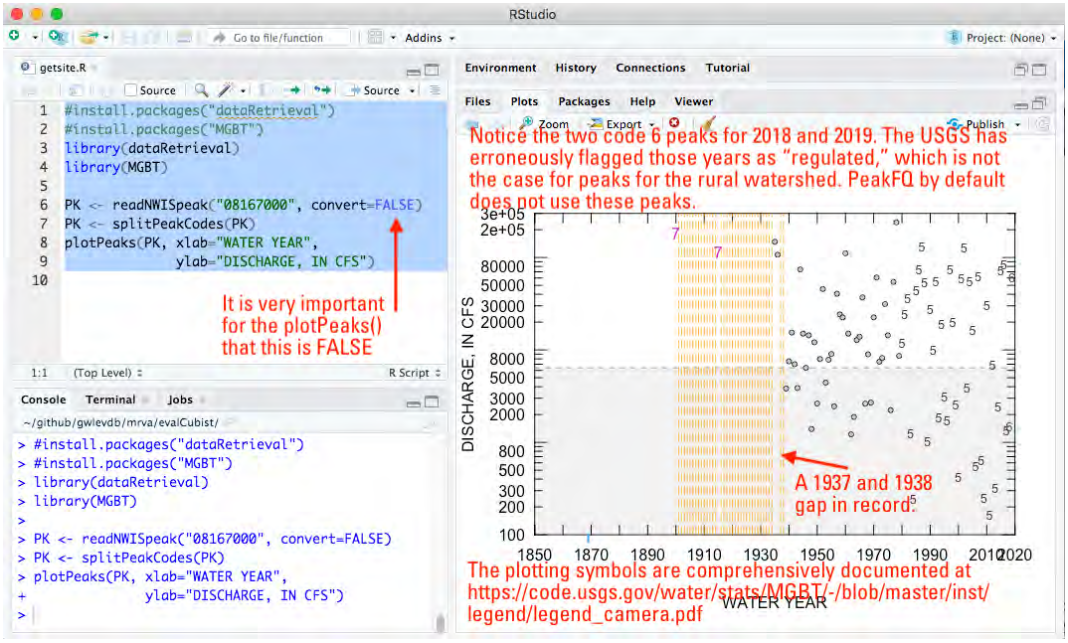


Figure 5.15. Using *RStudio* to dynamically retrieve USGS peak streamflow data and pass those data into the `plotPeaks()` function by Asquith, England, and Herrmann (2020) to illustrate the time series with sophisticated markup of the codes and other features of the data for the streamgage not visible in USGS PeakFQ software 7.3 graphics for USGS streamgage 08167000 Guadalupe River at Comfort, Texas. This method of visualizing annual peak streamflow data with features for viewing all discharge qualification codes and other features of the data is powerful and is a one-of-a-kind capability developed from this research project.

In the current example, water years 2018 and 2019 are coded as "6," which is considered as a flag for regulated discharge, which is incorrect for this location and the interest in the flood distribution and demonstrates potential weaknesses in USGS communications in NWIS concerning regulation. More details about discharge qualification codes are in Wagner et al. (2017) and chapter 8 of this report. Figure 5.16 shows the same result in the `.prt` file in a tabular format.

The designer can either edit the local copy of the NWIS file (generally not recommended) or select the **Urban/Reg Peaks** option on the **Station Specifications** tab in PeakFQ as shown in figure 5.17. Upon re-running of the analysis, the urban/regulated peak streamflows are included. Those that are above the low-outlier threshold are identified by a different marker type and color, as shown in figure 5.19.

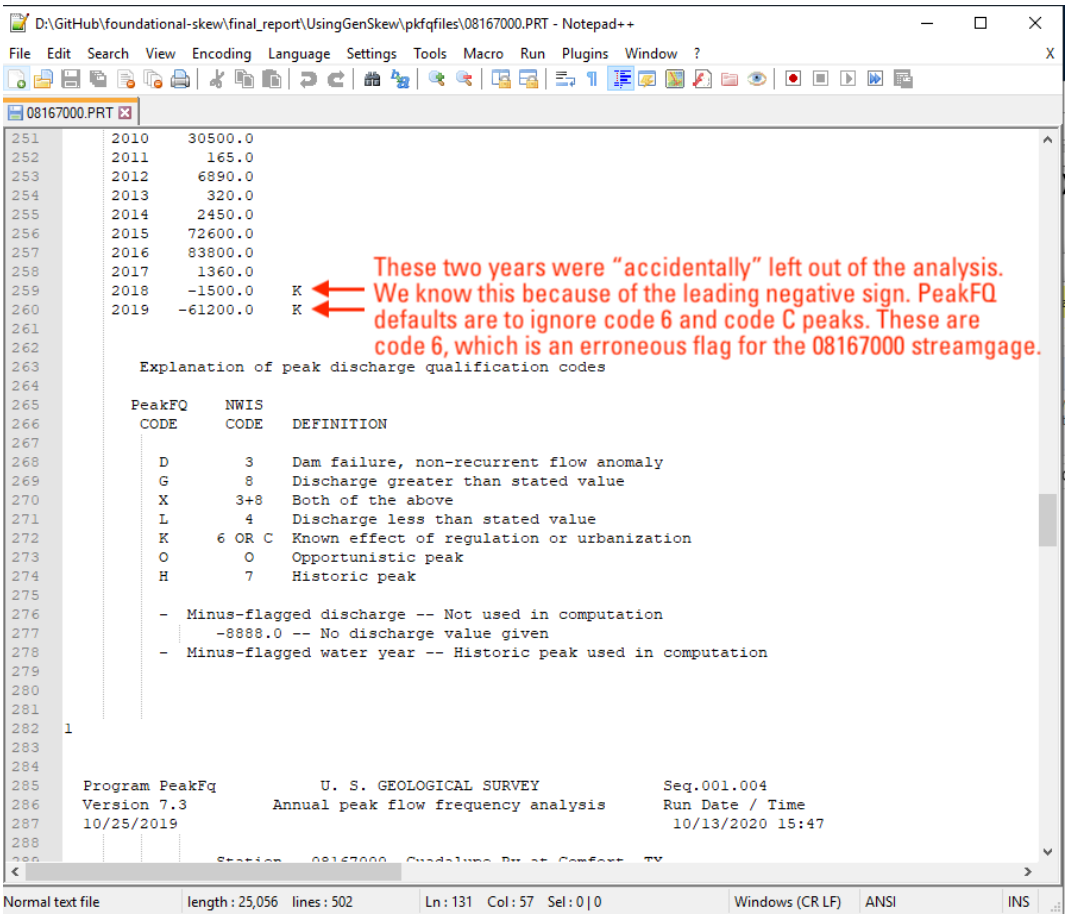


Figure 5.16. Inspection of the PRT output (text) file of USGS PeakFQ software 7.3 showing that the peak streamflows for water years 2018 and 2019 were not used in the analysis on "accident" because PeakFQ default settings are to ignore discharge qualification code 6 (these two years) or discharge qualification code C peak streamflows (see fig. 5.15) for USGS streamgage 08167000 Guadalupe River at Comfort, Texas.

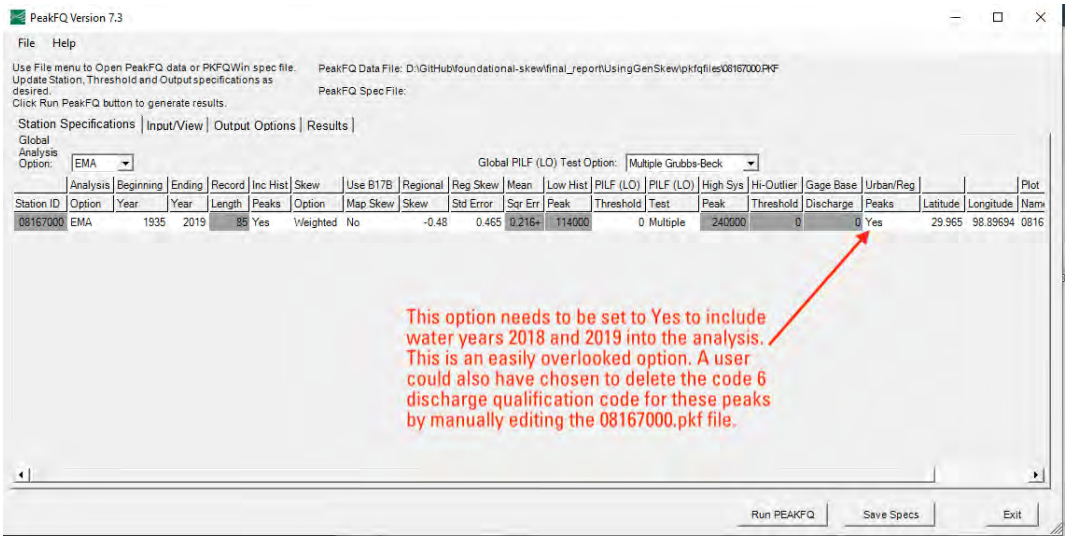


Figure 5.17. Station specifications screen of USGS PeakFQ software 7.3 showing change of setting to Yes to use regulated and(or) urban peak streamflows in the analysis for USGS streamgage 08167000 Guadalupe River at Comfort, Texas.

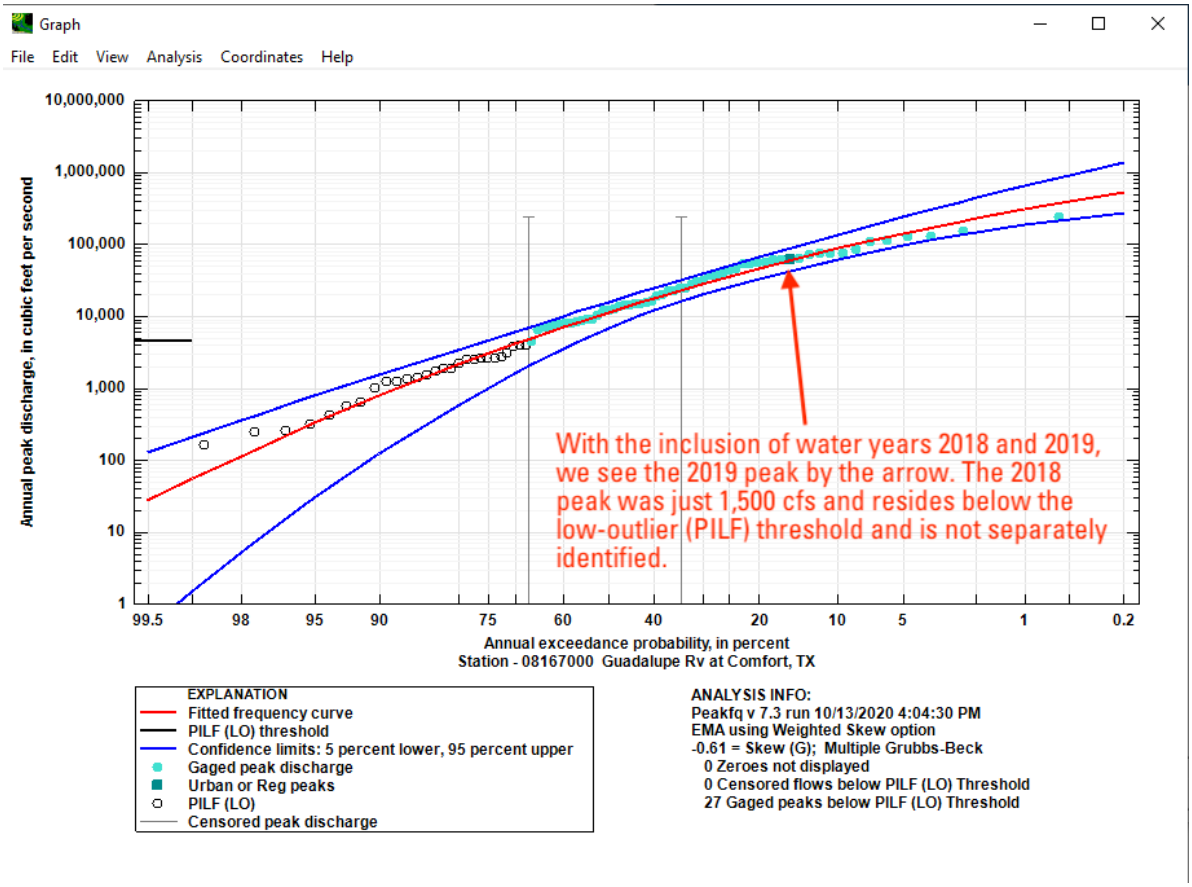


Figure 5.18. Results of flood-frequency analysis for USGS streamgage 08167000 Guadalupe River at Comfort, Texas, using 81 peak streamflows where the 2018 and 2019 water year peak streamflows were used in contrast to results in figure 5.12.

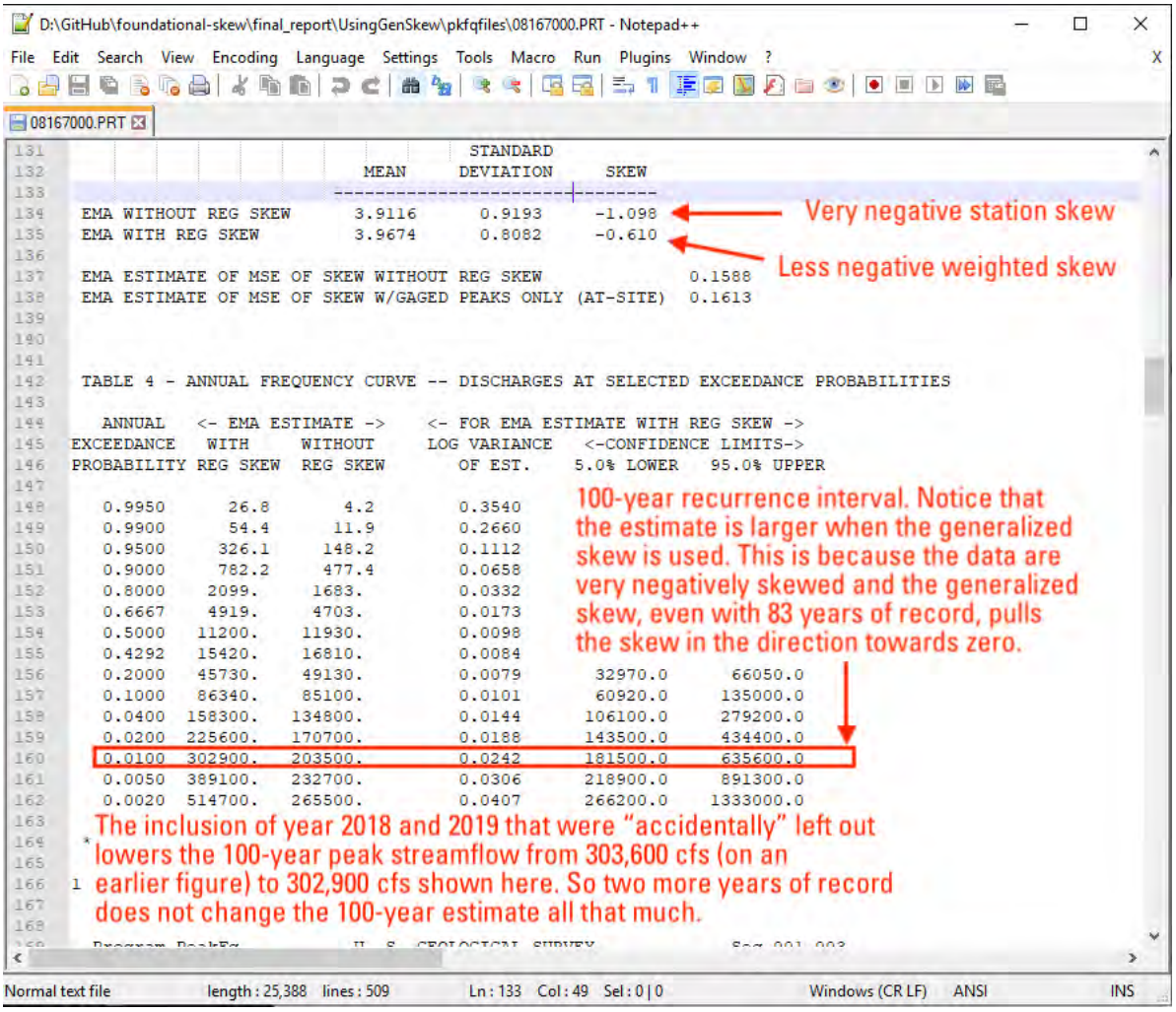


Figure 5.19. Inspection of the PRT output (text) file of USGS PeakFQ software 7.3 for USGS streamgage 08167000 Guadalupe River at Comfort, Texas, showing the flood-frequency values (see fig. 5.18) with emphasis on the 100-year return period with water years 2018 and 2019 included in contrast to results in figure 5.14 with those years omitted by accident.

5.2. Prototype On-Line Training Environment

A prototype on-line implementation of PeakFQ software 7.3 was developed as a training tool, as well as a usable map lookup tool (directly extracts regional skew estimates from figure 4.6 using designer supplied latitude and longitude values). The pre-configured training prototype is a Windows-like desktop running on an Amazon Web Services virtual private server, presented here as an example common tool. To the practicing engineer it functions like an ordinary workstation, but it is already populated with the relevant data files from this research.[2]

5.2.1. Connecting to the Remote Tools

As presented here, the on-line implementation uses Microsoft Remote Desktop client, which is available in the Windows operating system by default.[3] The necessary credentials that will be requested in a designer's connection using RDP are:

- PC Name :: `kittyinthewindow.ddns.net`, which is a pre-configured Windows 2019 server running on Amazon Web Services.
- User Name :: `texas-skew`
- Password :: `peakfq73$hare`

Figure 5.20 is a screen capture depicting a user search for the remote desktop connection client. Simply search for RDP (remote desktop protocol) and the OS returns the link to the installed software. Upon finding the application, the user starts the application and will be immediately prompted for the remote server name; in this example it is `kittyinthewindow.ddns.net`.[4] Figure 5.21 is a screen capture depicting the interaction with the RDP client where the remote name has been entered.

Next, the user will supply the remaining credentials to the remote server as depicted in figure 5.22.

[2] Moving files to/from the remote workstation is elaborate; it is intended for training in a preconfigured environment. For routine use, a design engineer would find the various tools would work faster and files immediately accessible if the USGS software (PeakFQ) are downloaded and installed on their local workstation, and similarly the scripts (and entire repository) downloaded from Cleveland and Fang (2021).

[3] Other operating systems; MacOS, Chromium, Linux will require installation of Microsoft RDC which is freely available from Microsoft. The authors believe that the tools can be adapted to work through a web interface so the engineer could train/use the tools from any device with a browser and internet connection.

[4] This server is not long-term persistent, however the authors maintain it for other experiments, and it can be expected to operate until calendar year 2023. The virtual machine files are transferrable (or buildable quite quickly) to the department (TxDOT) for long-term training use or even routine use.

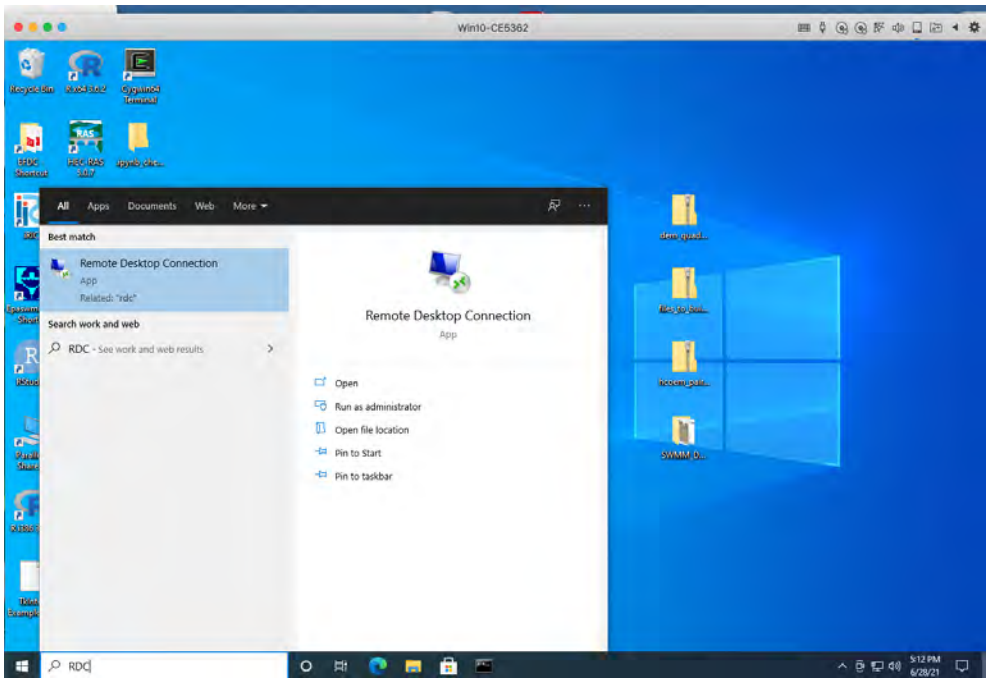


Figure 5.20. Find the RDP client in host operating system.

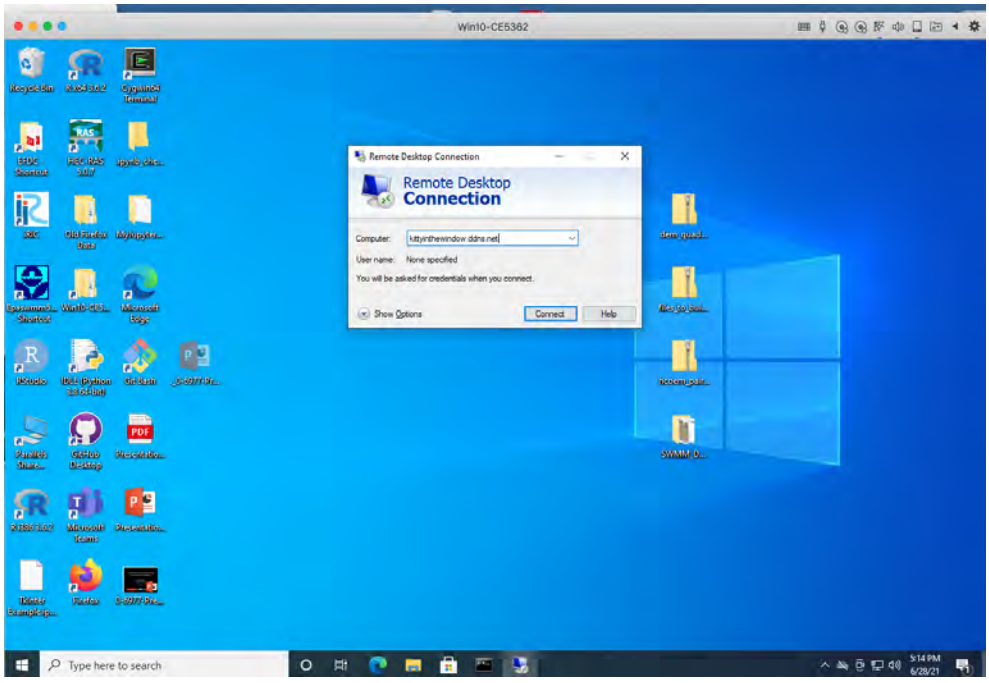


Figure 5.21. Start the RDP connection to the remote (AWS) server.

