science for a changing world

Prepared in cooperation with the U.S. Environmental Protection Agency, Region V, and the Indiana Department of Environmental Management

Estimation of Regional Flow-Duration Curves for Indiana and Illinois

Scientific Investigations Report 2014–5177
Version 2.0, April 2022

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

Cover image. The view upstream from CR3400E bridge, station 05554000, North Fork Vermilion River, near Charlotte, Illinois, at low flow. Photograph taken by Teresa Halfar, U.S. Geological Survey, July 12, 2007. http://il.water.usgs.gov/proj/nvalues/db/sites/05554000.shtml.

Estimation of Regional Flow-Duration Curves for Indiana and Illinois

By Thomas M. Over, James D. Riley, Mackenzie K. Marti, Jennifer B. Sharpe, and Donald Arvin

Prepared in cooperation with the U.S. Environmental Protection Agency, Region V, and the Indiana Department of Environmental Management

Scientific Investigations Report 2014–5177
Version 2.0, April 2022

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

U.S. Department of the Interior
SALLY JEWELL, Secretary

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

U.S. Geological Survey, Reston, Virginia: 2014
First release: 2014
Revised: April 2022 (ver. 2.0)

Suggested citation:
Over, T.M., Riley, J.D., Marti, M.K., Sharpe, J.B., and Arvin, D., 2014, Estimation of regional flow-duration curves for Indiana and Illinois (ver. 2.0, April 2022): U.S. Geological Survey Scientific Investigations Report 2014–5177, 24 p. and additional downloads, tables 2–5, 8–13, and 18, https://doi.org/10.3133/sir20145177.

ISSN 2328-0328 (online)

Contents

Figures

Tables

Tables available for download at https://doi.org/10.3133/sir20145177.

Conversion Factors

Inch/Pound to SI

Multiply	By	To obtain
	Length	
inch (in.)	2.54	centimeter (cm)
inch (in.)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
	Area	
square mile (mi^2)	2.590	square kilometer (km^2)
	Flow rate	
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)
inch per hour (in/h)	0 .0254	meter per hour (m/h)

Horizontal coordinate information is referenced to the North American Datum of 1927 (NAD 27).

Acknowledgments

Primary funding for this project was provided by the U.S. Environmental Protection Agency (EPA) Region V through Interagency Agreement DW-14-94818201-1. Christine Urban of EPA Region V served as project coordinator. Funding for the revisions leading to version 2.0 was provided by the Indiana Department of Environmental Management.

Two U.S. Geological Survey colleagues played key roles in this project: Dave Lorenz provided substantial guidance on the use of S+ for censored regression, and David Soong shared important information from his experience with peak-flow regionalization in Illinois.

Estimation of Regional Flow-Duration Curves for Indiana and Illinois

By Thomas M. Over,[2] James D. Riley,[1] Mackenzie K. Marti,[2] Jennifer B. Sharpe,[2] and Donald Arvin[2]

Abstract

Flow-duration curves (FDCs) of daily streamflow are useful for many applications in water resources planning and management but must be estimated at ungaged sites. One common technique for estimating FDCs at ungaged sites in a given region is to use equations obtained by linear regression of FDC quantiles against multiple basin characteristics that can be computed by means of a geographic information system (GIS) computer program. In this study, such regional regression equations for estimating FDC quantiles were computed at the 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 25, 30, 40, 50, 60, 70, 75, 80, 90, 95, 98, 99, 99.5, 99.8, and 99.9-percent exceedance probabilities for rural, unregulated streams in Indiana and Illinois with temporally stationary records, using data through September 30, 2007. The approach used accounts for censored values below 0.01 cubic feet per second, which are observed at exceedance probabilities as low as 70 percent (that is, occurring at least 30 percent of the time). The basin characteristics used are suitable for computation by the USGS Web-based application, StreamStats, and are available for all U.S. Environmental Protection Agency (EPA) Region V states and the larger Great Lakes area, with some specific local exceptions. Indiana and Illinois were each divided into three regions, and a different set of equations for estimating FDC quantiles was computed for each region.

