CIVIL ENGINEERING STUDIES
Illinois Center for Transportation Series No. 23-019
UILU-ENG-2023-2019
ISSN: 0197-9191

Estimating Peak-flow Quantiles for Selected Annual Exceedance Probabilities in Illinois

Prepared By
Thomas M. Over
Mackenzie K. Marti
Padraic S. O'Shea
Jennifer B. Sharpe
U.S. Geological Survey

Research Report No. FHWA-ICT-23-014

A report of the findings of

ICT PROJECT R27-181
Updated and Unified StreamStats Peak Discharges for Streams of Illinois

https://doi.org/10.36501/0197-9191/23-019

Illinois Center for Transportation

September 2023

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-ICT-23-014	N/A	N/A

4. Title and Subtitle	5. Report Date
Estimating Peak-Flow Quantiles for Selected Annual Exceedance Probabilities in Illinois	September 2023
	6. Performing Organization Code N/A

7. Authors	8. Performing Organization Report No.
Thomas M. Over, https://orcid.org/0000-0001-8280-4368 Mackenzie K. Marti, https://orcid.org/0000-0001-8817-4969 Padraic S. O'Shea, https://orcid.org/0000-0001-9005-8289 Jennifer B. Sharpe, https://orcid.org/0000-0002-5192-7848	ICT-23-019 UILU-2023-2019

9. Performing Organization Name and Address	10. Work Unit No.
Illinois Center for Transportation Department of Civil and Environmental Engineering University of Illinois at Urbana-Champaign 205 North Mathews Avenue, MC-250 Urbana, IL 61801	N/A **11. Contract or Grant No.** R27-181

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Illinois Department of Transportation (SPR) Bureau of Research 126 East Ash Street Springfield, IL 62704	Final Report 10/1/17–12/31/23 **14. Sponsoring Agency Code**

15. Supplementary Notes
Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration. https://doi.org/10.36501/0197-9191/23-019. Complementary data releases are available for this report: https://doi.org/10.5066/P9YIAUZQ, https://doi.org/10.5066/P9ZAMASB, https://doi.org/10.5066/P9XPWUMI, and https://doi.org/10.5066/P9XUH9SR

16. Abstract
This report presents the methods, results, and applications of an updated flood-frequency study for the State of Illinois. This study, which uses data through September 2017, updates two previous studies that used data through 1999 and 2009, respectively. Flood-frequency estimates are used for a variety of land-use planning and infrastructure design purposes, including for the hydraulic design of bridges. The flood frequencies presented are the stream discharges (or quantiles) that are exceeded each year with a given probability, called the annual exceedance probability (AEP). Estimates are provided for eight AEPs: 0.5, 0.2, 0.1, 0.04, 0.02, 0.01, 0.005, and 0.002. The basic data used are the annual maximum discharges (peak flows) at streamgages in and near Illinois and the measurable physical properties (basin characteristics) of the basins upstream from those streamgages. The logarithms of the peak-flow records at qualifying streamgages were fit to the Pearson Type 3 distribution using federally recommended methods, providing one of the two primary products of the study, which is the set of peak-flow quantiles at the streamgages and their uncertainties. The other primary product is sets of equations, determined by least-squares regression, that relate the peak-flow quantiles and the basin characteristics. There is one set of equations for each of Illinois' seven hydrologic regions, and, for each region, one equation for each AEP. These regional equations allow the estimation of peak-flow quantiles at ungaged locations throughout the state and their uncertainties. A set of illustrative example applications is provided, and tables of all results at all study streamgages and regions are provided in associated data releases. The study results are also being made available in a web application, StreamStats.

17. Key Words	18. Distribution Statement
Floods, Peak Discharge, Regression Analysis, Reservoirs, Statistical Analysis, Streamflow, Streamgages, Urbanization, Watersheds, Basin Characteristics, Pearson Type 3 Distribution, StreamStats	No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	72	N/A

ACKNOWLEDGMENT, DISCLAIMER, MANUFACTURERS' NAMES

This publication is based on the results of **ICT-R27-181: Updated and Unified StreamStats Peak Discharges for Streams of Illinois**. ICT-R27-181 was conducted in cooperation with the Illinois Center for Transportation; the Illinois Department of Transportation; and the U.S. Department of Transportation, Federal Highway Administration.

Members of the Technical Review Panel (TRP) were the following:

- Neil VanBebber, TRP Chair, Illinois Department of Transportation
- Dan Brydl, Federal Highway Administration
- Steve Ferguson, Illinois Department of Transportation
- Nicholas Jack, Illinois Department of Transportation
- Michelle Lewis, Illinois Department of Transportation
- Nick Lombardi, Federal Highway Administration
- Perry Masouridis, Illinois Department of Transportation
- Matthew O'Connor, Illinois Department of Transportation
- Megan Swanson, Illinois Department of Transportation

EXECUTIVE SUMMARY

This report presents the methods, results, and applications of an updated flood-frequency study for the State of Illinois. Flood-frequency estimates are used for a variety of land-use planning and infrastructure design purposes, including for the hydraulic design of bridges. The study, which used stream discharge data recorded through September 2017, updated two studies currently in use that used data through 1999 and 2009, respectively. The basic data used in the analyses were the annual maximum discharges (peak flows) at qualifying streamgages in and near Illinois and the measurable physical properties (basin characteristics) of the basins upstream from those streamgages.

The study presented in this report has three primary products: (1) flood-frequency estimates (and their uncertainties) at the study streamgages; (2) sets of equations (called regional regression equations) that relate the basin characteristics and flood-frequency estimates, allowing estimates of flood frequencies at ungaged locations throughout the state and the uncertainty of such estimates; and (3) implementation of the regional regression equations in the web application StreamStats, available at https://streamstats.usgs.gov/ss/.

At-site flood-frequency estimates computed in this study were the stream discharges that were exceeded each year with a given set of probabilities, called the annual exceedance probability (AEP). Such discharges are also called peak-flow quantiles. Estimates were provided for eight AEPs: 0.5, 0.2, 0.1, 0.04, 0.02, 0.01, 0.005, and 0.002. These AEPs were the reciprocal of the so-called return period of the floods. For example, the peak-flow quantile with an AEP of 0.01 is the flood with a 100-year return period, or simply the 100-year flood. These peak-flow quantiles were computed for all streamgages in the state with at least 10 years of record (and for one streamgage that had 9 years) that have roughly stationary drainage basin conditions.

The regional regression equations that relate the basin characteristics and flood-frequency estimates were developed using a subset of the qualifying streamgages. One set of equations was developed for each of seven conterminous hydrologic regions covering Illinois. Except for one region that includes the Chicago area, these equations did not use data that include substantial effects of regulation or urbanization, and therefore the equations were not directly applicable to such basins. However, a method was provided to adjust the results for the effect of urbanization. A set of illustrative example applications of these equations was provided, and tables of all results for all study streamgages and regions are provided in associated data releases.

The implementation of the study results in StreamStats will include access to the peak-flow quantiles at the study streamgages and a means of applying the regional regression equations to get peak-flow quantile estimates at ungaged locations throughout the state. The implementation will also include the improved, higher-resolution geographic base-data used for determining the basin characteristics used in study.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1

PURPOSE AND SCOPE 1

DESCRIPTION OF STUDY AREA 2

PREVIOUS STUDIES 4

CHAPTER 2: DATASET DEVELOPMENT 6

SITE SELECTION AND RECORD CLASSIFICATION 6

Northeastern Illinois (Region 2) 6

Outside of Northeastern Illinois (Non-Region 2) 11

Urbanization 11

Regulation 12

Statewide Dataset Properties 13

COMPUTATION OF OBSERVED PEAK-FLOW QUANTILES 16

Flow Intervals and Perception Thresholds for the Expected Moments Algorithm 16

Tests for Potentially Influential Low Floods 18

Regional Skew 18

Results of the Observed Peak-Flow Quantile Computations 24

BASIN CHARACTERISTICS 26

CHAPTER 3: DEVELOPMENT OF REGIONAL REGRESSION EQUATIONS 29

DETERMINATION OF HYDROLOGIC REGIONS 29

TESTING FOR REDUNDANCY 29

TRANSFORMATION OF BASIN CHARACTERISTICS 30

ORDINARY LEAST-SQUARES REGRESSION 31

GENERALIZED LEAST-SQUARES REGRESSION 31

SELECTED MODELS 33

ACCURACY OF FINAL SPATIAL REGRESSION EQUATIONS 38

CHAPTER 4: APPLICATIONS 44

APPLICATIONS OF REGIONAL REGRESSION EQUATIONS 44

At Ungaged Sites in Region 2 44

At Ungaged Sites Outside of Region 2 47
At Streamgages 52
Near Streamgages 53
INTERPOLATION BETWEEN STREAMGAGES 55
CHAPTER 5: LIMITATIONS 62
CHAPTER 6: STREAMSTATS 64
CHAPTER 7: SUMMARY 65
REFERENCES 67

LIST OF FIGURES

Figure 1. Map. Streamgages used in this study (a higher resolution version is available in Marti et al., 2023) 7

Figure 2. Map. Streamgages used in hydrologic region 2. 8

Figure 3. Map. Streamgages used in Cook, DuPage, and Southern Lake Counties, Illinois. 9

Figure 4. Chart. Histograms of properties of peak-flow data and streamgages used in this study in region 2: (A) annual peak-flows each water year, defined as the period from October 1 to September 30, designated by the year in which it ends; (B) drainage areas of streamgages, and (C) median urban fractions of streamgages. 14

Figure 5. Chart. Histograms of properties of peak-flow data and streamgages used in this study outside region 2: (A) annual peak-flows each water year, defined as the period from October 1 to September 30, designated by the year in which it ends; (B) drainage areas of streamgages; and (C) median urban fractions of streamgages. 15

Figure 6. Graph. Statistics of ratios of quantiles determined using weighted independent estimate skew (Q_{WIE_Skew}) to those using Bulletin 17C weighted skew (Q_{B17C_Skew}): (A) streamgages in region 2; (B) streamgages outside of region 2. 21

Figure 7. Graph. Flood-frequency curves from PeakFQ software for U.S. Geological Survey streamgage 05436900 Otter Creek Tributary near Durand, Illinois: (A) using Bulletin 17C weighted skew; (B) using weighted independent estimate (WIE) skew. 23

Figure 8. Equation. Definition of drainage area overlap fraction. 30

Figure 9. Equation. Functional form of cross-correlation between concurrent streamgage records. 32

Figure 10. Equation. Log-space variance of prediction. 42

Figure 11. Equation. Standard error of prediction in percent units. 42

Figure 12. Equation. Confidence interval for predicted discharge quantile. 42

Figure 13. Equation. Confidence interval for predicted discharge quantile transformed out of log space. 43

Figure 14. Equation. Calculating the 1% annual exceedance probability peak-flow quantile at an ungaged location in region 2. 45

Figure 15. Equation. Row vector (x_i) of basin characteristics at an ungaged location in region 2. 45

Figure 16. Equation. Covariance matrix $(X^t \Lambda^{-1} X)^{-1}$ at an ungaged location in region 2. 46

Figure 17. Equation. Calculating $(x_i (X^t \Lambda^{-1} X)^{-1} (x_i)^t)$ at an ungaged location in region 2. 46

Figure 18. Equation. Calculating the standard error of prediction in percent at an ungaged location in region 2. 46

Figure 19. Equation. Calculating the confidence intervals at an ungaged location in region 2. 47

Figure 20. Equation. Transforming the confidence intervals out of log space at an ungaged location in region 2. 47

Figure 21. Equation. Urbanization-adjusted peak-flow quantile at an ungaged location outside of region 2. 47

Figure 22. Equation. Log-transformed urbanization-adjusted peak-flow quantile at an ungaged location outside of region 2. 47

Figure 23. Equation. Variance of prediction of the urbanization-adjusted peak-flow quantile at an ungaged location outside of region 2. 48

Figure 24. Equation. Calculating the unadjusted 1% Annual Exceedance Probability peak-flow quantile at an ungaged location in region 4. 49

Figure 25. Equation. Urbanization-adjusted peak-flow quantile at an ungaged location in region 4. .. 49

Figure 26. Equation. Log-transformed urbanization-adjusted peak-flow quantile at an ungaged location in region 4. 49

Figure 27. Equation. Row vector (x_{U0}) of basin characteristics at an ungaged location in region 4. ... 50

Figure 28. Equation. Covariance matrix ($X^t \Lambda^{-1} X)^{-1}$ at an ungaged location in region 4. 50

Figure 29. Equation. Computation of $\boldsymbol{x_{U0}(X^t \Lambda^{-1} X)^{-1}(x_{U0})^t}$ at an ungaged location in region 4. 50

Figure 30. Equation. Variance of prediction (V_u) of the urbanization-adjusted peak-flow quantile at an ungaged location in region 4. 51

Figure 31. Equation. Calculating the standard error of prediction (S_u) in percent at an ungaged location in region 4. 51

Figure 32. Equation. Calculating the confidence intervals of a peak-flow quantile in log units at an ungaged location in region 4. 51

Figure 33. Equation. Transforming the confidence intervals out of log space at an ungaged location in region 4. 51

Figure 34. Equation. Weighted independent estimate mean of the regression equation estimate and the at-site flood-frequency analysis. 52

Figure 35. Equation. Variance of prediction for the weighted mean of the regression equation estimate and the at-site flood-frequency analysis. 53

Figure 36. Equation. Near-gage adjustment weighting factor. 53

Figure 37. Equation. Near-gage adjustment equation. 54

Figure 38. Equation. Adjustment weighting factor for ungaged location in region 2 near a streamgage. 55

Figure 39. Equation. Near-gage adjustment for the 0.01 Annual Exceedance Probability at an ungaged location in region 2 near a streamgage. 55
Figure 40. Equation. Peak-flow quantile interpolation function for two streamgages on the same reach. 56
Figure 41. Equation. Example peak-flow quantile interpolation computation. 56
Figure 42. Diagram. Rock River interpolation illustration. 57
Figure 43. Diagram. Fox River interpolation illustration. 58
Figure 44. Diagram. Sangamon River interpolation illustration. 59
Figure 45. Diagram. Kaskaskia River interpolation illustration. 60
Figure 46. Diagram. Big Muddy River interpolation illustration. 61

LIST OF TABLES

Table 1. Streamgage Counts by Region [USGS, U.S. Geological Survey; RRE, regional regression equation] 10

Table 2. Variance Statistics of Observed Peak-Flow Quantiles by Annual Exceedance Probability (AEP) from O'Shea, 2023 25

Table 3. Example Basin Characteristics Considered for Streamgages Outside of Region 2 28

Table 4. Selected Regional Regression Models for Each Region [Q_p, peak-flow quantile at annual exceedance probability p] 34

Table 5. Information on Basin Characteristics for Selected Regional Regression Models 35

Table 6. Coefficients of Selected Regional Regression Models for Each Region at an Annual Exceedance Probability of 0.01 36

Table 7. Standard Errors and t-ratio of Selected Regional Regression Models for Each Region at an Annual Exceedance Probability of 0.01 36

Table 8. Accuracy of Selected Regional Regression Models for Each Region at an Annual Exceedance Probability of 0.01 40

Table 9. Definitions of Goodness-of-Fit Statistics 41

Table 10. Ranges of Basin Characteristics Used in Regional Regression Equation Development 44

Table 11. Urban Fraction Coefficients from Temporal Quantile Regression Analysis 48

Table 12. Baseline Urban Fractions 48

CHAPTER 1: INTRODUCTION

Estimates of the magnitude and frequency of floods, which in this report refers to the discharge corresponding to a given annual exceedance probability (AEP)—that is, the discharge that is exceeded with a given probability each year, are a critical ingredient to floodplain management, emergency response planning, and infrastructure design tasks such as sizing of bridges and culverts. At streamgages, such estimates can be obtained directly from observations, usually by fitting to a selected theoretical probability distribution, although direct estimates are often adjusted using regional information—that is, information obtained from multiple streamgages in a region. A variety of techniques have been used to provide estimates at ungaged locations (Rosbjerg et al., 2013). Those techniques include statistical methods, such as regression of observed flood-frequency information from the streamgages in a region on a set of basin characteristics, and the development and application of process-based models, which also use observed flood frequencies and basin properties but in a different modeling framework. This study determines flood-frequency estimates at streamgages using the federally recommended method presented in Bulletin 17C (England Jr. et al., 2019) and at ungaged locations using regression on basin characteristics, using techniques developed beginning in the 1980s (refer to Eng et al., 2009, or Farmer et al., 2019, for an overview) that are customized for use with flood-frequency estimates determined using Bulletin 17C.

PURPOSE AND SCOPE

This report presents the methods and results of a study that determined flood-frequency estimates (called here peak-flow quantiles) at streamgages throughout Illinois with AEPs of 0.5, 0.2, 0.1, 0.04, 0.02, 0.01, 0.005, and 0.002 (often referred to as 2-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year floods, respectively). The study used the methods described in Bulletin 17C (England Jr. et al., 2019) and applied a regional spatial regression approach to develop equations for estimating peak-flow quantiles at the same AEPs for a set of seven hydrologic regions that cover the State of Illinois. Both elements of the study use stream discharge data through water year (WY) 2017. A water year runs from October 1 to September 30 and is named for the year in which it is ends; thus, the most recent data used in this study are from September 2017. This report updates two previous studies: Soong et al. (2004) presented peak-flow quantiles at streamgages and regional regression equations (RREs) for ungaged rural locations throughout Illinois using data through WY 1999. Over et al. (2021) presented peak-flow quantiles at streamgages and RREs for ungaged locations inclusive of urbanized basins in a hydrologic region encompassing the Chicago area in northeastern Illinois using data through WY 2009. Outside of the northeastern Illinois region, the spatial regression equations were developed exclusive of basins that have substantial regulation or urbanization and, thus, are not directly applicable to such basins. However, a method for adjusting for the effect of urbanization developed by Over et al. (2021) is presented that enables estimation of quantiles at urbanized basins throughout Illinois. Despite the exclusion of most urbanized and regulated basins from use in the RREs, peak-flow quantiles at streamgages monitoring such basins were determined and are included in the products of this study. Large rivers that are not necessarily regulated but would constitute the only basin of such size in their associated region (for example, the Rock River in northwestern Illinois) were also excluded from use in the RREs, but their peak-flow quantiles are included in the products of this study. For selected periods of record at large or regulated rivers where it was deemed to be

appropriate, figures illustrating an interpolation technique that provides peak-flow quantiles at ungaged locations along such rivers are also provided.

The products of the study described in this report comprise three items. The first item is this report, which describes the methods as well as summarizes and illustrates how to use the results of the study. The second item is four U.S. Geological Survey data releases. The third item is the update and implementation of RREs and other study results into the online application StreamStats (https://streamstats.usgs.gov/ss). The update will enable delineation of basins upstream from points of interest, determination of the values of the characteristics of those basins and the relevant RREs, and the computation of the associated peak-flow quantiles. The four data releases are listed as follows. Schafer and Sharpe (2023) provide the geographic data used to update StreamStats, using a 10 m (32.8 ft) digital elevation model and 1:24,000-scale hydrography (river network). Schafer et al. (2023) provide the basin characteristics used in the selected RREs as digital geographic datasets. O'Shea (2023) provides the input and output files used in this study in the determination of the peak-flow quantiles at streamgages using the PeakFQ software (Veilleux et al., 2014) and provides summary tabulations of the input and results for each streamgage. Marti et al. (2023) provides tables of basin characteristics and peak-flow quantiles by streamgage and tables of RRE coefficients and statistics by AEP and hydrologic region.

DESCRIPTION OF STUDY AREA

The study area includes the State of Illinois and selected additional river basins in Wisconsin and Indiana that drain into Illinois (Figures 1–3). Large parts of the boundaries of Illinois are defined by major rivers, comprising the Mississippi to the west, the Ohio to the south, and the Wabash to the southeast, and although Illinois is subject to flooding from these rivers, they are not considered in this study. The river with the largest drainage area internal to Illinois is the Illinois River, which begins at the confluence of the Kankakee and Des Plaines Rivers in northeastern Illinois and discharges into the Mississippi River at Grafton, Illinois, a few miles upstream from St. Louis. Peak-flow quantiles at streamgages on the Illinois River are presented in this report, but because of its high degree of regulation, they were not considered for use in the RREs.

Although mediated by surface and near-surface landscape properties, floods are climatically driven. In the flood climate classification of Hayden (1988) most of Illinois lies in a flood region characterized by barotropic (i.e., nonfrontal or convective) and unorganized (i.e., without tropical cyclones) summer storms with potential for frontal storms throughout the year and ephemeral winter snow cover for 10–50 days per year that may contribute to flooding during winter when rain falls on existing snow. Northern Illinois lies at the border of this region and a region with similar characteristics except with snow cover of more than 50 cm (20 in.) that may contribute to substantial spring snowmelt flooding (Hayden, 1988). More broadly, Illinois has a continental climate, classified as "Dfa" in the Köppen-Geiger system, with warm, humid summers, cold winters, and precipitation common throughout the year (Frankson et al., 2022; Peel et al., 2007). Its location within the interior of the continent makes it susceptible to large temperature variations, contributed to by the convergence of cold arctic air masses and relatively hot, humid air masses from the Gulf of Mexico (Frankson et al., 2022). Mean temperatures increase by about 12°F from the north to the south, from about 47°F to 59°F, with Lake

Michigan helping to moderate the otherwise cold winters and hot summers of the nearby areas (Changnon et al., 2004). Due to the low relief throughout the state, topographical effects on climate are not substantial. Like temperature, precipitation follows a mostly north-south gradient within the state, ranging from about 32–35 inches mean annual precipitation in the north to about 48–50 inches in the south, depending on the years of record considered (Frankson et al., 2022). Thunderstorms are the largest contributor to total precipitation in Illinois, with snowfall contributing about 10%–15% of annual precipitation in northern Illinois and only 3%–5% in the most southern reaches of the state (Changnon et al., 2004). Precipitation occurs throughout the year across the state, with May–August the wettest months for northern Illinois, with most months from March–September exhibiting heavy rainfall in central Illinois, and January–June the wettest period for southern Illinois (Frankson et al., 2022).

With regard to physiography, as described by ecoregions, most of Illinois is covered by two U.S. Environmental Protection Agency level III ecoregions with similar coverage extent: the Central Corn Belt Plains in most of northern, central, and eastern Illinois and the Interior River Valleys and Hills in most of southern and western Illinois (Woods et al., 2006). The northernmost part of the state is covered by small areas of Southeastern Wisconsin Till Plains and the Driftless Area. The southern part of the state also has small areas of the Interior Plateau and Mississippi Alluvial Plain ecoregions. The Central Corn Belt Plains are composed of Wisconsinan-aged (33,000 to 12,000 years ago, Curry et al., 2010; Hansel & McKay, 2010) glaciated plains that were covered by extensive prairies until the early 19th century when large areas were tiled and ditched to drain the land and make it more suitable for agriculture and settlement, for which the region is primarily used today (Woods et al., 2006). This ecoregion has very little topographic variation and has a less extensive natural drainage system than the Interior River Valleys and Hills ecoregion (Woods et al., 2006). The Interior River Valleys and Hills ecoregion consists of Illinoian-aged (190,000 to 130,000 years ago, Curry et al., 2010; Hansel & McKay, 2010) glacial deposits and till plains, with sections of extensive drainage, valleys, and major rivers (Woods et al., 2006).

To assess the runoff potential of different areas, soils are grouped into four main hydrologic soil groups based on their permeability and runoff potential by the U.S. Department of Agriculture Natural Resources Conservation Service (2009). These groups range from group A soils, which are the most permeable and have the smallest runoff potential, to group D soils, which are the least permeable and have the highest runoff potential. Group B soils are found in the northern, western, and central parts of the state, in much of the Central Corn Belt Plains ecoregion and the northern part of the Interior River Valleys and Hills ecoregion, and have moderately low runoff potential, or moderately high permeability (U.S. Department of Agriculture Natural Resources Conservation Service, 2009). Some small areas of the Illinois River valley have group A soils. Soil groups C and D are found primarily in northeastern and southern Illinois; these soils have moderately high and high runoff potential, respectively. These lower permeability soils are associated with glacial moraines in northeastern Illinois and, in southern Illinois, with the southern part of the Interior River Valleys and Hills ecoregion and the extent of the Interior Plateaus ecoregion. In the southern Illinois portion of the Interior River Valleys and Hills ecoregion, these soils with higher runoff potential coincide with hilly terrain, increasing the flood potential. Some areas throughout the state have dual soil groups, like the dual C/D group present in southern Illinois in the hill and valley areas common in this region.

The dual soil groups represent soils that have higher runoff potential because of a high-water table and are placed in a soil group with lower runoff potential if the soil is adequately drained.

According to the 2016 National Land Cover Dataset (Dewitz, 2019; Yang et al., 2018), land cover in Illinois is generally dominated by crops, with about 63% of the total land devoted to cultivated crops, and which are especially prevalent in the Central Corn Belt Plains ecoregion. Also present are areas of mostly deciduous forest, at about 13% of total land cover, concentrated mostly in the southern and western parts of the state, in the Interior River Valleys and Hills ecoregion, and elsewhere along river corridors. Developed areas, which make up about 11% of the state land cover, are present as smaller urban areas throughout the state, but most notably in the northeast near the Chicago metropolitan area.