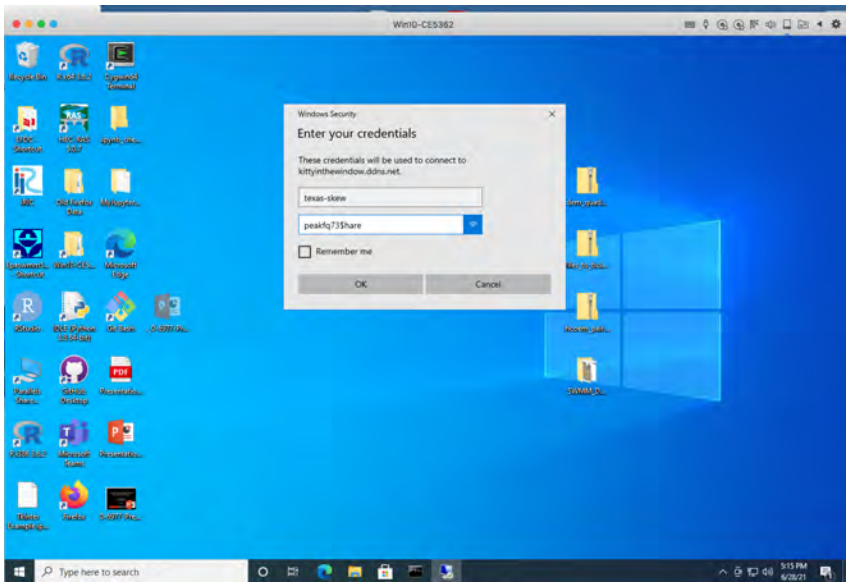


Figure 5.22. Send the credentials to the remote (AWS) server.

And then the user has to set an exception because the remote server does not use a third-party certificate authority.[5] The user will select "YES" to connect to the server as depicted in figure 5.23.

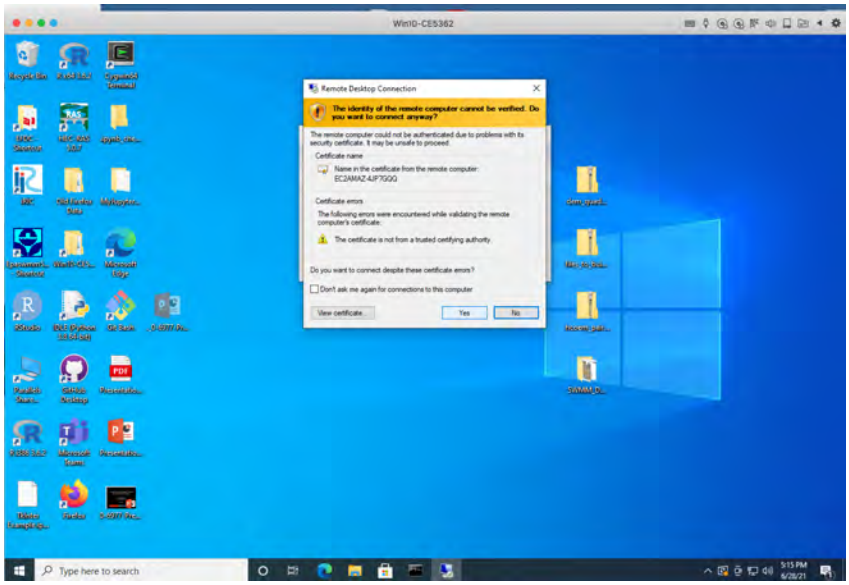


Figure 5.23. Accept the self-signed certificate on the remote (AWS) server.

[5] This step can be eliminated if the remote server subscribes to a certificate authority—however for this research prototype a self-signed certificate was used; hence the correct security warning.

Upon connection a bit of time elapses (usually a few seconds), and then the user is presented with a desktop that looks like figure 5.24 (without the annotations).

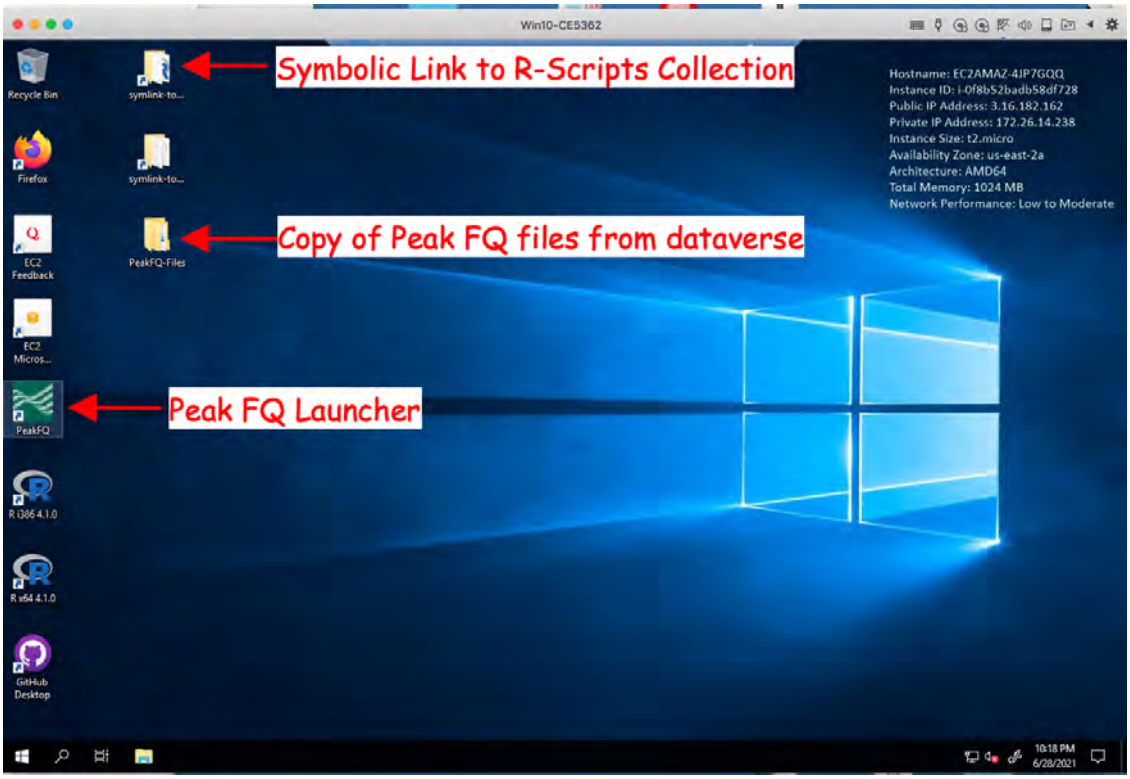


Figure 5.24. Windows desktop on the remote (AWS) server.

The annotations identify two important components that are pre-configured: (1) A symbolic link to the R-Scripts collection from Cleveland and Fang (2021) in the correct directory structure, (2) A folder named `PeakFQ-Files` that is a convenient place to store work, and has a copy of the PeakFQ files used on the 444 long-term streamgage list, and a link to start the PeakFQ software.[6]

5.2.2. Using the Remote Tools to Analyze a Streamgage

The next several pages present a use case to analyze streamgage 08070000 East Fork of San Jacinto River near Cleveland, Texas. The streamgage was part of the 444 long-record streamgages used in this study. The example will access the NWIS database and download a current streamgage annual peak streamflow file.

[6] The PeakFQ software has a restricted filename length; the file copy was an expedient solution for readers who wish to examine the behavior of the software using the identical files that the researchers employed.

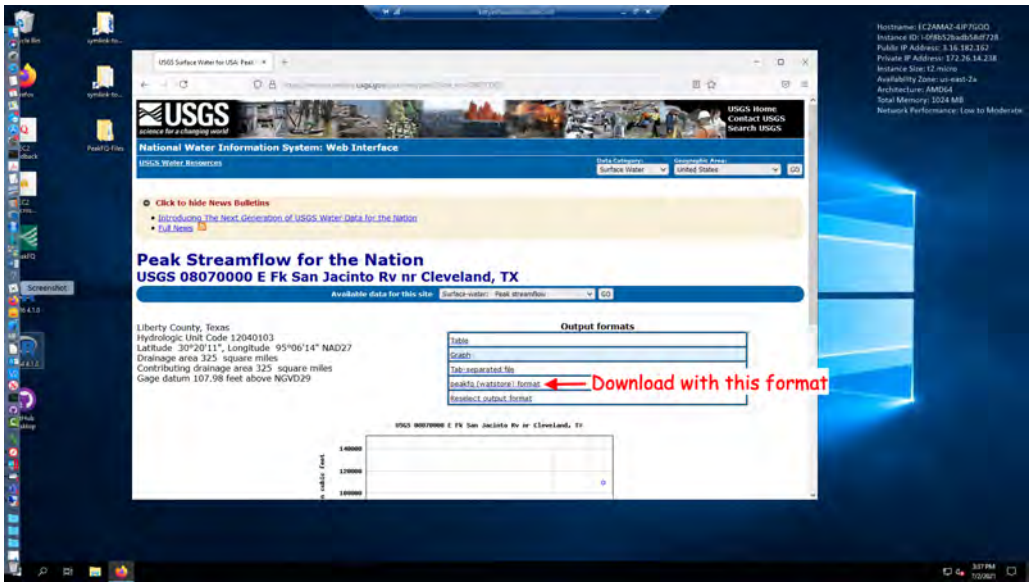


Figure 5.25. USGS NWIS interface after navigation to the streamgage and selection of peak streamflow data.

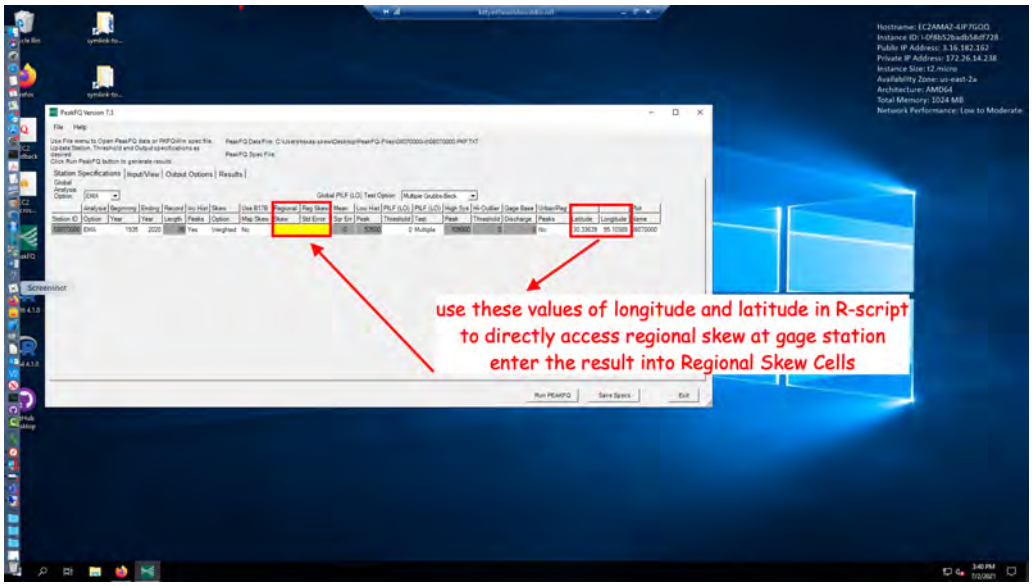


Figure 5.26. USGS PeakFQ software 7.3 interface but the user will need to use manual latitude and longitude (negative values are required) to look up the generalized (regional) skew from this study. To clarify, the PeakFQ software 7.3 does not have the updated skew from this project embedded into its subsystems.

On the remote server, open the web browser and navigate to the NWIS as in figure 5.25 and download streamgage data. Then start PeakFQ as in figure 5.26.

Next open the symbolic link to the R-scripts (or use figure 4.6 directly). Select the `demo03_latlongTOgenskew.R` and open the script using `RStudio` (**R** and **RStudio** are pre-installed in the training prototype environment).

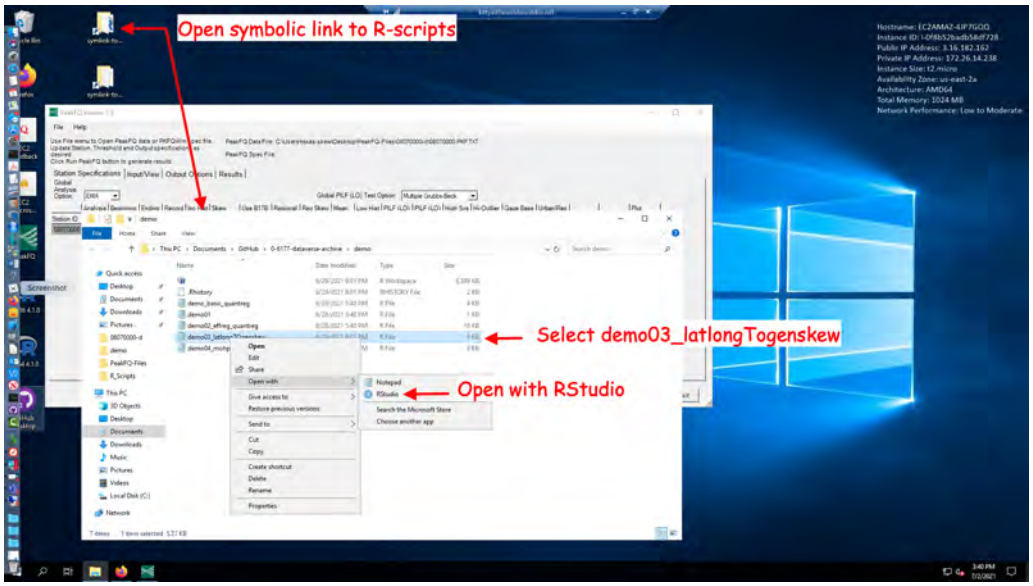


Figure 5.27. Launch script in R language to look up regional skew directly from a 1-kilometer gridded representation of predicted generalized skew. It takes awhile to load on the remote, so a short delay is typical.

The script takes a short time to start on the remote machine (about 10 seconds) and will appear similar to figure 5.28 (without the annotations or graphics). The script is run by selecting the **Source** icon to run the entire script that appears in the code window in the upper left panel.[7]

The console (lower left panel in RStudio) prompts for longitude and latitude, which are read/copied copied from PeakFQ. Input format is `longitude, latitude` in decimal degrees (the comma is important and is the delimiter for the input). Upon completion, pressing return instructs the script to access the GAM-built raster (predicted by the generalized additive model) and recover the values of skew based on location. These are then reported in the output (centroid value and streamgage location value as described in chapter 4, and a graphic is rendered that is useful to verify the location; the graphic is helpful to the engineer to verify that the geographic location used is the anticipated location.

[7] The prototype is pre-configured so the script traverses the correct file paths; users who download the entire dataverse should be able to use the script exactly as depicted in these figures.

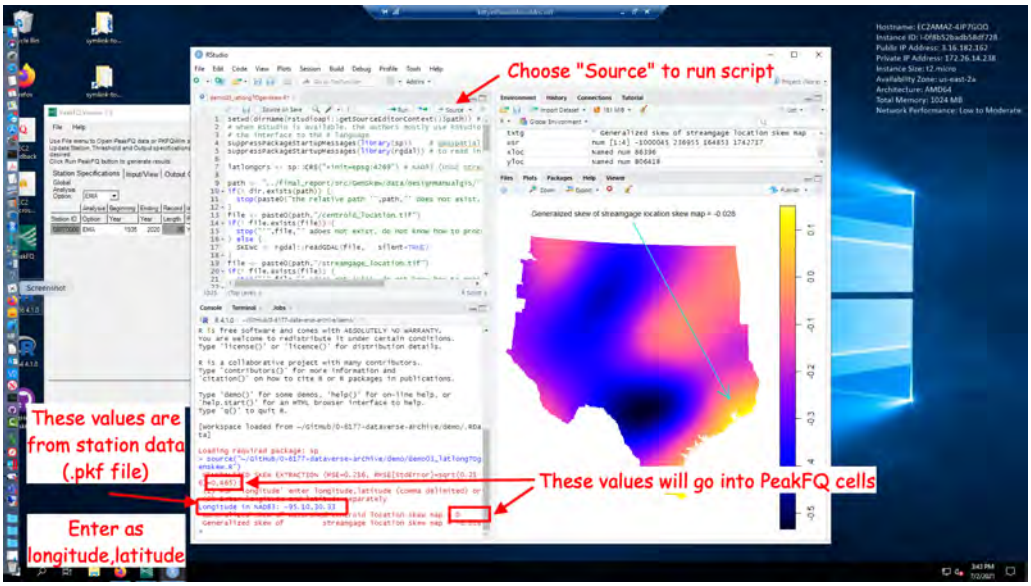


Figure 5.28. Input the decimal latitude and longitude into the console, copy-and-paste from USGS PeakFQ software 7.3 works fine, when inputs are complete, press return and the script `demo03_latlongTOgenskew.R` from Cleveland and Fang (2021) reads the 1-kilometer gridded representation of figure 4.6 in chapter 4 and a map is rendered to help confirm that the latitude and longitude (negative values are required) are in the user's anticipated location.

In this example, the values returned that will be put into the regional skew cells in PeakFQ software for streamgage 08070000 are generalized skew of watershed centroid of 0.0 and a standard error of 0.465.

Figure 5.29 depicts the appearance of the interface after the values from the script are supplied, the remainder of the configuration for the analysis is the same as the introductory example. After running the analysis the results can be displayed as in figure 5.30.

Inspection of PeakFQ output results for the station-skew option show that the 100-year return period estimate is 84,950 cubic feet per second with 90-percent prediction interval bounds of [55,180; 167,300], whereas weighting the station skew with the regional skew produce a 100-year estimate of 87,100 cubic feet per second with 90-percent prediction interval bounds of [57,080; 164,800].

For this particular streamgage, the 100-year return period estimate increases by about 2-percent when using the weighted skew estimate, and the multiple between the prediction intervals bounds decreases from 3.03 to 2.88 showing uncertainty contraction attributed to weighted skew being closer to zero than the station skew and about 25-percent smaller (station skew is -0.113; weighted skew is -0.085).

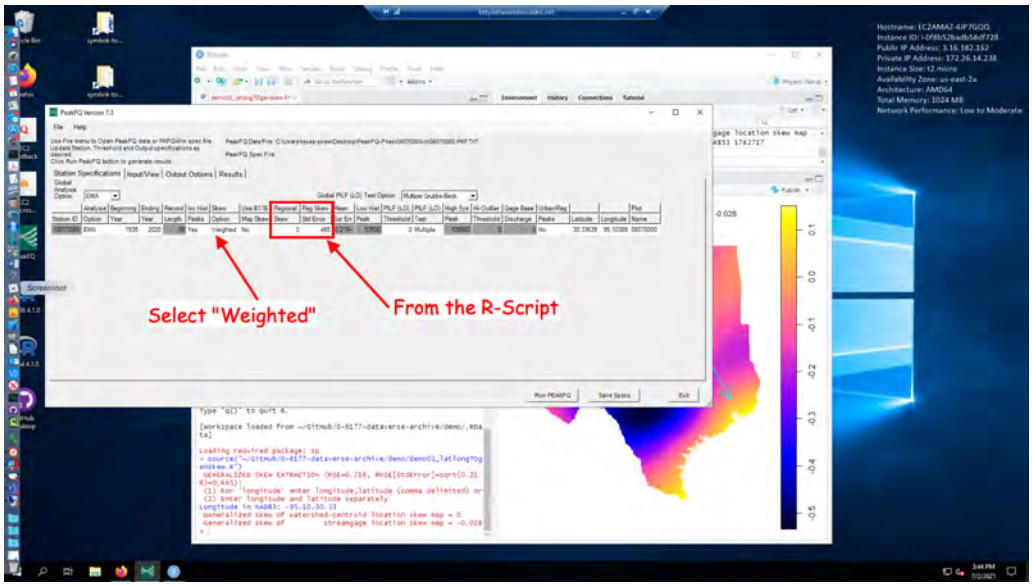


Figure 5.29. PeakFQ interface ready to analyze streamgage 08070000 East Fork of San Jacinto River near Cleveland, Texas.

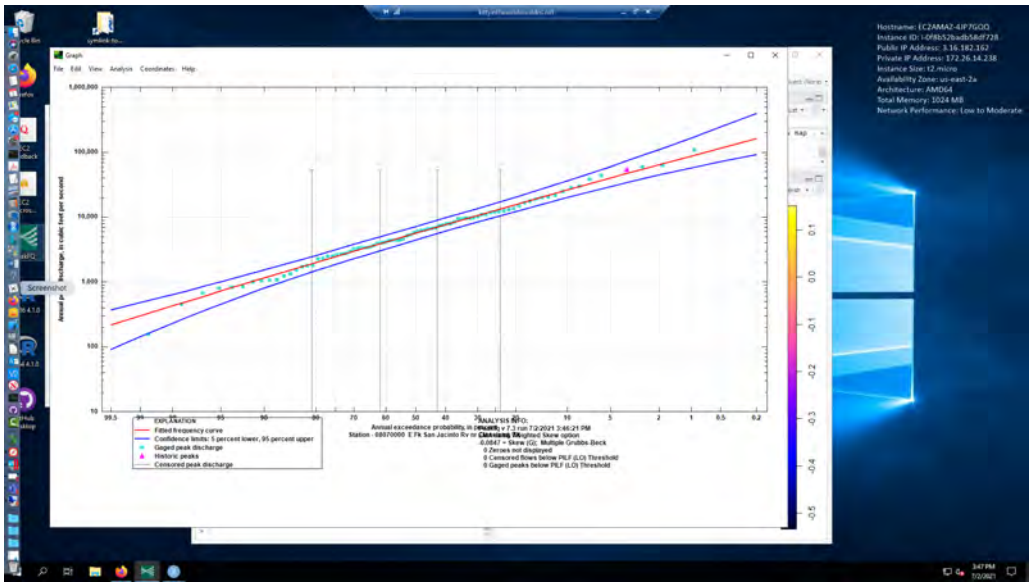


Figure 5.30. PeakFQ flood-frequency plot for streamgage 08070000 East Fork of San Jacinto River near Cleveland, Texas.

Nearby (in the sense of in same part of Texas) streamgage 08067000 Trinity River at Liberty, Texas is a streamgage with a positive regional skew value of 0.062, a station skew of -0.623, and weighted skew of -0.229. The resulting 100-year estimate using station skew is

128,100 cubic feet per second with 90-percent prediction interval bounds of [109,800; 165,600]. The resulting 100-year return period estimate using the weighted skew is 144, 700 cubic feet per second with 90-percent prediction interval bounds of [120,000; 193,900].