The error of estimation of the FDC quantiles, measured as the mean square residual in log space converted to a percentage of the quantile, varies somewhat among regions and varies strongly with exceedance probability, with a minimum error of 10 to 20 percent at an exceedance probability of 5 or 10 percent, but rises to 17 to 38 percent at the high-flow end of the FDCs (the 0.1-percent quantile) and 100 to 745 percent at the low-flow end. For comparison, errors of estimation also were computed for FDC quantiles estimated by linear regression on drainage area alone and by using the drainage-area ratio (DAR) method. Three criteria, the nearest basin centroid and two others termed “strict” and “broad”, were used to select index stations for the DAR method. The “strict” and “broad” criteria put conditions on the basin centroid distance and the range of their drainage-area ratios, and the errors were averaged for all index station pairs satisfying each criterion. The use of the simpler DAR method usually resulted in higher errors of estimation compared to the linear regression equations with multiple basin characteristics, except occasionally in the case of the DAR method with the strict index station selection criterion, a criterion that is rarely possible to satisfy in practice.

An example application of the estimated equations to one gaged and a few ungaged locations in a watershed in the study area is included to illustrate the steps required. These steps are the computation of the basin characteristics and, using those characteristics together with the estimated equations, the computation of the FDC quantiles and their uncertainties.

Introduction

Flow-duration curves (FDCs), which are the cumulative probability distributions of stream discharge values usually averaged during a daily time step, are used in a wide variety of water resources applications (see Searcy, 1959; and Vogel and Fennessey, 1995, for general reviews). By applying regionalization techniques, FDCs may be estimated for ungaged sites in a region (Fennessey and Vogel, 1990), and if combined with timing information from an index station, they can provide the basis for estimating continuous streamflow at ungaged sites (Fennessey, 1994; Smakhtin, 1999; Mohamoud, 2008; Archfield and others, 2010; Straub and Over, appendix A, 2010; Linhart and others, 2012; Stuckey and others, 2012).

The particular application of FDCs to the construction of contaminant load-duration curves (LDCs) in the support of the development of Total Maximum Daily Load (TMDL) estimates (Bonta and Cleland, 2003; Cleland, 2002; U.S. Environmental Protection Agency, 2007a; Stiles, 2001; Sullivan, 2002) was the original motivation for this project, which was carried out in cooperation with U.S. Environmental Protection Agency (EPA) Region V. Partially (Johnson and others, 2009) and fully (Kim and others, 2012) Web-based tools have been developed to facilitate development of LDCs at streamgages. The tool of Kim and others (2012) includes the option of inputting a drainage-area ratio to transfer the streamflow information from a nearby streamgage to an

[1]Department of Geology/Geography, Eastern Illinois University.

[2]U.S. Geological Survey.

ungaged site, but neither tool provides a general, state-of-the-art approach to estimating flow at ungaged locations. The equations presented here are suitable for implementation in StreamStats (Ries and others, 2008) or similar Web-based application, which could, in turn, be linked to a Web-based LDC tool to extend the use of such tools to ungaged locations.

Previous Studies

There are apparently no previously published methods for estimating FDCs in ungaged streams in Indiana. Recent regional low-flow studies in Indiana include Arihood and Glatfelter (1991), who presented a method of estimation of low-flow frequency statistics at ungaged basins in northern and central Indiana that uses drainage area and a flow-duration quantile ratio, and Fowler and Wilson (1996), who compiled low-flow frequency characteristics and flow-duration curves at continuous-record streamgages and estimated low-flow frequency characteristics at partial-record stations throughout Indiana.