PREVIOUS STUDIES

Statewide flood-frequency analyses for Illinois go back to Pickels (1937), who analyzed the annual maximum daily values at 24 streamgages with record lengths from 8 to 31 years and proposed a method for applying the results to ungaged basins of at least 200 square miles. The next statewide study was that of Mitchell (1954), who used a regional, Gumbel distribution-based peak-flow estimation technique applied to instantaneous maxima at 108 streamgages. Mitchell determined mean values of physiographic and climatic factors by region to facilitate estimation of the mean annual flood and regional frequency curves, which together were to be used to estimate floods at ungaged locations using an index flood approach (Benson, 1962). Using procedures established in Bulletin 15 (U.S. Water Resources Council, 1967), which uses the log-Pearson Type III distribution like subsequent federal guidelines, Carns (1973) estimated flood quantiles and developed regression equations for estimating flood quantiles at seven return periods from 1.25 to 100 years using 172 stations with peak-flow record through WY 1967 and four hydrologic regions. Curtis (1977) estimated flood quantiles with return periods from 2 to 500 years and developed RREs based on data from 241 streamgages, using the same basin characteristics as Carns (1973) and four similar regions. Curtis (1987) updated his 1977 frequency analysis through WY 1985 using procedures from Bulletin 17B (U.S. Interagency Advisory Council on Water Data, 1982), which comprised the federal guidelines from its publication in 1982 until the recent publication of Bulletin 17C (England Jr. et al., 2019). Curtis (1987) developed regression equations with the same four regions as his 1977 study using data from 268 streamgages. The most recent statewide peak-flow frequency study for Illinois is that of Soong et al. (2004), who divided the state into seven hydrologic regions and used methods from Bulletin 17B to develop regression equations using data through WY 1999 at 288 streamgages.

Statewide flood-frequency analyses in Illinois have generally focused on rural basins. The effects of urbanization on floods and the development of RREs to predict peak-flow quantiles at urbanized basins in Illinois were first presented by Allen and Bejcek (1979). They used peak flows through WY 1976 at 103 streamgages and developed regression equations for peak-flow quantiles as functions of drainage area, impervious fraction, and slope. The Allen and Bejcek (1979) study was updated by Over et al. (2021) that provided two sets of regression equations for estimating peak-flow quantiles based on data from 117 gages in northeastern Illinois with data through WY 2009 and peak-flow frequency estimates for 181 gages. One set of regression equations was developed using time series

of urbanization and precipitation as explanatory variables and was used in the study to adjust the annual maximum peak-flows from the study watersheds to 2010 urbanization conditions. The other set of regression equations was based on the adjusted peak-flows, which were assumed to have been made stationary by the adjustment, based on evidence from application of the same method to a subset of basins in Over et al. (2016). Two primary aspects of Over et al. (2021) were included in this study: the use of the peak-flow values adjusted to 2010 urbanization conditions at the 181 streamgages in the northeastern Illinois region, and the continued use of basin characteristics selected by Over et al. (2021) in the RREs in the northeastern Illinois region of this study.

CHAPTER 2: DATASET DEVELOPMENT

Dataset development for this study consisted of two sets of data values. The first set is observed peak-flow quantiles which are the statistical relations between flood magnitudes and frequencies based on observed floods at selected streamflow-gaging stations (hereafter, streamgages), the computation of which is called flood-frequency analysis. The second set is basin characteristics which are quantitative descriptors of characteristics of basins draining to the locations of the streamgages. Regression equations for estimating the peak-flow quantiles at ungaged locations were then developed using these two sets of data values, as described in the next chapter.

Dataset development began with the selection of streamgages in and near Illinois and the classification of their annual maximum discharge (here, peak flow) data regarding their eligibility for use in developing RREs.

SITE SELECTION AND RECORD CLASSIFICATION

Streamgages within the State of Illinois and in basins draining into Illinois within 50 miles of the Illinois border were evaluated for their suitability for at-site analysis and use in regional regression development. Periods of record at streamgages were deemed to be unsuitable for at-site analysis and use in regional regression because of excessive regulation or urbanization or because the associated drainage basins were large compared to any other streamgages in the region. Streamgages with at least 10 years of peak-flow record through WY 2017 with approximately stationary regulation and urbanization in the basins upstream from each streamgage were considered appropriate for computation of their at-site peak-flow quantiles and potential inclusion in regional regression (Figures 1–3). The period of record at these streamgages varied by site and was not limited to a common period of record. Because urbanization and some regulation were accepted in the study of hydrologic region 2 in northeastern Illinois by Over et al. (2021), the streamgages in that region were evaluated somewhat differently from the rest of the state.

Northeastern Illinois (Region 2)

For region 2, Over et al. (2021) used records from WYs 1940–2009 adjusted to 2010 urbanization conditions as described in Over et al. (2016). The peak-flow data in this study were updated to include unadjusted peak flows for WYs 2010 through 2017 if the levels of urbanization and regulation did not change substantially (including urban fraction increasing less than 10% post-2009); details on the urbanization and regulation analyses are given below. Streamgages in region 2 that had fewer than 10 years of record for the period of analysis considered in Over et al. (2021) but now had at least 10 years of record through WY 2017 were considered for addition to the study, also subject to having only minor changes in urbanization and regulation (including a change of less than 15% in urban fraction over the period of record). Seven additional streamgages in region 2 were only appropriate for at-site flood-frequency analyses (Table 1), which is indicated by an entry of “Not valid for RREs” in the “Usage” column of Marti et al. (2023, Table DR-5). This includes one streamgage on the Illinois River and five on the Fox River that had substantial regulation and one streamgage on the Kankakee River that had a drainage area more than double the next largest streamgage used in the development of RREs for region 2. Six streamgages were rejected for at-site analysis (and therefore

from use in developing RREs) for reasons including large complex regulation and poor annual peak-flow record (Table 1; Marti et al., 2023, Table DR-6). The complete records of the selected 187 streamgages were considered appropriate for use in the RREs (Table 1), which is indicated by an entry of "Valid for use in RREs" in the "Usage" column of Marti et al. (2023, Table DR-5).

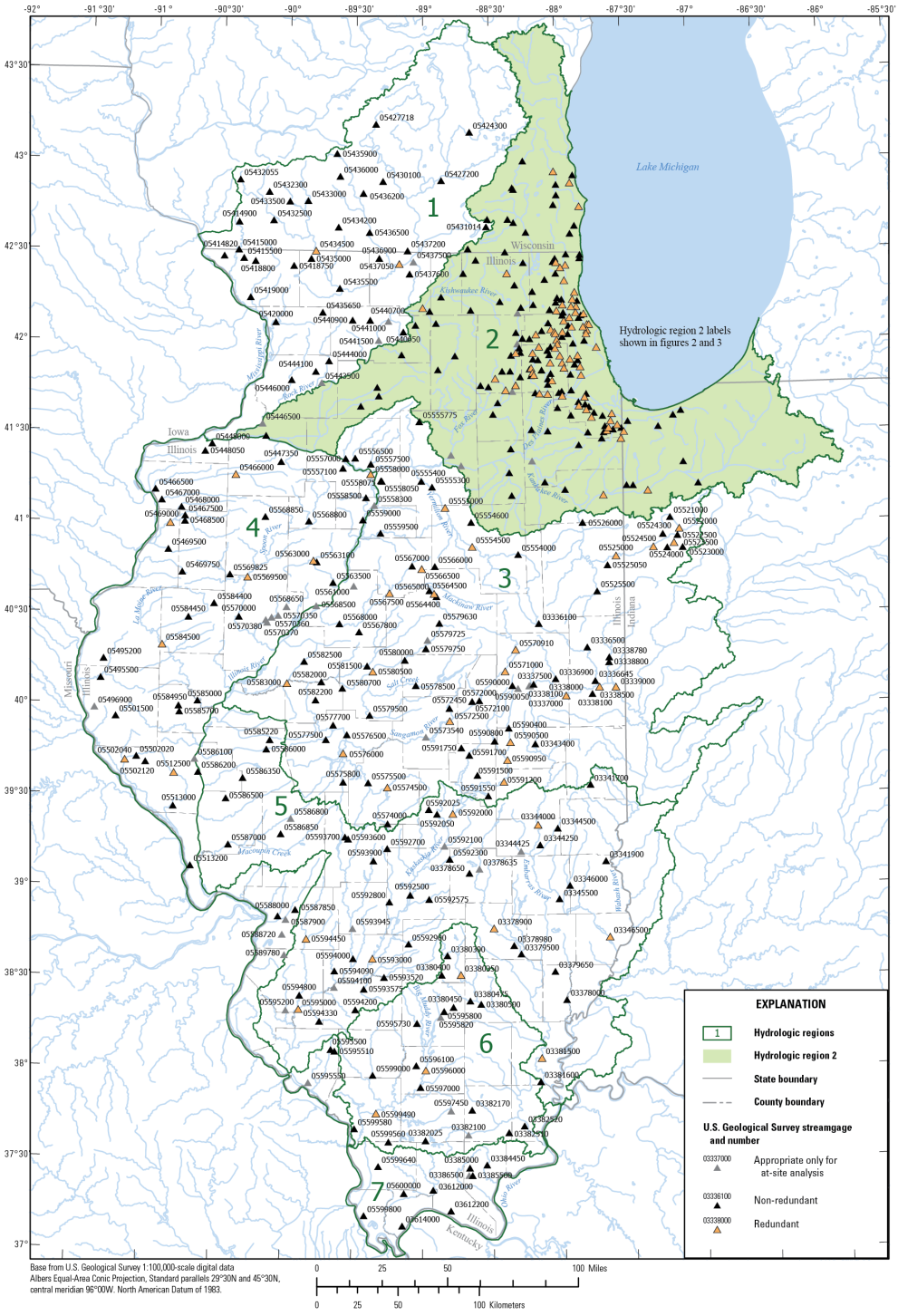

Figure 1. Map. Streamgages used in this study (a higher resolution version is available in Marti et al., 2023)

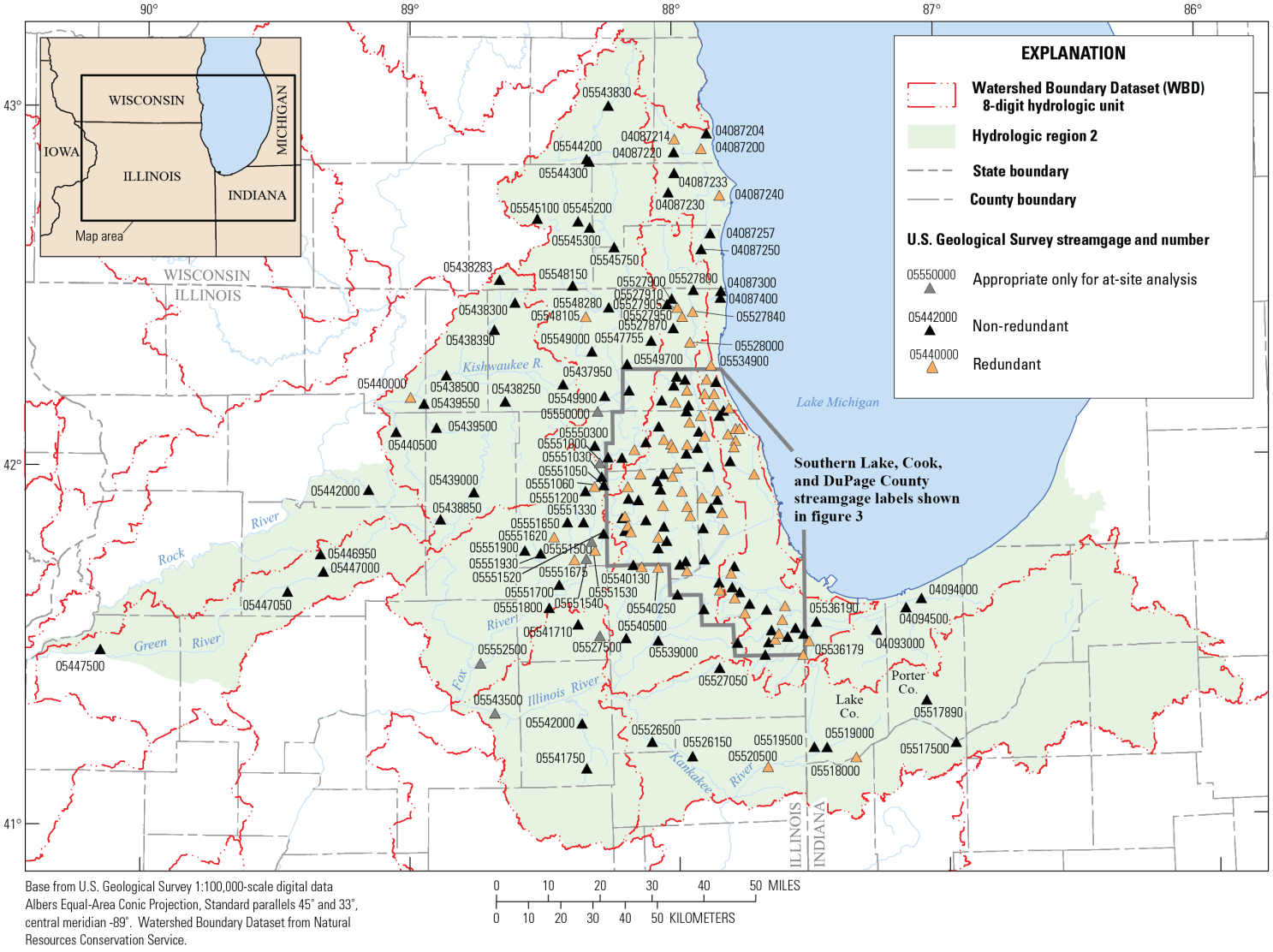

Figure 2. Map. Streamgages used in hydrologic region 2.

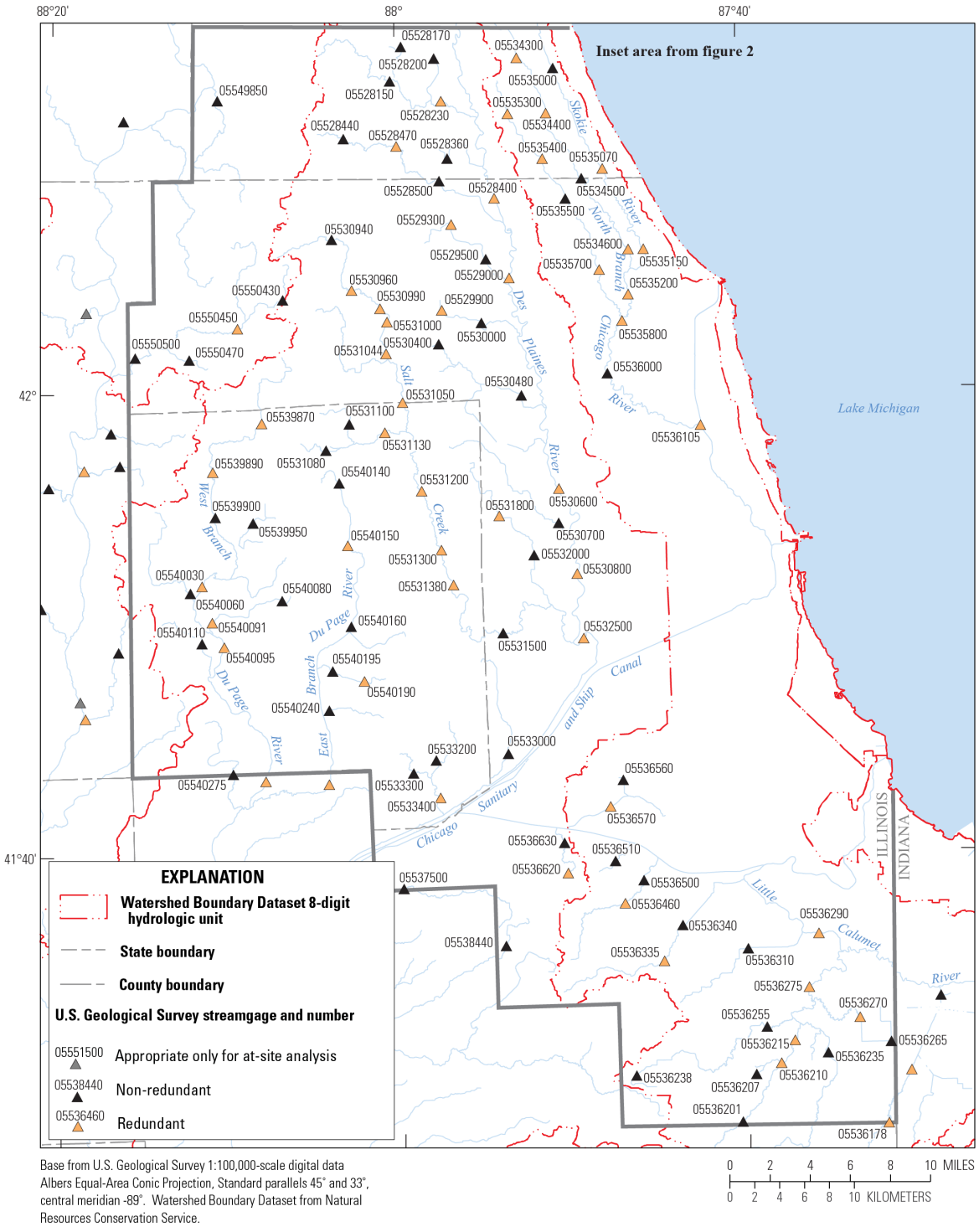

Figure 3. Map. Streamgages used in Cook, DuPage, and Southern Lake Counties, Illinois.

Table 1. Streamgage Counts by Region [USGS, U.S. Geological Survey; RRE, regional regression equation]

Region	Total USGS streamgages in region	Streamgages not appropriate for at-site analysis	Total streamgages considered in analysis	Streamgages only appropriate for at-site analysis	Total streamgages considered for RRE analysis	Redundant (Eligible for use in RREs but not used)	Non-redundant (used for RREs)	Streamgages with an urbanized or regulated period in addition to an RRE-eligible period	Total at-site periods of record
1	66	24	42	5	37	2	35	0	42
2	200	6	194	7	187	68	119	0	194
3	105	17	88	6	82	22	60	6	94
4	58	2	56	12	44	8	36	0	56
5	60	8	52	8	44	8	36	4	56
6	24	2	22	3	19	3	16	2	24
7	10	0	10	1	9	0	9	0	10
Total: regions 1, 3–7	323	53	270	35	235	43	192	12	282
Total: All regions	523	59	464	42	422	111	311	12	476

Outside of Northeastern Illinois (Non-Region 2)

Outside of hydrologic region 2—that is, in hydrologic regions 1 and 3–7—as defined in Soong et al. (2004; Figure 1), a total of 323 streamgages were evaluated for flood-frequency analysis (Table 1). Fifty-three streamgages in these regions were considered not appropriate for at-site analysis (Table 1; Marti et al., 2023, Table DR-6). They were rejected for having at least one of the following issues: excessive regulation, urbanization, or an inconsistent or broken period of record. A total of 270 streamgages were considered appropriate for flood-frequency analysis (Table 1), of which 8 are in Indiana and 19 are in Wisconsin (Marti et al., 2023, Table DR-5). At-site peak-flow quantiles associated with periods of record without substantial effects of urbanization (an urban fraction of less than 15%) or regulation were considered for the development of RREs.

For streamgages with an urbanized or regulated period, the pre- and post-urbanized (or regulated) periods were analyzed for changes in the level of urbanization and regulation. At-site peak-flow quantiles were computed for each period of at least 10 years of annual peak-flow record that had only minor changes in urbanization (changes in urban fraction of less than 15%) and regulation, which resulted in 12 streamgages having their records split in two segments (no record was split into three or more segments), for a total of 282 periods of record (Table 1). These streamgages are indicated by an additional row in Marti et al. (2023, Table DR-5) with “R” or “U” added to the end of the “Streamgage identifier,” where “R” indicates a regulated period and “U” an urbanized period. For example, refer to the South Fork Sangamon River near Rochester, Illinois Streamgage identifiers 05576000 and 05576000R in Marti et al. (2023, Table DR-5), where WYs 1950–64 were determined to be without substantial influence of urbanization or regulation, but WYs 1965–2017 were determined to be regulated. The urbanized or regulated segments at these streamgages were used only for reporting at-site quantiles and not for RRE development. An additional 35 streamgages were determined to have a single, at least 10-year period of record, with substantial influence of urbanization or regulation, or both, and only minor changes in that influence (Table 1). These streamgages are only suitable for reporting of at-site peak-flow quantiles, not regression development and are indicated in Marti et al. (2023, Table DR-5) with an entry of “Not valid for RREs” in the “Usage” column. This resulted in 235 streamgages outside region 2 (Table 1) remaining for potential use in regression analysis and are indicated in Marti et al. (2023, Table DR-5) by an entry of “Valid for use in RREs” in the “Usage” column.

As indicated, the selection of streamgages and the classification of their periods of record depends on the estimation of the effects of both urbanization and regulation. Descriptions of the estimation of these quantities are provided next.

Urbanization

The effect of urbanization was estimated using time series of basin-mean historical urbanization, as in Over et al. (2016) and Over et al. (2021). For the period from 1940–2001, these time series were computed using data from Theobald (2005), which can be found at https://doi.org/10.3133/sir20165049. These data comprise observed decadal housing density data from 1940–2000 based on the 2000 U.S. Census for 10 categories of housing density (including undeveloped) plus year 2000 nonhousing urbanized (commercial, industrial, and transportation) areas and housing density projections for 2010–2030 obtained using a spatially explicit land-use

change model (Theobald, 2005). The nonhousing urbanized area estimates, which are constant year 2000 values, for decades other than 2000 were adjusted (generally, going back in time before the year 2000 they were reduced, whereas going forward in time after the year 2000, they were increased), in proportion to changes in housing density for each basin according to the technique presented in Over et al. (2016), Appendix 2, with a minor modification. This modification was to the results of the fit of the initial estimate of the fraction of nonhousing urbanized $f_{C/I/T}(t)$ to the fraction of the densest housing density class (class 9, which is defined as those housing units having less than 0.6 acre per unit) $f_9(t)$, which was fit using least-squares regression using the fractions from each basin to a quadratic relation as $f_{C/I/T\_est}(t) = 0.139805 - 0.55922*(f_9(t) - 0.5)^2$, which has slightly different coefficients than those used in Over et al. (2016) because the basins used to fit the relation were different. With this adjustment, following Over et al. (2016) and Over et al. (2021) for each basin, the decadal fraction values of Theobald (2005) housing density classes 7–10, which encompass areas of housing densities of no more than 10 acres per unit (classes 7–9) plus the adjusted nonhousing urbanization, were linearly interpolated to annual values from 1940 to 2001, and these fractions constitute the basin-mean historical urbanization for that period through 2000.

To quantify urbanization from 2001 to 2017, rather than use the projected Theobald (2005)-based fractions due to issues about the accuracy of the projections, an urbanization fraction based on the developed land fractions (the sum of classes 22, 23, and 24, which comprise low-, medium-, and high-intensity developed land, respectively) for the four available years (2001, 2006, 2011, and 2016), with unique developed land values from the 2016 National Land Cover Database (NLCD) (Dewitz, 2019; Yang et al., 2018) were spliced onto the Theobald (2005)-based fractions for 1940–2001. To implement this splicing operation, the following steps were carried out. First, for the year 2001, a relation was fit between the 2001 Theobald (2005)-based urbanization fraction for each basin, $U_{Theo2001}$, and the NLCD-based fraction, $U_{NLCD2001}$, using nonlinear regression: $U_{Theo2001} = 1.11532(U_{NLCD2001})^{0.642} - 0.07292$. This relation was used to guide the conversion of changes in U_{NLCD} from 2001 to 2016 to an estimate of an equivalent Theobald (2005)-based value U_{Theo} such that increases in U_{NLCD} were converted to equivalent increases in U_{Theo}. These estimates of the Theobald (2005)-based urbanization fractions for the four years of 2016 NLCD data were then linearly interpolated to annual values from 2001 through 2016. Because NLCD data beyond 2016 were not available at the time of the analysis, 2017 urbanization fractions were assumed to be identical to the 2016 fractions. Splicing these urbanization fraction values extended from 2001 to 2017 using NLCD to the adjusted 1940–2001 Theobald (2005)-based fractions provided an annual time series from 1940–2017 of urbanization fractions for each streamgage.

Regulation

Basin regulation was analyzed using accumulated dam storage datasets from or associated with Wieczorek et al. (2018, 2021). Wieczorek et al. (2018) provide decadal information on the accumulated dam storage upstream from the flowlines of the conterminous United States river network from 1930 to 2013 using data published by the U.S. Army Corps of Engineers in the National Inventory of Dams (NID) (U.S. Army Corps of Engineers, 2018) and the National Hydrography Dataset (NHD) dataset NHDPlus Version 2 (U.S. Geological Survey, 2021). Wieczorek et al. (2021) provide two metrics that estimate the effect of reservoir storage on flowlines in the NHDPlus Version 2 from 1800 to 2018. Decadal storage values for total upstream storage from 1800 to 2018 that were used in the

computation of the published metrics were provided by Mike Wieczorek of the USGS (personal correspondence, March 23, 2021).