For this particular streamgage the 100-year return period estimate increases by about 12-percent when using the weighted skew estimate, and the multiple between the prediction intervals bounds increases from 1.51 to 1.61 showing some uncertainty expansion.

Whereas section was presented simply to illustrate use of the tools, a few more examples follow that illustrate interesting behavior and impacts of using weighted skew.

5.3. Selected Flood-Frequency Analyses Demonstrating Impact of the Updated Generalized Skew

5.3.1. Streamgage 08080750 Callahan Draw near Lockney, Texas

Streamgage 08080750 Callahan Draw near Lockney, Texas, was selected for another example of flood-frequency analysis using the generalized skew developed from this study to demonstrate potentially impacts on the 100-year return period streamflow in particular. The input data for this streamgage are shown in figure 5.31 and a gap of about 30 years is evident. This is a streamgage that was operated by the USGS for TxDOT (then Texas Highway Department) in the late 1960s and early 1970s (just 9 years of record). The streamgage was reactivated and again operated for TxDOT beginning in 2006 and ongoing to the present (2021) (Asquith and Harwell, 2018; Asquith and others, 2018; Harwell and Asquith, 2011). Reactivation of a streamgage such as this capitalizes on the earlier record so that now (2021) the record length is 24 years and growing. This streamgage was not part of the 444 long-record streamgages used in the development of generalized skew in this study; however, the streamgage is present in other analyses described in this report.

The flood-frequency results for a station-skew option operation of the PeakFQ software and then again using the weighted-skew option is shown in the top and bottom, respectively, of figure 5.32. The generalized skew was set to -0.355 (watershed-centroid location skew map). The differences in the figure curves are subtle and the differences progressively increase deep into the upper tail (the right side of the plots). It is important to note that the lower part of the upper tail (say annual exceedance probabilities from 0.5 to 0.10) has effectively the same predictions from the two curves.

Inspection of PeakFQ output results for the station-skew option show that the 100-year estimate is 648 cubic feet per second with 90-percent confidence bounds of [384; 2,410] cubic feet per second. Conversely, inspection of PeakFQ output results for the weighted-skew option show that the 100-year estimate is 534 cubic feet per second with 90-percent confidence bounds

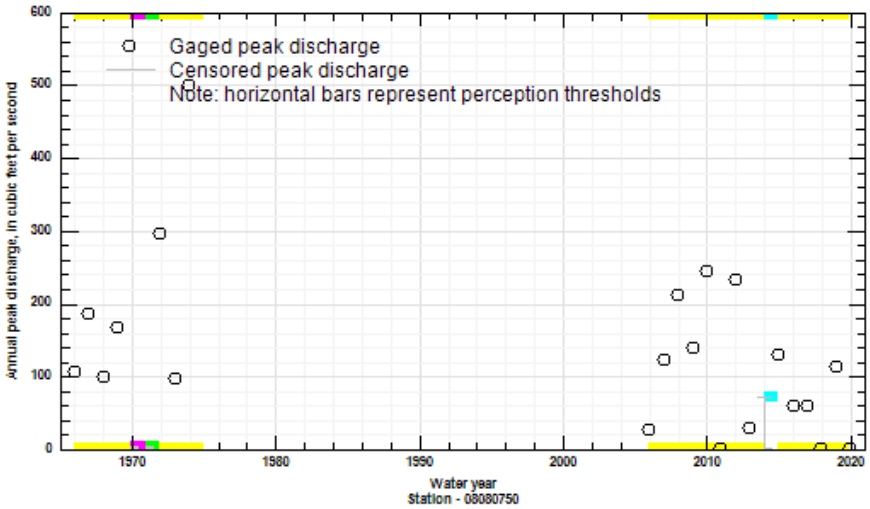


Figure 5.31. Input data plot from USGS PeakFQ software 7.3 for USGS streamgage 08080750 Callahan Draw near Lockney, Texas.

of [351; 1,900] cubic feet per second. So, for this example, the 100-year estimate decreases and the multiple between the confidence bounds goes from 6.27 to 5.41 (or 2,410/384 and 1,900/351, respectively), which shows contraction of the uncertainty bounds when information from generalized skew is incorporated into the analysis. However, part of the decrease in the 100-year estimate and the contraction of the bounds is also attributable to the fact that the skew in the computations of the frequency curve went from slightly positive for the station-skew option to negative for the weight-skew option as shown by the change in curvature direction of the frequency curves between the plots.

5.3.2. Streamgage 08148500 North Llano River near Junction, Texas

Streamgage 08148500 North Llano River near Junction, Texas, was selected for another example of flood-frequency analysis using the generalized skew developed from this study to demonstrate potentially impacts on the 100-year return period streamflow in particular. The input data for this streamgage are shown in figure 5.33. This streamgage was used in the development of the generalize skew product of this report because it has more than 30 years of record for natural watershed conditions.

The flood-frequency results for a station-skew option operation of the PeakFQ software and then again using the weighted-skew option is shown in the top and bottom, respectively, of figure 5.34. The generalized skew was set to -0.444 (watershed-centroid location skew map). The differences in the figure curves are readily seen for annual exceedance probabilities less than about 5 percent.

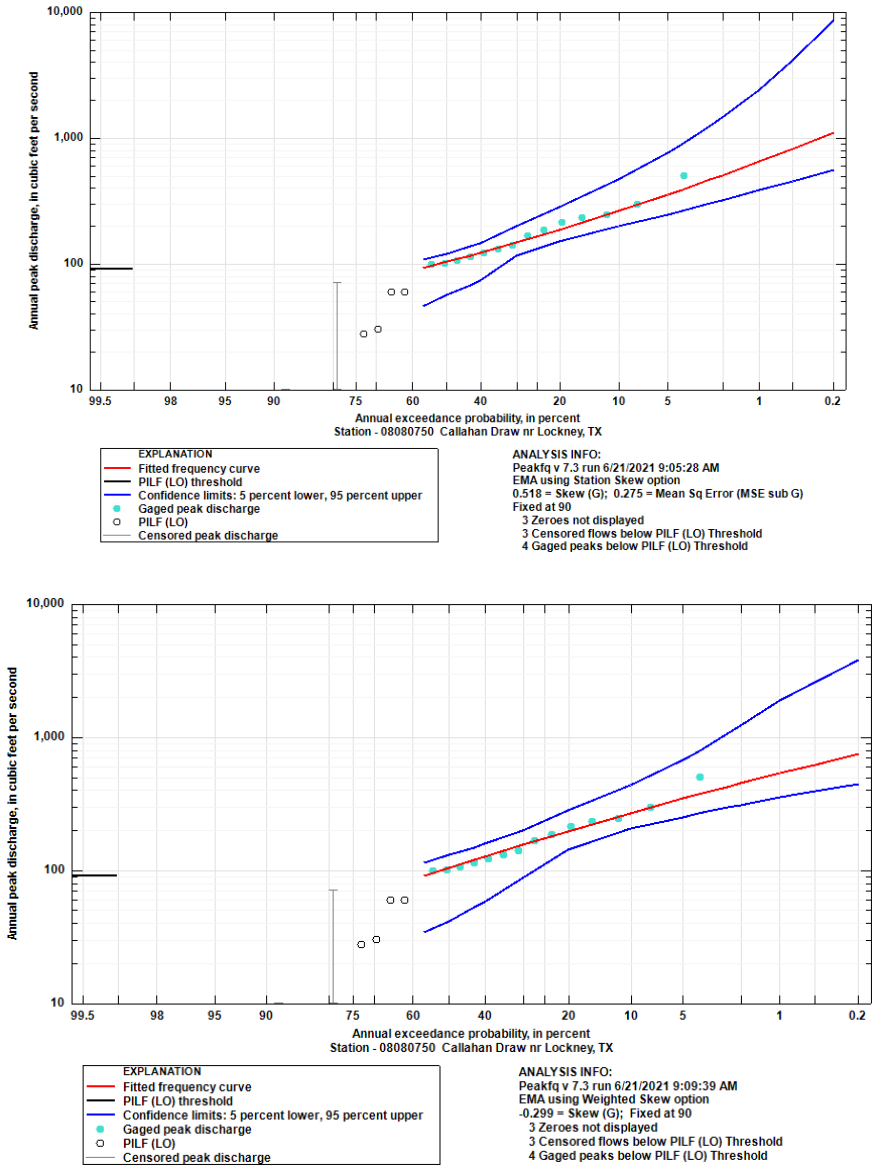


Figure 5.32. Flood-frequency plots using station skew (top) and weighted skew (bottom) based on a generalized skew value from this study from USGS PeakFQ software 7.3 for USGS streamgage 08080750 Callahan Draw near Lockney, Texas.

Inspection of PeakFQ output results for the station-skew option show that the 100-year estimate is 131,000 cubic feet per second with 90-percent confidence bounds of [96,300; 201,000] cubic feet per second. Conversely, inspection of PeakFQ output results for the weighted-skew option show that the 100-year estimate is 186,000 cubic feet per second with 90-percent confidence bounds of [120,000; 327,000] cubic feet per second. So, for this example, the

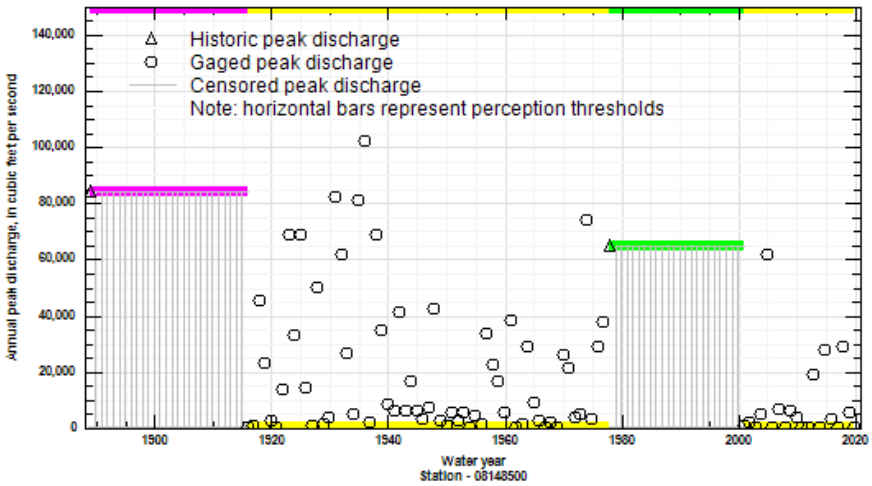


Figure 5.33. Input data plot from USGS PeakFQ software 7.3 for USGS streamgage 08148500 North Llano River near Junction, Texas.

100-year estimate increase and the multiple between the confidence bounds goes from 2.09 to 2.72 (or 201,000/96,300 and 327,000/120,000, respectively), which shows slight expansion of the uncertainty bounds when information from generalized skew is incorporated into the analysis. However, part of the increase in the 100-year estimate and the contraction of the bounds is also attributable to the fact that the skew in the computations of the frequency curves become less negative with the incorporation of the generalized skew.

5.4. Chapter Conclusions

The generalized skew suggested for updating the Hydraulic Design Manual is readily used in USGS PeakFQ software for flood-frequency analysis. All that is needed is for the user to select the generalized skew applicable to the streamgage. There is a slight author preference for use of the watershed-centroid location generalized skew map of chapter 4 and figure 4.6. The user manually inserts the generalized skew into the PeakFQ interface and inserts as well the root-mean-square error (RMSE) of 0.465. PeakFQ brands the generalized skew as **Regional Skew** in the interface and the brands the RMSE as **Regional Skew St[an]d[ard] Error**. PeakFQ will then automatically compute in the PeakFQ interface the square of the RMSE (mean-square error) but brands this value as **Mean Sq[ua]r[e] Error**. There is acknowledged confusion in the interface by both errors being stated, but only the error in native skew units (though dimensionless) of 0.465 requires entry. The user use chose the "Weighted" option for the **Skew Option** in order to use a weighted skew in the computations or "Regional" option for only the generalized skew to be used. Logic for reading a 1-kilometer gridded representation of the generalized skew is provided within Cleveland and Fang (2021)

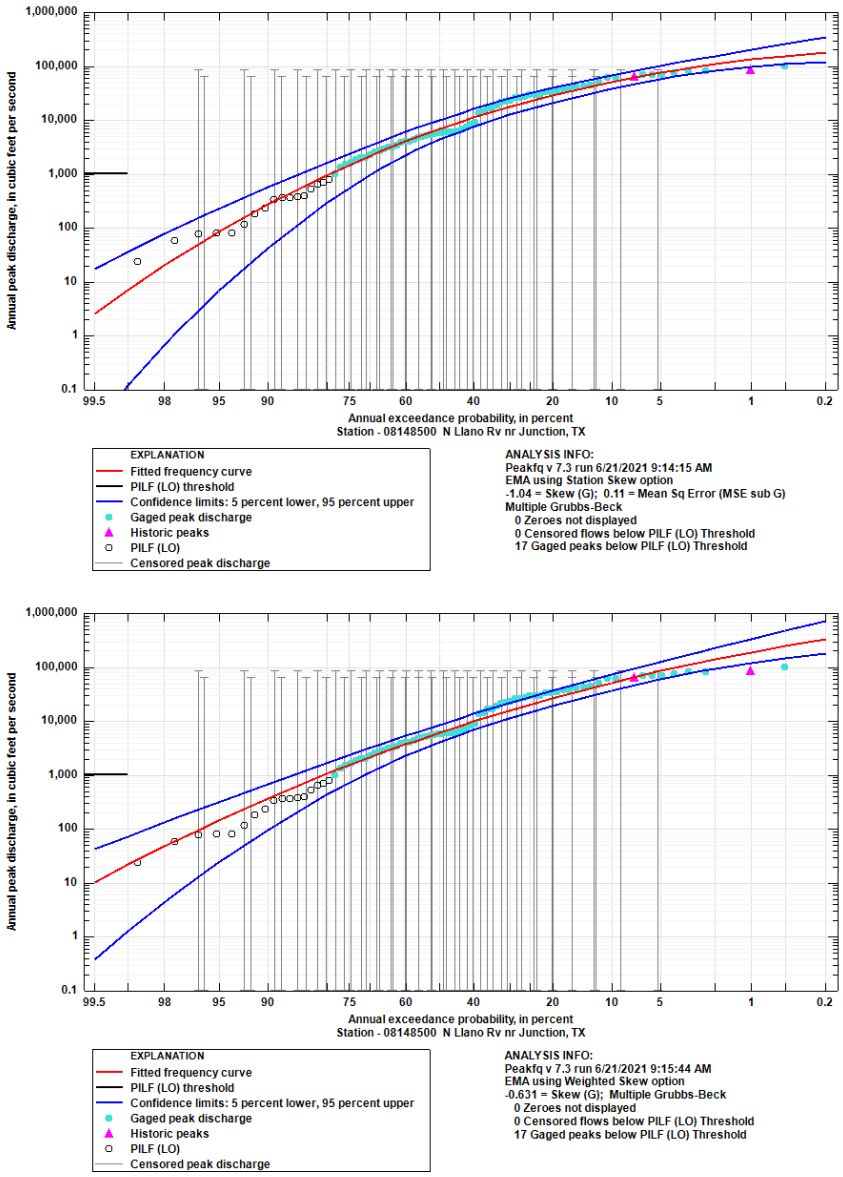


Figure 5.34. Flood-frequency plots using station skew (top) and weighted skew (bottom) based on a generalized skew value from this study from USGS PeakFQ software 7.3 for USGS streamgage 08148500 North Llano River near Junction, Texas.

along the file path: `0-6177-dataverse-archive/demo/demo03_latlongTOgenskew.R`. Finally, several examples are presented illustrating the use of the updated generalized skew, as is a prototype on-line training environment, and a tutorial-like presentation of its use.

Chapter References

Asquith, W.H., England, J.F., and Herrmann, G.R., 2020, *MGBT*—Multiple Grubbs–Beck low-outlier test: U.S. Geological Survey software release, R package, Reston, Va., accessed July 27, 2020, at https://doi.org/10.5066/P9CW9EF0.

Asquith, W.H., and Harwell, G.R., 2018, Database of peak streamflow derived from interpretations of indirect measurements for a crest-stage gage network in Texas through water year 2015: U.S. Geological Survey data release, https://doi.org/10.5066/F7057D39.

Asquith, W.H., Harwell, G.R., and Winters, K.E., 2018, Annual and approximately quarterly series peak streamflow derived from interpretations of indirect measurements for a crest-stage gage network in Texas through water year 2015: U.S. Geological Survey Scientific Investigations Report 2018–5107, 24 p., https://doi.org/10.3133/sir20185107.

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., https://doi.org/10.3133/sir20085084.

Asquith, W.H., Roussel, M.C., and Vrabel, J., 2006, Statewide analysis of the drainage-area ratio method for 34 streamflow percentile ranges in Texas: U.S. Geological Survey Scientific Investigations Report 2006–5286, 34 p., https://doi.org/10.3133/sir20065286.

Cleveland, T.G., and Fang, Z.N., 2021, Texas-Skew-Update-2021: Texas Data Repository, https://doi.org/10.18738/T8/SVLCOQ.

Cohn, T.A., Barth, N.A., England, J.F., Jr., Faber, B.A., Mason, R.R., Jr., and Stedinger, J.R., 2019, Evaluation of recommended revisions to Bulletin 17B: U.S. Geological Survey Open-File Report 2017–1064, 141 p., https://doi.org/10.3133/ofr20171064.

Cohn, T.A., England, J.F., Berenbrock, C.E., Mason, R.R., Stedinger, J.R., and Lamontagne, J.R., 2013, A generalized Grubbs–Beck test statistic for detecting multiple potentially influential low outliers in flood series: Water Resources Research, v. 49, no. 8, pp. 5047–5058, https://doi.org/10.1002/wrcr.20392.

England, J.F., Cohn, T.A., Faber, B.A., Stedinger, J.R., Thomas Jr., W.O., Veilleux, A.G., Kiang, J.E., and Mason, R.R., 2018, Guidelines for determining flood flow frequency Bulletin 17C: U.S. Geological Survey Techniques and Methods, book 4, chap. 5.B, 148 p., https://doi.org/10.3133/tm4B5.

Harwell, G.R., and Asquith, W.H., 2011, Annual peak streamflow and ancillary data for small watersheds in central and western Texas: U.S. Geological Survey Fact Sheet 2011–3082, 4 p., https://pubs.usgs.gov/fs/2011/3082/.

R Core Team, 2020, R—A language and environment for statistical computing: R Foundation for Statistical Computing, Vienna, Austria, version 4.0.2, accessed July 4, 2020, at https://www.r-project.org.

Texas Department of Transportation, 2020, Hydraulic design manual—September 2019: online resource accessed on July 26, 2020, at https://onlinemanuals.txdot.gov/txdotmanuals/hyd/index.htm and subsection https://onlinemanuals.txdot.gov/txdotmanuals/hyd/statistical_analysis_of_stream_gauge_data.htm.

U.S. Geological Survey, 2018, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed April 9, 2018, at https://doi.org/10.5066/F7P55KJN.

U.S. Geological Survey (USGS), 2020, PeakFQ—Flood frequency analysis based on Bulletin 17B and recommendations of the Advisory Committee on Water Information (ACWI) Subcommittee on Hydrology (SOH) Hydrologic Frequency Analysis Work Group (HFAWG), version 7.2, accessed January 29, 2019, at https://water.usgs.gov/software/PeakFQ/, version 7.3, accessed February 1, 2020, at https://water.usgs.gov/software/PeakFQ/.

Veilleux, A.G., Cohn, T.A., Flynn, K.M., Mason, R.R., Jr., and Hummel, P.R., 2014, Estimating magnitude and frequency of floods using the PeakFQ 7.0 program: U.S. Geological Survey Fact Sheet 2013–3108, 2 p., https://doi.org/10.3133/fs20133108.

Wagner, D.M., Kiang, J.E., and Asquith, W.H., 2017, U.S. Geological Survey streamgaging methods and annual peak streamflow, appendix 1 *of* Asquith, W.H., Kiang, J.E., and Cohn, T.A., 2017, Application of at-site peak-streamflow frequency analyses for very low annual exceedance probabilities: U.S. Geological Survey Scientific Investigations Report 2017–5038, 93 p., https://doi.org/10.3133/sir20175038.

6. Special Studies—Wet or Dry Classification of Annual Peak Streamflows

6.1. Background

The sensitivities of annual peak streamflows in a Texas, Oklahoma, and New Mexico east of the Great Continental Divide to measures of climate state are evaluated in this chapter. The working premise is that the peak streamflows are coupled to climate state, as expressed by the well-known Palmer Drought Severity Index (PDSI). The PDSI is one of many climate indices that measure the relative dryness or wetness of a region; however, the PDSI was explicitly the only index extensively reviewed for this study.

A national perspective of PDSI for May 2015 is shown in figure 6.1 that was acquired on July 3, 2019 at `https://www.ncdc.noaa.gov/sotc/drought/201505#det-pdi`, and at the time much of Texas was in extreme wet conditions and many floods were produced that month. The various regions shown throughout the United States and Texas are the climatic regions for which monthly climate indices are available.

Climate indices are aggregated on monthly time steps and traditionally are available for climatic regions throughout the United States dating back to about January 1895. There has been some recent (circa 2019) interest in sensitivity of Texas peak streamflows to climate indices within the Interagency Flood Risk Management initiative (`https://webapps.usgs.gov/infrm/` [accessed July 3, 2019]).

Experimental efforts with Texas peak streamflows (not otherwise reported here) and drought indices show that PDSI in the month of an annual peak had higher correlation with the peak streamflows than other climate indices for the same month. PDSI and other climate indices (Heddinghaus and Sabol, 1991; Heim, 2002; Palmer, 1965) include, and quoting from `https://www.ncdc.noaa.gov/temp-and-precip/drought/historical-palmers/overview` (accessed July 3, 2019), the following:

- **Palmer Z Index**—Measures short-term drought on a monthly scale;

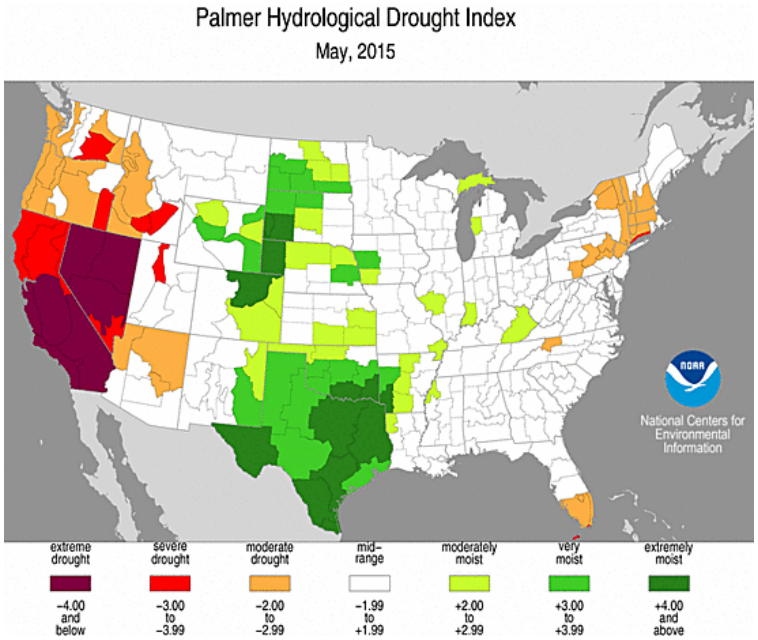


Figure 6.1. Example of the Palmer Drought Severity Index for May 2015, which is a month known to have produced many historically large peak streamflows at many streamgages in Texas.