Previous methods of estimating FDC quantiles for ungaged streams in Illinois based on drainage area and mapped flow characteristics were developed by Mitchell (1957) and Singh (1971). Singh (1971) considered exceedance probabilities from 1 to 95 percent and divided the State into fourteen hydrologic divisions. Mitchell (1957) considered exceedance probabilities from 0.01 to 99.99 percent and assumed the State was homogeneous except for the continuously variable flow characteristics for which he created maps.

In a series of reports beginning with Knapp and others (1985), under the rubric of the Illinois Streamflow Assessment Model (ILSAM), the Illinois State Water Survey has created a database of streamflow statistics for each reach, gaged or ungaged, of selected river basins in Illinois, including period-of-record and seasonal flow-duration curves for “virgin” and “present” conditions and various low-flow frequency statistics. The ILSAM effort is in one sense more broad than the present report, as it covers more statistics and includes regulated and natural conditions, but in another sense is narrower, in that it covers a smaller fraction of the State of Illinois and of course does not include Indiana.

Several studies have reported development of regionalized daily FDC quantile equations for regions of the United States outside Indiana and Illinois (Fennessey and Vogel, 1990; Ries and Friesz, 2000; Flynn, 2003; Koltun and Whitehead, 2002; Perry and others, 2004; Mohamoud, 2008; Risley and others, 2008; Archfield and others, 2010; Esralew and Smith, 2010; Linhart and others, 2012); this study relied on these previous studies for the general approach. Regionalized equations for instantaneous flood-peak quantiles for Indiana and Illinois were developed most recently by Knipe and Rao (2005) and Soong and others (2004), respectively; the regions defined in those studies were used as the starting point for the regions selected in this study.

The drainage-area ratio (DAR) method has been previously compared with other methods including FDC-based daily streamflow estimation by, for example, Hirsch (1979), Emerson and others (2005), and Asquith and others (2006). Hirsch (1979) determined that his method of reconstruction of streamflow time series based on regional moments was superior to the drainage-area ratio method. Emerson and others (2005) and Asquith and others (2006) recommend the use of non-unit exponents in the drainage-area ratio method.

Purpose and Scope

This report presents the methods and results of estimation of FDC quantiles of daily streamflow at rural, unregulated streams in Indiana and Illinois by using multiple linear regression of FDC quantiles as a function of basin characteristics such as drainage area and other descriptors of basin properties. Additionally, FDC quantile estimates computed by linear regression with drainage area as the only basin characteristic and by using the DAR method of streamflow estimation are presented to provide a comparison of the errors of estimation of the proposed regional regression equations with simpler methods. An example application is included to illustrate the use of the reported equations.

Methods

Computing Basin Characteristics

Basin characteristics were derived in a manner to ensure consistency with USGS StreamStats (Ries and others, 2008) using Arc Hydro Tools in ArcMap (Maidment, 2002). Primary data types and sources used in this study are listed in table 1. For each data type in table 1, several basin characteristics were computed for each basin used in this study and tested for use in predicting the FDCs. The values of selected basin characteristics for each basin used in this study, including those selected for use in the multiple regression equations are given in tables 2 and 3 (available at https://doi.org/10.3133/sir20145177). The values of the characteristics in tables 2 and 3 that are not used in the multiple regression equations are provided for general information.

Some basin characteristics were derived by combining qualitative or categorical descriptions of a property into a continuous numerical index, in a manner analogous to certain characteristics developed by Knipe and Rao (2005). In particular, the parameter “Drain.Index” used in the proposed equations for Illinois region 1 (tables 3 and 10) and “PermBXThick” used in the proposed equations for Indiana region 3 (tables 2 and 9) and Illinois region 2 (tables 3 and 11) were computed following this approach. The equations used to compute these parameters are

Drain.Index = fraction “very poorly drained” + 2*fraction “poorly drained” + 4*fraction “somewhat poorly drained”+ 8*fraction “moderately well-drained” + 16*fraction “well-drained” + 32*fraction “excessively drained”, (1)

where the fractions are decimal fractions and were computed with ArcHydro Tools in ArcMap from the “dominant” condition as specified in the State Soil Geographic (STATSGO) data (Schwarz and Alexander, 1995; see also table 1),

and

PermBXThick = QSS_PermB*QSS_Thick, (2a)

where QSS_PermB is an index of the permeability of surficial Quaternary sediments computed as