Streamgages in region 2 that were included in Over et al. (2021) were evaluated for changes in regulation post-2009 using the storage values provided by Mike Wieczorek that were used in the Wieczorek et al. (2021) dataset. If streamgages had substantial change in regulation post-2009, the peak-flow record beginning with the year of the regulation change was dropped from the record to be added. Streamgages in region 2 that were not included in Over et al. (2021) and streamgages outside of region 2 were evaluated for regulation across their entire periods of record using the Wieczorek et al. (2018) dataset. Streamgages with a storage depth (dam storage capacity divided by basin drainage area) of 1.5 inches or more were considered as substantially regulated, and basins were also visually analyzed using historical topographic maps to identify the presence of impoundments not captured by the NID. If the unregulated period of record for a given streamgage was at least 10 years in length, that period was used for regional regression analysis. If the regulated period of record was at least 10 years in length, the regulation over that period of record was assessed for stationarity. If the change in storage depth was less than 15% over the regulated period, the station was included in at-site flood-frequency analysis but was excluded from regional regression development.

Statewide Dataset Properties

The properties of the region 2 and non-region 2 datasets are illustrated in Figures 4 and 5, respectively. As described in Over et al. (2021), there was a major effort to operate streamgages on small urban basins in the 1960s and 1970s, which is reflected in the large number of peak flows during those years (Figure 4-A), in the drainage area distribution discharges centered on a value of about 10 square miles (Figure 4-B), and in the wide range of urban fraction values (Figure 4-C). The peak flows that are eligible for use in the RREs begin in 1940 with the beginning of the Theobald (2005) urbanization data, as discussed above. The few streamgages that were not eligible for use in the RREs are shown here for reference and tended to have larger basins, longer records, and smaller urban fractions (Figure 4-A through Figure 4-C).

Like region 2, the peak-flow data outside of region 2 also has a maximum during the 1960s and 1970s (Figure 5-A), reflecting a statewide program to install crest-stage gages on smaller basins throughout the state during that period (Knapp and Markus, 2003). The resulting distribution of drainage areas (Figure 5-B) is wider than that in region 2 and bimodal, with a maximum near 1 square mile and another near 100 square miles. The distribution of urban fractions is concentrated on small values as expected, except for a few streamgages that were eligible only for at-site analysis (Figure 5-C).

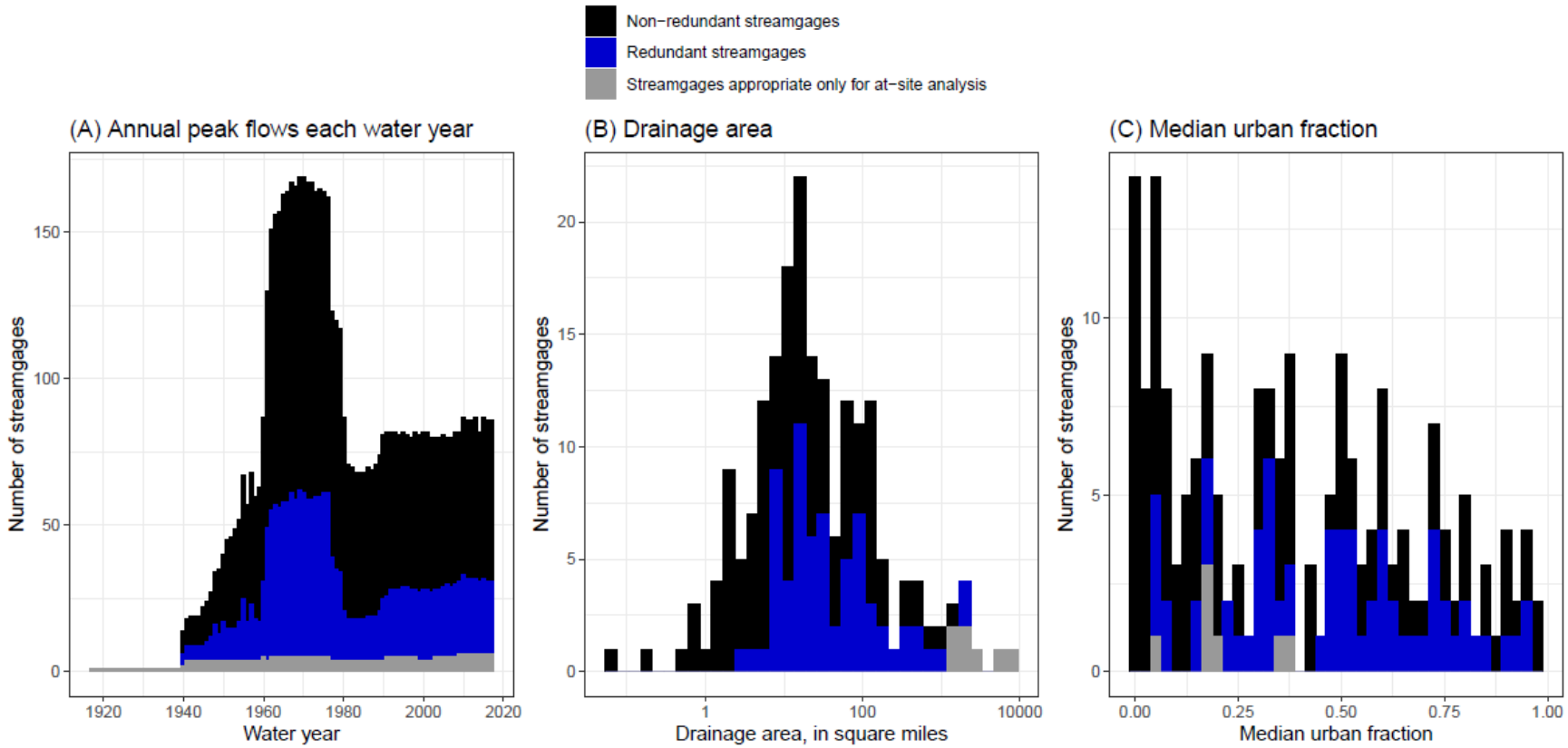

Figure 4. Chart. Histograms of properties of peak-flow data and streamgages used in this study in region 2: (A) annual peak-flows each water year, defined as the period from October 1 to September 30, designated by the year in which it ends; (B) drainage areas of streamgages, and (C) median urban fractions of streamgages.

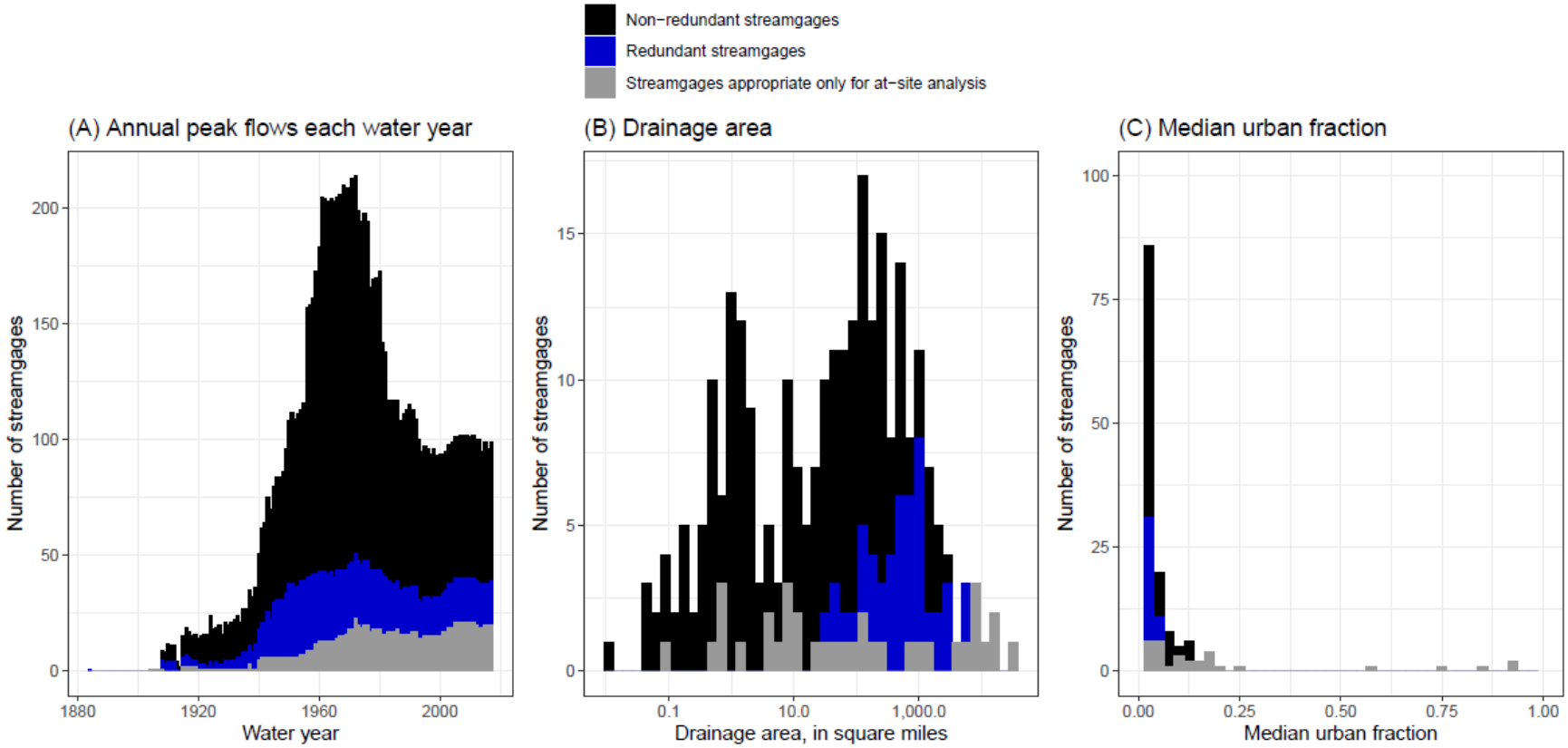

Figure 5. Chart. Histograms of properties of peak-flow data and streamgages used in this study outside region 2: (A) annual peak-flows each water year, defined as the period from October 1 to September 30, designated by the year in which it ends; (B) drainage areas of streamgages; and (C) median urban fractions of streamgages.

COMPUTATION OF OBSERVED PEAK-FLOW QUANTILES

The classification and selection process of streamgages was followed by the computation of peak-flow quantiles at selected AEPs for each selected streamgage. The peak-flow quantiles in this study were computed using the U.S. Geological Survey (USGS) software package *PeakFQ version 7.4* (Veilleux et al., 2014; https://water.usgs.gov/software/PeakFQ/, accessed on June 28, 2022), which, following the guidelines of Bulletin 17C (England Jr. et al., 2019), implements the expected moments algorithm (EMA) to fit the log-Pearson type 3 (LP3) distribution to the peak-flow data of each of the selected streamgages. The application of PeakFQ for flood-frequency analysis is a multi-step process involving the determination of flow intervals and perception thresholds, the application of tests for potentially influential low floods, weighting of at-site and regional skew, and the application of the EMA. The selection of streamgages and the classification of their periods of record along with the steps of the flood-frequency analysis and its results are described in this section. The complete inputs to and outputs from PeakFQ version 7.4 are available in the data release by O'Shea (2023).

Flow Intervals and Perception Thresholds for the Expected Moments Algorithm

The expected moments algorithm (EMA) (Cohn et al., 1997, 2001), as implemented in PeakFQ version 7.4, was used in this study to obtain estimates of the at-site peak-flow quantiles for 464 study streamgages (Table 1; O'Shea, 2023, Table DR-1) by fitting the observations to the LP3 distribution. The EMA generalizes the method-of-moments procedure of Bulletin 17B from the U.S. Interagency Council on Water Data (1982) and uses annual peak-flow data including historic and systematically recorded data, threshold-exceedance information, and regional skew information (England Jr. et al., 2019). EMA also adjusts for the presence of potentially influential low floods (PILFs), missing values from an incomplete period of record, and zero flood years when fitting the LP3 distribution (England Jr. et al., 2019; Stedinger & Griffis, 2008).

The EMA can account for various forms of uncertainty through the accommodation of interval estimates of peak flow, censored estimates of peak flow, and multiple thresholds of observation. For every year in the historic period, which is the continuous time period between and including the first and last observation and thus includes missing periods or gaps in the systematic record, a flow interval defined by a lower and upper bound must be provided. For most peak flows in the systematic period, the lower and upper bounds equal the observed peak flow. For WYs when no information is available, the lower and upper bounds are zero and infinity, respectively. Some peak flows at streamgages are flagged as maximum daily means and can differ substantially from the maximum instantaneous peak flow. Because maximum daily peaks are biased low, a bias correction using a method based on Chen et al. (2017) was used to define a flow interval for these peak flows.

Along with flow intervals, EMA also requires a defined lower and upper perception threshold for each year in the historic period. Perception thresholds identify the range of flows that would have been measured had they occurred. Systematic periods of record have perception thresholds defined as zero to infinity, because all flows are assumed to have been measured during these periods. At some streamgages, such as crest-stage gages (CSGs), flows can be determined only when the water-surface elevation in a stream, also known as a stage, reaches a certain minimum recordable elevation. In these cases, the lower perception threshold was set at the flow associated with that minimum

elevation, as defined by a stage-discharge rating curve. These flows are referred to as a minimum recordable discharge.

Perception thresholds were also set for ungaged or missing periods within the historic period. For periods with no information, the lower and upper perception thresholds were both set to infinity, which indicates that regardless of the flows that occurred for that period, no information about that peak would have been recorded. For certain periods at certain streamgages, the lower threshold was set to the smallest flow that the researchers believed would have been recorded even though the streamgage was not in operation; the upper threshold was still set to infinity in this case. The value of the smallest flow threshold was taken from prior historic flood peak flows. Flow intervals associated with these historic perception thresholds had a lower bound equal to zero and an upper bound equal to the lower perception threshold. This implies that the annual peak flow for that WY was somewhere between zero and the historic peak-flow value.

Historic peak-flow records are therefore an important type of peak, as they can inform multiple years of flow for a streamgage and extend the period of record back from the formal start of peak-flow collection to the earliest historic peak flow available. However, an important assumption about historic peak flows is that they capture the largest peak flow that occurred over an extended period of time, sometimes many years. A historic peak flow that was collected for a reason other than it being the largest peak flow of the WY is referred to as an opportunistic peak. The use of an opportunistic peak to inform a single year or multiple years in the analysis at a streamgage can bias the quantile estimates to be lower. Historic peak-flow records at streamgages included in this study were evaluated on a case-by-case basis to ensure they represent substantial peak flows. If a historic peak flow was smaller than most recorded peak flows at the streamgage, it was classified as opportunistic and was excluded from the at-site analysis.

For some peak flows, only the maximum annual peak stage was recorded with no associated peak flow. For these peak flows, the associated peak flow was calculated according to the following system:

1. If a stage-discharge rating was available for the period of the peak flow, it was used to estimate the associated peak flow. A plus and minus 10% flow interval around the estimate peak flow was used.

2. If a stage-discharge rating was not available for the period of the peak, maximum annual peak stage values associated with other annual peaks in the dataset that bracket the stage-only peak were identified. A log-linear interpolation (log of the peak, linear in the stage) was then used to derive an estimated flow for the stage-only peak. Finally, a plus and minus 10% flow interval around the estimated flow was used.

3. If an appropriate set of annual peaks that bracket the stage-only peak was not available, then the peak flow associated with the next highest annual peak stage was used to set an interval extending from that next peak-flow value to infinity.

For the first two options, the lower perception threshold for the WY associated with the peak stage was set using the lower bound of the estimated flow interval. The upper bound of the perception threshold was set to infinity. For missing years of record between the peak stage WY and the next (in time) peak value in the record, the mean value of the flow interval was used for the lower bound of the perception threshold rather than the lower bound of the flow interval. This indicates greater uncertainty in the possible range of flows that occurred in these additional years. Unlike the first two options, the associated perception threshold for option 3 matches the flow interval exactly. In other words, for this option, the lower bound of the perception threshold is set using the discharge associated with the next highest annual peak stage and the upper perception threshold was set to infinity. Because a mean value of the flow interval could not be calculated for this method due to the upper bound of the flow interval being set to infinity, subsequent missing years of record used the same perception threshold bounds (O'Shea, 2023, PeakFQ_Data.zip).

Tests for Potentially Influential Low Floods

In a general data analysis context, potentially influential points are data points that significantly depart from the trend of the remaining data. Flood-frequency analyses are most typically focused on large floods—that is, floods with small AEPs, which are near the upper end of the peak-flow distribution. Sometimes peak-flow records contain low-magnitude peaks that depart significantly from the trend of the higher-flow values. These low-magnitude peaks, referred to as potentially influential low floods (PILFs), can have a large influence on the fit of the flood-frequency curve in the upper end of the peak-flow distribution. The Multiple Grubbs-Beck test (MGBT; Cohn et al., 2013) was used to detect and censor PILFs as part of this flood-frequency analysis as recommended in Bulletin 17C (England Jr. et al., 2019). Peak flows that were detected as PILFs were recoded as flow intervals extending from zero to the MGBT PILF threshold. All perception thresholds with a lower bound less than the MGBT PILF threshold were changed to match the MGBT PILF threshold. All MGBT PILF thresholds at streamgages included in this study are reported in O'Shea (2023, Table DR-2).

Regional Skew

The EMA relies on estimates of the mean, variance, and skewness of the logarithms of the annual maximum peak flow data at each streamgage to fit the LP3 distribution (Cohn et al., 1997, 2001). Of these moments, the skewness, because it involves raising the data values to the third power, is the most sensitive to extreme values and thus has a large uncertainty when estimated from the peak flows at a single site. To reduce the uncertainty associated with skew, it is recommended that the at-site skew be combined with a regional skew estimate by weighting on the inverse proportions of their respective uncertainties to obtain a weighted skew estimate for a given watershed (England Jr. et al., 2019). However, if the absolute difference of the at-site skew and regional skew estimates at a streamgage is greater than 0.5, a careful examination of the data and flood-producing characteristics of the watershed should be made (England Jr. et al., 2019). A large difference in the station and regional skew raises the issue that the regional skew may not be representative of the streamgage, specifically that it may not correctly inform the flood-producing characteristics of that specific watershed (England Jr. et al., 2019).

For periods of record deemed to be inappropriate for use in regional regression (for example, during regulated or urbanized periods of record at streamgages outside of region 2), at-site skews were not

weighted with regional skew—that is, the unadjusted at-site skew was used in the estimation of the at-site peak-flow quantiles. The at-site peak-flow quantiles for all other periods of record were developed using a weighted skew. The source of an appropriate regional skew and how station and regional skew differences greater than 0.5 were handled is described next.

Regional skew estimates for region 2 were developed through Bayesian generalized least-squares (GLS) analysis as part of Over et al. (2021) as described in appendix 1 of that report. The regional skew γ_R was determined to increase with urbanization as measured by the fraction of 2011 NLCD developed land classes 22, 23, and 24, denoted here as *NLCD*_22_23_24, as $\gamma_R = -0.39 + 0.19NLCD\_22\_23\_24^{1/2}$. The mean squared error of the regional skew for region 2 was 0.19 (the average variance of prediction of the selected regional skew model from Over et al., 2021, Appendix 1). For use in estimating regional skew in region 2 in this study, values of *NLCD*_22_23_24 for each basin were updated to 2016 values from 2016 NLCD data. In doing so, it was discovered that *NLCD*_22_23_24 values using 2016 NLCD for the study basins were usually somewhat smaller than the values obtained using 2011 NLCD, which would make the regional skew values smaller too. To adjust for this effect, a no-intercept linear regression was fit between the 2016 and 2011 versions of the NLCD, which produced the relation $NLCD\_22\_23\_24_{2011} = 1.035067NLCD\_22\_23\_24_{2016}$. Using this relation, the 2016 *NLCD*_22_23_24 values were scaled up by a factor of 1.035067 before applying the regional skew relation.

Outside of region 2, regional skew estimates from Soong et al. (2004, Figure 4) were used. These regional skew estimates, as with the rest of the analyses in Soong et al. (2004), use peak-flow data through WY 1999 and were developed using kriging, as described in Soong et al. (2004, Appendix 4). The mean squared error of the regional skew outside of region 2 was 0.14 (Soong et al., 2004).

Absolute differences between station and regional skew greater than 0.5 were identified at 65 streamgages within region 2 and at 63 streamgages in the other regions. An additional review of the at-site data did not identify any common characteristics of the streamgages with differences greater than 0.5 that could explain why the regional skew may not be appropriate at these locations. No patterns were identified in drainage area, period of record, record length, or basin characteristics that differed substantially from the larger dataset. The weighted skew method described in England Jr. et al. (2019, equations 7–10), hereafter referred to as Bulletin 17C weighted skew, uses the mean squared error of the station and regional skews and an iterative process to estimate the three moments (mean, variance, and skew) of the EMA until the parameter estimates converge (England Jr. et al., 2019). Although the Bulletin 17C weighted skew method can converge after a single iteration, additional iterations always tend toward a solution closer to the regional skew. However, when the absolute difference in the station and regional skew estimates is greater than 0.5, it is considered more reasonable to give greater weight to the at-site skew than the Bulletin 17C weighted skew method provides (England Jr. et al., 2019). An alternative skew method that gives greater weight to the at-site skew was explored at the streamgages where station and regional skew differed by greater than 0.5.

The alternative method, hereafter referred to as the weighted independent estimate (WIE) skew, used a simple inverse uncertainty (mean-squared error) weighted mean following equation 7–20 of

England Jr. et al. (2019). The WIE skew method differs from the Bulletin 17C weighted skew method, equation 7–10 of England Jr. et al. (2019), in that there is no iterative process. In the instances where the Bulletin 17C weighted skew converges after the first iteration, the Bulletin 17C weighted skew and WIE skew are approximately equal. The WIE skew estimate of a streamgage is bounded by the at-site skew and the Bulletin 17C weighted skew, and, for streamgages where the Bulletin 17C weighted skew does not converge after a single iteration, the WIE skew gives more weight to the at-site skew than the Bulletin 17C weighted skew does.

The mean and median of the quantile ratios using WIE skew divided by those using the Bulletin 17C weighted skew for the 65 streamgages in region 2 with absolute differences in station and regional skew greater than 0.5 are shown in Figure 6-A. The median and mean quantile ratios are plotted on this figure for the eight selected AEPs from 0.5 to 0.002. Of the 65 streamgages with differences greater than 0.5, the WIE skew and Bulletin 17C weighted skew had the same weighted skew at 32 streamgages. The median quantile ratio is equal to 1 for each of the eight AEPs. The difference is very small in the mean quantile ratio of these two skew methods starting at AEP 0.04 that gets gradually larger for the smaller AEPs. This indicates that the WIE skew will provide smaller quantile estimates at the smallest AEPs than the Bulletin 17C weighted skew. However, even for the smallest AEP (0.002) the mean quantile ratio differs from 1.0 by less than 1%.

The quantile ratios of WIE skew divided by the Bulletin 17C weighted skew for the 63 streamgages outside of region 2 with absolute differences in station and regional skew greater than 0.5 is shown in Figure 6-B. Of these streamgages, 22 had the same weighted skew as calculated by the WIE skew and Bulletin 17C weighted skew. The median quantile ratios for all 63 of these streamgages is equal to 1 for each of the eight AEPs. Similar to the differences observed for the region 2 streamgages, the mean quantile ratio gets smaller for the WIE skew method beginning at the 0.04 AEP and gets gradually larger for the smaller AEPs. For AEP 0.002, the quantile with the largest difference, the mean quantile ratio differs from 1.0 by just over 2%. This finding matches what was observed in the region 2 streamgages with station and regional skew differences greater than 0.5 in that the WIE skew method will usually provide smaller quantile estimates at the smallest AEPs than the Bulletin 17C weighted skew.