- **Palmer Drought Severity Index**—Attempts to measure the duration and intensity of the long-term drought-inducing circulation patterns. Long-term drought is cumulative, so the intensity of drought during the current month is dependent on the current weather patterns plus the cumulative patterns of previous months. Because weather patterns can change almost literally overnight from a long-term drought pattern to a long-term wet pattern, the PDSI can respond fairly rapidly;
- **Palmer Modified Drought Index**—Operational version of the PDSI (Heddinghaus and Sabol, 1991); and
- **Palmer Hydrological Drought Index**—Measures hydrological impacts of drought (*e.g.*, reservoir levels, groundwater levels, and others) that take longer to develop and longer to recover from. This long-term drought index was developed to quantify these hydrological effects, and it responds more slowly to changing conditions than the PDSI.

Preparatory analyses were organized around a data linkage that included assigning to each peak the various drought indices of the month of the peak. The climatic region in which the streamgage exists was used as the climatic region on which to perform table joins between

the drought indices and the annual peak streamflows. Study of lagging an index back from the month of a peak streamflow was not studied.

The analyses indicated that using the PDSI assigned to the peak streamflows for a streamgage, can be chosen such that a PDSI threshold (magnitude) bifurcates the peak streamflows into two classifications, referred to respectively as wet and dry peak streamflows that produces the maximal separation in the classification-specific mean of the logarithms of the peak streamflows.[1] The resulting mean of the logarithms of wet-classified peak streamflows is almost universally maximally larger than the mean of the logarithms of the dry-classified peak streamflows. The premise for PDSI to produce a maximum separation in peak streamflows is that the "PDSI can respond fairly rapidly" to changes in climate state, which is consistent with the time scales on which annual peak streamflows are generated by watersheds and precipitation inputs.

A brief synopsis of key strengths and weaknesses of the PDSI have been summarized by Dai and National Center for Atmospheric Research (2017), and listed verbatim below. The key strengths of the PDSI are

- Effective in determining long-term drought, especially over low and middle latitudes;
- By using surface air temperature and a physical water balance model, the PDSI takes into account the basic effect of global warming through potential evapotranspiration; and
- Takes precedent (prior month) conditions into account.

In converse, the key weaknesses of the PDSI are

- Not as comparable across regions as the Standardized Precipitation Index (SPI), but this can be alleviated by using the self-calibrating PDSI;
- Lacks multi-timescale features of indices like the SPI, making it difficult to correlate with specific water resources like runoff, snowpack, reservoir storage, and other; and
- Does not account for snow or ice (delayed runoff); assumes precipitation is immediately available

Given this background, in early stages of this research, it was decided to exclusively retain use of PDSI for the month of the annual peak streamflow and proceed with data processing and analysis as described in this chapter.

[1] PSDI outperforms the other indices with respect bifurcation, and maximal separation is important in classification to avoid ambiguity over short intervals.

6.1.1. Climate Index Data Sources and Preprocessing

The data sources and preprocessing are comparatively short in the number of steps achieved. Using various Internet portals of the National Oceanic and Atmospheric Administration (NOAA), the aforementioned climate indices were acquired. Subsequently for each streamgage in the greater database of this project, the climatic regions for each streamgage in which the latitude and longitude of the streamgage exist were identified. This step implies that the centroid of a polygon matching or approximating the delineation of the watershed was used, which also implies that a weighted mean representation of the watersheds should they span multiple climatic regions was not used. The final stage in preprocessing for this study was matching to each of the peak streamflows possessing the year and month of the peak, the PDSI for the corresponding year and month.

These data reside are archived by Cleveland and Fang (2021) in file: `data/1703gages/pdsi_with_peak_flows_1698.feather`. The streamgage count therein does not match the parent count of 1,703 stated elsewhere in this report because of lack of data availability for a few streamgages. There are 55,737 peak streamflows in the listed file.

6.1.2. Bifurcation of the Peak Streamflows by the Palmer Drought Severity Index

Figure 6.2 shows a box plot of the distribution of the PSDI values at bifurcation. For each of the streamgages (1,698 streamgages had PDSI assignments from which 1,884 site tags exist), the PDSIs amongst those for the systematic record that approximately splits the record of a given site tag into halves were computed. A site tag is a colon delimited extension to the USGS streamgage number to indicate yes/no for code 6 and code C. This coding has an effect of potentially isolating distinct gaged record within gaged record. These codes are further discussed in chapters 4 and 8.

Annual peak streamflows that are part of the systematic record, but lack a month associated with some peak streamflows make such streamflows implicitly eliminated from the analysis reported here. To clarify the structure of the data, each site tag has its own streamgage-specific PDSI value that split the record. Algorithmically, the computation is simple; the median PDSI was computed for each site tag. The mean record length for the site tags is about 29.5 years and the quartiles are 11 and 45 years. The bifurcation results in a classification of each peak as either wet or dry.

The PDSI by definition has a central tendency of zero and hence negative PDSI are associated with drought and positive PDSI are associated with periods of abundant rainfall. The bifurcation of the peak streamflows by PDSI does not imply that the median PDSI will be zero or approximately so. In general, the researchers have found that the PDSI bifurcating peak streamflow records has about a median of 1.2 and a quartile range of PSDI 0.54–1.64.

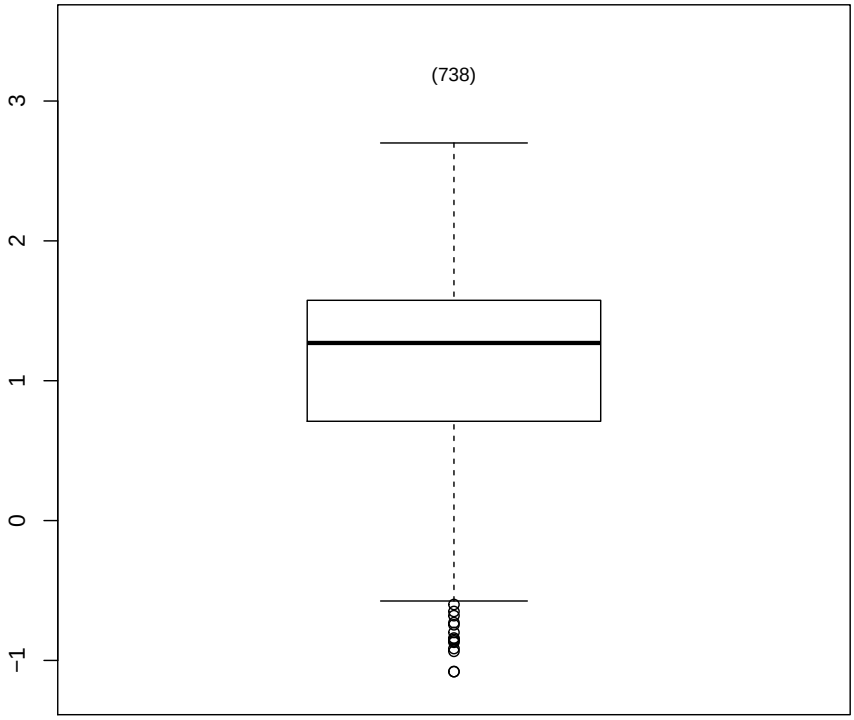


Figure 6.2. Distribution of PDSI values that bifurcate the systematic record for all 738 streamgages in the study area that have at least 15 peak streamflows in the wet and 15 peak streamflows in the dry classification. The results show that a median PDSI of about 1.2 for the month of occurrence of peak streamflows would split the records in half.

6.1.3. Spatial Distribution of the PDSI Bifurcating the Record

Spatial distribution was inferred from a surrogate that is by elevation bands. Figure 6.3 is a box plot by elevation band of the bifurcation PDSI for the study area. The PDSI is above 1.0 for elevations up to 2,000 feet, then declines to below 1.0 at higher elevations. Interpretations of this result is unclear, perhaps higher elevations are normally drier anyway and with somewhat less variability.

6.1.4. Frequency Curve Changes between Wet and Dry Classes

Frequency curves for each study station were examined to interpret the climatic influence, a few representative examples follow. Figure 6.4 (top) is an example of a streamgage where the wet or dry classification has little predictive value for the frequency behavior; the flood

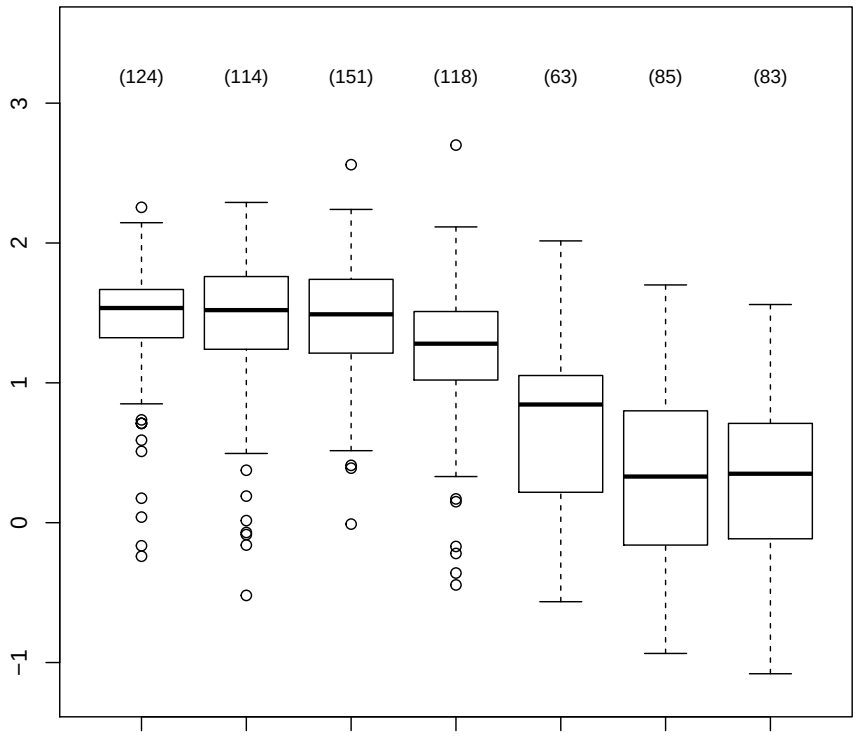


Figure 6.3. Distribution of PDSI values that bifurcate the systematic record for all 738 streamgages in the study region with distinction according to seven elevation classes from left to right (0–200; 200–500; 500–1,000; 1,000–2,000; 2,000–4,000; 4,000–6,000; and 6,000+feet).

frequency curves are essentially the same. The dry classification (red) has lower magnitudes at higher probability (left side of chart), and projected high magnitudes a bit larger than the equivalent wet (blue) classification.

These lower magnitudes are further identified as below the low-outlier threshold, whereas the climate classified wet values have no low outliers in this example. The large circles on the plot represent the projected (fitted log-Pearson type III distribution using station skew) 2-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year return period streamflow estimates for this streamgage.

Figure 6.4 (bottom) is an example of a streamgage where the wet and dry classification has predictive value; the wet climate state flood frequency curve lies above and nearly parallel to the dry curve. The wet classification (blue) has a low outliers identified, whereas the dry state all the values are useable for the flood frequency curve construction. For this streamgage, climate state matters, and wet climate streamflows are greater than dry state for the same probability. In the following maps, this streamgage would plot as a red marker for the mean values, and red or brown for the standard deviation. The skew difference is not

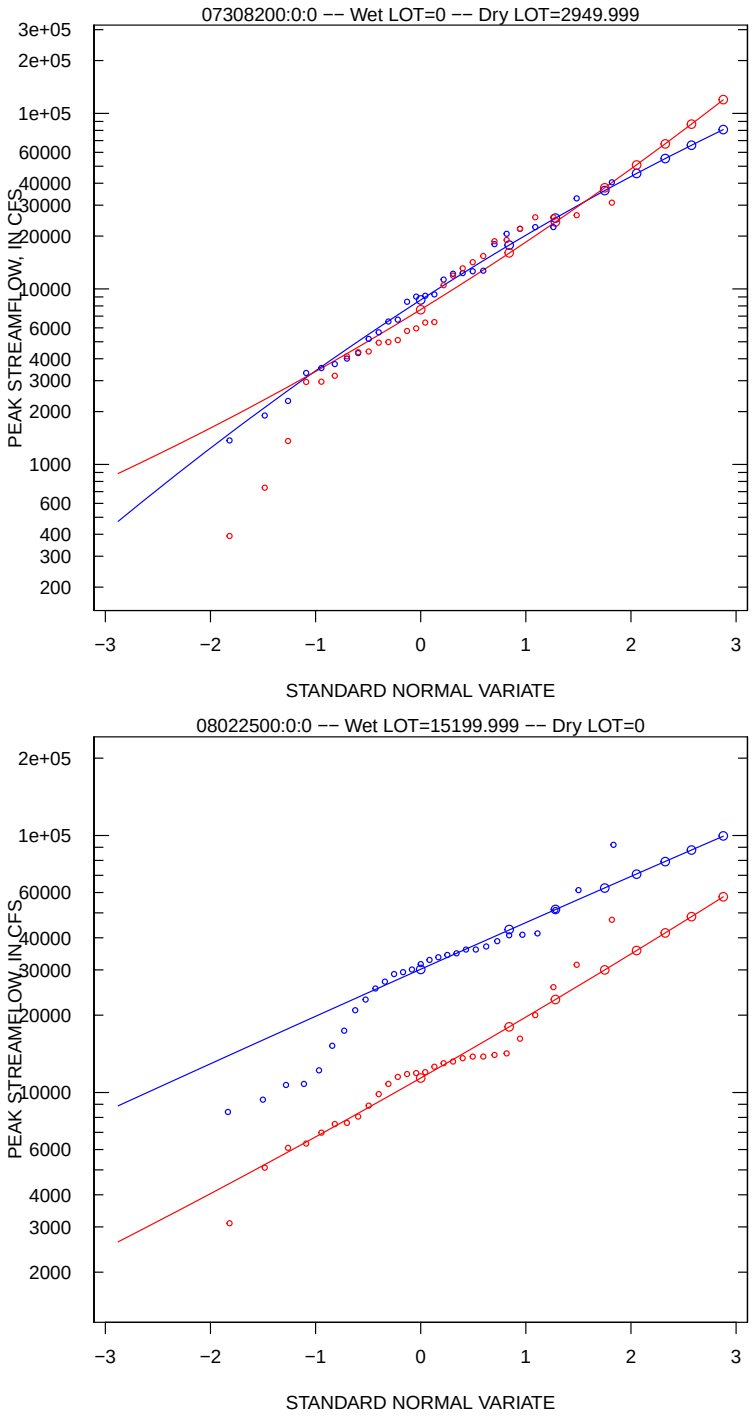


Figure 6.4. Example frequency curves for wet (blue) and dry (red) classified peak streamflows for U.S. Geological Survey streamgage 07308200 Pease River near Vernon, Texas (top) and 08022500 Sabine River at Logansport, Louisiana (bottom).

easily inferred from this plot. The large circles on the plot represent the projected (fitted log-Pearson type III using station skew) 2-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year return period streamflow estimates for this streamgage.

Figure 6.5 (top) is an example of a streamgage where the wet and dry classification shows a cross-over; the wet climate state flood frequency curve lies above the dry curve at sub 2-year return period, and crosses over at a 5-year return period. Neither classification has low outliers identified. In the following maps, this streamgage would plot as a green marker for the mean values, and green or olive for the standard deviation. The skew difference is not easily inferred from this plot. The large circles on the plot represent the projected (fitted log-Pearson type III using station skew) 2-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year return period streamflow estimates for this streamgage.

Figure 6.5 (bottom) is an example of a streamgage where the wet/dry state has some predictive value. Both classification states have low outliers identified. In the following maps, this streamgage would plot as a red or brown marker for the mean values, and dark green or green for the standard deviation. The skew difference will likely plot as a red marker in this case. The large circles on the plot represent the projected (fitted log-Pearson type III using station skew) 2-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year return period streamflow estimates for this streamgage.

Figures 6.6 and 6.7 provide four additional examples of wet and dry flood-frequency curves for four selected USGS streamgages. These are included to simply provide examples of the types of offsets and differences of frequency curves by the PDSI wet and dry peak streamflow classification.

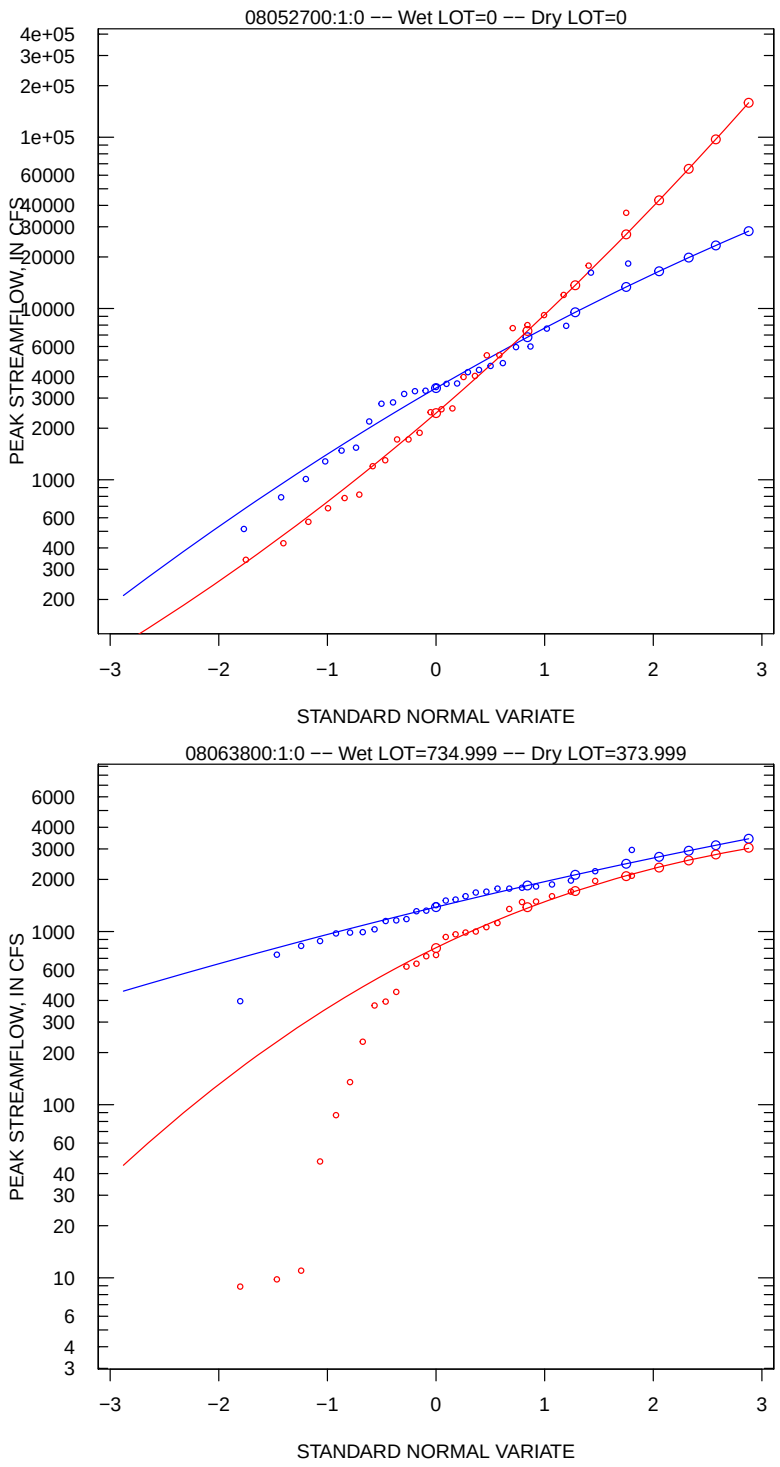


Figure 6.5. Example frequency curves for wet (blue) and dry (red) classified peak streamflows for U.S. Geological Survey streamgage 08052700 Little Elm Creek near Aubrey, Texas (top) and 08063800 Waxahachie Creek near Bardwell, Texas (bottom).

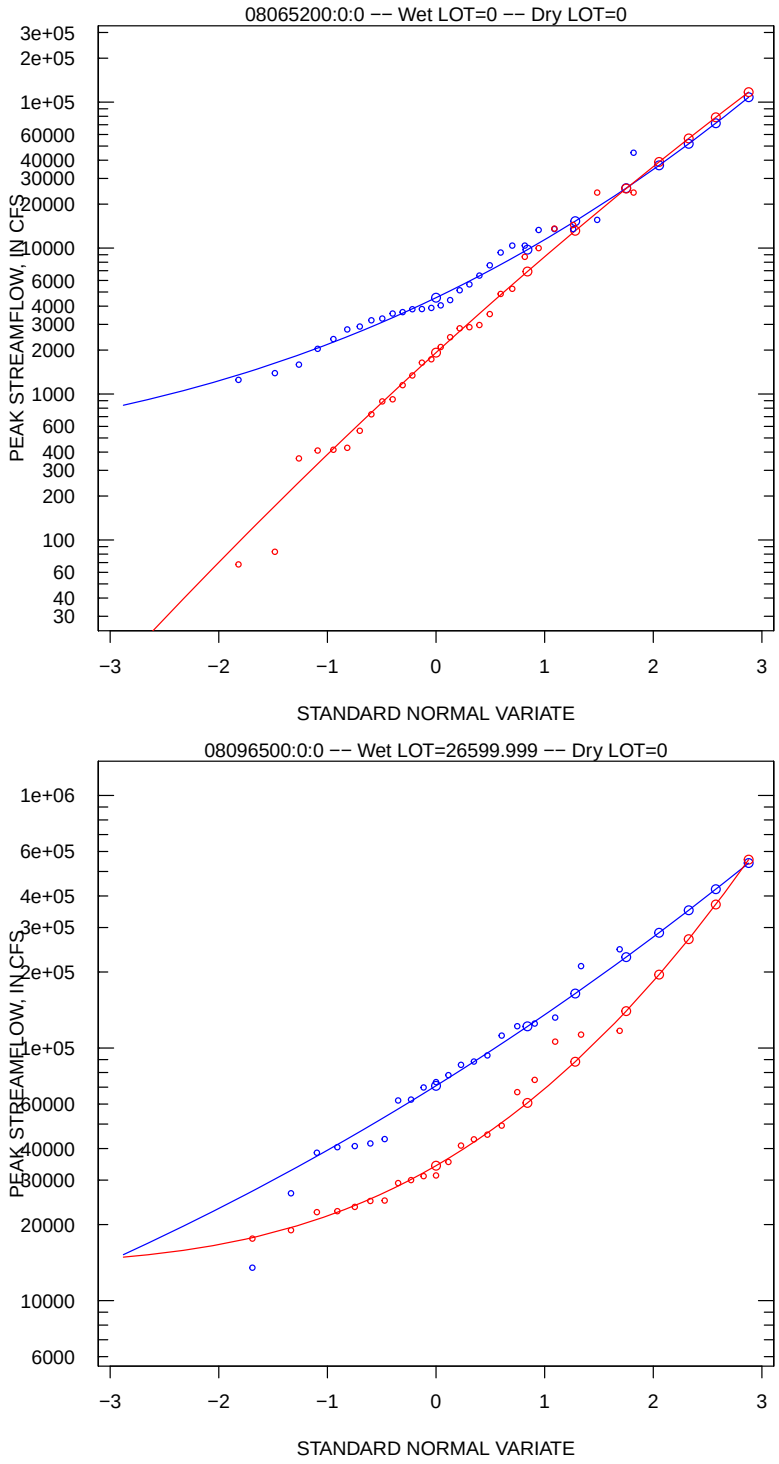


Figure 6.6. Example frequency curves for wet (blue) and dry (red) classified peak streamflows for streamgages 08065200 Upper Keechi Creek near Oakwood, Texas (top) and 08096500 Brazos River at Waco, Texas (bottom).