QSS_PermB = 100*fraction coarse-grained stratified sediment + 1*fraction fine-grained stratified sediment + 1*fraction glacial till + 0.1*fraction exposed bedrock or sediment not of glacial origin, (2b)

where the fractions are decimal fractions and the weights (100, 1, and 0.1) are estimated relative hydraulic conductivity values (Soller and Berg, 1992), and QSS_Thick is a weighted average of the thickness of the surficial Quaternary sediments, computed as

QSS_Thick = 25*fraction 0-50 feet thick + 75*fraction 50-100 feet thick + 150*fraction 100–200 feet thick + 300*fraction 200–400 feet thick + 500*fraction 400–600 feet thick, (2c)

where the fractions of surficial Quaternary sediment types and thicknesses were computed in ArcMap from USGS Digital Data Series DDS 38 (Soller and Packard, 1998; see also table 1). Notice the weights in equation 2c (25, 75, and so on) are the midpoints of the thickness intervals; because the thickness intervals are in units of feet, these weights are as well, and thus QSS_Thick also has units of feet.

Table 1. Primary sources of GIS data used in this study and example derived basin characteristics.

[NED, National Elevation Dataset; NHD, National Hydrography Dataset; NOAA, National Oceanic and Atmospheric Administration; PRISM, Parameter-elevation Regressions on Independent Slopes Model; USGS, U.S. Geological Survey; WRD, Water Resources Discipline; NSDI, National Spatial Data Infrastructure; STATSGO, State Soil Geographic Database; CONUS-SOIL, Conterminous United States multi-layer SOIL characteristics dataset; NLCD, National Land Cover Database; NRGDC, Natural Resources Geospatial Data Clearinghouse; DDS: Digital Data Series; WWW, World Wide Web.]

Data type	Source	WWW reference	Example basin characteristics
Morphometric	NED (Gesch and others, 2002), NHD (Simley and Carswell, 2009)	http://ned.usgs.gov/, http://nhd.usgs.gov/	Basin area.
Climatic – precipitation frequency	NOAA Atlas 14 (Bonnin and others, 2006)	http://hdsc.nws.noaa.gov/hdsc/pfds/pfds_gis.html	100-year, 24-hour storm depth.
Long-term (1971–2000) average precipitation and temperature	PRISM Climate Group (Daly and others, 2008)	http://www.prism.oregonstate.edu/	Annual precipitation, December-January-February minimum temperature.
Mapped hydrologic properties	USGS-WRD NSDI node	http://water.usgs.gov/lookup/getgislist	Mean annual runoff.
Soil properties	Soils data derived from STATSGO (Schwarz and Alexander, 1995).	http://water.usgs.gov/GIS/metadata/usgswrd/XML/ussoils.xml	Hydrologic soil group, drainage class.
Soil properties	CONUS-SOIL (Miller and White, 1998)	http://www.soilinfo.psu.edu/index.cgi?index.html	Available water content, fraction sand-silt-clay by layer.
Soil properties	Soils data derived from STATSGO (Wolock, 1997)	https://doi.org/10.3133/ofr97656	Soil permeability.
Land use/land cover	NLCD 2001 (Homer and others, 2007)	http://www.mrlc.gov/nlcd.php	Fraction forested land, fraction imperviousness.
Wetlands	National Wetlands Inventory (Cowardin and others, 1979)	http://www.fws.gov/wetlands/	Fraction palustrine wetlands.
Illinois geology	Illinois NRGDC	http://www.isgs.uiuc.edu/nsdihome/	Thickness of glacial drift.
Indiana geology/hydrology	IndianaMap	http://maps.indiana.edu/LayerGallery.html	Fraction sinkhole area.
Glaciated Eastern U.S. Quaternary geology	USGS DDS38 (Soller and Packard, 1998)	http://pubs.usgs.gov/dds/dds38/	Thickness and coarseness of Quaternary sediments.