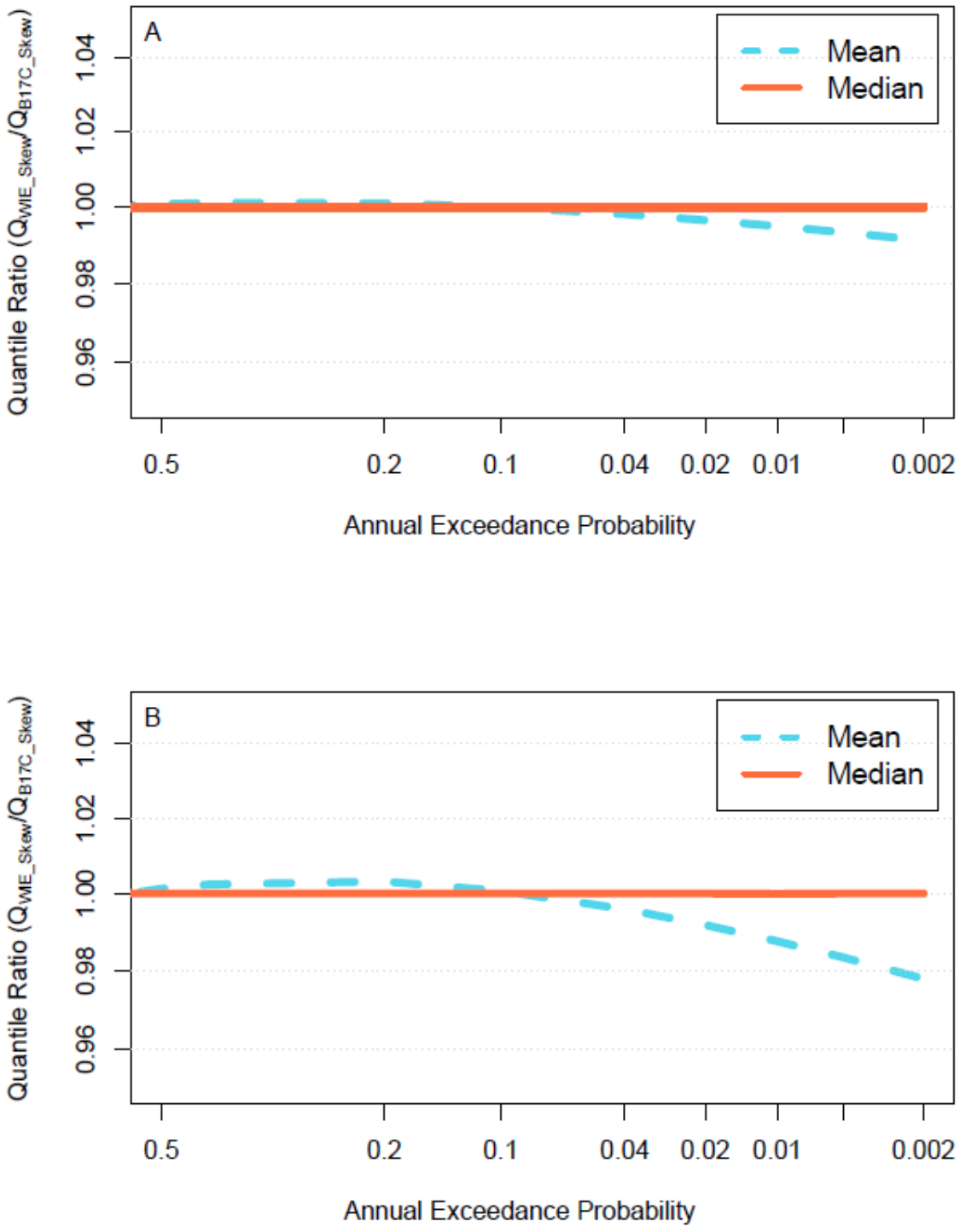

Figure 6. Graph. Statistics of ratios of quantiles determined using weighted independent estimate skew (Q_{WIE_Skew}) to those using Bulletin 17C weighted skew (Q_{B17C_Skew}): (A) streamgages in region 2; (B) streamgages outside of region 2.

In cases where the WIE skew and Bulletin 17C weighted skew are equal, the AEP quantile estimates will also be equal. However, the confidence intervals associated with the AEP quantile estimates are not equal between these two skew methods. The WIE skew method uses only the mean squared error of the station and regional skews to estimate a weighted skew. The Bulletin 17C weighted skew, as calculated by EMA, incorporates additional information including the number of annual peak flows, the number of perception threshold values used in the at-site analysis, and the upper and lower perception threshold values. The result is that the Bulletin 17C weighted skew results in a smaller variance at most quantiles, and therefore smaller confidence intervals. An example of the alternative confidence intervals based on these two methods is shown in Figure 7 for USGS streamgage 05436900 Otter Creek Tributary near Durand, Illinois. The station and regional skew differed by greater than 0.5 at this streamgage. The confidence intervals, particularly for the upper 95% confidence limit, are noticeably larger with the WIE skew (panel B) for AEPs of 0.05 (5%) and smaller. At the 0.002 (0.2%) AEP, the upper confidence limit of the WIE skew is almost double the upper confidence limit of the Bulletin 17C weighted skew. In addition to larger confidence intervals, the larger variance associated with the WIE skew will result in streamgages using this skew method getting less weight in the development of the RREs. Therefore, in cases where the WIE skew and Bulletin 17C weighted skew are approximately equal, the Bulletin 17C weighted skew was considered preferable and was selected.

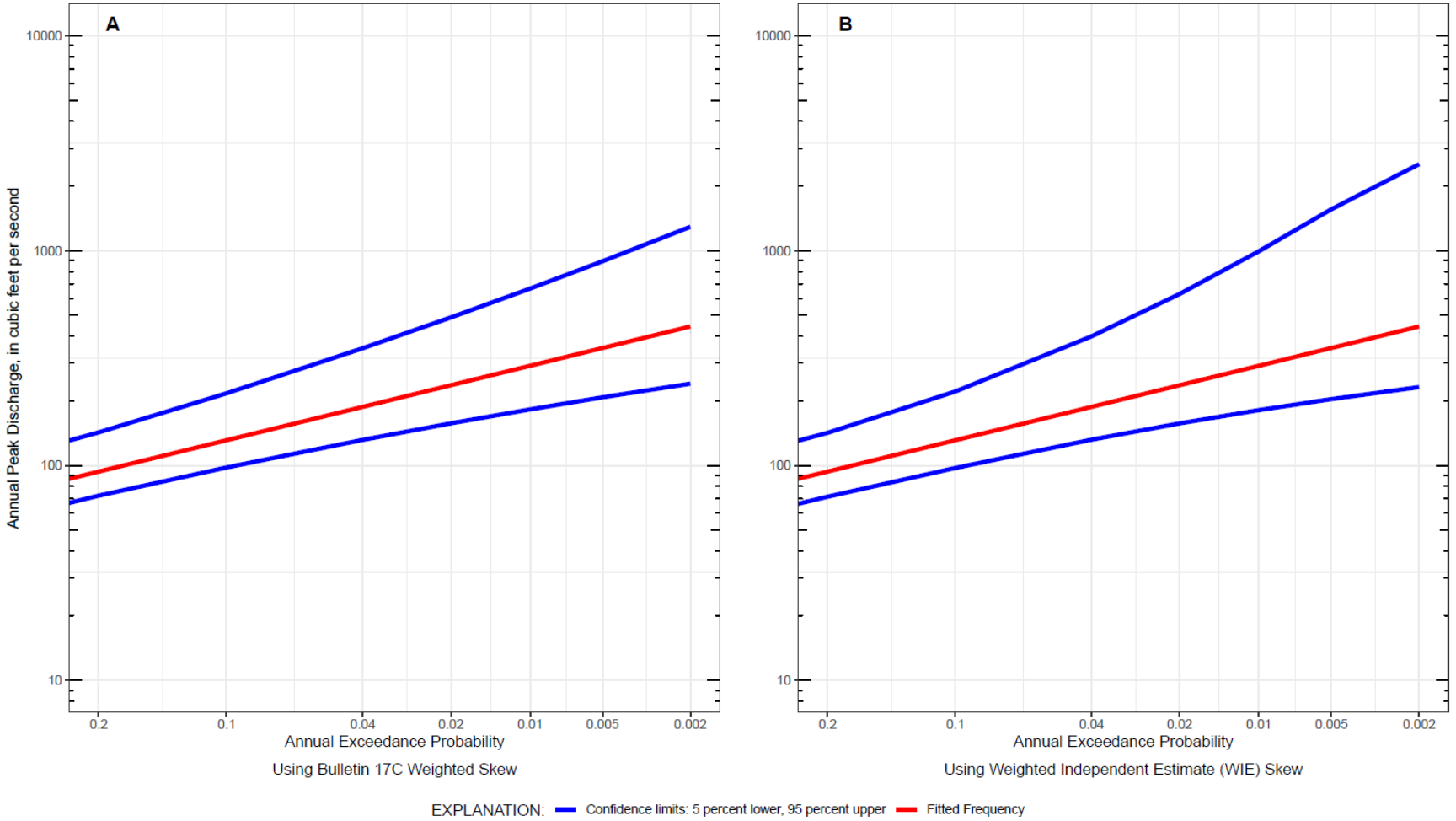

Figure 7. Graph. Flood-frequency curves from PeakFQ software for U.S. Geological Survey streamgage 05436900 Otter Creek Tributary near Durand, Illinois: (A) using Bulletin 17C weighted skew; (B) using weighted independent estimate (WIE) skew.

The skew methods used for the estimation of at-site peak-flow quantiles at USGS streamgages in this study can be summarized as follows. For periods of record deemed to be inappropriate for use in regional regression, as described in the "Site Selection and Record Classification" section, unadjusted at-site skew was used. For other periods of record, when station and regional skew differed by greater than 0.5 and the Bulletin 17C weighted skew and WIE skew were not approximately equal, the WIE skew was used in the estimation of the at-site peak-flow quantiles. There were 33 streamgages in region 2 and 41 streamgages outside of region 2 that used the WIE skew method. The peak-flow quantiles for all other periods of record deemed suitable for use in regional regression, including those with a difference between station and regional skew greater than 0.5 but with WIE skew approximately equal to Bulletin 17C weighted skew, used the Bulletin 17C weighted skew in the estimation of the at-site peak-flow quantiles. The type of skew used for each period of record at each streamgage in this study is given in Marti et al. (2023, Table DR-5, Regional skew weighting column) and in O'Shea (2023, Table DR-2).

Results of the Observed Peak-Flow Quantile Computations

The input data in O'Shea (2023, PeakFQ_Data.zip) include all flow intervals and perception thresholds for the period of record at every streamgage, the selected skew methods and associated regional skews, and the identification of a PILF threshold test for each streamgage included in the study. The output data in O'Shea (2023, PeakFQ_Data.zip) include the resulting at-site peak-flow quantiles and their uncertainties (variances) for the eight selected AEPs from 0.5 to 0.002 (O'Shea, 2023, Tables DR-3 and DR-4, respectively). The at-site peak-flow quantiles and other PeakFQ outputs (primarily the fitted LP3 moments) are used in the regression analyses to develop the RREs described in Chapter 3, and the at-site peak-flow quantile variances are used in weighting the at-site and regional regression quantile estimates at streamgages to obtain a more accurate and less uncertain weighted estimate.

Because of the importance of the uncertainties of the at-site peak-flow quantiles and the value of knowing how these uncertainties compare to the prediction uncertainties of the RREs, statistics of their variances are provided in Table 2. The first two rows of the table present the mean and median of the variances of all 476 periods of record for which at-site peak-flow quantiles were determined (O'Shea, 2023, Table DR-4). The next two rows present these variances in percent units. The variances increase steadily from AEP = 0.5 to AEP = 0.002, as expected, ranging in the mean from about 12.5% to 36% and in the median from about 10% to about 30%. These variances are substantially less than those associated with predictions from the RREs.

Table 2. Variance Statistics of Observed Peak-Flow Quantiles by Annual Exceedance Probability (AEP) from O'Shea, 2023

Statistic	Units	AEP=0.5	AEP=0.2	AEP=0.1	AEP=0.04	AEP=0.02	AEP=0.01	AEP=0.005	AEP=0.002
Mean of Variances	$(\log_{10}(ft^3/s)^2$	0.002958	0.003485	0.004670	0.007149	0.009744	0.012999	0.016923	0.023171
Median of Variances	$(\log_{10}(ft^3/s)^2$	0.0019	0.0023	0.0033	0.0052	0.0069	0.0092	0.012	0.0162
Mean of Variances	Percent	12.57	13.66	15.83	19.65	23.03	26.71	30.64	36.15
Median of Variances	Percent	10.06	11.08	13.29	16.72	19.30	22.36	25.63	29.95

BASIN CHARACTERISTICS

Quantitative values of hydrologically relevant basin characteristics computed from geospatial data available throughout the study region are needed to develop RREs, which express statistically determined relations between peak-flow quantiles and basin characteristics. Because such basin characteristics can be computed for any basin in the study area, the RREs developed using them can be used to compute predictions of the peak-flow quantiles at ungaged locations.

The computation of basin characteristics began with the delineation of the boundaries of the gaged basins to be used in the RRE development. For this study, candidate basin delineations for basins with streamgages in Illinois were determined using StreamStats (https://streamstats.usgs.gov/) basin delineation tools applied to the updated high-resolution StreamStats geographic information system base-data layers (Schafer & Sharpe, 2023) that will be implemented in the public version of StreamStats as part of this study. Basins with streamgages in Indiana and Wisconsin were delineated using the public StreamStats applications for those states. As a check, the drainage areas of the basins delineated using StreamStats were compared to the drainage areas from the USGS National Water Information System database (U.S. Geological Survey, 2018), which in Illinois are based on hand-drawn boundary delineations created when the associated streamgage was established. For basins with differences larger than about 10%, the hand-drawn boundary delineation was used. These basins are usually small urban basins with records that ended around 1980, for which it is difficult to determine basin boundaries from recent data because of urban development and the related potential for later modifications to the hydrography and topography.

Using the delineated basins, candidate basin characteristics were then computed from geospatial data. For hydrologic region 2, basin characteristic selection was not performed as part of this study. RREs were developed for region 2 by Over et al. (2021) and additional peak-flow data from 2011–2017 were available at less than half of streamgages included in that study. Therefore, the basin characteristics used by Over et al. (2021) were retained for use in this study. However, the basin characteristics were recomputed because the basins were redelineated because of the updated geographic information system base-data layers. Outside of region 2, a new basin characteristic selection process was initiated. A list of potential basin characteristics was developed based on previously used basin characteristics for hydrologic regression equation development in Illinois, especially in Soong et al. (2004) and Over et al. (2014, 2021). The available precipitation frequency characteristic data for Wisconsin are not consistent with that for Illinois and Indiana, so that characteristic was not considered for hydrologic region 1, which includes parts of Wisconsin. Basin characteristics from the following categories were considered for streamgages outside of region 2: morphometric, geologic, land-use/land-cover, soils, and climatic characteristics (Table 3).

Morphometric basin characteristics were derived from the basin delineation or the digital elevation model (DEM) of the study area (Schafer & Sharpe, 2023). Geologic basin characteristics were computed as in Over et al. (2014) based on data from Soller et al. (2012). Land-use/land-cover basin characteristics were computed using the NLCD (version 2016) values from year 2016 (Dewitz, 2019). Soils' basin characteristics were computed using soils data from the National Resources Conservation Service State Soil Geographic Data Base (STATSGO) derived by USGS (Schwarz & Alexander, 1995) and

STATSGO soils data derived by Wolock (1997). Climatic characteristics were computed using data from the Parameter-Elevation Regressions on Independent Slope Model (PRISM) Climate Group (PRISM Climate Group, 2014) and the National Oceanic and Atmospheric Administration's (NOAA) Atlas 14 precipitation frequency estimates (Bonnin et al., 2006; National Oceanic and Atmospheric Administration, 2006). The geographic information system layer for snow percent of precipitation characteristic was provided by James Falcone of the USGS (personal communication, May 16, 2019); the data in this layer were used for this characteristic in the Geospatial Attributes of Gages for Evaluating Streamflow, version II, dataset (Falcone, 2011).

Table 3. Example Basin Characteristics Considered for Streamgages Outside of Region 2

Data Type	Source	URL	Example Basin Characteristics
Morphometric	StreamStats, digital elevation model (DEM)	https://streamstats.usgs.gov/ss/	Basin drainage area, DEM-based basin slope, mean basin elevation
Geology	Soller et al. (2012)	https://pubs.usgs.gov/ds/656/	Quaternary sediment thickness
Land use	National Land Cover Database (NLCD) 2016	https://www.mrlc.gov/data/nlcd-2016-land-cover-conus	Fraction open water, fraction forest
Soils	Schwarz & Alexander (1995), Wolock (1997)	https://water.usgs.gov/GIS/metadata/usgswrd/XML/ussoils.xml; https://water.usgs.gov/GIS/metadata/usgswrd/XML/muid.xml	Permeability, soil slope, texture permeability index
Climate	PRISM Climate Group (2014); NOAA Atlas 14 (National Oceanic and Atmospheric Administration, 2006)	https://prism.oregonstate.edu/normals/; https://hdsc.nws.noaa.gov/hdsc/pfds/pfds_gis.html; https://hdsc.nws.noaa.gov/hdsc/pfds/meta/na14_vol2_orb_grid_metadata.xml	Seasonal precipitation, precipitation frequency

CHAPTER 3: DEVELOPMENT OF REGIONAL REGRESSION EQUATIONS

The peak-flow quantiles at qualifying streamgages and the basin characteristics characterizing the associated drainage basins (computed as described in Chapter 2) provide the dependent and independent variables, respectively, for the development of RREs that allow the estimation of peak-flow quantiles at ungaged locations. This chapter describes the methods and results of developing these RREs. It begins with the definitions of the regions, describes the testing for and removal of streamgages that are redundant due to their proximity to other streamgages on the same rivers or streams, the transformation of the basin characteristics to help ensure homoscedastic and Gaussian residuals, the search for candidate regional regression models using ordinary least-squares regression, the selection of final models using generalized least-squares regression, and the accuracy of the selected models. For the equations found in the tables and figures of this report, instances of $\log_{10}()$ are used to represent a common logarithm or a log base 10, $\log_2()$ represents a log of base 2, and *ln*() represents the natural logarithm or a log with base e.

DETERMINATION OF HYDROLOGIC REGIONS

The seven hydrologic regions for Illinois from Soong et al. (2004) were used for initial RRE development. Combinations and other adjustments of the regions were considered, but ultimately the regions from Soong et al. (2004), with region 2 expanded to include the basins in Illinois that drain into Lake Michigan as in Over et al. (2021), are used for the final results of this study (Figure 1).

TESTING FOR REDUNDANCY

Regression models in this analysis were fit using a generalized least-squares analysis that allows for cross-correlation among observed peak-flow quantiles. Cross-correlation is introduced through factors such as the same weather affecting proximal streamgages and resulting in the annual maximum peak flows at both streamgages. However, the GLS approach does not account for correlation that is caused by two nested basins with considerable overlap in drainage area (Veilleux, 2009). In cases where this overlap is substantial enough, one of the nested basins is considered redundant and not incorporated into regional regression modeling (Veilleux, 2009). An analysis to evaluate redundancy was implemented to establish a non-redundant subset of the streamgages appropriate for regional regression analysis. The redundancy analysis was carried out separately for streamgages inside and outside of region 2.

A given streamgage basin may be nested within one or more other gaged basins. An algorithm for how to remove streamgages for optimal results was developed by Over et al. (2021). The algorithm operates by assigning a score to each basin according to its record length and drainage-basin properties relative to other basins. The score of a basin is higher the more drainage-area overlap it has with other basins, the shorter its record length, and more common its drainage area is compared to that of other basins in the collection of basins being considered. The basin with the highest (worst) score is removed, the scores for the remaining basins are recomputed, and the basin with the highest

remaining score is removed until no remaining overlap fractions between basins are larger than 0.215. The overlap fraction, DAR_{ij}, for two nested basins *i* and *j* is defined as the smaller of the two possible drainage-area ratios,

$$DAR_{ij} = \min\left[\frac{DA_i}{DA_j}, \frac{DA_j}{DA_i}\right],$$

Figure 8. Equation. Definition of drainage area overlap fraction.

where DA_i and DA_j are the drainage areas of basins *i* and *j*, respectively. The R script (R Core Team, 2022) used by Over et al. (2021) was adapted to implement this algorithm in this study.

In region 2, 187 streamgages were considered for use in regression development, and 68 of those were removed due to redundancy, leaving 119 non-redundant streamgages (Table 1). Outside of region 2, 235 streamgages were considered for use in regression development, and 43 of those were removed due to redundancy, leaving 192 non-redundant streamgages (Table 1). The larger fraction of redundant streamgages in region 2 is likely due to higher density of streamgages in that region. A total of 311 non-redundant streamgages were used for the development of RREs. The non-redundant streamgages that were ultimately used in the development of regression equations and redundant streamgages are distinguished in the study area maps (Figure 1), in the histograms displaying basic streamgage properties (Figure 4 and Figure 5), and in the table listing the basic properties of the streamgage records and their basins (Marti et al., 2023, Table DR-5).

TRANSFORMATION OF BASIN CHARACTERISTICS

Transformations are often applied to both the dependent and independent variables in linear regression analysis, to linearize their relations, make the variance of the residuals constant, and make the distribution of the residuals Gaussian (Helsel et al., 2020, section 9.2). In this study, the dependent variables, the LP3-fitted peak-flow quantiles, are always $\log_{10}$-transformed because they are positive and positively skewed. The transformation applied to the independent variables, the basin characteristics, was varied depending on their properties. A set of transformations, including those such as $\log_{10}$ and square root that reduce the skew and squaring that increases skew, were tested on each characteristic to determine the transformation that reduced skewness as near to zero as reasonably possible. For example, drainage area, which is positive and positively skewed (like the peak-flow quantiles), was determined to have its skewness reduced nearest to zero by use of a $\log_{10}$-transformation, the same as the peak-flow quantiles. In addition, variables with relatively narrow distributions far from zero were centered—that is, the mean values were subtracted—and values of variables whose maximum values were numerically large (greater than 1,000) were divided by their maximum values. These operations reduce the likelihood of numerical issues in the regression analysis. The transformations of the basin characteristics were selected and applied separately for basin characteristic values inside region 2 and outside of region 2. Based on the decision to use the basin characteristics for region 2 that were previously selected in Over et al. (2021), the same transformations used in that study were reviewed and selected for this analysis.

ORDINARY LEAST-SQUARES REGRESSION

Ordinary least-squares (OLS) linear-regression techniques were applied to search for the best combination of basin characteristics for use as independent variables in regression equations for regions outside of region 2. Combinations of basin characteristics were investigated for each region and at each of the eight AEPs, and in doing so, drainage area was always included by assumption. This analysis was accomplished using several R scripts (R Core Team, 2022), beginning with the use of the function *regsubsets* from the *leaps* package (Lumley, Thomas, based on Fortran code by Alan Miller, 2020) to search among all combinations of basin characteristics. The best combinations returned from *regsubsets* were further investigated to test for the statistical significance of the coefficients (the t-ratio—that is, the ratio of the absolute value of the coefficient to its standard error—was constrained to be at least 2) and for variance inflation factor (VIF) values to be at most 5. The maximum number of basin characteristics used in a given model was determined by the number of streamgages in the region, with at least 10 streamgages required per explanatory variable (Farmer et al., 2019). Basin characteristics for which 10% of streamgages had identical values or where the range of values was smaller than the median divided by 3 were not considered for a given region. Examination of regression results included the evaluation of various goodness-of-fit measures and summary plots. The adjusted coefficient of determination (R^2; Helsel et al., 2020) and Akaike information criterion (AIC; Akaike, 1974) were evaluated for individual AEPs for a given model and the mean values of adjusted R^2 and AIC across all AEPs were compared between models. The models whose corresponding model fits have the largest magnitude of adjusted R^2 and smallest magnitude of AIC, along with the sign of the coefficient being physically reasonable and smaller VIFs prioritized, were selected as candidate variables for development of the final spatial regression equations.

GENERALIZED LEAST-SQUARES REGRESSION

Final regression equations were determined using GLS regression. GLS regression accounts for correlations among the values of the dependent data and for variations in the uncertainty of those data by use of a covariance matrix. Specialized methods have been developed to implement GLS regression, including the estimation of the covariance matrix among the dependent variable values, in the context of peak-flow quantiles fitted using the method of moments to the LP3 distribution (Eng et al., 2009; Farmer et al., 2019; Griffis & Stedinger, 2007, 2009; Stedinger & Tasker, 1985; Tasker & Stedinger, 1989).These methods have been implemented into the function *WREG.GLS* from the R package *WREG* (Farmer, 2021), and R scripts (R Core Team, 2022) were written and used to implement *WREG.GLS* for this study.

The inputs for the *WREG.GLS* function comprise the peak-flow quantiles (O'Shea, 2023, Table DR-4), which were $\log_{10}$-transformed; basin characteristics for each gaged basin in each region (Marti et al., 2023, Table DR-5), which were transformed as discussed previously; statistics describing the results of the fitting of the LP3 distribution by PeakFQ (the standard deviate, standard deviation, at-site skew, and regional skew); and data needed to estimate the cross-correlations of the peak-flow quantiles, which comprise a matrix of concurrent streamgage record lengths and a table of streamgage locations (latitude and longitude). The latter two datasets were extracted using an R script from the .PRT and .EXP files generated by PeakFQ and published in O'Shea (2023).

The cross-correlations, ρ_{ij}, among streamgage records were estimated by selecting streamgage record pairs with a minimum number of concurrent years and fitting a function of the form

$$\rho_{ij} = \theta^{d_{ij}/(\alpha d_{ij}+1)}$$

Figure 9. Equation. Functional form of cross-correlation between concurrent streamgage records.

where

- d_{ij} is the distance between streamgages *i* and *j* in miles.
- α and θ are dimensionless parameters, where α is nonnegative and $0 < \theta < 1$.

For the specified ranges of α and θ, the correlation ρ decreases with increasing distance *d* and lies between 0 and 1. Typically good fits α and θ are found for α near zero (but still positive) and θ slightly less than 1 (Eng et al., 2009). The parameters α and θ are selected by examining plots of $\rho_{ij}(d_{ij})$ for different values of α and θ and considering a goodness-of-fit statistic. The Nash-Sutcliffe efficiency goodness-of-fit statistic was considered for this analysis. For this study the minimum number of concurrent years of record was taken as 45, and α and θ were taken as 0.040 and 0.920 for region 2 and as 0.006 and 0.975 outside of region 2, respectively.

The 'regSkew' input parameter for the *WREG.GLS* function was set to "TRUE," which implements an adjustment to the regression for the effect of the uncertainty of regional skews and requires a 'MSEGR' input value. The 'MSEGR' value used is the mean squared error of the regional skew for stations with weighted skew. The value used for region 2 was 0.19 (the average variance of prediction of the selected region skew model from Over et al. 2021, Appendix 1); for outside of region 2 the value used was 0.14 (the mean squared error of the selected regional skew method from Soong et al. 2004, which was also used in this study). The parameters of *WREG.GLS* not described here were set to their default values.