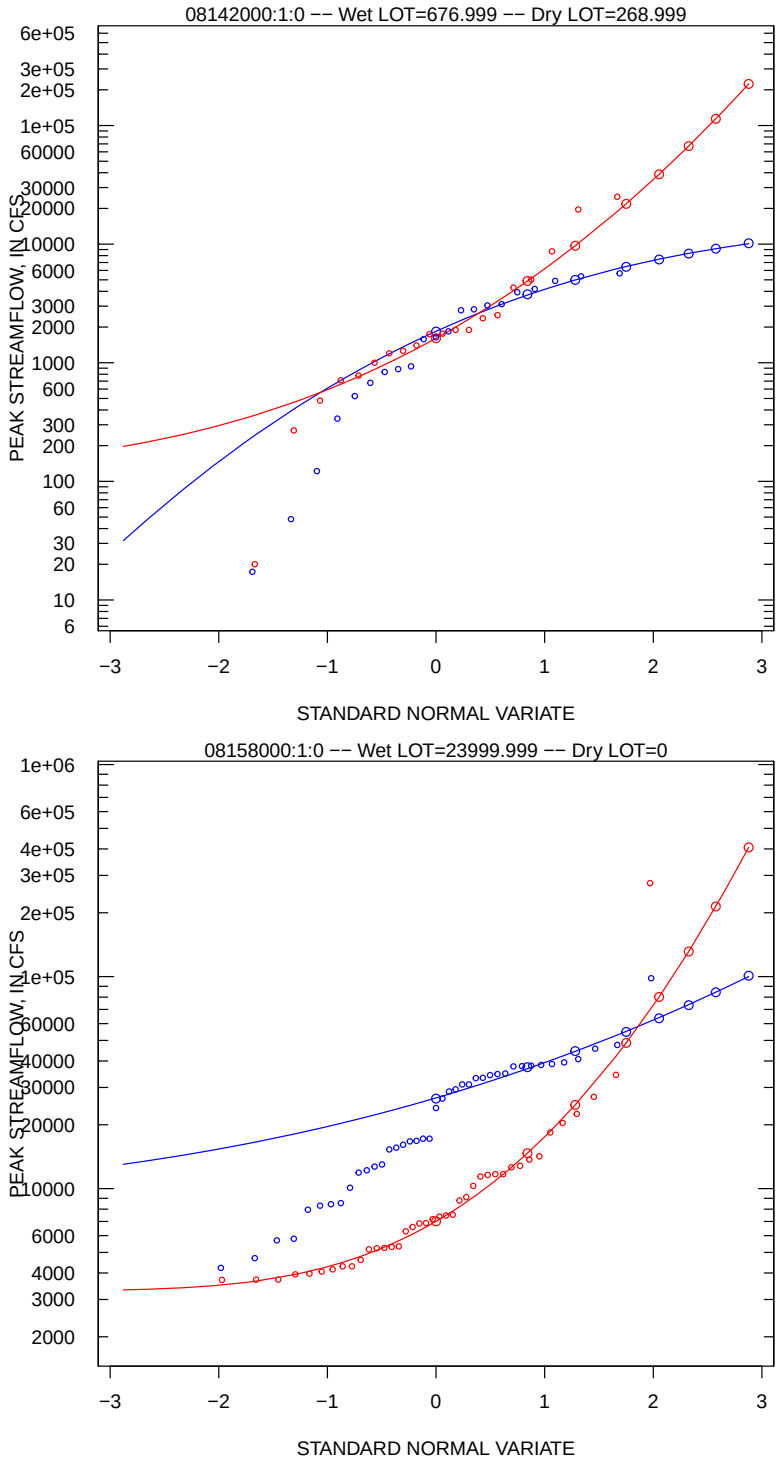


Figure 6.7. Example frequency curves for wet (blue) and dry (red) classified peak streamflows for streamgages 08142000 (top) Hords Creek near Coleman, Texas and 08158000 Colorado River at Austin, Texas (bottom).

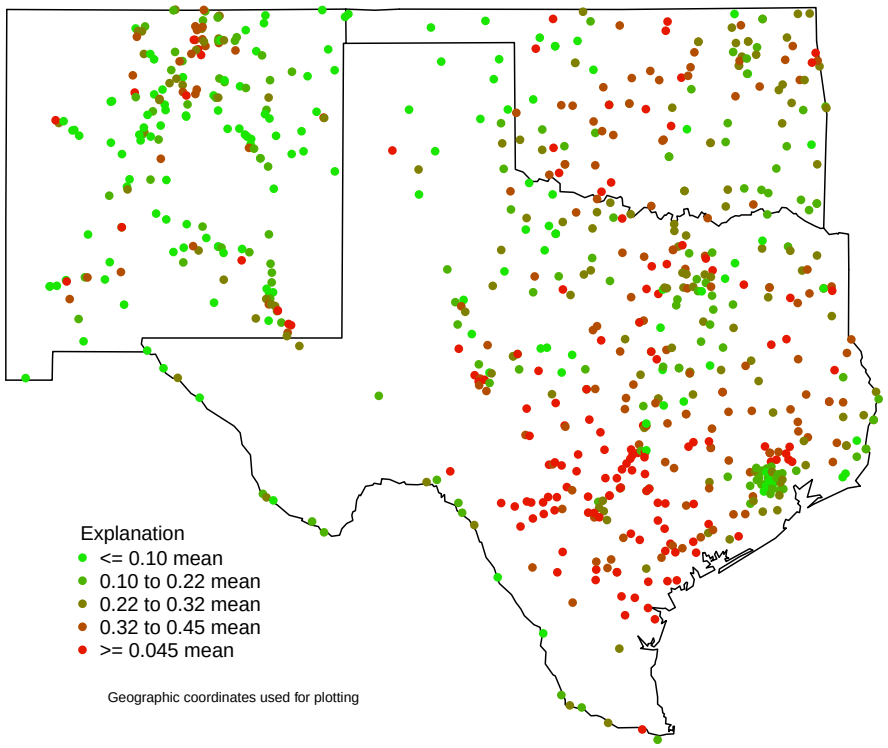


Figure 6.8. Spatial distribution of the difference of the wet to dry means of the log-Pearson type III from PDSI bifurcation for 738 streamgages with at least 30 years of systematic record, where color break points chosen from quintiles (0.10 [green], 0.22, 0.32, and 0.45 [red]) from the overall dataset (all elevations).

6.1.5. Spatial Effects on Distribution Parameter Differences

Exploratory analysis of the effects of the wet and dry classifications by PDSI bifurcation and the change in the moments (parameters) of the log-Pearson type III distribution are shown in this section. In particular, the wet or dry bifurcation of mean, standard deviation, and station skew as a function of location is displayed in figures 6.8–6.10. The reasoning for showing this information is to assess how distribution geometry (shape) changes in semi-quantified sense across the study area. Efforts to interpret these figures are made.

Comparison of the Means—The color scheme in figure 6.8 shows differences in the means. Red locations show a comparatively large difference (wet compared to dry is quite different in the mean) whereas green locations the differences in means are small, implying wet minus dry differences are not evident in the peak flows. In terms of location, the Texas coastal bend and central Texas are sensitive to wet and dry classification, as are central Oklahoma, and Northern New Mexico.

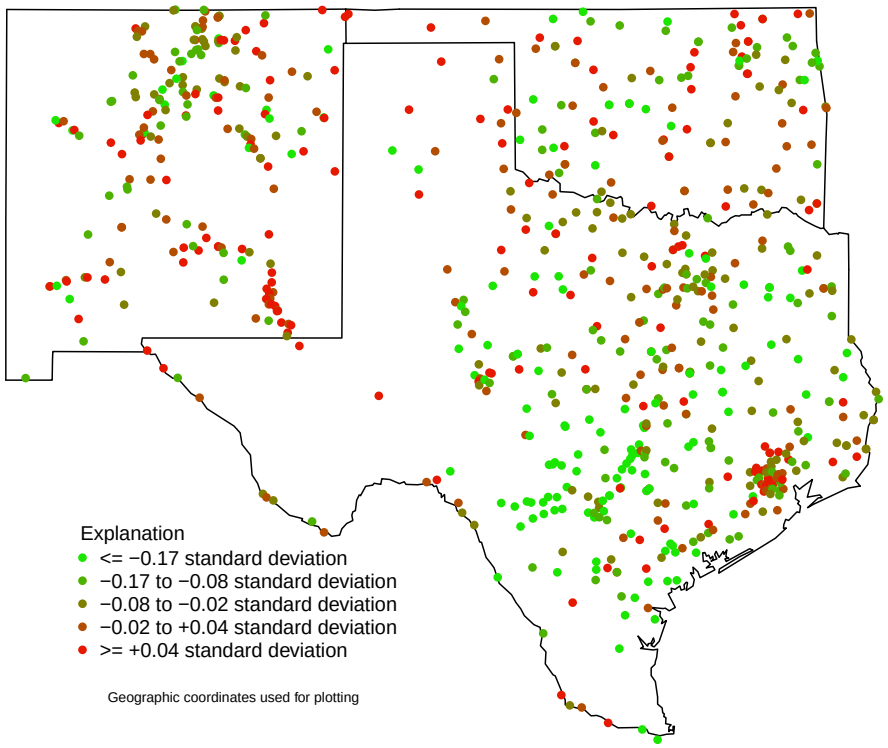


Figure 6.9. Spatial distribution of the difference of the wet to dry standard deviations of the log-Pearson type III from PDSI bifurcation for 738 streamgages with at least 30 years of systematic record present, where color break points chosen from quintiles (-0.17 [green], -0.08, -0.02, and $+0.04$ [red]) from the overall dataset (all elevations).

So, for figure 6.8, the redder colors indicate that the peak distribution associated with large PDSI (greater than median PDSI as seen in fig. 6.1), from a rule-of-thumb, is about 1/3 log10-cycles greater and the greener colors shown that the difference is only about 1/10 log10-cycles greater. Perhaps of most interest to hydrologic engineering practice in Texas is the preponderance of a large climate effect in the mean of the flood frequency distribution in central Texas southeast to the coastal bend.

Comparison of Standard Deviations—The color scheme in figure 6.9 is that red markers indicate variability differences are positive, that is variability (expressed as standard deviation) in wet conditions is greater than variability in dry conditions for the same location. The steepness of the frequency curve increases from the dry to the wet classification for positive values of the differences shown in the figure. The remaining colors, brown to green are progressively decreasing magnitudes of negative differences, that is variability (again, expressed as standard deviation) in wet conditions is smaller than variability in dry conditions for the same location.

The map shows that positive differences in variability between wet to dry are detected over the entire study area. The identifiable areas with large positive differences (variability is greater in wet classification) are the Harris County area in Texas, North Central Texas, most of Oklahoma, the Guadalupe Mountains in South East New Mexico, and Northern New Mexico. Perhaps of most interest to hydrologic engineering practice in Texas is the preponderance of a large climate effect then in the mean of the flood frequency distribution in central Texas southeast to the coastal bend. However, the effect at first glance might be counterintuitive to that for the mean in figure 6.8. The effect is negative, that is, the standard deviation relatively decreases from dry to wet conditions, in central Texas, but this in the context of the mean having increased in the region.

Comparison of Skews—The color scheme in figure 6.10 is that red markers indicate skew differences that are positive, that is skew in wet conditions is greater than skew in dry conditions for the same location. Green markers indicate skew differences are negative, that is skew in wet conditions is smaller than skew in dry conditions for the same location. The remaining colors, brown to dark green are progressively decreasing magnitudes of skew differences. The olive markers represent locations where skew differences are small and include zero, and these locations are not sensitive to wet and dry classification. There does not appear to be much of a spatial signal or trend to highlight. There is much inter-scattering of the redder colors with the greener colors.

6.2. Chapter Conclusions

The wet and dry classification of annual peak streamflows by the PDSI bifurcation shown by a streamgage-to-streamgage comparisons that many interesting patterns and different frequency curves exist. The flood distribution is clearly conditionally influenced by PDSI. Spatially as a general rule, the wet and dry classification shows considerable differences in the first two parameters of the log-Pearson type III distribution. Skew is not sensitive to the wet and dry classification.

Whereas the results shown in this chapter are certainly of curiosity to flood hydrology and the fact that the frequency curves seem to increase for wetter conditions as conditioned by PDSI for the month of a peak streamflow, it is not clear how this information could be used for purposes of hydrologic engineering design. The researchers suggest that the content of the analysis in this chapter would be difficult to integrate into the TxDOT Hydraulic Design Manual (September 2019) (Texas Department of Transportation, 2020), and at least, refinement of the generalized skew discussed in chapter 4 by the wet and dry classification, is not likely to improve design practice appreciably. There might be a place for climate state as a covariate in an algorithm discussed in chapter 8.

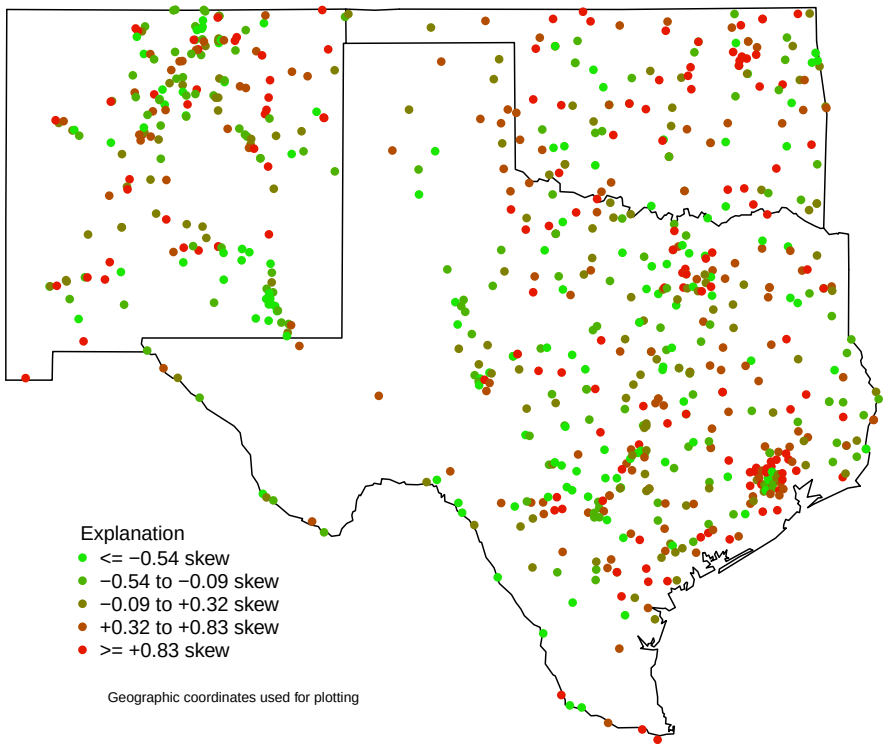


Figure 6.10. Spatial distribution of the difference of the wet to dry skews of the log-Pearson type III from PDSI bifurcation for 738 streamgages with at least 30 years of systematic record present, where color break points chosen from quintiles (-0.54 [green], -0.09, $+0.32$, and $+0.83$ [red]) from the overall dataset (all elevations).

Last remarks are needed. The analysis herein clearly shows that annual peak streamflow magnitudes are sensitive to the "state of climate" as represented by PDSI. This means that the conditional probability of peak streamflows is a function of PDSI. From a perspective of bivariate joint probability, months of high PDSI in contemporaneous time could be indicative that a watershed is potentially primed for larger than usual floods than in months of low PDSI (drought conditions). It was outside the scope of this research to define joint probability behavior, such as by bivariate copula (Salvadori et al., 2007), between peak streamflow and PDSI. Further statistical study as shown here could be useful for enhancing operational hydrologic or public safety concerns as a contrast to traditional needs and stakeholders in hydraulic design from regional statistical models.

Chapter References

Cleveland, T.G., and Fang, Z.N., 2021, Texas-Skew-Update-2021: Texas Data Repository, https://doi.org/10.18738/T8/SVLCOQ.

Dai, Aiguo, and National Center for Atmospheric Research Staff (Eds), Last modified 12 July 2017. The climate data guide—Palmer Drought Severity Index (PDSI): accessed on July 3, 2019 at https://climatedataguide.ucar.edu/climate-data/palmer-drought-severity-index-pdsi.

Heddinghaus, T.R., and Sabol, P., 1991, A review of the Palmer Drought Severity Index and where do we go from here?: Proceedings, 7th Conf. on Applied Climatology, September 10–13, 1991, Boston: American Meteorological Society, 242–246.

Heim, Jr., R.R., 2002, A review of twentieth-century drought indices used in the United States: Bulletin of the American Meteorological Society, vol. 83, pp. 1149–1165, https://doi.org/10.1175/1520-0477-83.8.1149.

Palmer, W.C., 1965, Meteorological drought: U.S. Weather Bureau Research Paper No. 45, NOAA Library and Information Services Division, Washington, D.C., 20852.

Salvadori, G., De Michele, C., Kottegoda, N.T., and Rosso, R., 2007, Extremes in nature—An approach using copulas: Dordrecht, Netherlands, Springer, Water Science and Technology Library 56, 292 p.

Texas Department of Transportation, 2020, Hydraulic design manual—September 2019: online resource accessed on July 26, 2020, at https://onlinemanuals.txdot.gov/txdotmanuals/hyd/index.htm and subsection https://onlinemanuals.txdot.gov/txdotmanuals/hyd/statistical_analysis_of_stream_gauge_data.htm.

7. Special Studies—Multiorder Hydrologic Position (MOHP) Analysis

7.1. Introduction

Increasing the number of potential predictor variables for annual peak streamflow statistical analyses is necessary to improve upon predictive hydrologic flood models in the future. An example being considered is the use of the U.S. Geological Survey (USGS) Multiorder Hydrologic Position (MOHP), which includes lateral position and distance from stream to divide (Yesildirek et al., 2021). Lateral position (dimensionless) is the relative position of a point between the stream and its watershed divide. Distance from stream to divide (units of length) is an indicator of position within a watershed: generally small near a confluence and generally large in headwater areas. MOHP was incorporated into the Research Project 0–6977 because of the predictive value of MOHP for groundwater-flow and groundwater-quality modeling (Belitz et al., 2019) is favorable, and it appears to have value in surface water settings. MOHP metrics are used as explanatory factors in random forest machine learning models, an emerging technology to create useable prediction and classification engines. Whereas investigation into the importance or usefulness of MOHP for annual peak streamflows needs further exploration, early investigation for this project and previous use by others within groundwater-flow and groundwater-water quality statistical models indicate there is value in using machine learning to evaluate these relations for potential use in predictive models of surface-water hydrology including floods.

7.2. Previous MOHP Research

The MOHP is the ensemble of hydrologic positions on the landscape within the drainage network pattern including distance from stream to divide (DSD) and lateral position (LP). MOHP raster datasets including DSD and LP have been produced nationally for the 48 contiguous United States (CONUS) at 30-meter and 90-meter cell resolution for stream orders 1 through 9. For a given hydrologic order, DSD and LP are defined based on horizontal distance to the nearest stream and horizontal distance to the nearest divide. The detailed

steps of computing the hydrologic position ensemble can be found in Belitz et al. (2019) and the raster datasets can be acquired from Moore et al. (2019). Hydrologic order 1 consists of all streams, hydrologic order 2 includes streams with order 2 and higher, and hydrologic order n includes streams with order n and higher, with stream order being earlier defined by Strahler (1957). Instead of only considering the nearest stream, MOHP provides the possibility that the position of a point within a high-order watershed can be as relevant to the prediction of hydrologic characteristics as the position of that point within its local first-order watershed.

Several studies have used MOHP data as predictor variables for groundwater level tree-based machine learning. In case studies of Central Valley geomorphic provinces and U.S. physiographic provinces (Fenneman and Johnson, 1946; U.S. Geological Survey, 2020), MOHP metrics were found informative in predicting landscape classes using the Random Forest Classification model (Kuhn and Johnson, 2016) and the higher-order MOHP metrics were more influential than the lower order metrics (Belitz et al., 2019). For case studies in the Fox–Wolf–Peshtigo area in Wisconsin, the higher-order MOHP metrics were found not necessarily more influential than the lower order metrics on the application of the simulated/observed depth to the water table using the Random Forest Regression model (Belitz et al., 2019). Overall, results indicate that MOHP metrics have prediction utility on groundwater-level machine learning applications. Whereas groundwater studies have shown predictive value using the MOHP (Knierim and others, 2020), the use in flood peak and regionalization analyses have not been thoroughly investigated.

7.3. Development of MOHP Covariates

For this report, a variety of hydrologic positions were explored including LP and DSD for stream orders 1 through 9. The values for DSD and LP were extracted using the location of the streamgage along with the means and standard deviations for DSD and LP of the main flow line within a watershed (Yesildirek et al., 2021). The addition of DSD and LP statistics for flow lines is a novel data preparation step unique to this project. The MOHP datasets (DSD and LP at 90-meter cell resolution for stream orders 1 through 9) were obtained from Moore et al. (2019). In total, 18 rasters (`MOHP_DSDn` for `n` $\in 1 \cdots 9$ and `MOHP_LPn` `n` $\in 1 \cdots 9$) were used to assign values to the 1,703 streamgaging stations in a geographical information system (Yesildirek et al., 2021). The aforementioned "18" means that there are 18 MOHP covariates associated with each streamgage and with the flow-line statistics, another 36 MOHP covariates.[1] These MOHP assignments by streamgage are a major product of Research Project 0–6977.

[1] The term "covariate" is basically a term for a "column" of a potential predictor variable in a input table for a statistical method. The most important covariate in surface-water problems like floods is the contributing drainage area, for example.