The candidate RRE models (collections of basin characteristics) indicated by the OLS regression analyses for each region were analyzed using *WREG.GLS*. For region 2, only the basin characteristics used in Over et al. (2021) were considered. The WREG results were evaluated by comparing the following properties:

- Goodness-of-fit statistics: These statistics measure how well the predictions from the fitted model match the fitting data—that is, the observed LP3-fitted quantiles. The statistics considered were the pseudo R^2, root mean square error (RMSE), average variance of prediction, and average standard error of prediction (refer to Table 8 for definitions). These statistics were considered for individual AEPs for each model, and the mean values across all AEPs were compared between models.
- The significance and physical meaning of the regression coefficients: The ratio of the absolute value of the coefficient to its standard error was required to be at least 2, which is the approximate cutoff for statistical significance at a *p*-value of 0.05 (Helsel et al.,

2020). Further, the sign of the coefficient should be physically meaningful: for example, peak flows should increase with drainage area and slope but decrease with soil permeability. Requiring physical meaningfulness helps prevent spurious models.

- The maximum VIF among the basin characteristics: Models with maximum VIFs up to 5 were considered acceptable but smaller VIFs were preferred (the minimum is 1).

- Regression diagnostic plots: These comprise various residual plots, including residuals versus fitted values, residuals versus drainage area, and absolute residuals versus fitted values. These plots help to determine the presence of trends in the residuals and indicate whether the residuals have a normal distribution. Also considered were plots of leverage and influence values, which test how far a data point's explanatory values are from those of others in the model and how much effect a data point has on the regression, respectively (Eng et al., 2009).

- Maps of residuals and of leverage and influence values: Such maps show whether spatial trends in these quantities exist.

GLS models were developed for regions individually and for multiple combinations of regions with and without the inclusion of a dummy variable for the region. Such regional dummy variables were called "region factors" in Soong et al. (2004).

SELECTED MODELS

Although various combinations of regions were considered, and upon review of the maps of residuals and leverage and influence values, it was decided to retain the regions from Soong et al. (2004) and to develop separate models for each region. The selected model equations for each region and AEP are presented in Table 4, and the explanations of the selected basin characteristics are provided in Table 5. The regional regression equations shown in Table 4 include two forms of the same equation for each region. The first equation displays the $\log_{10}$-transformed discharge equation and is the form in which the equations were developed. The second equation is the real-space regression equation obtained by exponentiating with base 10. For all AEPs in each region, the same basin characteristics were used, but the coefficients ($a_{0,p}$, $a_{1,p}$, $a_{2,p}$, $a_{3,p}$, $a_{4,p}$) will vary by AEP (p). An example of the coefficients for a single AEP (0.01) for all regions is provided in Table 6 and their corresponding standard errors and t-ratios are presented in Table 7. The coefficients and standard errors for all AEPs are included in Marti et al. (2023, Table DR-7).

Table 4. Selected Regional Regression Models for Each Region [Q_p, peak-flow quantile at annual exceedance probability p]

Region	Model
1	$log_{10}(Q_p) = a_{0,p} + a_{1,p} log_{10}(DA) + a_{2,p}(log_{10}(TPI) - 3.2259)$ $Q_p = 10^{(a_{0,p} - 3.2259 a_{2,p})} (DA)^{a_{1,p}} (TPI)^{a_{2,p}}$
2	$log_{10}(Q_p) = a_{0,p} + a_{1,p} log_{10}(DA) + a_{2,p}\sqrt{NLCD\_22\_23\_24} + a_{3,p}\sqrt{DrainageClass1a} + a_{4,p} log_{10}(DEM\_1\_0\_P)$ $Q_p = 10^{a_{0,p}} (DA)^{a_{1,p}} 10^{a_{2,p}\sqrt{NLCD\_22\_23\_24}} 10^{a_{3,p}\sqrt{DrainageClass\ 1a}} (DEM\_1\_0\_P)^{a_{4,p}}$
3	$log_{10}(Q_p) = a_{0,p} + a_{1,p} log_{10}(DA) + a_{2,p}(DEM\_Slope - 55.329) + a_{3,p}((TPI - 1823.4)/4758.3)$ $Q_p = 10^{a_{0,p}} (DA)^{a_{1,p}} 10^{a_{2,p}(DEM\_Slope - 55.329)} 10^{a_{3,p}((TPI-1823.4)/4758.3)}$
4	$log_{10}(Q_p) = a_{0,p} + a_{1,p} log_{10}(DA) + a_{2,p} log_{10}(Soil\_Slope)$ $Q_p = 10^{a_{0,p}} (DA)^{a_{1,p}} (Soil\_Slope)^{a_{2,p}}$
5	$log_{10}(Q_p) = a_{0,p} + a_{1,p} log_{10}(DA) + a_{2,p}(DEM\_Slope - 55.329)$ $Q_p = 10^{a_{0,p}} (DA)^{a_{1,p}} 10^{a_{2,p}(DEM\_Slope - 55.329)}$
6	$log_{10}(Q_p) = a_{0,p} + a_{1,p} log_{10}(DA)$ $Q_p = 10^{a_{0,p}} (DA)^{a_{1,p}}$
7	$log_{10}(Q_p) = a_{0,p} + a_{1,p} log_{10}(DA)$ $Q_p = 10^{a_{0,p}} (DA)^{a_{1,p}}$

Note: Definitions of basin characteristics are provided in Table 5.

Table 5. Information on Basin Characteristics for Selected Regional Regression Models

Name	StreamStats Name	Definition	Units	Regions	Reference for Data
DA	DRNAREA	Basin drainage area	square miles	All	Schafer & Sharpe (2023)
TPI	TPISTATSGO	100*percent_sand + 10*percent_silt + percent_clay using soil texture fractions from STATSGO	weighted percent	1, 3	Wolock (1997)
DEM_Slope	DEMSLX100	Mean of slopes in degrees of 10-meter digital elevation model (DEM) pixels with 100 times vertical exaggeration in each delineated basin, computed using Slope tool in ArcMap.	Degrees	3, 5	Schafer & Sharpe (2023)
Soil_Slope	STATSSLPWT	Area-weighted mean of soil slopes of dominant component of STATSGO soil units.	Percent	4	Wolock (1997)
NLCD_22_23_24	FLC16DVLMH	Fraction NLCD 2016 classes 22, 23, and 24 (low, medium, and high intensity developed)	decimal fraction	2	Dewitz (2019); Yang et al. (2018)
DrainageClass1a	FSSURGDC78	Sum of SSURGO fractions “very poorly drained” and “unknown (likely water)”	decimal fraction	2	Soil Survey Staff (n.d.)
DEM_1_0_P	RELRELF	Basin elevation range divided by basin perimeter	feet per mile	2	Schafer & Sharpe (2023)

Table 6. Coefficients of Selected Regional Regression Models for Each Region at an Annual Exceedance Probability of 0.01

Region	Annual exceedance probability	Number of streamgages used in regional regression development	Intercept $a_{0,p}$	Basin characteristic A1	Coefficient a_1	Basin characteristic A2	Coefficient $a_{2,p}$	Basin characteristic A3	Coefficient $a_{3,p}$	Basin characteristic A4	Coefficient $a_{4,p}$
1	0.01	35	2.755	DA	0.5691	TPI	−1.320	NA	NA	NA	NA
2	0.01	119	2.010	DA	0.7595	NLCD_22_23_24	0.2167	DrainageClass1a	−1.028	DEM_1_0_P	0.4778
3	0.01	60	2.919	DA	0.5813	DEM_Slope	0.01216	TPI	−0.7065	NA	NA
4	0.01	36	2.573	DA	0.5680	Soil_Slope	0.5199	NA	NA	NA	NA
5	0.01	36	2.969	DA	0.5913	DEM_Slope	0.004526	NA	NA	NA	NA
6	0.01	16	2.898	DA	0.6239	NA	NA	NA	NA	NA	NA
7	0.01	9	3.138	DA	0.5090	NA	NA	NA	NA	NA	NA

Note: NA = not applicable. Definitions of basin characteristics are provided in Table 5; transformations of basin characteristics are provided in Table 4.

Table 7. Standard Errors and t-ratio of Selected Regional Regression Models for Each Region at an Annual Exceedance Probability of 0.01

Region	Annual exceedance probability	Number of streamgages used in regional regression development	SE of intercept	t-ratio of intercept	SE of coefficient $a_{1,p}$	t-ratio of coefficient $a_{1,p}$	SE of coefficient $a_{2,p}$	t-ratio of coefficient $a_{2,p}$	SE of coefficient $a_{3,p}$	t-ratio of coefficient $a_{3,p}$	SE of coefficient $a_{4,p}$	t-ratio of coefficient $a_{4,p}$
1	0.01	35	0.08318	33.12	0.04430	12.85	0.2274	5.803	NA	NA	NA	NA
2	0.01	119	0.1860	10.80	0.05722	13.27	0.09644	2.247	0.1931	5.321	0.1468	3.256
3	0.01	60	0.06822	42.79	0.02495	23.30	0.002680	4.538	0.1098	6.432	NA	NA
4	0.01	36	0.1383	18.60	0.02436	23.31	0.1562	3.329	NA	NA	NA	NA
5	0.01	36	0.05503	53.95	0.02227	26.55	0.002781	1.627	NA	NA	NA	NA
6	0.01	16	0.09073	31.94	0.05204	11.99	NA	NA	NA	NA	NA	NA
7	0.01	9	0.08724	35.97	0.05736	8.875	NA	NA	NA	NA	NA	NA

Note: SEs = standard errors; NA = not applicable

As planned, the basin characteristics used for the region 2 model are those used in Over et al. (2021), except that urban fraction characteristic is updated from using 2011 NLCD data to using that from 2016 NLCD. These characteristics comprise drainage area (DA); urban fraction, computed using the National Land Cover Database (NLCD) year 2016 classes 22 (developed, low intensity), 23 (developed, medium intensity), and 24 (developed, high intensity) (NLCD_22_23_24); DEM_1_0_P, basin elevation range divided by basin perimeter, which is a measure of slope; and DrainageClass1a, a sum of fractions of Soil Survey Geographic (Soil Survey Staff, n.d.) fractions "very poorly drained" and "unknown (likely water)." The coefficients (Table 6; Marti et al., 2023, Table DR-7) are little changed from those in Over et al. (2021, Table 11) which is as expected, given the minor changes in streamgages included, record length, and basin characteristics. The biggest difference is that the urbanization coefficient has become somewhat smaller for AEPs of 0.04 and smaller. Physically, peak flows increase with drainage area, urban fraction, and slope, whereas they decrease with the fraction of poorly drained soils—that is, usually wet, and areas that are likely water. Further, the magnitudes of the coefficients of the slope measure (DEM_1_0_P) and of the fraction of poorly drained and unknown (likely water) areas both increase substantially with decreasing AEP (higher peak flows), indicating that these factors are more important at these higher flows.

Basin characteristics selected for models used in regions outside of region 2 comprise drainage area (DA), slope (DEM_Slope), soil slope (Soil_Slope), and texture permeability index (TPI) (Table 5). Log_{10}-transformed drainage area is used in all regression equations and has a positive value that ranges from about 0.5 to 0.75. Drainage area is the sole basin characteristic selected for regions 6 and 7 because of the small number of non-redundant streamgages available in those regions. Of the selected basin characteristics, drainage area is the basin characteristic with the most explanatory power, as indirectly seen from its large t-ratio (the ratio of the absolute value of the coefficient to its standard error). For regions 1, 2, 3, 4, and 7, the drainage area coefficient generally decreases from a larger value at high AEPs (low flows) to a smaller value at low AEPs (high flow), whereas it is approximately constant with AEP in region 5 and increases slightly in region 6 (Table 6). The drainage area coefficient is highest for region 2 (northeastern Illinois).

In addition to drainage area, TPI appears in both the region 1 and region 3 model, albeit with different transformations (Table 4). TPI, a measure of permeability that is a weighted mean of the fractions of sand, silt, and clay, has negative coefficient values for all AEPs in both region 1 and region 3 (Table 6), indicating an inverse relationship with increasing peak flows, which is as expected because higher permeability implies higher infiltration of surface water. Further, the magnitude of the TPI coefficients in both regions increases with decreasing AEP (higher flows).

The other basin characteristic included in the model for region 3 is DEM_Slope (Table 6), which is also the second basin characteristic in the region 5 model. The slope coefficient magnitude increases with decreasing AEP (larger flows) in both regions, although the increase in region 5 is small. The magnitudes of the slope coefficients in both regions appear to be small, but that is because the values are numerically large, being in degree units and based on DEM data with vertical exaggeration (Table 5) and transformed only by centering; their significance is better indicated by their t-ratios (Table 7).

A different measure of slope, Soil_Slope, which is based on the slopes of the soils making up the basins and in percent units, is used along with drainage area in the region 4 model (Table 6). Its coefficients decrease with decreasing AEP (larger flows), ranging from 0.6043 at the largest AEP (smallest flow) to 0.4878 at the smallest AEP (largest flow).

ACCURACY OF FINAL SPATIAL REGRESSION EQUATIONS

Several measures of the accuracy of the final GLS spatial regression equations are presented in Table 8 (for AEP = 0.01; refer to Marti et al., 2023, Table DR-8 for the complete results) and their definitions in Table 9. All these measures are means for all streamgages in the region. The RMSE, which is the square root of the mean square error (MSE) converted to percent, starts, for most regions, in the 40s, for AEP = 0.5 and increases with decreasing AEP, to the 70s for region 1, 2, and 6, but less for other regions. The Pseudo_R2, which measures the fraction of the variability explained by the model after removing the time sampling error and is expressed in percent, is in the high 80s to the high 90s for all regions except region 2, which has values from the middle 70s to the high 80s, similar to but slightly higher that than the values in Over et al. (2021). The lower values of Pseudo_R2 in region 2 are presumably due to inclusion of urbanized basins in that region. The average variance of prediction (AVP) is the mean over all the streamgages in a region of the prediction variance for the given quantile; it is computed as the sum of the model error variance (MEV) and the time sampling error. The model error and the time sampling error, are, respectively, the error that arises from an imperfect model that does not explain the observations (here, the regression model of the observed flood quantiles using the basin characteristics) and the error that comes from imperfect estimates of the values of the dependent variable (here, the uncertainty of the estimates of the observed flood quantiles arising from a finite sampling time) (Farmer et al., 2019). The GLS methods implemented in WREG were designed to estimate the variances of these two errors separately. The conversion of AVP to percent units yields the standard error of prediction (Sp), which has values similar to the RMSE for larger AEPs but somewhat smaller values at smaller AEPs. The values of Sp rise only to the middle to high 60s for the regions with the highest values (regions 2 and 6). The difference between the RMSE and the Sp is that the latter considers the properties of the GLS regression, in particular its covariance matrix, whereas the RMSE, which is based simply on the sum of the squared errors or residuals, does not. As mentioned, the MEV is one of two terms of the AVP (both of which are positive), and as such it is less than the AVP. For most AEPs in most regions, the MEV is between 80% and 90% of the AVP, and it decreases slightly with decreasing AEP, which reflects the larger time-sampling error of rarer floods. The MEV/AVP fraction is smallest in region 7, reflecting the larger time-sampling error in the region with fewest streamgages, and is largest in region 2, which has the most streamgages. As noted, the mean and median variances of the at-site quantiles (Table 2) are much less than the corresponding AVP and Sp values obtained from the RREs, although the latter do not trend as strongly with AEP as do the uncertainties of the at-site quantiles.

For sake of comparison, RMSE values from Soong et al. (2004) and standard error of prediction (Sp) values from Over et al. (2021) for region 2 and from Soong et al. (2004) for all regions are provided in Table 8 for AEP = 0.01 and in Marti et al. (2023, Table DR-8) for all AEPs. The region 2 Sp values from Over et al. (2021) are similar to the ones from this study, which is as expected, due to the similarity of the data and equations. The Sp values from Soong et al. (2004) are within the range of the values

obtained from this study but are much less variable among the regions. This is in large part because Soong et al. (2004) combined regions 2, 6, and 7 into one set of GLS equations and combined regions 1, 3, and 5 into another; in that sense, only the region 4 values are directly comparable. Because of the lack of direct comparability of Sp values and as part of the analysis done to decide on grouping of regions in this study, RMSE values by region (except in region 2) were computed as part of this study using the equations and data of Soong et al. (2004). By this measure, the present study shows some improvement over results from Soong et al. (2004). This is particularly true in regions 6 and 7, but not for region 5, which had a larger RMSE value in the present study. However, the Sp values for region 5 were relatively small in both studies.

Table 8. Accuracy of Selected Regional Regression Models for Each Region at an Annual Exceedance Probability of 0.01

Region	Annual exceedance probability	RMSE (%)	Pseudo_R 2 (%)	AVP ((log_{10} (ft^3/s))^2)	Sp (%)	MEV ((log_{10} (ft^3/s))^2)	Sm (%)	MEV/AVP (unitless)	Soong et al. (2004) RMSE (%)	Soong et al. (2004) Sp (%)	Over et al. (2021) Sp (%)
1	0.01	62.17	89.90	0.05093	55.68	0.04393	51.21	0.8627	72.54	49.0	NA
2	0.01	63.68	78.24	0.05468	57.99	0.05106	55.76	0.9339	NA	49.2	59.5
3	0.01	51.17	94.57	0.03114	42.37	0.02745	39.58	0.8816	53.70	49.0	NA
4	0.01	38.71	97.04	0.01795	31.60	0.01466	28.43	0.8168	47.29	50.0	NA
5	0.01	42.05	97.31	0.02382	36.69	0.02032	33.72	0.8528	33.50	49.0	NA
6	0.01	66.58	91.79	0.06628	64.89	0.05626	58.95	0.8487	170.3	49.2	NA
7	0.01	46.77	94.07	0.03369	44.22	0.02420	37.00	0.7184	80.49	49.2	NA

Note: Definitions of goodness-of-fit statistics are provided in Table 9.

Table 9. Definitions of Goodness-of-Fit Statistics

Statistic	Abbreviation	Explanation	Units	Report equation number
Mean squared error	MSE	Mean of squared model-observation residuals e_i.	Square of the dependent variable units: $(\log_{10}(ft^3/s))^2$	$MSE = \frac{1}{(n-k-1)}\sum_{i=1}^{n}(e_j)^2$, where $e_j = \hat{y}_j - \hat{y}_{jr}$, where $\hat{y}_j$ is the observed estimate and $\hat{y}_{jr}$ is the regression estimate of the peak-flow quantile
Root mean squared error	RMSE	MSE converted to percent.	Percent	$RSME(\%) = 100\left\{e^{[\ln(10)^2 MSE} - 1\right\}^{1/2}$
Pseudo coefficient of determination	Pseudo_R2	Fraction of error variance (not including sampling error) explained by regression model.	Percent	$\text{Pseudo\_R2} = 1 - \frac{\sigma_\delta^2(k)}{\sigma_\delta^2(0)}$, where $\sigma_\delta^2(k)$ is the MEV for generalized least-squares (GLS) regression model with *k* basin characteristics (BCs) and $\sigma_\delta^2(0)$ is the MEV for a model with 0 BCs.
Average variance of prediction	AVP	Sum of model error variance and sampling error variance.	Square of the dependent variable units: $(\log_{10}(ft^3/s))^2$	$AVP = \sigma_\delta^2 + \frac{1}{n}\sum_{i=1}^{n} x_i (X^{\mathrm{t}} \Lambda^{-1} X)^{-1} x_i^{\mathrm{t}}$
Standard error of prediction	Sp	AVP converted to percent using the same formula as used for RMSE.	Percent	$S_p = 100\left\{e^{[\ln(10)^2 AVP} - 1\right\}^{1/2}$
Model error variance	MEV	Total variance minus sampling error variance.	Square of the dependent variable units: $(\log_{10}(ft^3/s))^2$	There is no equation for MEV as such; it is determined as part of the iterative GLS estimation process performed by WREG.
Standard model error	Sm	MEV converted to percent using the same formula as used for RMSE.	Percent	$S_m = 100\left\{e^{[\ln(10)^2 MEV} - 1\right\}^{1/2}$

Techniques for computing accuracy estimates at individual locations (streamgages or ungaged locations) are also available, including the variance and standard error of prediction and confidence interval values. According to Farmer et al. (2019, eq. 49), the variance of prediction V_i at an individual location based on a GLS regression analysis is,

$$V_i = \gamma^2 + x_i(X^{\mathrm{t}}\Lambda^{-1}X)^{-1}x_i^{\mathrm{t}}$$

Figure 10. Equation. Log-space variance of prediction.

where

- γ^2 is the model error variance (values in Marti et al., 2023, Table DR-8).
- x_i is a row vector specifying the basin characteristics (values in Marti et al., 2023, Table DR-5) of the individual location augmented with a 1.
- X is a ($n \times p$) matrix whose rows are the transformed basin characteristics for each streamgage used in the GLS model augmented by a 1, where n is the number of streamgages used in the spatial regressions for the region (values in Marti et al., 2023, Table DR-7, column C) and p is the number of basin characteristics for the region plus 1.
- Λ^{-1} is the matrix inverse of Λ, the ($n \times n$) covariance matrix used in the GLS regression analysis.
- X^t is the matrix transpose of X.
- x_i^t is the matrix transpose of x_i.

The $(X^{\mathrm{t}}\Lambda^{-1}X)^{-1}$ matrices, which give the covariance matrices for the regression coefficients for the selected AEPs, are given in Marti et al. (2023, Table DR-9, columns E to the end).

The variance of prediction V_i can be converted to a standard error of prediction in log units by taking the square root—that is, $S_i = V_i^{1/2}$, and in percent units by using the following formula (Eng et al., 2009, eq. 33):

$$S_{pi} = 100\{\exp[(ln(10))^2 V_i] - 1\}^{1/2}$$

Figure 11. Equation. Standard error of prediction in percent units.

The confidence intervals of the predicted discharge quantile Q_i are computed as (Farmer et al., 2019, eq. 48)

$$\log_{10} Q_i \pm t_{\alpha/2,n-p} S_i$$

$$= \left[log_{10}Q_i - t_{\alpha/2,n-p}S_i < log_{10}Q_i < log_{10}Q_i + t_{\alpha/2,n-p}S_i\right]$$

Figure 12. Equation. Confidence interval for predicted discharge quantile.

where

- $t_{\alpha/2,n-p}$ is the critical value of the *t* distribution at the alpha level α (for example, $\alpha = 0.05$ for 90% confidence intervals).

- $n - p$ are the degrees of freedom, where n is the number of streamgages used in the spatial regressions for the region (Marti et al., 2023, Table DR-7, column C) and p is the number of basin characteristics for the region plus 1.

After inverting the logarithmic transformation, the interval is

$$\left[Q_i 10^{-t_{\alpha/2,n-p} S_i} < Q_i < Q_i 10^{t_{\alpha/2,n-p} S_i}\right]$$

Figure 13. Equation. Confidence interval for predicted discharge quantile transformed out of log space.

where

- Q_i is the median prediction.

- $Q_i 10^{t_{\alpha/2,n-p} S_i}$ and $Q_i 10^{-t_{\alpha/2,n-p} S_i}$ are the upper and lower 1 - 2α confidence limits, respectively.

For example, for $\alpha = 0.05$ (therefore, 90% confidence limits), according to the data and methods used in this study, the probability that the true value of the discharge quantile exceeds the upper limit $Q_i 10^{t_{\alpha/2,n-p} S_i}$ is $\alpha = 0.05$; likewise the probability that the true value of the discharge quantile is less than the lower limit $Q_i 10^{-t_{\alpha/2,n-p} S_i}$ is $\alpha = 0.05$, so that the probability that the true value lies between the limits is 90%.

CHAPTER 4: APPLICATIONS

Previous chapters have described how estimates of peak-flow quantiles at the study streamgages and RREs for Illinois' seven hydrologic regions were determined and presented the results of doing so. This chapter, which has two sections, describes how to apply the results to estimate quantiles at qualifying ungaged locations. The first section presents applications of the RREs, which are applicable at ungaged locations that satisfy the streamgage selection procedures described in the previous chapters. The second section presents an interpolation method for estimating peak-flow quantiles in certain reaches of certain rivers that did not qualify for use in the RREs because of regulation or drainage area.

APPLICATIONS OF REGIONAL REGRESSION EQUATIONS

At Ungaged Sites in Region 2

For an ungaged location far (such that the drainage area ratio differs by more than a factor of 2 from the nearest gage; refer to the "Near Streamgages" section) from a streamgage in region 2, the method is unchanged from Over et al. (2021) except for the use of updated values of various statistics. To start, the basin characteristic values (Marti et al., 2023, Table DR-5) for the real-space regression equation (Table 4) for region 2 need to be determined. These values should lie within the bounds of the values used to develop the equations (Table 10). The equations are applied using coefficients from Marti et al. (2023, Table DR-7) (refer to Table 6 for AEP = 0.01), which yields the expected median quantile at the selected AEP. To estimate the uncertainty of the median estimate, the log-space variance of prediction is obtained from the equation of Figure 10, with the model error variance coming from Marti et al. (2023, Table DR-9, column D) and the $(X^t \Lambda^{-1} X)^{-1}$ matrices from Marti et al. (2023, Table DR-9, columns E to end). This log-space variance of prediction can be converted to percent using the equation in Figure 11. To obtain a confidence interval, the equation in Figure 12 or Figure 13 is used, which uses the log-space standard error of prediction (S_i) and the t statistic $t_{\alpha/2,n-p}$, where $\alpha = 0.05$ for the 90% confidence interval, n is the number streamgages used in the spatial regressions for region 2 (119 from Marti et al., 2023, Table DR-7, column C), and p is the number of estimated parameters for the region 2 regression equations plus one (5, from the four basin characteristics and the intercept).

Table 10. Ranges of Basin Characteristics Used in Regional Regression Equation Development

Basin Characteristic	Region	Minimum	Maximum
DA	1	0.1773	1325
TPI	1	988	4793
NLCD_22_23_24	2	0.002045	0.9692
DrainageClass1a	2	0	0.2506
DA	2	0.07031	1352
DEM_1_0_P	2	0.8122	35.97
DA	3	0.0378	2618
DEM_Slope	3	22.12	77.92

Basin Characteristic	Region	Minimum	Maximum
TPI	3	988	6582
DA	4	0.0461	1636
Soil_Slope	4	2.04	13.37
DA	5	0.01327	1943
DEM_Slope	5	24.52	79.35
DA	6	0.07158	791.8
DA	7	0.08691	243.7

Note: Definitions of basin characteristics are provided in Table 5.

Consider a hypothetical ungaged basin with a drainage area (*A*) of 50 square miles (mi^2), urbanized land-use fractions (*U*) NLCD_22_23_24 of 20%, water and wetland fraction (*W*) DrainageClass1a of 15%, and slope (*S*) DEM_1_0_P of 3.67 feet per mile. These values usually will be obtained by StreamStats after delineation of the basin. First, note that the basin characteristic values are well within their corresponding ranges (Table 10). If the basin is within region 2, the peak-flow quantiles should be computed by using the region 2 spatial regression equations in Table 4, with coefficient values given in Marti et al. (2023, Table DR-7). Taking the peak-flow quantile with AEP = 0.01 as an example, the quantile is computed as follows:

$$Q_{0.01} = 10^{a_{0,0.01}} A^{a_{1,0.01}} 10^{a_{2,0.01}\sqrt{U}} 10^{a_{3,0.01}\sqrt{W}} S^{a_{4,0.01}}$$

$$Q_{0.01} = 10^{2.01} 50^{0.7595} 10^{0.2167\sqrt{0.20}} 10^{-1.028\sqrt{0.15}} 3.67^{0.4778}$$

$$Q_{0.01} = 1{,}858 \text{ ft}^3/\text{s}.$$

Figure 14. Equation. Calculating the 1% annual exceedance probability peak-flow quantile at an ungaged location in region 2.

Notice the values of *W* and *U* enter the equation as decimal fractions, whereas *S* enters in feet per mile.

The uncertainty of this estimate is computed by using the variance of prediction equation in Figure 10, where

- y^2 = 0.05106 is the model error variance from Marti et al., 2023, Table DR-8.
- x_i is shown in Figure 15:

$$x_i = \left[1, \quad log_{10}A_i, \quad {U_i}^{1/2}, \quad {W_i}^{1/2}, \quad log_{10}S_i\right]$$

$$x_i = [1, \quad log_{10}50, \quad 0.20^{1/2}, \quad 0.15^{1/2}, \quad log_{10}3.67]$$

$$x_i = [1, \quad 1.699, \quad 0.4472, \quad 0.3873, \quad 0.5647]$$

Figure 15. Equation. Row vector (x_i) of basin characteristics at an ungaged location in region 2.

- The matrix $(X^t \Lambda^{-1} X)^{-1}$ from Marti et al. [2023], Table DR-9, columns E to the end) is shown in the equation of Figure 16:

$$(X^t \Lambda^{-1} X)^{-1} = 0.001 \begin{bmatrix} 34.60 & -9.180 & -7.487 & -6.675 & -24.39 \\ -9.180 & 3.275 & 1.267 & -1.495 & 6.707 \\ -7.487 & 1.267 & 9.301 & -1.245 & 2.409 \\ -6.675 & -1.495 & -1.245 & 37.30 & 1.023 \\ -24.39 & 6.707 & 2.409 & 1.023 & 21.54 \end{bmatrix}$$

Figure 16. Equation. Covariance matrix $(X^t \Lambda^{-1} X)^{-1}$ at an ungaged location in region 2.

With these values, the second term of the equation in Figure 10 is computed as:

$$x_i (X^t \Lambda^{-1} X)^{-1} (x_i)^t$$

$$= 0.001[1, 1.699, 0.4472, 0.3873, 0.5647] \begin{bmatrix} 34.60 & -9.180 & -7.487 & -6.675 & -24.39 \\ -9.180 & 3.275 & 1.267 & -1.495 & 6.707 \\ -7.487 & 1.267 & 9.301 & -1.245 & 2.409 \\ -6.675 & -1.495 & -1.245 & 37.30 & 1.023 \\ -24.39 & 6.707 & 2.409 & 1.023 & 21.54 \end{bmatrix} \begin{bmatrix} 1 \\ 1.699 \\ 0.4472 \\ 0.3873 \\ 0.5647 \end{bmatrix}$$

Figure 17. Equation. Calculating $(x_i(X^t \Lambda^{-1} X)^{-1}(x_i)^t)$ at an ungaged location in region 2.

Therefore, the variance of prediction V_i = 0.05106 + 0.001831 = 0.05289.

Given the V_i value, the standard error of prediction in log units is $S_i = V_i^{1/2}$ = 0.23, and from the equation in Figure 11, the standard error of prediction in percent is

$$S_{pi} = 100\{\exp[(ln(10))^2 V_i] - 1\}^{1/2}$$

$$S_{pi} = 100\{\exp[(ln(10))^2 0.05289] - 1\}^{1/2}$$

$$S_{pi} = 56.9\%$$

Figure 18. Equation. Calculating the standard error of prediction in percent at an ungaged location in region 2.

The equation for the confidence interval of a predicted quantile is given in the equation in Figure 12. The $t_{\alpha/2,n-p}$ is the critical value of the t distribution at the alpha level α and $n–p$ degrees of freedom, where n = 119 (Marti et al., 2023, Table DR-7, column C) is the number of streamgages used in the spatial regression for region 2, and p = 5 is the number of basin characteristics plus 1. Here, for the 90% confidence intervals $t_{\alpha/2,n-p} = t_{0.95,114}$ = 1.658, so the confidence interval in log units is:

$$\log_{10} Q_i \pm t_{\alpha/2,n-p} S_i$$

$$= \log_{10}(1858) \pm 1.658 * 0.23$$

$$= [2.888,\ 3.650]$$

Figure 19. Equation. Calculating the confidence intervals at an ungaged location in region 2.

The equation in Figure 13 provides the confidence intervals for a predicted discharge transformed out of log space:

$$[Q_i 10^{-t_{\alpha/2,n-p} S_i}, Q_i 10^{t_{\alpha/2,n-p} S_i}] = 10^{[2.888,\ 3.650]}$$

$$= [773,\ 4{,}467]\ \text{ft}^3/\text{s}$$

Figure 20. Equation. Transforming the confidence intervals out of log space at an ungaged location in region 2.

Summing up, the estimated 1% AEP peak-flow quantile $Q_{0.01}$ for this hypothetical ungaged basin on a stream in region 2 far from a streamgage is 1,860 ft^3/s (rounded to three significant figures) with a standard error of prediction of 56.9% and a 90% confidence interval of [773, 4,470] ft^3/s (rounded to three significant figures).

At Ungaged Sites Outside of Region 2

For an ungaged location far (such that the drainage area ratio differs by more than a factor of 2 from the nearest gage; refer to the "Near Streamgages" section) from a streamgage for regions other than region 2, the approach is the same as that described in example 1 for a region 2 location, with one major additional element: after the applicable equation from Table 4 is applied using coefficients from Marti et al. (2023, Table DR-7) and the variance of prediction is determined using the equation in Figure 10, these are adjusted for urbanization as in Over et al. (2021, eq. 21, 22, and 23).

$$Q_p(U) = Q_p(U_0) 10^{(b_U)_p (U - U_0)}$$

Figure 21. Equation. Urbanization-adjusted peak-flow quantile at an ungaged location outside of region 2.

$$\log_{10} Q_p(U) = \log_{10} Q_p(U_0) + (b_U)_p (U - U_0)$$

Figure 22. Equation. Log-transformed urbanization-adjusted peak-flow quantile at an ungaged location outside of region 2.

where

- $Q_p(U)$ is the urbanization-adjusted peak-flow quantile for urban fraction U.

- $Q_p(U_0)$ is the peak-flow quantile with AEP p computed from the applicable equation from Table 4.

- U is the present or hypothetical future urban fraction.

- U_0 is the baseline urban fraction for the region of interest—that is, the urban fraction that is already accounted for in $Q_p(U_0)$ because of the urban fraction in the basins used to develop the RREs used to compute $Q_p(U_0)$—computed as the median of the median values at each streamgage for each region (Table 12).

- $(b_u)_p$ is the temporal urbanization coefficient for AEP p from Table 11.

The variance of prediction V_u of the urbanization-adjusted peak-flow quantile in log units, $\log_{10}Q_p(U)$, can be derived from the equation in Figure 22 by computing the variance of the sum of the two terms on the right-hand side assuming they are independent as

$$V_U = V_{U_0} + V_{b_U}\left(U - U_0\right)^2$$

Figure 23. Equation. Variance of prediction of the urbanization-adjusted peak-flow quantile at an ungaged location outside of region 2.

where

- V_{U_0} is the variance of prediction of the unadjusted peak flow in log units, $\log_{10}Q_p(U_0)$.

- V_{b_U} is the variance of the urbanization coefficient $(b_u)_p$, which is the square of the standard error given in Table 11.

Table 11. Urban Fraction Coefficients from Temporal Quantile Regression Analysis

Annual Exceedance Probability	0.5	0.2	0.1	0.04	0.02	0.01	0.005	0.002
Coefficient $(b_u)_p$	0.487	0.405	0.391	0.359	0.331	0.312	0.301	0.294
Standard error of $(b_u)_p$	0.075	0.080	0.078	0.085	0.086	0.134	0.112	0.115

Source: Over et al. (2021)

Table 12. Baseline Urban Fractions

Region	Number of Streamgages	Median Urban Fraction
1	35	0.005952
2	119	0.3029
3	60	0.01037
4	36	0.01033
5	36	0.008841
6	16	0.002571
7	9	0.0

The standard error of prediction of $\log_{10}Q_p(U)$, S_u, can be computed from V_u as $S_U = V_U^{1/2}$, and then S_u can be used in the equations from Figures 11–13 to obtain a standard error in percent and confidence intervals.

As an example, consider a hypothetical ungaged basin located in hydrologic region 4 with a drainage area (*A*) of 100 square miles, mean soil slope (*Soil_Slope*) of 6.0%, and 2010 Theobald (2005) urban fraction (*U*) of 20%. Using Table 10 it can be confirmed that the basin characteristic values are within their corresponding range. The peak-flow quantiles should be computed by using the region 4 spatial regression equations in Table 4, with coefficient values given in Marti et al. (2023, Table DR-7). Taking the peak-flow quantile with AEP = 0.01 as an example, the quantile is computed as follows:

$$Q_{0.01}(U_0) = 10^{a_{0,0.01}} A^{a_{1,0.01}} Soil\_Slope^{a_{2,0.01}}$$

$$Q_{0.01}(U_0) = 10^{2.573} 100^{0.5680} 6.0^{0.5199}$$

$$Q_{0.01}(U_0) = 12{,}989 \text{ ft}^3/\text{s}$$

Figure 24. Equation. Calculating the unadjusted 1% Annual Exceedance Probability peak-flow quantile at an ungaged location in region 4.

Now that the unadjusted peak-flow quantile estimate, $Q_{0.01}(U_0)$, has been estimated, the equation in Figure 21 can be used to calculate the urbanization-adjusted peak-flow quantile at the ungaged location as:

$$Q_{0.01}(U) = Q_{0.01}(U_0) 10^{(b_U)_{0.01}(U-U_0)}$$

$$Q_{0.01}(U) = 12989\ 10^{0.312(0.20-0.01033)}$$

$$Q_{0.01}(U) = 14{,}885 \text{ ft}^3/\text{s}$$

Figure 25. Equation. Urbanization-adjusted peak-flow quantile at an ungaged location in region 4.

The log-transformed urbanization-adjusted peak-flow quantile estimate, $\log_{10}Q_p(U)$, will be needed to calculate the variance of prediction of the urbanization-adjusted peak-flow quantile. This quantile can be estimated now using the equation from Figure 23:

$$\log_{10} Q_{0.01}(U) = \log_{10} Q_{0.01}(U_0) + (b_U)_{0.01}(U - U_0)$$

$$\log_{10} Q_{0.01}(U) = \log_{10} 12989 + 0.312(0.20 - 0.01033)$$

$$\log_{10} Q_{0.01}(U) = 4.17275$$

Figure 26. Equation. Log-transformed urbanization-adjusted peak-flow quantile at an ungaged location in region 4.

To estimate the confidence interval associated with the urbanization-adjusted peak-flow quantile estimate, the variance of prediction of the urbanization-adjusted peak-flow quantile, V_u, needs to be estimated. The estimation of V_u is given in the equation in Figure 23 and requires that V_{U_0}, which is equivalent to V_i from the equation in Figure 10 be calculated first. Using the variance of prediction equation in Figure 10, V_{U_0} for the ungaged location in region 4 is calculated as follows:

where

- y^2 = 0.01466 is the model error variance from Marti et al. (2023, Table DR-8).
- x_{U_0} (row vector) is determined using Figure 27:

$$x_{U_0} = [1,\ log_{10}A_{U_0},\ log_{10}Soil\_Slope_{U_0}]$$

$$x_{U_0} = [1,\ log_{10}100,\ log_{10}6.0]$$

$$x_{U_0} = [1,\ 2,\ 0.7782]$$

Figure 27. Equation. Row vector (x_{U_0}) of basin characteristics at an ungaged location in region 4.

- The matrix $(X^t \Lambda^{-1} X)^{-1}$ from Marti et al. (2023, Table DR-9, columns E to the end) is shown in the equation of Figure 28:

$$(X^t \Lambda^{-1} X)^{-1} = 0.001 \begin{bmatrix} 19.13 & -1.604 & -20.03 \\ -1.604 & 0.5936 & 0.9224 \\ -20.03 & 0.9224 & 24.38 \end{bmatrix}$$

Figure 28. Equation. Covariance matrix $(X^t \Lambda^{-1} X)^{-1}$ at an ungaged location in region 4.

With these values, the second term of the equation in Figure 10 is computed as:

$$x_{U_0}(X^t \Lambda^{-1} X)^{-1}\left(x_{U_0}\right)^t = 0.001[1,\ 2,\ 0.7782] \begin{bmatrix} 19.13 & -1.604 & -20.03 \\ -1.604 & 0.5936 & 0.9224 \\ -20.03 & 0.9224 & 24.38 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 0.7782 \end{bmatrix}$$

$$= 0.001549$$

Figure 29. Equation. Computation of $x_{U_0}(X^t \Lambda^{-1} X)^{-1}(x_{U_0})^t$ at an ungaged location in region 4.

Therefore, the unadjusted variance of prediction V_{U_0} = 0.01466 + 0.001549 = 0.01621.

The variance of prediction, V_u, of the urbanization-adjusted peak-flow quantile from the equation in Figure 23 is then calculated as

$$V_U = V_{U0} + V_{bU}(U - U_0)^2$$

$$V_U = 0.01621 + 0.01796(0.20 - 0.01033)^2$$

$$V_U = 0.0169$$

Figure 30. Equation. Variance of prediction (V_u) of the urbanization-adjusted peak-flow quantile at an ungaged location in region 4.

Given the V_u value, the standard error of prediction in log units is $S_u = V_U^{1/2} = 0.130$, and from the equation in Figure 11, the standard error of prediction in percent is

$$S_{pU} = 100\{\exp[(ln(10))^2 V_U] - 1\}^{1/2}$$

$$S_{pU} = 100\{\exp[(ln(10))^2 0.0169] - 1\}^{1/2}$$

$$S_{pU} = 30.6\%$$

Figure 31. Equation. Calculating the standard error of prediction (S_u) in percent at an ungaged location in region 4.

The equation for the confidence interval of the predicted quantile is shown in Figure 12. The quantity $t_{\alpha/2,n-p}$ is the critical value of the t distribution at the alpha level α and $n–p$ degrees of freedom, where $n = 36$ is the number of streamgages used in the spatial regression for region 4 (Marti et al., 2023, Table DR-7, column C), and $p = 3$ is the number of basin characteristics plus 1. Here, for the 90% confidence intervals $t_{\alpha/2,n-p} = t_{0.95,33} = 1.692$, so the confidence intervals in log units are

$$\log_{10} Q_{0.01}(U) \pm t_{\alpha/2,n-p} S_U$$

$$= \log_{10}(14885) \pm 1.692 * 0.130$$

$$= [3.953,\ 4.393]$$

Figure 32. Equation. Calculating the confidence intervals of a peak-flow quantile in log units at an ungaged location in region 4.

The equation in Figure 13 provides the confidence intervals for a predicted discharge transformed out of log space.

$$[Q_{0.01}(U)10^{-t_{\alpha/2,n-p}S_U}, Q_{0.01}(U)10^{t_{\alpha/2,n-p}S_U}] = 10^{[3.953,\ 4.393]}$$

$$= [8{,}974,\ 24{,}720]\ \text{ft}^3/\text{s}$$

Figure 33. Equation. Transforming the confidence intervals out of log space at an ungaged location in region 4.

Summing up this example, the estimated urbanization-adjusted peak-flow for the 1% AEP peak-flow quantile, $Q_{0.01}$, for this hypothetical ungaged basin on a stream in region 4 far from a streamgage is 14,900 ft^3/s (rounded to three significant figures) with a standard error of prediction of 30.6% and a 90% confidence interval of [8,970, 24,700] ft^3/s (rounded to three significant figures).

Finally, if an estimate at an ungaged location, either in region 2 or outside, with a hypothetical future urbanization, such as 100% urbanization, is desired, then the future urbanization value could be used in replacement of *U* in the equations found in Figures 21–23. As long as the future urbanization is expected to have similar hydrologic effects as the basins used in Over et al. (2021), the temporal urbanization coefficients (Table 11) would be applicable. A future increase in urbanization does not mean the urban fraction needs to be completely built out—that is, *U* = 1.0. Any value of *U* greater than the current median urban fraction from Table 12 for a given region could be used to estimate the effects of future increases in urbanization.

At Streamgages

Because of the finite length of streamgage records, the regional equations can improve the accuracy of the peak-flow quantile estimates at streamgages by incorporating regional information. The procedure recommended by Cohn et al. (2012) is to compute this peak-flow quantile from the weighted mean of the regression equation estimate and the result of the at-site flood-frequency analysis of the streamgage record, by using the weighted independent estimates (WIE) method where the weights are the inverses of the variance of each of the discharge estimates (Tasker, 1975). The WIE quantiles are computed with the following equation:

$$\log_{10}\left(Q_p\right)_{g,w} = \frac{\left(V_p\right)_{g,r}\log_{10}\left(Q_p\right)_{g,s} + \left(V_p\right)_{g,s}\log_{10}\left(Q_p\right)_{g,r}}{\left(V_p\right)_{g,r} + \left(V_p\right)_{g,s}}$$

Figure 34. Equation. Weighted independent estimate mean of the regression equation estimate and the at-site flood-frequency analysis.

where

- $(Q_p)_{g,w}$ is the WIE peak-flow quantile at the streamgage (*g*) for an AEP of *p*.
- $(Q_p)_{g,s}$ is the peak-flow quantile for an AEP of *p* computed from the streamgage record as described in the "Computation of Observed Peak-Flow Quantiles" section.
- $(Q_p)_{g,r}$ is the peak-flow quantile for an AEP of *p* obtained from the spatial regression equation in Table 4 for the region where the streamgage is located.
- $(V_p)_{g,r}$ is the variance of prediction of $(Q_p)_{g,r}$ for the AEP of *p* computed by using the equation in Figure 10.
- $(V_p)_{g,s}$ is the variance of prediction of $(Q_p)_{g,s}$ for the selected AEP computed by PeakFQ as part of the at-site frequency analysis.

This method, because it combines at-site and regional estimates, is only valid for streamgages that were deemed eligible for developing the RREs, and it is the same for streamgages in region 2 and outside of region 2. Values of $(Q_p)_{g,s}$, $(Q_p)_{g,r}$, and $(Q_p)_{g,w}$, for both the redundant and nonredundant streamgages in this study are tabulated in Table DR-10 of Marti et al. (2023), in the first, second, and third rows, respectively, for each streamgage.

The equation for the variance of prediction $(V_p)_{g,w}$ for the weighted discharge $(Q_p)_{g,w}$ is given by (Tasker, 1975)

$$(V_p)_{g,w} = (V_p)_{g,s}(V_p)_{g,r} / \left[(V_p)_{g,s} + (V_p)_{g,r}\right]$$

Figure 35. Equation. Variance of prediction for the weighted mean of the regression equation estimate and the at-site flood-frequency analysis.

The values of $(V_p)_{g,s}$, $(V_p)_{g,r}$, and $(V_p)_{g,w}$ for both the redundant and nonredundant streamgages in this study are tabulated in Table DR-11 of Marti et al. (2023), in the first, second, and third rows, respectively, for each streamgage. With this value of $(V_p)_{g,w}$, the standard error of prediction $(S_p)_{g,w}$ can be computed as $(S_p)_{g,w} = (V_p)_{g,w}^{1/2}$, and then the equations in Figures 10–13 can be used to compute the standard error of prediction and the confidence interval in the same way as shown in previous examples.

Near Streamgages

If an ungaged location of interest is near to (as defined in Figure 36) and on the same stream as a streamgage, the accuracy of the peak-flow quantile estimate at the ungaged location can be improved if the estimate from the regional equation is combined with the estimate at the streamgage (Ries III, 2007). A few different methods are available in the literature for this adjustment; the method used here is a modified version of the method presented in Over et al. (2021). According to this method, the near-gage adjustment has an effect only if the ratio A_u/A_g of the drainage area of the ungaged basin of interest A_u to that of a gaged basin A_g is between 0.5 and 2 (see the equation in Figure 36); this constraint on the effect of the adjustment method defines being "near" a streamgage.

First define the adjustment weighting factor w_a, which is given by

$$w_a = \begin{cases} \left|\log_2\left(A_u/A_g\right)\right| \; if \; 0.5 < \left(A_u/A_g\right) < 2 \\ 1 \text{ otherwise} \end{cases}$$

Figure 36. Equation. Near-gage adjustment weighting factor.

where

- A_u is the drainage area at the ungaged location of interest.
- A_g is the drainage area at the streamgage.

The near-gage adjustment equation can then be written as follows:

$$(Q_p)_{u,w} = w_a(Q_p)_{u,r} + (1 - w_a)(Q_p)_{g,w}(A_u / A_g)$$

Figure 37. Equation. Near-gage adjustment equation.

where

- $(Q_p)_{u,w}$ is the gage-adjusted peak-flow quantile estimate for an AEP of p at the ungaged location of interest.

- $(Q_p)_{u,r}$ is the peak-flow quantile estimate for an AEP of p at the ungaged location of interest from the spatial regression equation in Table 4 for the region where the ungaged location is located.

- $(Q_p)_{g,w}$ is the weighted peak-flow quantile estimate for an AEP of p at the streamgage (obtained using equation in Figure 34 and tabulated in Marti et al. (2023, Table DR-10).

- w_a, A_u, and A_g are as defined in the equation for Figure 36.

According to the equations in Figures 36 and 37, the near-gage adjustment has no effect when $A_u/A_g \leq 0.5$ or $A_u/A_g \geq 2$, because $w_a = 1$ and therefore the near-gage adjusted value $(Q_p)_{u,w}$ is identical to the value $(Q_p)_{u,r}$ from the RREs. At a streamgage, $A_u = A_g$ and $w_a = 0$, so the near gage adjusted value reduces to the weighted peak-flow at the gage $(Q_p)_{g,w}$. When $0.5 < (A_u/A_g) < 2$, then both $(Q_p)_{u,r}$ and $(Q_p)_{g,w}$ contribute to the value of $(Q_p)_{u,w}$ according to the values of the weights w_a and $1 - w_a$.

This near-gage adjustment is the same for a streamgage in region 2 as it is for a streamgage located outside of region 2. In both cases, because the adjustment accounts for differences in the basin properties through the drainage area ratio A_u/A_g, other basin properties, such as urbanization and other land-use/land-cover properties, need to be similar for the basin of interest and the nearby gaged basin. Note that the use of the drainage area ratio A_u/A_g, for this adjustment assumes that peak flow varies with the drainage area with an exponent of 1, whereas typically the exponent is somewhat less; for example, in this study for AEP = 0.01, it ranges from about 0.5 to 0.7 (Table 6). The error arising from the use of the unit exponent is small in this application; however, because the range of drainage area ratios where it is applied is limited and because the farther apart the two locations are, the smaller the weight on the peak-flow quantile at the nearby streamgage ($(Q_p)_{g,w}$, Figure 37).