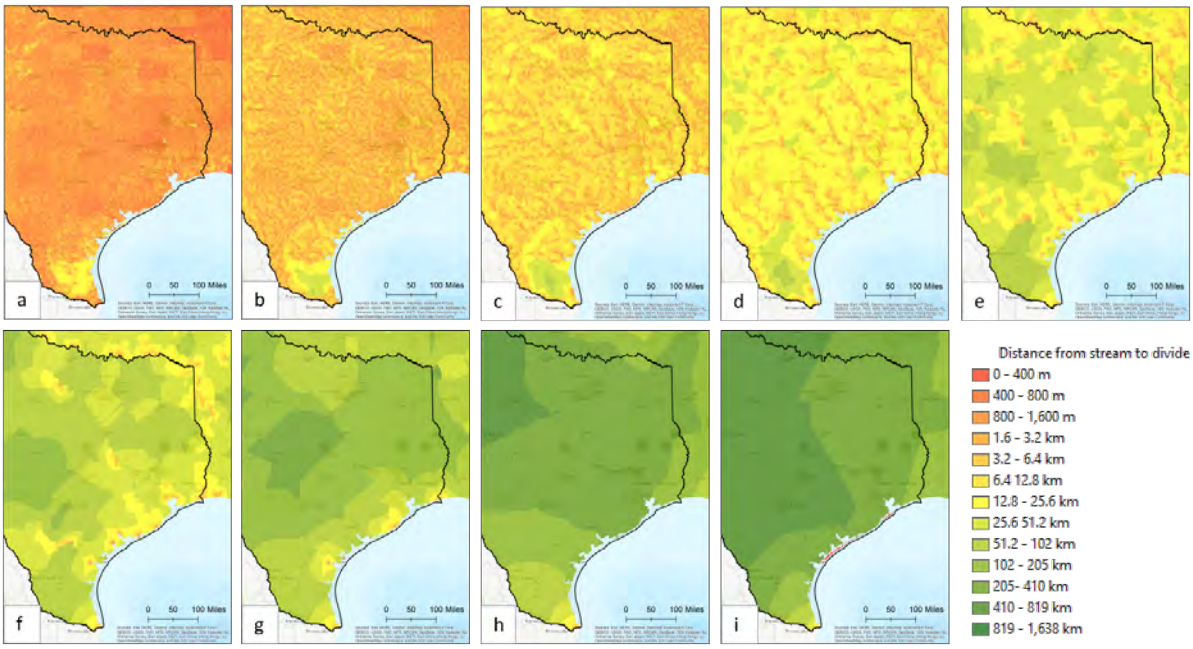


Figure 7.1. Maps of distance from stream to divide (DSD): (a) DSD_1, (b) DSD_2, (c) DSD_3, (d) DSD_4, (e) DSD_5, (f) DSD_6, (g) DSD_7, (h) DSD_8, (i) DSD_9, and the numeric subscript indicates hydrologic order.

Example maps of DSD for just part of the study area focuses on the eastern half or so of Texas including the Gulf of Mexico, from order 1 through order 9, are shown in figure 7.1. This spatial extent was chosen so that intricacies of each order can be seen. The DSD unit is in length and values are relatively small near stream confluences. For hydrologic orders from 1 to 5, values of DSD range from 0 to 100 kilometers (km) with the shades of red and yellow in most areas (figs. 7.1a–7.1e). For hydrologic orders from 6 to 9, most areas have DSD values on the order of 100 to 1,000 km with the shades of green. To aid in potential meaning or usefulness of the DSD, inspection of the maps in the figure suggests that DSD of different orders will function similar to the "[Texas] hydrologic regions" in the spirit of that concept used in Asquith and Slade (1997) for regional equations to estimate flood frequency for Texas.

Example maps of LP for the partial of Texas including the Gulf of Mexico, from order 1 through order 9, are shown in figure 7.2. LP is dimensionless with values ranging from zero at a stream reach to one at a drainage divide. It can be seen that the number of watersheds decreases as hydrologic order increases. The pattern of LP9 (fig. 7.2i) near the Gulf of Mexico can be taken as a general indicator of distance from the coast (Belitz et al., 2019). For example, the pattern in LP7 (fig. 7.2g) is nearly perpendicular to LP9 (fig. 7.2i) near the

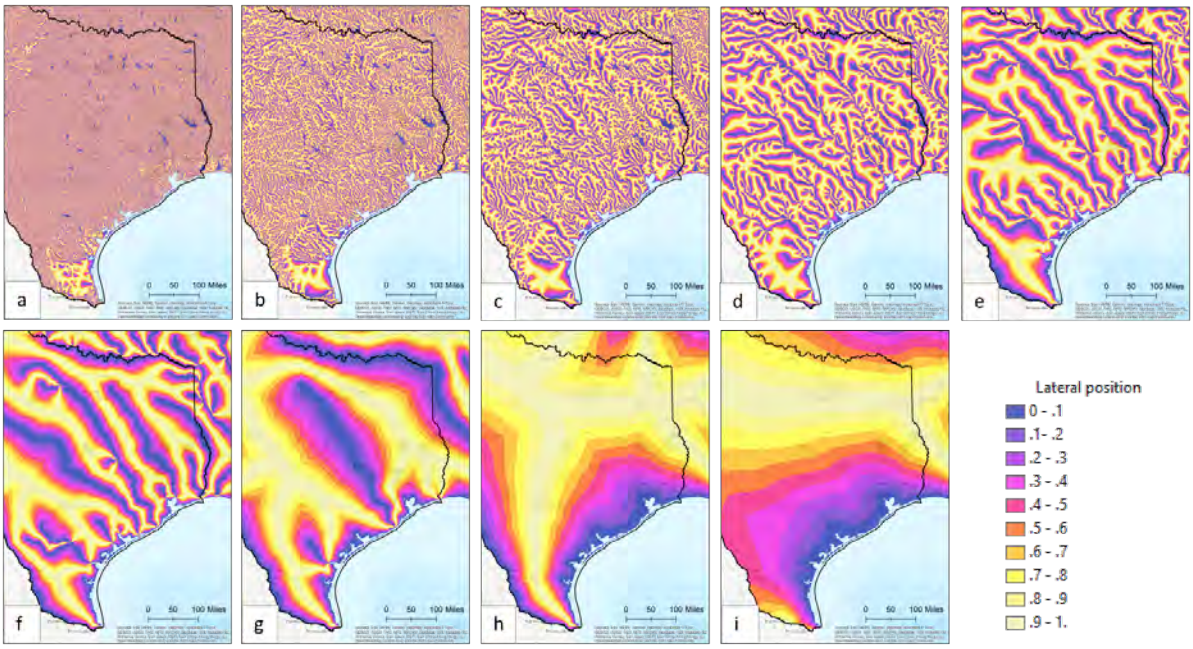


Figure 7.2. Maps of lateral position (LP): (a) LP_1, (b) LP_2, (c) LP_3, (d) LP_4, (e) LP_5, (f) LP_6, (g) LP_7, (h) LP_8, (i) LP_9, and the numeric subscript indicates hydrologic order.

Gulf of Mexico. Belitz et al. (2019) mentioned that LP7 can be taken as an indicator of distance from streams which are aligned approximately orthogonal to the gulf coast. So, LP9 and LP7 together can serve as an indicator of distance from the coast and distance from a major river. As with DSD, inspection of the figure suggests too that LP could function as a type of hydrologic region in statistical modeling.

7.4. Data for MOHP Flood Regionalization Studies

The specific fields related to MOHP for streamgages in Yesildirek et al. (2021) are

- `site_no`—The USGS streamgage identification number;
- `dec_lat_va`—The decimal latitude in the North American (horizontal) Datum of 1983 (NAD83) from the USGS National Water Information System (NWIS) (U.S. Geological Survey, 2018);
- `dec_long_va`—The decimal longitude in the North American (horizontal) Datum of 1983 (NAD83) from the USGS NWIS (U.S. Geological Survey, 2018);

- `GAGE_DSDn` (`n` is from 1 to 9)—The cell values equal to the sum of the shortest distance to the stream with hydrologic order `n` plus the shortest distance to the matching Thiessen divide;
- `GAGE_LPn` (`n` is from 1 to 9)—The cell values equal to the shortest distance to the stream with hydrologic order `n` divided by the DSD;
- `FL_DSDn_MEAN` (`n` is from 1 to 9)—The mean of DSD cell values along the flow line by hydrologic order `n`. These covariates are not evaluated in this chapter;
- `FL_LPn_MEAN` (`n` is from 1 to 9)—The mean of LP cell values along the flow line by hydrologic order `n`. These covariates are not evaluated in this chapter;
- `FL_DSDn_SD` (`n` is from 1 to 9)—The standard deviation of DSD cell values along the flow line by hydrologic order `n`. These covariates are not evaluated in this chapter; and
- `FL_LPn_SD` (`n` is from 1 to 9)—The standard deviation of LP cell values along the flow line by hydrologic order `n`. These covariates are not evaluated in this chapter.

7.5. Exploratory Evaluation of MOHP in a Regional Statistical Model

The watershed properties, MOHP, other watershed features associated with each peak streamflow originally sourcing from U.S. Geological Survey (2018) are in the project archive for this report in Cleveland and Fang (2021) along the file path: `0-6177-dataverse-archive/data/peaks_props_NID_1703.feather.zip`. This file contains over 58,000 years of systematic annual peak streamflows, watershed properties, and temporally integrated reservoir storages on a water-year by water-year basis. This file contains, therefore, data especially suitable for the exploratory research of MOHP as an explanatory variable on distributional properties of floods in the study area. The watershed properties and MOHP canonically source from Yesildirek et al. (2021).

A Cubist machine learning model (Kuhn and Johnson, 2016; Kuhn et al., 2020) using 33 covariates (18 are MOHP values for the streamgages) was assembled and is in Cleveland and Fang (2021) along the file path: `0-6177-dataverse-archive/demo/demo04_mohp.R`. This script uses the regulation covariates discussed in chapter 8 as means in this case to adjust for the effects of regulation within the Cubist model.

A Cubist model is a tree-based form of machine learning in which branches of the tree are rolled into a concept known as "rules," which logically are just a sequence of "if" and "and" statements using some or all of the covariates. The simplest description for say the first

rule would be "Rule 1: if drainage area (A) exceeds 1,000 square miles, use Linear Model 1." At the end of each rule, is a linear model (in concept, a linear regression equation); these each-rule-specific linear models use some or all of the covariates to make a prediction and need not match those covariates used in the rule itself. Continuing the simplest description, "Linear Model 1: $\log_{10}(Q) = 0.5 \times \log_{10}(A) + 0.8 \times \log_{10}(S) - 0.3$" for streamflow ($Q$), drainage area ($A$) and main-channel slope (S) and the listed numbers are coefficients and intercept.

The Cubist model made by the script is simple and only for reporting on some exploratory results. The model has 58,000 years of peaks for 1,634 streamgages. The objective of the Cubist model is to estimate the mean annual peak in log10 space given basic watershed properties, the conditional inclusion of the regulation covariates (for example, cumulative flood storage in the watershed for a given annual peak streamflow), and the 9 DSD and 9 LP MOHP covariates. Spatial location of the streamgage is not involved, which means that projected coordinates of the streamgage are not used as predictor variables.

The Cubist model summarized in this chapter is not tuned up. For sake of documentation, the number of "rules" was set to 500, the number of "committees" was set to 10, and the number of "neighbors" in prediction phase was set to zero. These are difficult to succinctly explain control features of Cubist but are set so that an instinctive feel for the what the model might be formally capable of is available to a first order approximation.

7.5.1. Standard Model Diagnostics and Comparison to Related Work

The normalized Nash–Sutcliffe Efficiency (NNSE) and root-mean-square error (RMSE) of the Cubist model fully training on the entire dataset, respectively, are NNSE=0.654 and RMSE=0.558 (log10 cubic feet per second). The RMSE in particular is a statement that the standard error for an estimated mean annual peak streamflow at least 1/2 log10-cycle and such error then in real-space is approximately a multiple or divisor of three.

The mean annual peak and 2-year peak are phenomenologically not the same but for rough comparison are loosely in the same realm of the flood regime. The 1/2 log10 error from the Cubist model can be compared to about the 1/3 log10 error for say the 2-year return period flood equation using on drainage area in Asquith, and Thompson (2008, tables 6 and 9). By itself in isolation, error comparisons are not reflective of either model superiority at this time and are provided in order to satiate curiosity. The 1/3 log10 error stems from Asquith, and Thompson (2008) using 656 streamgages in Texas and peak streamflow records restricted to natural conditions (meaning unregulated and unurbanized). The 1/2 log10 error from the Cubist is from a massively larger data set—almost 1,000 more streamgages—and unrestricted in watershed conditions (natural, regulated, urban). In essence, machine learning, such as Cubist, permits wholesale consumption of peak streamflow information into a general model.

Of interest to the exploratory evaluation of MOHP usefulness for the purposes of this chapter, is the question: "How relatively important are the MOHP for flood regionalization?" The Cubist model provides for a review of relative impact or "information presence" of each of the 33 covariates in the model on a scale from 0 to 100. For example, the top three covariates in decreasing order of importance are drainage area, mean elevation of the watershed, and mean annual precipitation. The fourth most important variable is MOHP DSD9 (fig. 7.1i). An interpretation then is that DSD9 is functioning somewhat like a hydrologic region—study closely the color banding in figure 7.1i, for example. The next five most important covariates are related to regulation covariates (see chap. 8). Further evaluation of the Cubist model itself is left unstated hereinafter.

7.5.2. Importance of MOHP in a Machine Learning Model

This chapter concerns the usefulness of MOHP for statistical flood hydrology. To this end, the DSD and LP covariates in terms of the 0 to 100 scale of relative importance were isolated and their order number extracted. Figure 7.3 shows that the importance of MOHP is nontrivial and providing some type of information into the regionalization of peak streamflows.

More specific interpretations can be made using figure 7.3. The DSD is relatively more important than LP because on the LP data plotting to the left of the DSD. The higher the MOHP order, the relatively more informative those data are than the very local scales presented by MOHP order 1. The results in the figure, in order for simplification of a production model, of peak streamflow, suggest that the MOHP DSD 7, 8, and 9 are quite important. Recalling from early discussion here, the DSD 9 is the fourth most important variable in the model, so retaining DSD 9 in further model development seems logical.

The LP is a bit more difficult to interpret because the importance measures are much lower and cluster in the 10 to 30 range. It is important to now that in general both DSD and LP show increasing contribution of information in the regionalization process as the order increases. Perhaps for LP, the MOHP orders to focus on would also be 7, 8, and 9 (even though fig. 7.3 shows a "wobble" in the third most important LP order). A figure of relative importance such as shown naturally changes as some orders, clearly the lower orders of MOHP, were removed as part of model simplification expected in more rigorous modeling efforts than reported here.

7.6. Chapter Conclusions

Advances in watershed processing and machine learning create an opportunity to explore a variety of explanatory variables for distributional properties of floods, including MOHP. Prior studies have indicated the importance of MOHP as groundwater-flow and groundwater quality

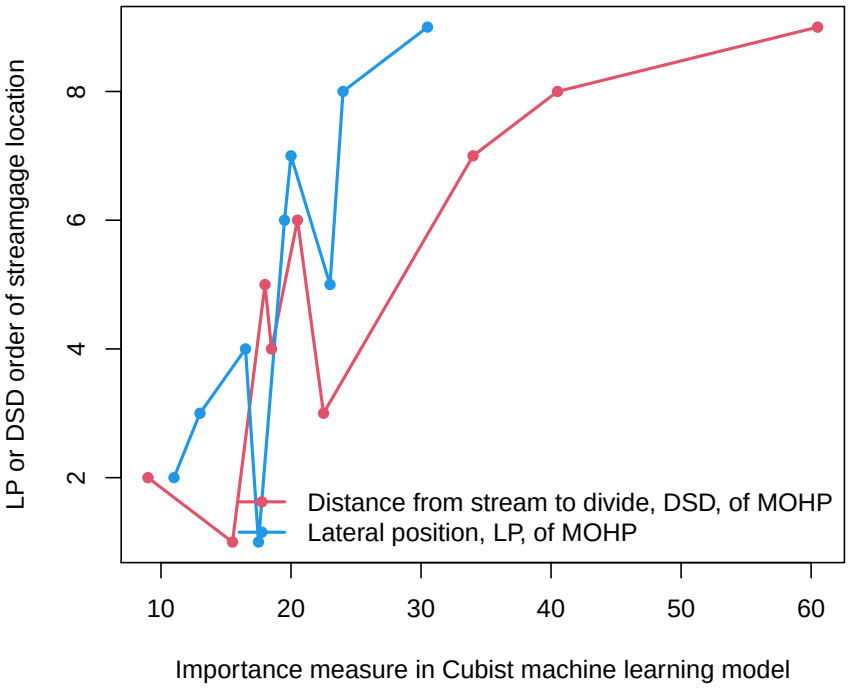


Figure 7.3. Relation between the MOHP order number and the relative importance metric for a Cubist model predicting base-10 logarithm of mean annual peak streamflow from 1,634 U.S. Geological Survey streamgages; 58,000 years of annual peak streamflows; and 33 covariates including the 9 distance from stream to divide (DSD) and 9 lateral position metrics from the multiorder hydrologic position (MOHP) data.

predictors. In this report, 18 MOHP covariates for streamgage location are investigated for their predictive value potential of mean annual peak streamflow (logarithmic). Results of an exploratory study are presented based on a Cubist machine learning model. Of particular interest from Cubist, is the relative importance of the 18 MOHP covariates and a diagnostic plot created. The results show that the higher order MOHP, in particular, orders 7, 8, and 9, are more influential than the lowest orders of MOHP. The distances to stream divide (DSD7, DSD8, and DSD9) and likely these same orders for LP (LP7, LP8, and LP9) are the most influential in statistical modeling of peak streamflows. MOHP is therefore of interest in regional models of flood hydrology in Texas, Oklahoma, and eastern New Mexico because indicated modeling diagnostics because MOHP seamlessly, in gridded form (an attractive attribute), represents fundamental stream network and terrain relations. Further study of MOHP in peak streamflow statistical analyses is justified.

Chapter References

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., https://doi.org/10.3133/wri964307.

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., https://doi.org/10.3133/sir20085084.

Belitz, K., Moore, R.B., Arnold, T.L., Sharpe, J.B., and Starn, J.J., 2019, Multiorder hydrologic position in the conterminous United States—A set of metrics in support of groundwater mapping at regional and national scales: Water Resources Research, v. 55, no. 12, pp. 11188–11207, https://doi.org/10.1029/2019WR025908.

Cleveland, T.G., and Fang, Z.N., 2021, Texas-Skew-Update-2021: Texas Data Repository, https://doi.org/10.18738/T8/SVLCOQ.

Fenneman, N.M., and Johnson, D.W., 1946, Physical divisions of the United States: U.S. Geological Survey, 1 sheet, scale 1:7,000,000, https://doi.org/10.3133/70207506.

Knierim, K.J., Kingsbury, J.A., Haugh, C.J., Ransom, K.M., 2020, Using boosted regression tree models to predict salinity in Mississippi Embayment aquifers, central United States: Journal of the American Water Resources Association, v. 56, no. 6, pp. 1010–1029, https://doi.org/10.1111/1752-1688.12879.

Kuhn, M., and Johnson, K., 2016, Applied predictive modeling: New York, Springer, ISBN 978–1–4614–6848–6, https://doi.org/10.1007/978-1-4614-6849-3.

Kuhn, M., Weston, S., Keefer, C., Coulter, N., Quinlan, R., and Rulequest Research, 2020, CubistRule- and instance-based regression modeling: R package version 0.2.3, dated January 10, 2020, accessed on October 6, 2020, at https://CRAN.R-project.org/package=Cubist.

Moore, R., Belitz, K., Arnold, T.L., Sharpe, J.B., and Starn, J.J., 2019, National Multi Order Hydrologic Position (MOHP) predictor data for groundwater and groundwater-quality modeling: U.S. Geological Survey data release, https://doi.org/10.5066/P9HLU4YY.

Strahler, A.N., 1957, Quantitative analysis of watershed geomorphology: Transactions of the American Geophysical Union, v. 38, pp. 913–920, https://doi.org/10.1029/TR038i006p00913.

U.S. Geological Survey, 2018, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed April 9, 2018, at https://doi.org/10.5066/F7P55KJN.

U.S. Geological Survey, 2020, Physiographic divisions of the conterminous U.S., accessed July 24, 2020, at https://water.usgs.gov/GIS/metadata/usgswrd/XML/physio.xml.

Yesildirek, M.V., McDowell, J.S., Zhang, J., and Asquith, W.H., 2021, Geospatial data of watershed characteristics for select U.S. Geological Survey streamgaging stations in New Mexico, Oklahoma, and Texas useful for statistical study of annual peak streamflows in and near Texas: U.S. Geological Survey data release, https://doi.org/10.5066/P9A91W4Z

8. Special Studies—Accommodation of Effects of Regulation

8.1. Introduction

Whereas the primary objective of this study was to update (chap. 4) the generalized skew map in the TxDOT Hydraulics Design Manual (Texas Department of Transportation, 2020), an opportunity was afforded to study the potential effects or impacts of regulation on the distributional properties of annual peak streamflows. The opportunity arose because of the extensive watershed property computations provided in Yesildirek et al. (2021) in conjunction with specialized software (*scNIDaregis*) by Asquith et al. (2021) that were both recently (2021) published products of Research Project 0–6977. The watershed properties listed by Yesildirek et al. (2021) identify 1,703 U.S. Geological Survey (USGS) streamgages that are presentative of most of the annual peak streamflow information of both current (2021) and historical USGS data collection in the study area. Peak streamflow data are provided by the USGS National Water Information System (NWIS) U.S. Geological Survey (2018). The study area includes Oklahoma, Texas, and New Mexico east of the Great Continental Divide. The aforementioned term "regulation" is used here in represent aggregate information from the U.S. Army Corps of Engineers National Inventory of Dams (NID) database (U.S. Army Corps of Engineers, 2020), for which greater context of these data to this study are within the *scNIDaregis* software documentation of Asquith et al. (2021).

Background to USGS peak-streamflow data and associated discharge qualification codes are summarized by Wagner et al. (2017). Discharge qualification codes "6" and "C" are the most germane with peak streamflow analyses. The code 6 definition states "Streamflow is affected by regulation or diversion," and the code C definition states "All or part of the record is affected by urbanization, mining, agricultural changes, channelization, or other anthropogenic activity." The USGS PeakFQ software 7.3 (U.S. Geological Survey, 2020) by default ignores peaks with code 6 and(or) code C—though settings in the software can be used to permit such data to enter into analysis. There is also a code 5 that states "Streamflow affected to an unknown degree by regulation or diversion." These statements are statements of qualification alone. Urbanization (development) (code C) is outside the scope of this chapter and a side note is made about development in a later section within this chapter.

A summary of the code 6 and its relation to the study of generalized skew is described in another chapter of this report (chap. 4). Suffice to say here that presence or absence and(or) of code 6s and Cs by the USGS for the study are not statements of individual peaks not being suitable or contrarily not suitable for statistical study. However, the USGS in Texas and is seems for the broader region of the study area has tried to make the code 6 as a flag, as it were, that careful and deliberate application should be made by the engineer studying flood hydrology.