For an example of this application, assume that the hypothetical ungaged location from the "At Ungaged Sites in Region 2" section was located upstream from the USGS 05527950 Mill Creek at Old Mill Creek, Illinois example, from the "At Streamgages" location. The near-gage adjustment described in this section could then be considered. The first step is to check if the drainage ratio of the ungaged location and the streamgage is between 0.5 and 2. USGS 05527950 Mill Creek has a drainage area of 62.61 square miles (Marti et al., 2023, Table DR-5). In this case the value of the ratio A_u/A_g is 50/62.61 = 0.7986, which is between 0.5 and 2. Therefore, from the equation in Figure 36, the adjustment weighting factor w_a takes the value

$$w_a = \left|log_2\left(A_u/A_g\right)\right| = \ |log_2(0.7986)|$$

$$w_a = 0.3245$$

Figure 38. Equation. Adjustment weighting factor for ungaged location in region 2 near a streamgage.

From Marti et al. (2023, Table DR-10), the AEP 0.01 weighted peak-flow quantile, $(Q_{0.01})_{g,w}$ at streamgage 05527950 is 1,860 ft^3/s, and as previously computed in Figure 14, the regional regression estimate at the ungaged site of interest $(Q_p)_{u,r}$ is 1,858 ft^3/s. Therefore, the weighted estimate at the ungaged site of interest is

$$(Q_{0.01})_{u,w} = w_a(Q_{0.01})_{u,r} + (1 - w_a)(Q_{0.01})_{g,w}\left(A_u/A_g\right)$$

$$(Q_{0.01})_{u,w} = 0.3245 * 1858 + (1 - 0.3245) * 2150 * (50/62.61)$$

$$(Q_{0.01})_{u,w} = 1{,}763 \text{ ft}^3\text{/s}.$$

Figure 39. Equation. Near-gage adjustment for the 0.01 Annual Exceedance Probability at an ungaged location in region 2 near a streamgage.

Summing up, the estimated 1% AEP peak-flow quantile $Q_{0.01}$ for this hypothetical ungaged basin on a stream in region 2 near to streamgage 05527950 (A_u/A_g = 0.799) is 1,763 ft^3/s, compared to 1,858 ft^3/s without the near-gage adjustment.

INTERPOLATION BETWEEN STREAMGAGES

Periods of record at streamgages were deemed to be unsuitable for use in regional regression because of regulation or urbanization or because the associated drainage basins are large compared to any other streamgages in the region. Peak-flow quantiles were computed for these periods of record if the regulation or urbanization conditions, if any, were approximately stationary and long enough (at least 10 years), and the values of these quantiles and their uncertainties appear in Tables DR-10 and DR-11, respectively, of Marti et al. (2023). If the associated streamgages are on the same river and their periods of record approximately agree, interpolation of the peak-flow quantiles along the river reach between the streamgages was considered as a means of providing quantile estimates for ungaged locations along those reaches.

Six reaches where interpolation was deemed to be appropriate were identified on five rivers: the Rock and the Fox Rivers, which were unsuitable for use in regional regression due to the size of the drainage basins; as well as the Sangamon, the Kaskaskia (two reaches), and the Big Muddy Rivers which were unsuitable due to significant regulation. These interpolated reaches are illustrated in Figures 42–46. The method of interpolation is log-log linear versus drainage area—that is, the log-discharge at AEP p, Q_p, is assumed to vary linearly with log of the drainage area A between the drainage area at the upstream streamgage A_{us} and the drainage area at the downstream streamgage A_{ds}, which is expressed by the following equation:

$$log_{10}Q_p(A) - log_{10}Q_p(A_{ds}) = \frac{log_{10}Q_p(A_{us}) - log_{10}Q_p(A_{ds})}{log_{10}A_{us} - log_{10}A_{ds}} (log_{10}(A) - log_{10}(A_{ds}))$$

$$= \frac{log_{10}[Q_p(A_{us})/Q_p(A_{ds})]}{log_{10}[A_{us}/A_{ds}]} log_{10}\left(\frac{A}{A_{ds}}\right) + log_{10}Q_p(A_{ds})$$

Figure 40. Equation. Peak-flow quantile interpolation function for two streamgages on the same reach.

The lines in the interpolation figures illustrate this interpolation method for each selected reach in these rivers. Note that in these figures the streamgages labeled with black lettering and whose quantiles are indicated with black circles are those used for the interpolation. The streamgages labeled with gray lettering and have gray squares for the quantiles are for reference only and were not used for interpolation because their data are inconsistent in some way with the other streamgages (for example, because their periods of record differ substantially). The uncertainties (as standard errors) of the quantile values at the streamgages are indicated by the error bars; these uncertainties are included in the figures to provide qualitative guidance, but a method for estimating the uncertainty of interpolated quantiles has not been developed.

To implement the method, the user should identify upstream and downstream discharges for a given AEP (p), $Q_p(A_{us})$ and $Q_p(A_{ds})$ and drainage areas, A_{us} and A_{ds}, and the drainage area at the point of interest A and solve the equation in Figure 40 for $log_{10}Qp(A)$. For example, if the point of interest is on the Rock River and has a drainage area of 7,000 square miles and the AEP of interest is p = 0.01, then the upstream streamgage is 05437500 and the downstream gage is 05443500 (Figure 42). Then, from Marti et al. (2023, Table DR-10), the discharges $Q_p(A_{us})$ and $Q_p(A_{ds})$ are 35,500 and 55,900 ft^3/s, respectively, and from Marti et al. (2023, Table DR-5), the drainage areas A_{us} and A_{ds} are 6,365 and 8,755 square miles, respectively, so from the equation in Figure 40, $log_{10}Q_p(A)$ is given by:

$$log_{10}Q_p(A) = \frac{log_{10}(35{,}500/55{,}900)}{log_{10}(6{,}365/8{,}755)} log_{10}\left(\frac{7{,}000}{8{,}755}\right) + log_{10}(55{,}900) = 4.609$$

Figure 41. Equation. Example peak-flow quantile interpolation computation.

So $Q_p(A) = 10^{4.609}$= 40,600 ft^3/s (rounded to three significant figures).

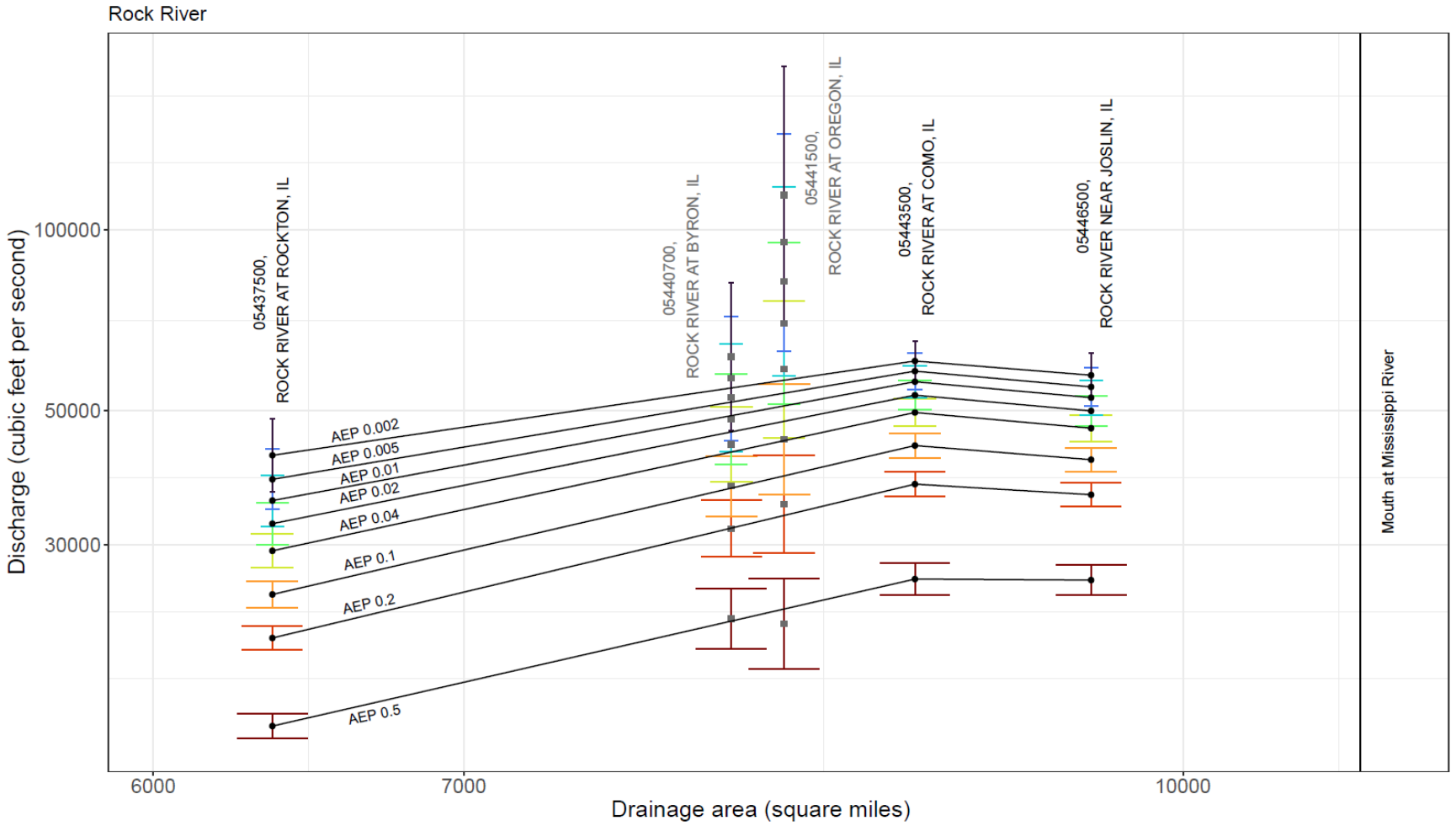

Figure 42. Diagram. Rock River interpolation illustration.

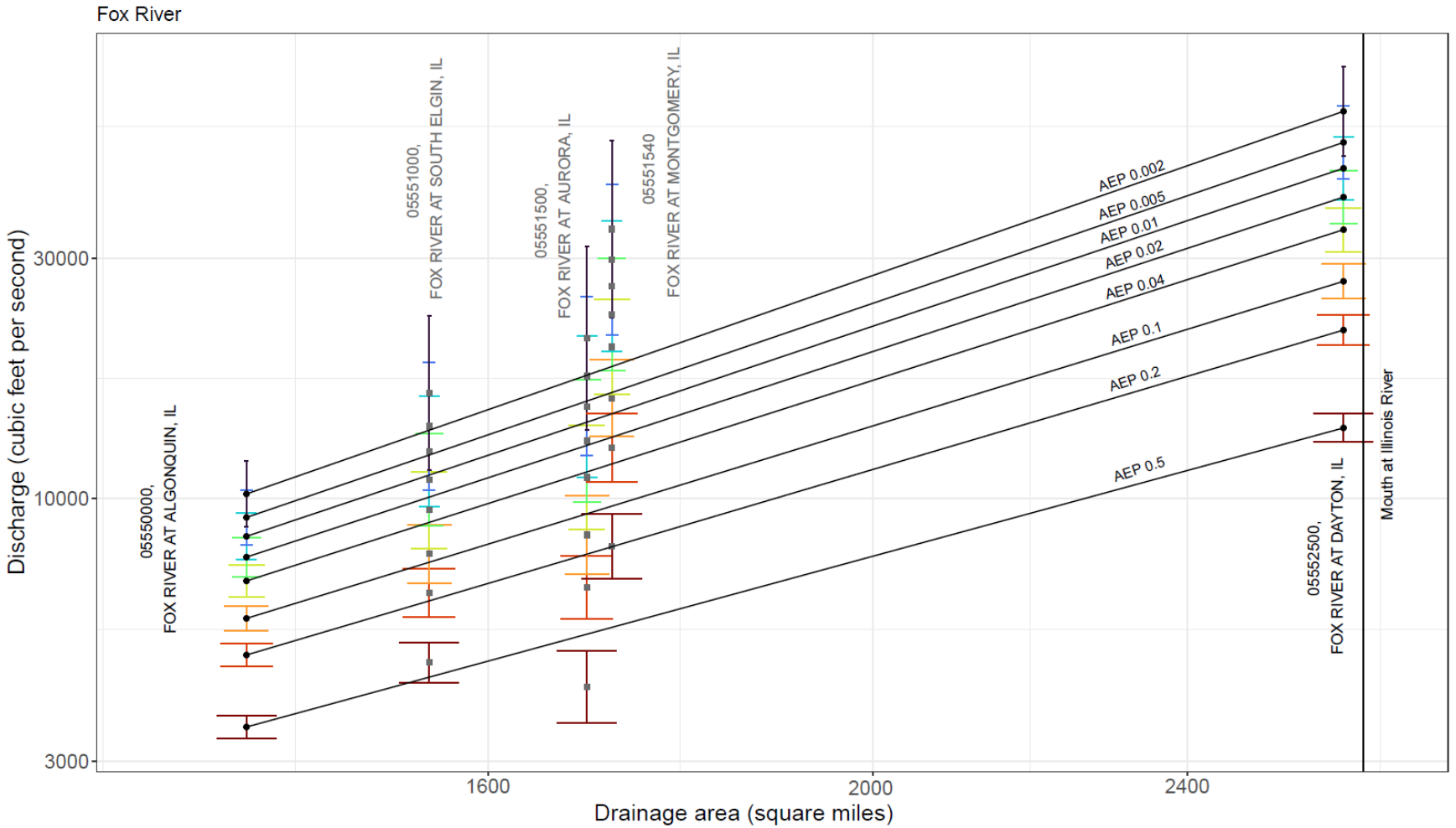

Figure 43. Diagram. Fox River interpolation illustration.

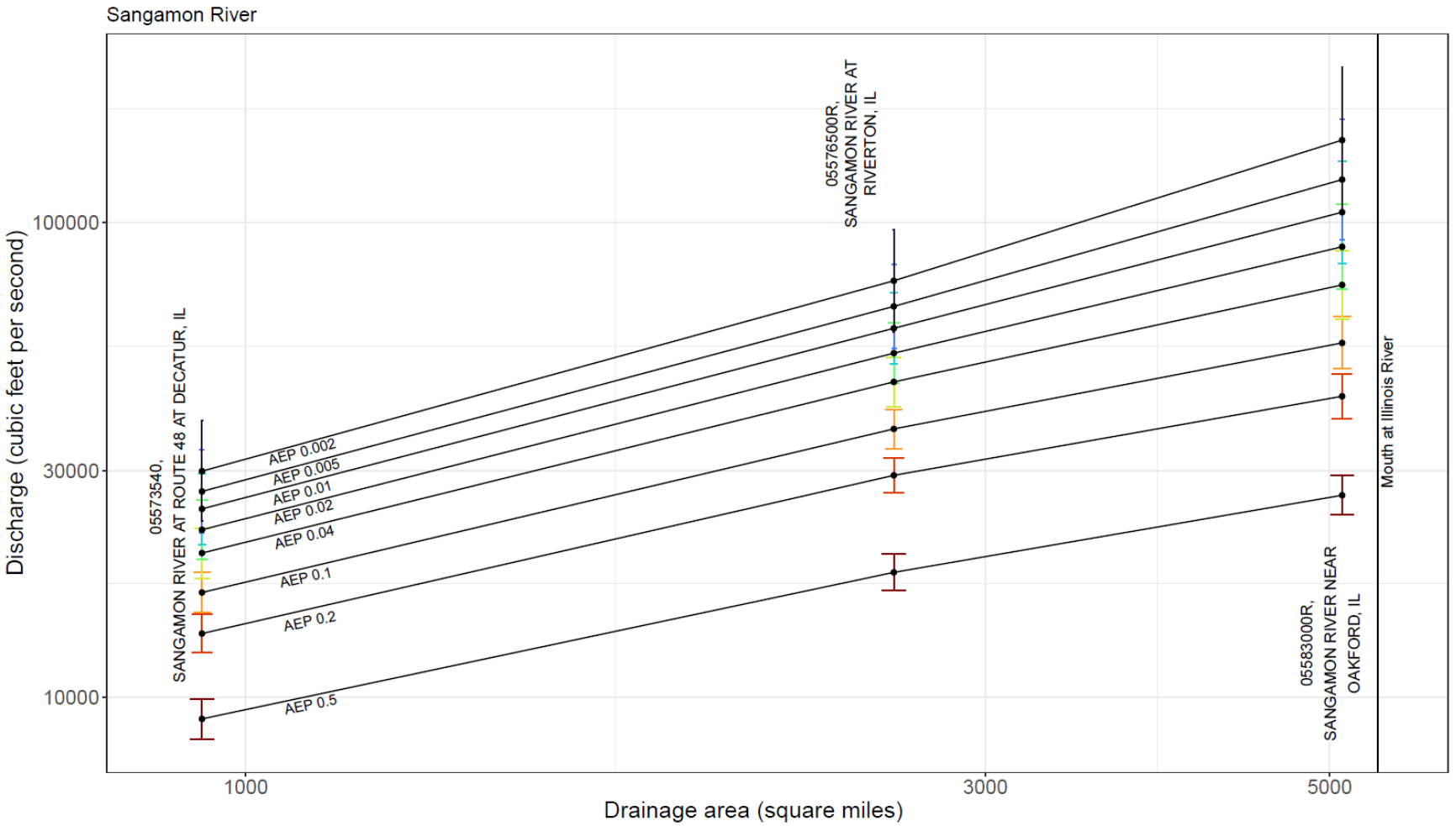

Figure 44. Diagram. Sangamon River interpolation illustration.

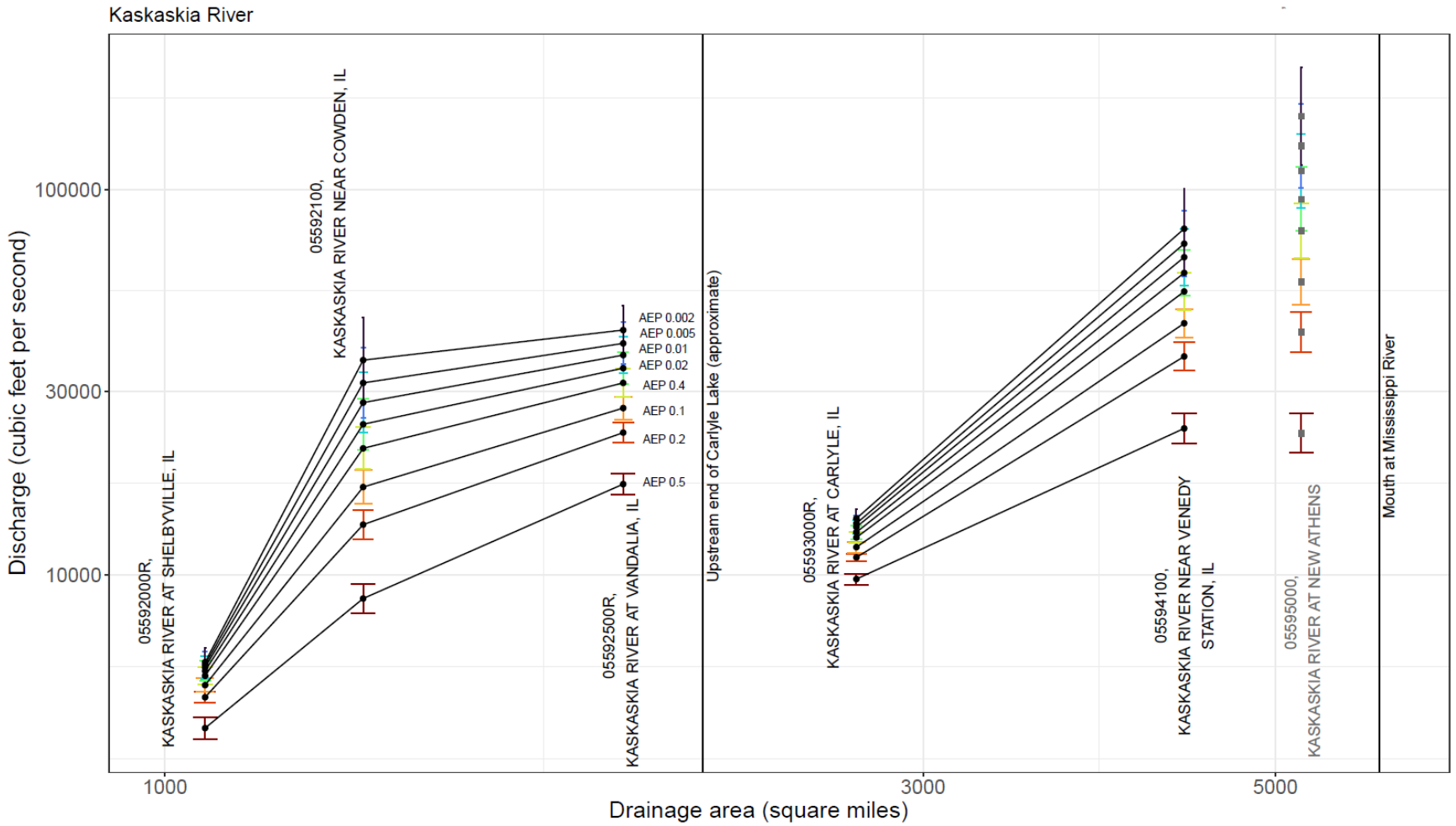

Figure 45. Diagram. Kaskaskia River interpolation illustration.

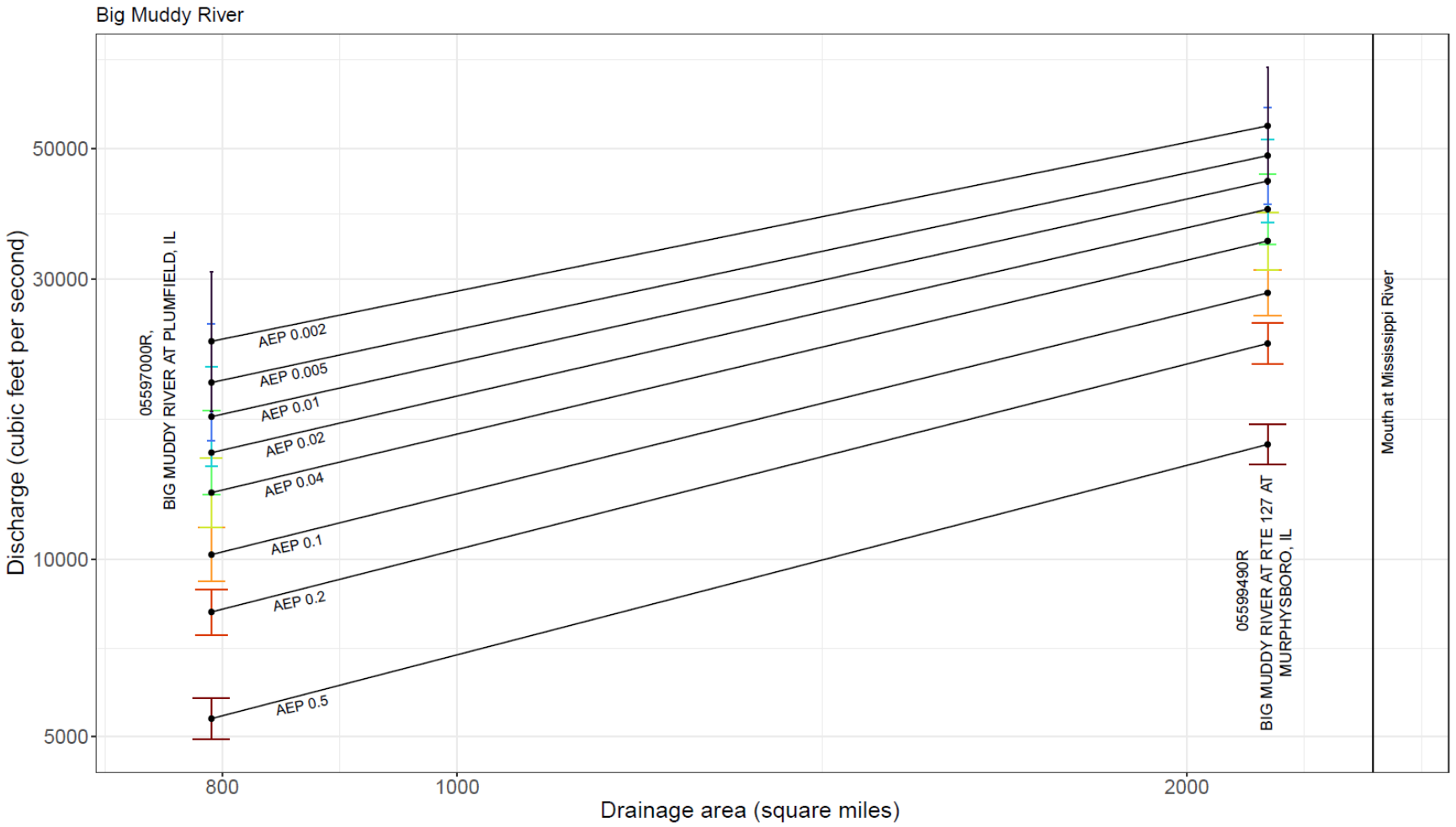

Figure 46. Diagram. Big Muddy River interpolation illustration.

CHAPTER 5: LIMITATIONS

Many assumptions are involved in the conduct and application of a flood-frequency analysis, and the validity of these assumptions governs the appropriateness of applications. Some assumptions are regarding the quantile estimates at the streamgages and others the RREs; many apply to both. A related fact is that the peak-flow quantile estimates at streamgages and from RREs are uncertain; for most applications of these techniques, an estimate of the uncertainty or a means of obtaining such an estimate is provided in this report. The primary assumptions and related uncertainties are as follows:

- The data used here come from streamgages, which require considerable expense to install and operate and, thus, are mostly located for purposes unrelated to getting the best regional information (Kiang et al., 2013; Knapp & Markus, 2003), and they are not always operated for periods that are long relative to the return periods of the estimates required in applications (Illinois Water Plan Taskforce, 2022). According to Knapp and Markus (2003), small basins and rural areas in general are underrepresented in the Illinois gaging network (refer also to Figures 1–5 of this report). For example, in this study, regions 6 and 7 had only 16 and 9 streamgages respectively, available for use in developing RREs. Therefore, uncertainties in those regions are larger than they would be otherwise.

- In this analysis, peak flows were assumed to be stationary—that is, their statistical properties were assumed to be unchanging in time, apart from effects of urbanization and regulation, for which efforts were made to avoid or adjust for. According to this assumption, it does not matter that different periods of record are available for each streamgage (for example, Figures 4 and 5). In the context of applications, the assumption of stationarity implies that the distribution of peak-flow values in the historical record is a reasonable approximation of the distribution applicable to the period of time under consideration in the use of the results of this study, which is usually in the future. This stationarity assumption includes that the historical record and the design period are similar in those aspects of climate that control floods.

- The distribution of peak flows was assumed to follow a log-Pearson type 3 distribution, and it was assumed to be accurately fitted using the method of moments as generalized by the expected moments algorithm and with regional skew. These assumptions are as recommended in Bulletin 17C (England Jr. et al., 2019), and the reasons for these recommendations are discussed there. Quantiles with small AEPs (extreme floods) are the most sensitive to the distributional assumption, and although long-term information such as historic floods was included in the analysis, such quantile estimates are inherently uncertain, as reflected in the uncertainty statistics provided (Marti et al., 2023, Table DR-11; O'Shea, 2023, Table DR-4).

- The selection of streamgages for use in the RRE development depended on various assumptions. Outside of region 2, the most important are that the effects of urbanization and regulation are minor, whereas within region 2, basins with a range of urbanization and

regulation were included. That the effects of regulation are minor continues to be a necessary assumption for applying the results of this study outside region 2, but a numerical measure of regulation was not available for all basins used in the study nor is a means of computing one for any site of interest. A method of addressing the effect of urbanization outside region 2 is provided; the applicability of this method depends on the hydrologic effects of urbanization in the basin of interest being similar to the mean effects of urbanization in region 2. The selected streamgages and regression equations in each region also define a range of basin characteristics (Table 10); the equations are not known to be applicable outside of these ranges.

- In developing the RREs, the statistical properties of the peak flows were also assumed to be homogeneous, given the values of the selected basin characteristics used in the equations, within each selected hydrologic region of the state. Deviations from such homogeneity are reflected in the goodness-of-fit statistics of the regression equations (Table 8; Marti et al., 2023, Table DR-8) and in the uncertainties of their predictions (refer to the "Accuracy of Final Spatial Regression Equations" section and the "Applications" chapter). Further, the effects of urbanization are assumed to be homogeneous across the state: the urbanization adjustment being applied outside of region 2 is based on mean urbanization effects determined from basins within region 2; however, note that the uncertainty arising from non-homogeneity in these effects within region 2 is incorporated into urbanization adjustment outside of region 2 (refer to the "Applications" chapter) through the uncertainty of the coefficients in Table 11.

CHAPTER 6: STREAMSTATS

StreamStats (https://streamstats.usgs.gov/ss/) is a web-based application developed and maintained by the U.S. Geological Survey to support the provision of streamflow statistics at streamgages and at ungaged locations (Ries III et al., 2017). StreamStats enables the delineation of drainage basins upstream from points on the river network and the determination of the values of basin characteristics characterizing the delineated basin along with their use in computing flow statistics for which RREs have been developed, among other capabilities. The RRE computations are also available from the web-based National Streamflow Statistics program application (https://streamstats.usgs.gov/nss/).

Peak-flow quantiles in Illinois were first implemented in StreamStats in 2007 (Ishii et al., 2010) using the results of Soong et al. (2004) with base data consisting of a 30 m (98.4 ft) digital elevation model resampled to 10 m (32.8 ft) resolution and 1:100,000 hydrography. Peak-flow quantiles for hydrologic region 2 (northeastern Illinois) were revised in 2016 based on the results of Over et al. (2021). As part of this study, the peak-flow quantiles throughout Illinois and the RREs for estimating them are being updated in StreamStats, and the base data being used in StreamStats is being upgraded to the true 10 m digital elevation model and 1:24,000 hydrography used to determine the basin boundaries and basin characteristics used in the results in this report.

The peak-flow quantiles and associated confidence intervals calculated by StreamStats for an ungaged location outside of region 2 will be the urbanization-adjusted values. These estimates will be based on the difference in the urban fraction at the ungaged location relative to the baseline urban fraction of the streamgages used in the development of the RREs shown in Table 12 as described in the section “At Ungaged Sites Outside of Region 2.”

CHAPTER 7: SUMMARY

Flood-frequency estimates are a critical ingredient in a variety of land-use planning, emergency response planning, and infrastructure design purposes, including in the hydraulic design of bridges. Statewide estimates of flood-frequency for most of Illinois have not been updated since a previous study that used data through 1999, and for northeastern Illinois, since a study that used data through 2009. This study, which uses data recorded through September 2017, updates the results of those two studies, and in doing so, it makes use of higher-resolution geographic base data, using a 10 m digital elevation model and 1:24,000-scale hydrography.

The study has three primary products: (1) flood-frequency estimates (and their uncertainties) at the study streamgages; (2) sets of equations (called regional regression equations, RREs) that relate the basin characteristics and flood-frequency estimates, and thereby allow estimates of flood frequencies at ungaged locations throughout the state and the uncertainty of such estimates; and (3) implementation of the RREs in the web application StreamStats.

Streamgages throughout Illinois as well as streamgages within 50 miles of the Illinois border in basins that flow into Illinois were considered for use in this study. Criteria for inclusion were somewhat different between the northeastern region and the rest of the state. All 181 streamgages from the previous study in the northeastern region, which included urbanized basins with records adjusted to 2010 urbanization conditions, were included in this study, and their records were extended if the additional record had little to no further change in drainage basin conditions. Streamgages were added that had at least 10 years of record with approximately stationary drainage basin conditions, for a total of 194. Outside of the northeastern region, streamgages were required to have one or more periods of record of at least 10 years with stationary basin conditions; 270 such streamgages and 282 such periods of record were found, for a total of 464 streamgages and 476 periods of record statewide.

For all periods of record at these 464 streamgages, peak-flow quantiles and their uncertainties were computed for eight AEPs, 0.5, 0.2, 0.1, 0.04, 0.02, 0.01, 0.005, and 0.002, by fitting to the log-Pearson Type 3 distribution using the expected moments algorithm and censoring of potentially influential low floods by the multiple Grubbs-Beck test. For streamgages deemed to be suitable for regional regression, regional skew values were obtained from previous studies, and the regional and at-site skews were weighted with weights inversely related to their uncertainties. The median uncertainties across all 476 periods of record of the estimated at-site peak-flow quantiles range from about 10% for AEP = 0.5 to about 30% for AEP = 0.002.

The periods of record were further winnowed for use in developing RREs that relate the basin characteristics and the at-site peak-flow quantiles, leaving a total of 422 streamgages across the state. After a redundancy analysis, 311 non-redundant streamgages remained for the development of the RREs. One set of equations was developed for each of seven conterminous hydrologic regions covering the state. The equations use one to four basin characteristics, depending on the region, with basin drainage area always included. The other characteristics used were measures of slope, soil permeability, and, in the northeastern Illinois region, a measure of urbanization and another of water

and wetlands. The prediction uncertainty of these equations is somewhat larger than that of the at-site peak-flow quantiles; for example, for AEP = 0.01, uncertainty ranges from about 32% to 65%, depending on the region.

Except for the northeastern Illinois region, these equations did not use data that include substantial effects of regulation or urbanization, and therefore the equations are not directly applicable to such basins. However, a method was provided to adjust the results for the effect of urbanization. A set of illustrative example applications of the equations is provided, and tables of all results at all the study streamgages and regions are provided in associated data releases.

The implementation of the study results in the web application, StreamStats, which is available at https://streamstats.usgs.gov/ss/, is ongoing. StreamStats will include access to the peak-flow quantiles at the study gages and a means of applying the RREs to get peak-flow quantile estimates at ungaged locations throughout the state. The implementation includes the improved, higher-resolution geographic base-data used for determining the basins used in study.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Allen, H. E., & Bejcek, R. M. (1979). *Effects of urbanization on the magnitude and frequency of floods in northeastern Illinois* (Water-Resources Investigations Report 79-36). U.S. Geological Survey. https://doi.org/10.3133/wri7936

Benson, M. A. (1962). *Evolution of methods for evaluating the occurrence of floods* (Water Supply Paper 1580-A). U.S. Geological Survey. https://doi.org/10.3133/wsp1580A

Bonnin, G. M., Martin, D., Lin, B., Parzybok, T., Yekta, M., & Riley, D. (2006). *Precipitation-frequency atlas of the United States—NOAA Atlas 14*. National Oceanic and Atmospheric Administration. https://www.weather.gov/media/owp/oh/hdsc/docs/Atlas14_Volume2.pdf

Carns, J. M. (1973). *Magnitude and frequency of floods in Illinois*. Illinois Department of Transportation, Division of Water Resources Management.

Changnon, S. A., Angel, J. R., Kunkel, K. E., & Lehmann, C. M. B. (2004). *Climate atlas of Illinois*. Illinois State Water Survey.

Chen, B., Krajewski, W. F., Liu, F., Fang, W., & Xu, Z. (2017). Estimating instantaneous peak flow from mean daily flow. *Hydrology Research*, *48*(6), 1474–1488. https://doi.org/10.2166/nh.2017.200

Cohn, T. A., Berenbrock, C., Kiang, J. E., & Mason Jr., R. R. (2012). *Calculating weighted estimates of peak streamflow statistics.* (Fact Sheet 2012–3038). U.S. Geological Survey. https://doi.org/10.3133/fs20123038

Cohn, T. A., England, J. F., Berenbrock, C. E., Mason, R. R., Stedinger, J. R., & Lamontagne, J. R. (2013). A generalized Grubbs-Beck test statistic for detecting multiple potentially influential low outliers in flood series. *Water Resources Research*, *49*(8), 5047–5058. https://doi.org/10.1002/wrcr.20392

Cohn, T. A., Lane, W. L., & Baier, W. G. (1997). An algorithm for computing moments-based flood quantile estimates when historical flood information is available. *Water Resources Research*, *33*(9), 2089–2096. https://doi.org/10.1029/97WR01640

Cohn, T. A., Lane, W. L., & Stedinger, J. R. (2001). Confidence intervals for expected moments algorithm flood quantile estimates. *Water Resources Research*, *37*(6), 1695–1706. https://doi.org/10.1029/2001WR900016

Curry, B. B., Wang, H., Panno, S. V., & Hackley, K. C. (2010). Quaternary paleoclimate. In D. R. Kolata, & C. K. Nimz (Eds.), *Geology of Illinois*. Illinois State Geological Survey.

Curtis, G. W. (1977). *Technique for estimating magnitude and frequency of floods in Illinois* (Water-Resources Investigations Report 77-117). U.S. Geological Survey. https://doi.org/10.3133/wri77117

Curtis, G. W. (1987). *Technique for estimating flood-peak discharges and frequencies on rural streams in Illinois* (Water-Resources Investigations Report 87-4207). U.S. Geological Survey. https://doi.org/10.3133/wri874207

Dewitz, J. (2019). *National Land Cover Database (NLCD) 2016 Products (ver. 2.0, July 2020)* [dataset]. U.S. Geological Survey. https://doi.org/10.5066/P96HHBIE

Eng, K., Chen, Y.-Y., & Kiang, J. E. (2009). *User's guide to the weighted-multiple-linear regression program (WREG version 1.0)* (Techniques and Methods, Vols. 4-A8). U.S. Geological Survey. https://doi.org/10.3133/tm4A8

England Jr., J. F., Cohn, T. A., Faber, B. A., Stedinger, J. R., Thomas Jr., W. O., Veilleux, A. G., Kiang, J. E., & Mason Jr., R. R. (2019). *Guidelines for determining flood flow frequency—Bulletin 17C* (Techniques and Methods, Vols. 4-B5). U.S. Geological Survey. https://doi.org/10.3133/tm4B5

Falcone, J. A. (2011). *GAGES-II: Geospatial Attributes of Gages for Evaluating Streamflow* [dataset]. https://doi.org/10.3133/70046617

Farmer, W. H. (2021). *WREG: Weighted Least Squares Regression for Streamflow Frequency Statistics* (3.00 [Computer software]) [Computer software]. https://doi.org/10.5066/P9ZCGLI1

Farmer, W. H., Kiang, J. E., Feaster, T. D., & Eng, K. (2019). *Regionalization of surface-water statistics using multiple linear regression* (Techniques and Methods, Vols. 4-A12). U.S. Geological Survey. https://doi.org/10.3133/tm4A12

Frankson, R., Kunkel, K. E., Champion, S. M., Stewart, B. C., Easterling, D. R., Hall, B., Angel, J. R., & Timlin, M. S. (2022). *Illinois State Climate Summaries* (Technical Report 150-IL). National Oceanic and Atmospheric Administration. https://statesummaries.ncics.org/downloads/Illinois-StateClimateSummary2022.pdf

Griffis, V. W., & Stedinger, J. R. (2007). Incorporating climate change and variability into Bulletin 17B LP3 Model. In *World Environmental and Water Resources Congress 2007: Restoring Our Natural Habitat*. American Society of Civil Engineers. http://ascelibrary.org/doi/abs/10.1061/40927%28243%2969

Griffis, V. W., & 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*, *14*(2), 121–130. https://doi.org/10.1061/(ASCE)1084-0699(2009)14:2(121)

Hansel, A. K., & McKay, E. D., III. (2010). Quaternary Period. In D. R. Kolata & C. K. Nimz (Eds.), *Geology of Illinois*. Illinois State Geological Survey.

Hayden, B. P. (1988). Flood climates. In V. R. Baker, R. C. Kochel, & P. C. Patton (Eds.), *Flood geomorphology*. John Wiley and Sons. https://www.ltrr.arizona.edu/~katie/kt/pubs/Flood-Climates-Hayden.1988.pdf

Helsel, D. R., Hirsch, R. M., Ryberg, K. R., Archfield, S. A., & Gilroy, E. J. (2020). *Statistical methods in water resources* (Techniques and Methods, Vols. 4-A3). U.S. Geological Survey. https://doi.org/10.3133/tm4A3

Illinois Water Plan Taskforce. (2022). *2022 State Water Plan*. https://dnr.illinois.gov/content/dam/soi/en/web/dnr/waterresources/statewaterplantaskforce/documents/swptf_report_dec2022.pdf

Ishii, A. L., Soong, D. T., & Sharpe, J. B. (2010). *Implementation and evaluation of the Streamflow Statistics (StreamStats) Web application for computing basin characteristics and flood peaks in Illinois* (Scientific Investigations Report 2009-5197). U.S. Geological Survey.

https://pubs.usgs.gov/sir/2009/5197/

Kiang, J. E., Stewart, D. W., Archfield, S. A., Osborne, E. B., & Eng, K. (2013). *A national streamflow network gap analysis* (Scientific Investigations Report 2013-5013). U.S. Geological Survey. https://doi.org/10.3133/sir20135013

Knapp, H. V., & Markus, M. (2003). *Evaluation of the Illinois streamflow gaging network* (ISWS Contract Report CR 2003-05). Illinois State Water Survey. http://hdl.handle.net/2142/94208

Lumley, Thomas, based on Fortran code by Alan Miller. (2020). *leaps: Regression Subset Selection* (3.1) [Computer software]. https://CRAN.R-project.org/package=leaps

Marti, M. K., Over, T. M., & O'Shea, P. S. (2023). *Data for estimating peak-flow quantiles for selected annual exceedance probabilities in Illinois* [Data Release]. U.S. Geological Survey. https://doi.org/10.5066/P9XPWUMI

Mitchell, W. D. (1954). *Floods in Illinois: Magnitude and frequency: Prepared in cooperation with the U.S. Geological Survey and Division of Waterways, Department of Public Works and Buildings*. https://babel.hathitrust.org/cgi/pt?id=mdp.39015037920652&seq=5

National Oceanic and Atmospheric Administration. (2006). *Precipitation frequency for Ohio River Basin, USA - NOAA Atlas 14 Volume 2* [dataset]. Office of Hydrologic Development, Hydrometeorological Design Studies Center. http://hdsc.nws.noaa.gov/hdsc/pfds/pfds_gis.html

O'Shea, P. S. (2023). *Peak-flow frequency analysis for 464 U.S. Geological Survey streamgages in Illinois, Indiana, and Wisconsin, based on data through water year 2017* [Data Release]. U.S. Geological Survey. https://doi.org/10.5066/P9XUH9SR

Over, T. M., Riley, J. D., Marti, M. K., Sharpe, J. B., & Arvin, D. V. (2014). *Estimation of regional flow-duration curves for Indiana and Illinois (ver. 2.0, April 2022)* (Scientific Investigations Report 2014-5177). U.S. Geological Survey. https://doi.org/10.3133/sir20145177

Over, T. M., Saito, R. J., & Soong, D. T. (2016). *Adjusting annual maximum peak discharges at selected stations in northeastern Illinois for changes in land-use conditions* (Scientific Investigations Report 2016–5049). U.S. Geological Survey. https://doi.org/10.3133/sir20165049

Over, T. M., Saito, R., Veilleux, A., O'Shea, P., Sharpe, J., Soong, D., & Ishii, A. (2021). *Estimation of peak discharge quantiles for selected annual exceedance probabilities in northeastern Illinois* (Illinois Center for Transportation Report FHWA-ICT-16-013). Illinois Center for Transportation. https://doi.org/10.36501/0197-9191/16-014

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, *11*(5), 1633–1644. https://doi.org/10.5194/hess-11-1633-2007

Pickels, G. W. (1937). *Magnitude and frequency of floods on Illinois streams* (University of Illinois Bulletin 296). University of Illinois. https://www.ideals.illinois.edu/items/4893

PRISM Climate Group. (2014). *Parameter-elevation regressions on independent slope model normals* [dataset]. Oregon State University. https://prism.oregonstate.edu

R Core Team. (2022). *R: A language and environment for statistical computing.* (4.0.2) [Computer

software]. R Foundation for Statistical Computing. https://www.R-project.org/

Ries III, K. G. (2007). *The National Streamflow Statistics Program: A computer program for estimating streamflow statistics for ungaged sites* (Techniques and Methods, Book 4, Section A, Chapter 6; p. 37). U.S. Geological Survey. https://doi.org/10.3133/tm4A6

Ries III, K. G., Newson, J. K., Smith, M. J., Guthrie, J. D., Steeves, P. A., Haluska, T., Kolb, K. R., Thompson, R. F., Santoro, R. D., & Vraga, H. W. (2017). *StreamStats, version 4* (Fact Sheet 2017-3046). U.S. Geological Survey. https://doi.org/10.3133/fs20173046

Rosbjerg, D., Blöschl, G., Burn, D. H., Castellarin, A., Croke, B., Di Baldassarre, G., Iacobellis, V., Kjeldsen, T. R., Kuczera, G., Merz, R., Montanari, A., Morris, D., Ouarda, T. B. M. J., Ren, L., Rogger, M., Salinas, J. L., Toth, E., & Viglione, A. (2013). Prediction of floods in ungauged basins. In A. Viglione, G. Blöschl, H. Savenije, M. Sivapalan, & T. Wagener (Eds.), *Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales* (pp. 189–226). Cambridge University Press. https://doi.org/10.1017/CBO9781139235761.012

Schafer, L. A., & Sharpe, J. B. (2023). *Elevation, flow accumulation, flow direction, and stream definition data in support of the Illinois StreamStats upgrade to the Basin Delineation Database* [Data Release]. U.S. Geological Survey. https://doi.org/10.5066/P9YIAUZQ

Schafer, L. A., Sharpe, J. B., & Marti, M. K. (2023). *Geographic data for the estimation of peak flow statistics for Illinois* [Data Release]. U.S. Geological Survey. https://doi.org/10.5066/P9ZAMASB

Schwarz, G. E., & Alexander, R. B. (1995). *Soils data for the conterminous United States from the NRCS State Soil Geographic (STATSGO) Data Base* (Open-File Report 95–449). U.S. Geological Survey. https://water.usgs.gov/GIS/metadata/usgswrd/XML/ussoils.xml

Soil Survey Staff. (n.d.). *Soil Survey Geographic (SSURGO) Database for Illinois* [dataset]. Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. https://websoilsurvey.nrcs.usda.gov/

Soller, D. R., Packard, P. H., & Garrity, C. P. (2012). *Database for USGS Map I-1970 — Map showing the thickness and character of Quaternary sediments in the glaciated United States east of the Rocky Mountains* [USGS Data Series 656]. U.S. Geological Survey. https://pubs.usgs.gov/ds/656/

Soong, D. T., Ishii, A. L., Sharpe, J. B., & Avery, C. F. (2004). *Estimating flood-peak discharge magnitudes and frequencies for rural streams in Illinois* (Scientific Investigations Report 2004-5103). U.S. Geological Survey. https://doi.org/10.3133/sir20045103

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

Stedinger, J. R., & Tasker, G. D. (1985). Regional hydrologic analysis: 1. Ordinary, weighted, and generalized least squares compared. *Water Resources Research*, *21*(9), 1421–1432. https://doi.org/10.1029/WR021i009p01421

Tasker, G. D. (1975). Combining estimates of low-flow characteristics of streams in Massachusetts and Rhode Island. *Journal of Research of the U.S. Geological Survey*, *3*(1), 107112. https://pubs.er.usgs.gov/publication/70232675

Tasker, G. D., & Stedinger, J. R. (1989). An operational GLS model for hydrologic regression. *Journal of Hydrology*, *111*(1), 361–375. https://doi.org/10.1016/0022-1694(89)90268-0

Theobald, D. (2005). Landscape patterns of exurban growth in the USA from 1980 to 2020. *Ecology and Society*, *10* (1), 29. https://www.jstor.org/stable/26267722

U.S. Army Corps of Engineers. (2018). *National inventory of dams: U.S. Army Corps of Engineers*. https://nid.sec.usace.army.mil/#/

U.S. Department of Agriculture Natural Resources Conservation Service. (2009). Chapter 7 Hydrologic Soil Groups. In *Part 630 Hydrology—National Engineering Handbook*. U.S. Department of Agriculture. https://directives.sc.egov.usda.gov/OpenNonWebContent.aspx?content=22526.wba

U.S. Geological Survey. (2018). USGS Water Data for the Nation. U.S. Geological Survey National Water Information System database, accessed January 9, 2018, at https://doi.org/10.5066/F7P55KJN

U.S. Geological Survey. (2021). *National Hydrography Dataset*. https://www.usgs.gov/national-hydrography/national-hydrography-dataset

U.S. Interagency Advisory Council on Water Data. (1982). *Guidelines for determining flood flow frequency—Hydrology Subcommittee, Bulletin 17B*. U.S. Geological Survey. https://water.usgs.gov/osw/bulletin17b/dl_flow.pdf

U.S. Water Resources Council. (1967). *A uniform technique for determining flood-flow frequencies: U.S. Water Resources Council Bulletin 15*. U.S. Water Resources Council. https://water.usgs.gov/osw/bulletin17b/Bulletin_15_1967.pdf

Veilleux, A. G. (2009). *Bayesian GLS regression for regionalization of hydrologic statistics, floods, and Bulletin 17 skew* [M.S.]. Cornell. https://ecommons.cornell.edu/bitstream/handle/1813/13819/Veilleux,%20Andrea.pdf?sequence=1

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

Wieczorek, M. E., Jackson, S. E., & Schwarz, G. E. (2018). *Select attributes for NHDPlus version 2.1 reach catchments and modified network routed upstream watersheds for the conterminous United States: National Inventory of Dams (NID) storage and construction by decade, 1930 to 2010* [Data Release]. U.S. Geological Survey. https://doi.org/10.5066/F7765D7V

Wieczorek, M. E., Wolock, D. M., & McCarthy, P. M. (2021). *Dam impact/disturbance metrics for the conterminous United States, 1800 to 2018* [Data release]. U.S. Geological Survey. https://doi.org/10.5066/P92S9ZX6

Wolock, D. M. (1997). *STATSGO soil characteristics for the conterminous United States* (Open-File Report 97–656). U.S. Geological Survey. https://doi.org/10.3133/ofr97656

Woods, A. J., Omernik, J. M., Pederson, C. L., & Moran, B. C. (2006). *Level III and IV Ecoregions of Illinois*. Environmental Protection Agency. https://gaftp.epa.gov/EPADataCommons/ORD/Ecoregions/il/il_report.pdf

Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S. M., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Granneman, B., Liknes, G. C., Rigge, M., & Xian, G. (2018). A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies. *ISPRS Journal of Photogrammetry and Remote Sensing*, *146*, 108–123. https://doi.org/10.1016/j.isprsjprs.2018.09.006


Illinois Department
of Transportation


ILLINOIS