The code 6 in NWIS, incidentally, is assigned during the data acquisition and review process by USGS hydrographers and not influenced by the needs of end practitioners of flood hydrology. For some streamgages, the code 6 truly reflect some major change in the watershed because of one or more major reservoirs self-evidently change the distributional properties of streamflow. Such streamgages then might have a type of "before" and "after" subset of data in the collective record (Asquith, 2001). Other times the effects of regulation by reservoirs, inclusive of small impoundments and exclusive of major flood control activities, are much more nuanced. For example, perhaps only the lower end of peaks are effects but the upper end, the right tail or flood tail, which is of interest to the engineer, are unaffected by the impoundments. Alternatively, major flood control reservoirs could be so far upstream from a streamgage or watershed outlet that the effects of regulation are no longer a factor because of copious unregulated intervening drainage area (Herrmann, 2013). As can be seen, understanding and applying what the USGS is communicating to the data consumer by the peak discharge qualification coding is difficult.

8.1.1. Data for Study of Effects of Regulation

The advent of the *scNIDaregis* permits the "binding" of each annual peak streamflow to a temporal aggregation of cumulative storages by overlaying a polygon representative of the watershed for a given streamgage. The polygons for 1,703 streamgages are available from Yesildirek et al. (2021), the NID available in Asquith et al. (2021), and peak streamflows from U.S. Geological Survey (2018). These collective data are combined into the project archive for this report in Cleveland and Fang (2021) along the file path: `0-6177-dataverse-archive/data/peaks_props_NID_1703.feather.zip`. This file contains over 58,000 years of systematic annual peak streamflows, watershed properties, and temporally integrated reservoir storages on a water-year by water-year basis. This file contains, therefore, data especially suitable for the exploratory research of the effects of regulation on distributional properties described hereinafter in this chapter. These data afford a chance to greatly extend from the earlier TxDOT sponsored research described by Asquith (2001).

8.1.2. Previous Research and Interests of the Engineer

The TxDOT Research Program has had historical interest in the statistical nature of regulation of peak streamflows in Texas. More than 20 years ago, Asquith (2001) investigated, in cooperation with TxDOT, for about 325 streamgages in Texas, the changes in the L-moments (distribution summary statistics including mean, variation, skewness) by regulation. The 2001 study focused on hand-picked streamgages with major jumps (change points) in the temporal aggregation (accumulation) of upstream normal, maximum, and flood storage capacities of reservoirs. In short, the study depicted percent changes in the mean by cumulative flood capacity per unit drainage area and demonstrated that higher dimensionless moments of the peak streamflow distribution were unaffected. The study though used periods or plateaus of static cumulative flood capacity and statistically summarized changes in the peaks. Flood capacity was defined as the summation of maximum storages minus the summation of normal storages of reservoirs as listed in the NID.

There are several remarks to make in the interests of the engineer involved in flood frequency analysis for streamgages and (or) using regional methods to estimate flood frequency at ungaged (unmonitored) locations in Texas. Using the aforementioned data file in the previous section, there are 1,333 annual peak streamflows that appear as systematic record with the water year equalling or exceeding 2019. These data stem from 722 streamgages. Some 547 of the 1,333 peaks are coded as 6. This means that the USGS is stating, in a difficult to interpret way, that about 40 percent of the USGS annual peak streamflow data collected today (circa 2021) within the study area (Texas, Oklahoma, and New Mexico east of the Great Continental Divide) is "regulated" to a degree that the USGS considers the peak streamflows as influenced.

The fact that about 40 percent of all currently acquired USGS peak streamflow data are said to be "Streamflow is affected by regulation or diversion" has major ramifications for the interpretation and uses of peak streamflow data in the region. Further details to this definition are provided in chap. 4. For this chapter, these interpretations directly influence how in many watersheds from the USGS could be used, how hydrology and hydraulic models for watersheds can be compared to USGS data, and how statistical regional methods and other aspects of the TxDOT Hydraulic Design Manual should be applied inclusive of uncertainties in computations. At present (2021), there is no standard of practice available for the engineering community formulated to quantify or accommodate the effects of regulation in statistical hydrology.

As stated at the chapter beginning, the primary purpose of this study was to update the generalized skew map for the TxDOT Hydraulics Design Manual, and repeating, this is accomplished in chapter 4. The data analyses leading to those results deliberately focused on early-in-time records though most applicable to "natural watershed" conditions (unregulated watersheds in other words) when code 6s or other qualitative questions in whether regulation

of the peak streamflow records are thought to contaminate the peak streamflow data for purposes for establishing a generalized skew for natural watershed conditions. Left unsaid in that chapter and ultimately also unresolved for this chapter is guidance how to accommodate effects of regulation from the very many thousands of reservoirs constructed in Texas (Asquith et al., 2021, documentation).

The objective of this chapter is to report on exploratory research towards understanding how the effects of regulation could be rigorously quantified with an eye towards end-user applications. This chapter ends with short discussion on potential future research directions with a perspective of implementation.

8.2. Exploratory Research on Effects of Regulation

A component of the greater study was to make evaluations of the effects of regulation on distributional properties of peak streamflows. To this end, exploratory research on the effect of regulation of annual peak streamflows was made, and the results reported in this chapter. The report includes description of an algorithm and identification of experimental implementation of the algorithm in the R programming language (R Core Team, 2021). A demonstration for a selected streamgage is provided from which visual depiction of results produce clarity that the algorithm might be viable and certainly an avenue for further investigation of Texas flood hydrology is presented.

8.2.1. Step-by-Step Algorithmic Description

The basic algorithm for comprehensive inclusion of reservoir storage is describe step-by-step in this section. This algorithm is absolutely unique to this study and represents the culmination of long-duration discernment on generally under studied and thus unimplemented accommodation of reservoir regulation in a statistical sense into regional hydrologic models in Texas and the greater study area. The algorithm so described is experimentally implemented in Cleveland and Fang (2021) along the file path: `0-6177-dataverse-archive/demo/demo02_effreg_quantreg.R` and an elementary introduction to quantile regression for at least a reader with basic R knowledge is along the file path: `0-6177-dataverse-archive/demo/demo_basic_quantreg.R`. The two figures provided in this section stem from the former script. Both of these scripts are designed "run out of the box" (circa summer 2021) given compatibility of R and external R libraries (packages). Steps of the algorithm are shown in the following enumerated list.

1. Start with an annual peak streamflow dataset for many streamgages of contributing drainage area, mean annual precipitation, main-channel slope, watershed mean

elevation, and other covariates that included water-year-specific temporal integration of normal, maximum, and flood storages (maximum storage minus normal storages as cumulated year over year). Definitions of the maximum and normal storage terms are available within file `zUSACE_NID_DataDictionary2018.pdf` within the *scNIDaregis* software (Asquith et al., 2021). However, these storages do not represent storages in the real-world contemporaneous with peak streamflows; these are "as constructed storages" for purposes of the inventorying dams in the United States. For the algorithm, these storages represent the regulation covariates and use an inverse distance weighted version also of these storages; For this demonstration, 1,634 streamgages with a combined total of 58,000 years of peaks (1,703 was the original count but some streamgages were removed for some missing information);

2. Select appropriate variable transformations (herein base-10 logarithmic transformation [log10] was used with one unit of positive offset to accommodate zero values for a covariate [here that would be the regulation]). Logarithmic transformation is common in regional study of flood frequency;

3. Compute a linear model (conventional regression, conditional response on the mean annual peak discharge) using a formula to be repeated in the quantile regression step described later. This model is simply a regional equation to estimate the mean annual peak streamflow conditioned on the aforementioned watershed properties including the regulation covariates;

4. Compute the residuals of the linear model in the previous step;

5. Compute a generalized additive model (GAM) on the residuals in the previous step using the projected coordinates of the streamgage locations in an Albers equal area projection on which a 2-dimensional smooth function by the GAM is the only predictor term. Predictions from this GAM are labeled in a vector "Omega." A map of this Omega, in kilometers of Albers projected coordinates, is shown in figure 8.1. This step effectively creates a map of a geospatial residual correction much in the same spirit as the "OmegaEM" parameter as a spatial residual correction described by Asquith and Roussel (2009);

6. Compute quantile regression for the 2-, 5-, 10-, 25-, and 50-year return period quantiles using the same model structure as used for the linear model of the annual mean peak streamflow as previously described but include another term, the Omega term of the previous step. For each quantile, a linear equation of conventional slopes (coefficients) and an intercept is created. Each of these five equations (for the five quantiles selected) produce, for a given combination of watershed properties and regulation covariates, five quantiles in the upper or right tail of the distribution of annual peak streamflows. Readers should notice that the distribution form is not named for at this step, the quantiles represent the flood distribution non-parametrically, and

the quantile regression is focused on predicting the upper tail only because this tail is of universal interest in flood hydrology;

7. Select the Pearson type III distribution as the assumed correct parent form of the upper tail. This distribution is selected for the primary reason of its basic familiarity and application in many peak streamflow analyses. Other distributions could be used;

8. Using the five quantile predictions stemming from the quantile regression step, fit the Pearson type III by the method of percentiles. (The method of percentiles is simply choosing distribution parameters by a multidimensional optimization by minimization of available quantiles of the fitted distribution to the specified quantiles of some sample or other prediction. For the circumstances here, a sample does not exist, but through quantile regression, parts of the distribution tail are predicted and whose percentiles are known—specified as part of the regression.) The key point to make is that the analyst does not have to have product moments or L-moments of a random sample to fit the distribution by conventional methods; and

9. Study the effects of regulation by adjusting the regulation covariates. The effects of regulation are expressed as a "logarithms of regulation effects" by taking for example the frequency curve from the previous step (in log10 space) and subtracting that curve from a new one using updated regulation covariates (see discussion that follows).

A demonstration of the results of the algorithm, named for this report as "QR-Pearson III," is shown visually later in this chapter. For the immediate present, it is important to make more statements about the algorithm itself before describing the demonstration figure in a following section.

The second to last step, in brief, is a regional scheme for estimation of flood frequency in Texas and the greater study area that $\star\star\star$fully digests nearly the entire USGS peak streamflow database and seamlessly accommodates the wide spectrum of peak streamflow regulation by the many thousands of reservoirs$\star\star\star$. This scheme is mostly a variable substitution computation in the same manner as regional regression equations for flood frequency already available in Texas with just a minor complication of numerical optimization to back fit the Pearson type III by the method of percentiles. The numerical aspects of such optimization are straightforward and implementation as a web page form and ancillary graphics should be readily achievable using long-standard, web-programming functionality (`https://en.wikipedia.org/wiki/LAMP_(software_bundle)`, accessed June 28, 2021).

The last step in the algorithm might or might not be of general interest. The last step involving "updated regulation covariates" permits changing of the regulation covariates and seeing how the regional frequency curve changes. The last step though does permit the

unwinding of "current" regulation conditions to some original or assumed "true" natural watershed conditions (zeros for all the regulation covariates).

Closing remarks about the chosen quantiles for the fitting are needed. Though the 2-, 5-, 10-, 25-, and 50-year return period are commonly used in applications of hydraulic design, the authors at present think that perhaps the 50-year period is too deep into the tail to be expecting the quantile regression to perform satisfactorily. Remembering that the 2-year return period is the median (50th percentile) and the 5-year return period is the 80th percentile, it could be prudent to pick a quantile between the 50th and 80th percentiles, such as the 65th percentile. Different choices in quantiles are easily used in the algorithm but not considered herein—more study is justified.

8.2.2. Regulation Covariates are Used in the Algorithm

To reiterate algorithm described is experimentally implemented in Cleveland and Fang (2021) along the file path: `0-6177-dataverse-archive/demo/demo02_effreg_quantreg.R`. That script uses specific notation for the inclusion of watershed properties and regulation covariates into the quantile regression. The purpose of this subsection is to name the regulation covariates used and provide definitions as provided within the Asquith et al. (2021) software. Six regulation covariates are in the following list.

- `NRM_SUM`—The summation up through the corresponding year of normal storage, in acre-feet per square mile, using the `MINMAXNOR` column of the NID from Asquith et al. (2021, script `makeNIDtrim.R`). The area comes from the polygon watershed area.
- `MAX_SUM`—The summation up through the corresponding year of maximum storage, in acre-feet per square mile, using the `MAXMAXNOR` column of the NID from Asquith et al. (2021, script `makeNIDtrim.R`). The area comes from the polygon watershed area.
- `FLD_SUM`—The summation up through the corresponding year of flood storage, in acre-feet per square mile, using the `FLD_STOR` column of the NID from Asquith et al. (2021, script `makeNIDtrim.R`). The drainage area is that of the polygon representing the drainage watershed.
- `NRM_WGT`—The inverse-distance weighted summation up through the corresponding year of normal storage, in acre-feet per square mile-kilometer, using the `MINMAXNOR` column of the NID from Asquith et al. (2021, script `makeNIDtrim.R`). The area comes from the polygon watershed area. Further unit conversion is made to cancel units and return the `NRM_WGT` as a dimensionless coefficient that is then multiplied by 1E6.

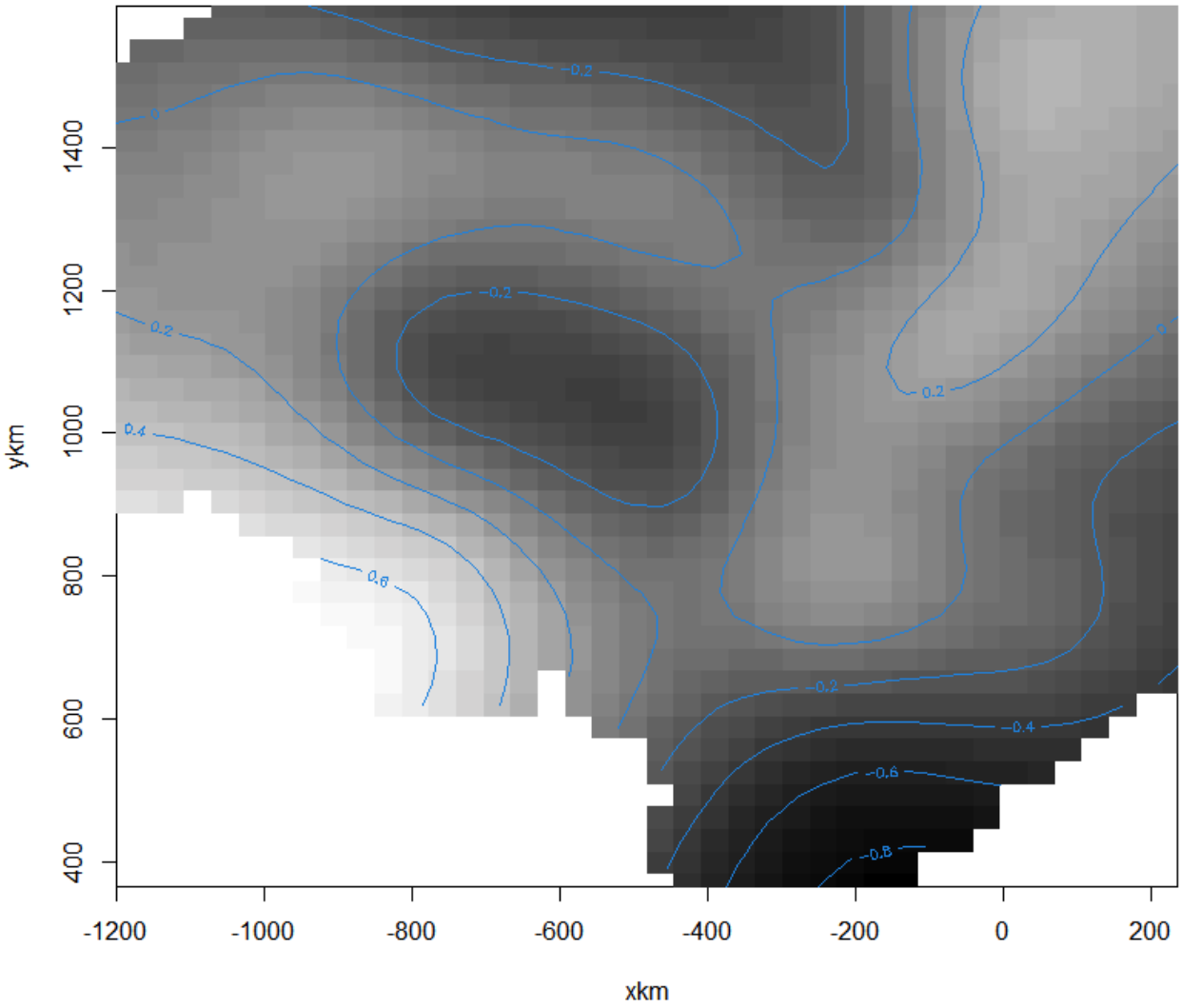


Figure 8.1. Map showing a residual adjustment from a 2-dimensional smooth within a generalized additive model (GAM) on the projected coordinates of streamgage from a linear model to estimate annual peak streamflow using the same structural form of the model later used in the quantile regression and this map is a general depiction of the Omega parameter described. Notes: The horizontal axis is an Albers equal area in standard USGS form of easting in kilometers, similarly, the vertical axis is northing in kilometers. The aspect ratio of this particular map is not geospatially correct in aspect ratio because the built-in diagnostic utility for GAM model plotting is not computationally aware that geospatial data are present. To help orient some readers, the lower right-hand corner (white region) is the Gulf of Mexico.

- `MAX_WGT`—The inverse-distance weighted summation up through the corresponding year of maximum storage, in acre-feet per square mile-kilometer, using the `MINMAXNOR` column of the NID from Asquith et al. (2021, script `makeNIDtrim.R`). The area comes from the polygon watershed area. Further unit conversion is made to cancel units and return the `MAX_WGT` as a dimensionless coefficient that is then multiplied by 1E6.

- `FLD_WGT`—The inverse-distance weighted summation up through the corresponding year of flood storage, in acre-feet per square mile-kilometer, using the `MINMAXNOR` column of the NID from Asquith et al. (2021, script `makeNIDtrim.R`). The area comes from the polygon watershed area. Further unit conversion is made to cancel units and return the `FLD_WGT` as a dimensionless coefficient that is then multiplied by 1E6.

For the purposes of this chapter, it is useful to note that the first three listed regulation covariates are computed without regard to the straight-line distance from the reservoir (dam) to the streamgage (or watershed outlet). The final three (weighted, `WGT`) storages involved inverse distance weighting of each dam's contribution to the whole. So, with these six covariates in hand, the algorithm is attempting to insert relative position of the storage as a context to effect of peak streamflow. For example, if a major reservoir is hundreds of river miles upstream from the streamgage (or watershed outlet) then its impact on annual peak streamflows is likely quite different if the major reservoir is just 1 kilometer upstream. The data available in the NID and implemented by the *scNIDaregis* have no operational rules or storage history of the reservoirs themselves available. Further study of inclusion or exclusion choices in the six regulation covariates is justified.

8.2.3. Remarks on Urbanization Covariates

It is informative to discuss urban development as an aside. If urbanization or watershed development, by one or more metrics, could be assembled and then temporally integrated or interpolated to bind such metrics to each annual peak as done for regulation by Asquith et al. (2021), then in principle, it would be possible to simultaneously accommodate within QR-Pearson III covariates for the effects of urbanization. As prototyped in this chapter, USGS streamgages that are in metropolitan areas were not removed. A perspective of a development land-use category is shown in figure 8.2.

8.2.4. QR-Pearson III Algorithm Demonstrated for a Streamgage

A demonstration of the algorithm for addressing the effects of regulation for USGS streamgage 08171000 Blanco River near Wimberley, Texas (fig. 8.3) is provided in this section. The

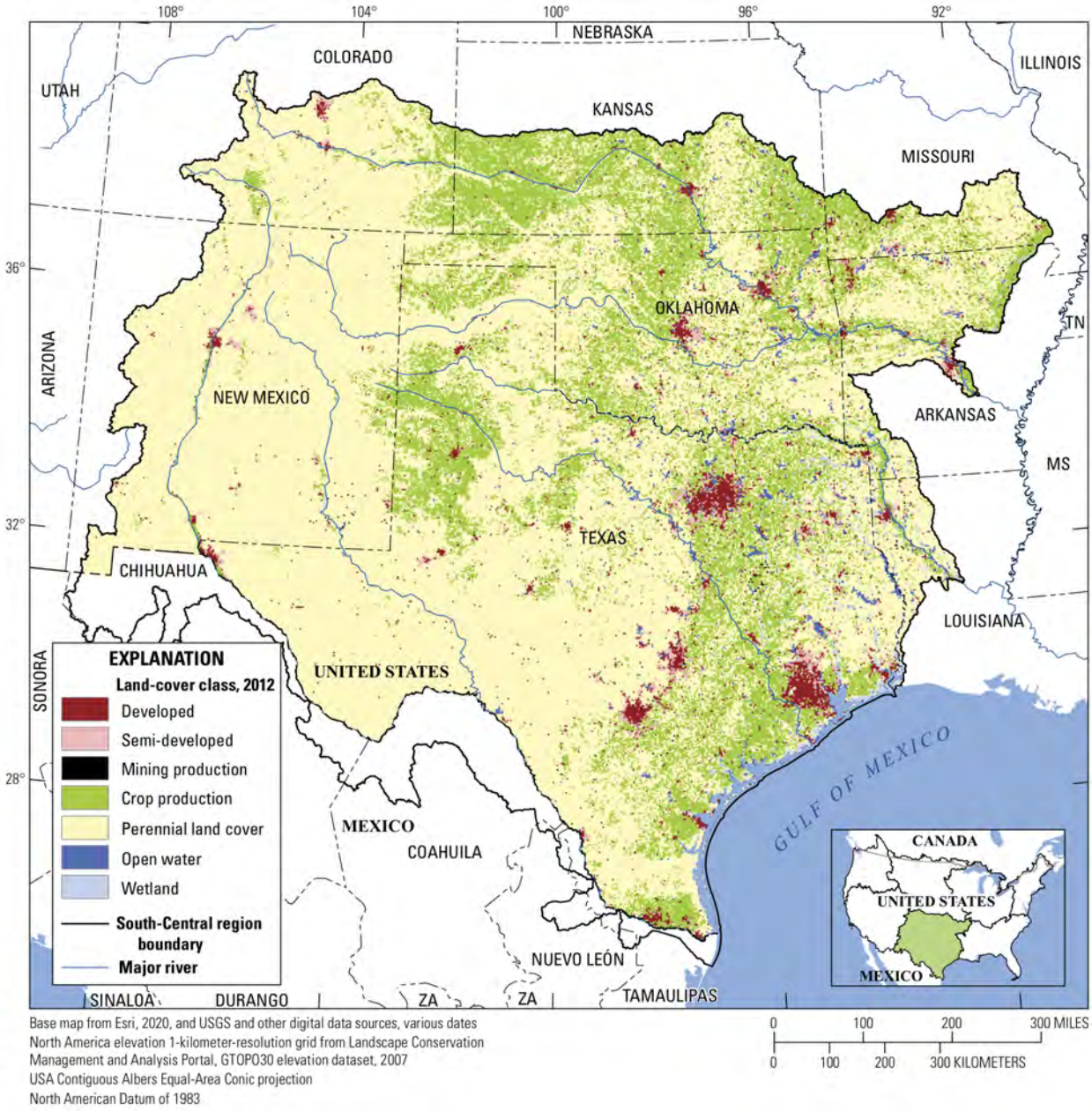


Figure 8.2. Map showing distribution of selected land-cover classifications including "Developed" that is especially pertinent to flood hydrology (annual peak streamflows) Texas, Oklahoma, and New Mexico east of the Great Continental Divide.

Figure 8.3. Remote camera photograph of location of USGS streamgage 08171000 Blanco River near Wimberley, Texas derived from USGS current conditions webpage for the streamgage (`https://waterdata.usgs.gov/nwis/uv?site_no=08171000`).

purpose not only is to described results of the algorithm but to also converse in details the demonstrate the future potential for implementation in practical circumstances. Various frequency curves from both the observational data from the streamgage and those from the algorithm are shown in figure 8.4.

Piece-by-piece, the content of the QR-Pearson III demonstration in figure 8.4 is described. First, the streamgage was selected because it has very long record from what it deemed a rural and undeveloped watershed—a natural watershed. The flood hydrology of the Blanco River is of general interested to students of Texas flood hydrology (see Burnett, 2008). The figure shows 94 annual peak streamflows from the observational record. These were used to fit a Pearson type III to the logarithms of the peaks using L-moments, and this distribution is depicted by the solid blue line. This will be the reference frequency curve for purposes of evaluation of the quantile regression method.

Second, for each of the 94 years of record (water years, 1925 to 2020 with 1927 and 1928 missing), there exist water-year-specific regulation covariates. Even though this is a rural watershed without either limited or major flood regulation, there have been small

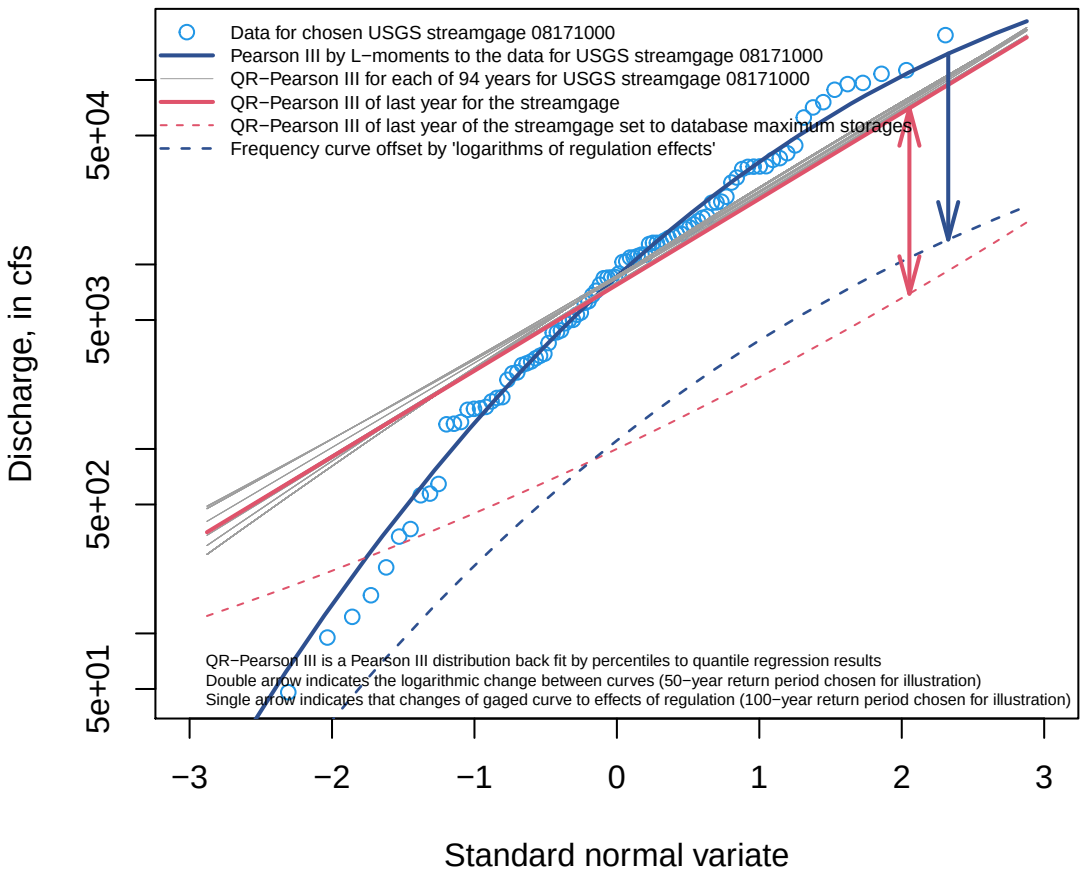


Figure 8.4. Demonstration of a flood-frequency curves, and in particular, using "logarithms of regulation effects," using quantile regression on 2-, 5-, 10-, 25-, and 50-year return period estimates along with a geospatial residual adjustment (fig. 8.1) and method of percentiles to back fit the Pearson type III distribution. Note: For design applications, interests are entirely for standard normal variates greater than zero.

impoundments or low-head run of the river dams constructed in the past 100 years or so. As a result, there are nonzero cumulative normal, maximum, and flood storages. For each of the years, the five quantiles for each year from the quantile regression model were computed and preserved for the next step.

For each of the years, a Pearson type III distribution was back fit to the water-year-specific quantiles and each of the distributions where drawn. These are the 93 solid grey frequency curves shown on the figure. Obviously, not all curves for 93 distributions can be visually counted and many years have static regulation covariates so the curves are plotted on top of each other. The 94th and last year (water year 2020) is plotted on top and plotted in a red color, and this red curve, therefore, represents the most contemporaneous. Some minor

reservoir construction has occurred; it is satisfactory, therefore, that the red curve appears to plot below or on top of the other 93 curves in the upper tail.

The first and second parts of the figure can be compared. The solid blue curve certainly represents a preferred site-specific estimate of flood frequency for this watershed because a USGS streamgage is monitoring the watershed. But in the vast majority of stream reach and the upstream watershed, streamgages do not exist with observational record. Hence, a regional statistical model is needed. The algorithm described in effect is such a regional model much in the same spirit as the equations in Asquith and Roussel (2009). The ability exists, therefore, with the quantile regression and Pearson type III back fit by the method of percentiles, to acquire regional flood-frequency curves. These curves are shown by the solid grey and solid red lines as previously described.

A comparison between the solid blue and solid red curves indicates that for the right half of the distribution (greater than the median or 2-year event that is the standard normal variate of zero) that the QR-Pearson III is quite similar. Much better than order of magnitude consistency exists. This streamgage is located in one of the highest unit discharge per unit watershed area regions in the United States; the fact that the blue curve does plot above the red curve is consistent with intuitive expectation of a regional model. At least for this one demonstration, therefore, it seems in principle that the QR-Pearson III is viable—a major "bug" in logic does not appear to exist.

Third, because the quantile regression has digested in effect all data for all time from the USGS and has the conditioning terms for regulation, scenarios can be studied. For example, if all the storages are turned to zero, then the upper, that is, largest most, solid grey curve would be representative of presumably true unregulated conditions.

Another experiment can now be conducted. What if all the reservoir storage terms in the quantile regression were increased to the non-joint maximums of each term from the database? If this were done, then hypothetically colossal amounts of storage in the watershed would have been emplaced.

The flood-frequency curve for such a hypothetical situation is depicted as the dashed red line, and this curve is by quick eyeball estimation about one order of magnitude (one log10 cycle) smaller than the solid red line. The doubled-headed red arrow highlights the differences between the two curves. This difference or more specifically the vector in log10 difference of the dashed red curve minus the solid red curve would result then in a scale-free representation of the effects of the regulation. This vector of differences between the two curves is termed herein the "logarithms of regulation effects," and for this example is uniformly negative for which the negative sign means regulation reduces the flood-frequency curve.

Fourth, with the aforementioned logarithms of regulation effects in hand, one can then offset other estimates of the flood frequency for the location. For example this vector is

subtracted from the solid blue curve to form then an estimate of the impact of colossal hypothetical regulation in the watershed yet be considered enhanced from just the QR-Pearson III estimate because the solid blue curve from the observation data formed a type of base line. This new frequency curve is the dashed blue line shown and the blue, downward-facing, arrow depicts this difference.

With then a visual description of the algorithm now described, general comments are offered. Recalling that about 40 percent of modern data collection (by streamgage) has the USGS reporting code 6s in NWIS and the fact that the USGS only has at best vague qualification that a given water year for a given streamgage has (or might have) regulated record, the QR-Pearson III scheme described could be immensely important for future considerations of Texas, Oklahoma, and eastern New Mexico flood hydrology. There exists little guidance, including Federal guidance (England et al., 2018), the TxDOT Hydraulic Design Manual (Texas Department of Transportation, 2020), or the USGS in Asquith et al. (2017) and particularly therein Wagner et al. (2017) on the treatment or accommodation of regulated peak streamflow records.

Complicating effect of regulation further is the fact that regulation of annual flood peaks itself exists on a broad spectrum. One watershed might seemingly have only the lowest of the low floods (the left tail, less than the 2-year event) modified by regulation of streamflow, and another watershed might have wholesale removal of what would be considered a "flood signal" from the watershed. When analysis is done on long record streamgages with record before reservoir construction, then potentially some site-specific interpretations of the effects of regulation could be attained from statistical or graphical trend analyses. Relatively young streamgages (a streamgage with relatively modest record lengths) will only provide annual peak streamflow data without deep temporal knowledge of what the natural flood frequency distribution might have looked like.

A comprehensive statistical model as experimented with for this chapter, provides a mechanism to incorporate all the modern data collection from the USGS along with all historical information as well into a single whole. Scenario exercises, as described for figure 8.4, provide for utility of function that currently (2021) is otherwise nonexistent within the TxDOT Hydraulic Design Manual as well as either prior academic or USGS studies of flood hydrology in the study area.

Lastly, nonstationarity as a rule is an assumption that is often made by practitioners of flood frequency for there are few alternatives. An experienced practitioner might chose to not include or include records based on local knowledge, but this assumes that a streamgage has record spanning periods of major change in a watershed. Speaking from a statistical perspective, implementation of nonstationary statistical methods in streamflow time series frequency analysis is difficult. For example, it is difficult to accommodate in site-specific analysis of data with trends inclusive of drift or change points along the real number line or changes in variance or shape (skewness) because the record lengths for a streamgage are

generally short for modeling much more complicated than a few parameters. A comprehensive statistical model as demonstrated could be useful for the nonstationarity question because additional covariates related to climate state could be included in the quantile regression.

8.3. Chapter Conclusions

The results in the chapter show that it might be possible to have a regional model of flood frequency purposed for Texas that accommodates the effects of regulation. The idea is to use as much of the USGS peak streamflow data as possible. Quantile regression using linear equations can be used to predict specific upper-tail quantiles of the frequency curve without yet making an assumption of distributional form. A geospatial residual correction term can be added to the quantile regression following the ideas of Asquith and Roussel (2009) to enhance the model. Then with the quantile predictions in hand along with the nonexceedance probabilities defining each, a Pearson type III distribution can be back fit to these quantiles to form a Pearson type III distribution specific to the watershed properties and regulation covariates provided. This is the reasoning for abbreviating the approach as as QR-Pearson III.

The QR-Pearson III after training on gaged watersheds can be applied for both gaged and ungaged watersheds. Scenarios of regulation covariates could be studied to establish statistically-based estimates in log10-cycle offsets of the effects of regulation. Key products of this project that were absolutely foundational to have reached this juncture are the availability of watershed properties for 1,703 streamgages by Yesildirek et al. (2021) and the *scNIDaregis* software that provides an engine for temporally integrating reservoir storages from the National Inventory of Dams and binds results to USGS annual peak streamflow series. These streamgages present over 58,000 years in aggregate of annual peak streamflow information from what constitutes nearly the entirety of the USGS streamgage network in Texas, Oklahoma, and eastern New Mexico. Extending a study area beyond just the geographic borders of Texas is important because the quantile regression is inherently data intensive. The data intensity topic is important because interest is in upper tail alone of the flood frequency distribution so half the data at each streamgage (the lower tail, usually nonfloods) are not really used in principle.

The authors are not aware of prior literature suggesting this type of statistical approach. The approach appears capable of accommodating the effects of regulation in flood frequency analyses and by extension effects of urbanization and(or) climate state. Though some of the most elementary concepts herein are a logical extension of results of Asquith (2001) and the spatial residual correction motivated by Asquith and Roussel (2009), the 2001 study, in particular, represents the full extent of prior involvement by TxDOT sponsored research into the realm of regulated streamflow.

An attractive aspect of the QR-Pearson III and demonstration in this chapter, though further study of the modeling form for the quantile regression is needed, is that implementation of the QR-Pearson III (once trained or fit) to a type of cloud-hosted service is plausible given the relatively straightforward technicals of the mathematics behind the scenes. The most complex numerical criteria are access to an incomplete and complete gamma function (or high quality linear-series approximations) and a multi-dimensional optimizer (root solver) to back fit a Pearson III distribution by method of percentiles. For concluding emphasis, the demonstration herein is documented in Cleveland and Fang (2021) along the file path: `0-6177-dataverse-archive/demo/demo02_effreg_quantreg.R`.

Chapter 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., `https://doi.org/10.3133/wri014243`.

Asquith, W.H., Cleveland, T.G., Yesildirek, M.V., Zhang, J., Fang, Z.N., and Otto, L.D., 2021, *scNIDaregis*—Geospatial processing of dams in the United States from the National Inventory of Dams with a state-level aggregation scheme, demonstrated for selected dams in eight states in south-central region of the United States, and post-processing features for basin-specific tabulation: U.S. Geological Survey software release, Reston, Va., `https://doi.org/10.5066/P90NJVB9`.

Asquith, W.H., Kiang, J.E., and Cohn, T.A., 2017, Application of at-site peak-streamflow frequency analyses for very low annual exceedance probabilities: U.S. Geological Survey Scientific Investigation Report 2017–5038, 93 p., `https://doi.org/10.3133/sir20175038`.

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 20095087, 48 p., `https://doi.org/10.3133/sir20095087`.

Burnett, J., 2008, Flash floods in Texas: Texas A&M Press, ISBN 978–1–58544–590–5.

Cleveland, T.G., and Fang, Z.N., 2021, Texas-Skew-Update-2021: Texas Data Repository, `https://doi.org/10.18738/T8/SVLCOQ`.

England, J.F., Cohn, T.A., Faber, B.A., Stedinger, J.R., Thomas Jr., W.O., Veilleux, A.G., Kiang, J.E., and Mason, R.R., 2018, Guidelines for determining flood flow frequency Bulletin 17C: U.S. Geological Survey Techniques and Methods, book 4, chap. 5.B, 148 p., `https://doi.org/10.3133/tm4B5`.

Herrmann, G.R., 2013, Conceptual, algorithmic, and statistical exploration of relations between runoff generation, stream geomorphology, and watershed topography in west Texas: Doctoral Dissertation, Texas Tech Univeristy, `https://ttu-ir.tdl.org/handle/2346/58209`.

R Core Team, 2021, R—A language and environment for statistical computing: R Foundation for Statistical Computing, Vienna, Austria, version 4.0.3, accessed March 1, 2021, at https://www.r-project.org.

Texas Department of Transportation, 2020, Hydraulic design manual—September 2019: online resource accessed on July 26, 2020, at https://onlinemanuals.txdot.gov/txdotmanuals/hyd/index.htm and subsection https://onlinemanuals.txdot.gov/txdotmanuals/hyd/statistical_analysis_of_stream_gauge_data.htm.

U.S. Army Corps of Engineers, 2020, CorpsMap—National Inventory of Dams, accessed on July 7, 2020 at https://nid.sec.usace.army.mil/ords/f?p=105:1::::::.

U.S. Geological Survey, 2018, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed April 9, 2018, at https://doi.org/10.5066/F7P55KJN.

U.S. Geological Survey (USGS), 2020, PeakFQ—Flood frequency analysis based on Bulletin 17B and recommendations of the Advisory Committee on Water Information (ACWI) Subcommittee on Hydrology (SOH) Hydrologic Frequency Analysis Work Group (HFAWG), version 7.2, accessed January 29, 2019, at https://water.usgs.gov/software/PeakFQ/, version 7.3, accessed February 1, 2020, at https://water.usgs.gov/software/PeakFQ/.

Wagner, D.M., Kiang, J.E., and Asquith, W.H., 2017, U.S. Geological Survey streamgaging methods and annual peak streamflow, appendix 1 *of* Asquith, W.H., Kiang, J.E., and Cohn, T.A., 2017, Application of at-site peak-streamflow frequency analyses for very low annual exceedance probabilities: U.S. Geological Survey Scientific Investigations Report 2017–5038, 93 p., https://doi.org/10.3133/sir20175038.

Yesildirek, M.V., McDowell, J.S., Zhang, J., Asquith, W.H., 2021, Geospatial data of watershed characteristics for select U.S. Geological Survey streamgaging stations in New Mexico, Oklahoma, and Texas useful for statistical study of annual peak streamflows in and near Texas: U.S. Geological Survey data release, https://doi.org/10.5066/P9A91W4Z.

9. Report Summary

The flood hydrology of Texas, Oklahoma, and eastern New Mexico (east of the Great Continental Divide) is complex because of a myriad of meteorological and physiographic factors, and flood hydrology is typically studied using the annual peak streamflow data collected by the U.S. Geological Survey (USGS) at streamgages. Hydraulic design engineers need standard of practice guidance for various tasks involving the analysis and application peak streamflow information. Analyses of this information materially influence bridge design, operational safety of drainage infrastructure, flood-plain management, and other decisions influencing society.

Common tasks for the Texas hydraulic/hydrologic engineer are flood frequency analyses using streamgage data. The Research Project 0–6977, which was performed in joint collaboration between researchers at Texas Tech University, the University of Texas at Arlington, and the USGS Oklahoma-Texas Water Science Center, was tasked with a primary objective of updating the Texas generalized skew map and its mean-square error (chap. 4) for the Texas Department of Transportation (TxDOT) Hydraulic Design Manual. The current (September 2019) generalized skew in the Design Manual was last updated for TxDOT in 1996. Justification for a 2021 update is the fact that since 1996: (1) the mathematics in Federal guidance (Bulletin 17C) (England and others, 2018, `https://doi.org/10.3133/tm4B5`) have been updated (2018) after about 40 years and result in more accurate estimation of skew and the attendant error of skew for streamgages and (2) about 22 years more annual peak streamflow data from USGS streamgages data are available.

Generalized skew is important because it helps to reliably shape the probability distribution used in flood frequency analyses and particularly improve flood estimation and narrow confidence limits for rare events beyond about the 25-year return period. This final project report provides for update of generalized skew in the Hydraulic Design Manual. The generalized skew though is most applicable (but not uniquely so) to watersheds that are considered unregulated and undeveloped (not urbanized). As a foundational basis, the update used at least 30 years of streamflows acquired by 444 USGS long-term (30 years or more of record) streamgages representing such watersheds in Texas, Oklahoma, and eastern New Mexico. Data from outside of Texas, in particular from Oklahoma and eastern New Mexico, greatly enhances statistical methods in general, and in particular, many drainages in or pertinent to Texas extend upstream into these states. A novel application of a generalized

additive model is used for 2-dimensional spatial regression on so-called station-skew values from the streamgages to create predictions of generalized skew on the 1-kilometer USGS National Hydrogeologic Grid clipped to the study area.

Secondary objectives of the project were to provide as shown in this report training materials (chap. 5) using Texas watershed examples on flood frequency analysis oriented around Bulletin 17C, the use and example impacts of the updated generalized skew, and describe nuances with using USGS peak streamflow data that are not well represented by the Design Manual. Tertiary objectives of the Research Project and reported on herein further extend understanding of flood hydrology in the region inclusive of (1) climate sensitivity of peak streamflows (chap. 6), (2) experimental evaluation of a USGS multiorder hydrologic position (MOHP) metric that is gridded at continental scale and expresses stream reach position on the landscape (chap. 7), and (3) experimental accommodation of effects of regulation (chap. 8). The MOHP evaluation results semi-quantitative description of the information that MOHP could bring to statistical estimation of flood-frequency at ungaged (unmonitored) locations. The MOHP evaluation is cross-linked, but separately executed study, to the experimental effort towards how stakeholders in flood hydrology might accommodate the effects of regulation by reservoirs into statistical methods. The approach therein is based on quantile regression of the upper tail of flood magnitudes and back fitting of Pearson type III distributions. The accommodation of regulation effects is designed with deliberate views towards implementation constraints of information-technology-mathematical structures suitable for end users (engineers and other practitioners).

The USGS published persistent, publicly-accessible, Federal archival of three products stemming from this project. First, Yesildirek and et al. (2021, `https://doi.org/10.5066/P9A91W4Z`) provide extensive watershed properties and ancillary metadata for 1,703 USGS streamgages (inclusive of the aforementioned 444) in Texas, Oklahoma, and eastern New Mexico. These streamgages have at least 6 years of peak streamflows and also include the 50 odd streamgages currently (2021) sponsored by TxDOT in other program funding activity (2006–present [2021]). Many of these streamgages, including the aforementioned those of sponsored by TxDOT, have never had systematic watershed properties available until completion of this project. Geospatial polygons of the watersheds (the "Yesildirek watershed polygons") are also published.

Second, Asquith, Cleveland, et al. (2021, `https://doi.org/10.5066/P90NJVB9`) published software to be used the Yesildirek watershed polygons, overlay them on the U.S. Army Corps of Engineers National Inventory of Dams (NID), and the temporally integrate cumulative storages of dams (reservoirs) in the watershed, and finally, bind (associate) on a year-by-year basis peak streamflows to cumulative reservoir storages. This NID-related software provides for a large leap forward from TxDOT sponsored research into effects of regulation of peak streamflows at the end of the 20th century.

Third, Asquith et al. (2020, https://doi.org/10.5066/P9CW9EF0) published software for computation of the multiple Grubbs–Beck test for "low outliers" in annual peak series following Bulletin 17C; this software could be useful in semi-automated, large-scale, statistical studies by conditionally truncating nonfloods from the peak streamflow data, which is an area particularly prone to drought signals entering peak streamflow databases.

Finally, this report along with its applicable data archival into the Texas Digital Library (tdl.org) Texas Data Repository (Cleveland and Fang, 2021, https://doi.org/10.18738/T8/SVLCOQ) and the three USGS publications resulting from the project establish a comprehensive and well-documented update of generalized skew and set a foundation for future statistical research into Texas, Oklahoma, and eastern New Mexico flood hydrology. Data preparation efforts, in particular, for future researchers are immensely streamlined by the publicly-accessible publications and information archival completed as part of TxDOT Research Project 0–6977.