Selection and Testing of Streamgage Records and Computation of Flow-Duration Curves

The streamgage records used were intended to be rural and unregulated, and streamflow data through water year 2007 were used, where a water year is defined as beginning the prior calendar year on October 1 and ending during the specified calendar year on September 30. For Illinois, the status of records as being rural and unregulated was ensured by starting with the records used by Soong and others (2004) in their regional rural flood-frequency study of Illinois (which used data through water year 1999) and then checking for recent additions to the streamgage network. For Indiana, the set of streamgages includes currently (2007) unregulated records and years of record at currently regulated streams before regulation began. In Illinois, portions of records affected by the construction of major dams also were removed from the analysis. Because of its high degree of regulation, the Illinois River in particular was excluded from the analysis. Streamgages with records shorter than 8 years, as well as streamgages whose influent watersheds had large fractions (greater than 20 percent) of impervious land according to the National Land Cover Dataset (NLCD) of 2001 (Homer and others, 2007), also were removed. Further, records at streamgages with basin areas greater than 5,000 square miles (mi^2) in Indiana and 6,500 mi^2 in Illinois were removed because these would include only a few rivers, with the result that a representative sample could not be obtained and the risk that the non-representative sample would have a substantial effect on the regression coefficients. A regression estimate at these large scales usually is not needed anyway, because such large rivers generally have several streamgages that could be used for interpolation or because the estimate that would be produced would be inapplicable because the river is no longer unregulated.

Streamgage records were tested for stationarity as follows. As a preliminary step, when a streamgage had a few years of data separated from the bulk of the record by a much longer period without data, and this few years of data apparently were causing a trend, the short period of data was removed. Then formal trend-testing was applied to the quantiles of annual FDCs with exceedance probabilities between 1 and 99 percent. These quantiles were computed from complete water years of record at each streamgage and tested for temporal nonstationarity not explainable by annual (water-year) variation in basin-average precipitation computed from the 4-kilometer gridded Parameter-elevation Regressions on Independent Slopes Model (PRISM) precipitation data (Daly and others, 2008) by means of the “adjusted variable Kendall test” proposed by Alley (1988), which is applied as follows (Helsel and Hirsch, 2002, p. 335). First, the residuals of a linear regression of the dependent variable (here, an annual FDC quantile of a certain exceedance probability) on one or more exogenous variables (here, the annual basin-average PRISM precipitation) are computed. Then a second set of residuals, those of the exogenous variable regressed compared to time, are computed. Finally, a Mann-Kendall test is used to test for a trend in the first set of residuals as a function of the second set of residuals. Using the time residuals as the time variable in the Mann-Kendall test removes the effect of a possible trend in the exogenous variables. Quantiles in a given record failing this trend test were removed from the analysis by the following criterion: if any “low-flow” quantile (defined as an exceedance probability of 20 percent or greater) failed this test at the 1-percent significance level, then all low-flow quantiles were removed; similarly if one “high-flow” quantile (defined as an exceedance probability of 10 percent or less) failed this test at the 1-percent significance level, then all high-flow quantiles were removed.

For the streamgage records passing the tests described above, FDCs were computed from complete water years of published daily USGS discharge data by sorting the daily values and assigning exceedance probabilities to each value by means of the plotting position formula

$$p_i=(i-a)/(n+1-2a), \tag{3}$$

where p_i is the non-exceedance probability, i is the rank (1 to n, smallest to largest), n is the number of values, and a is a constant, taken here as 0.4 (Helsel and Hirsch, 2002, p. 23).The 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 25, 30, 40, 50, 60, 70, 75, 80, 90, 95, 98, 99, 99.5, 99.8, and 99.9-percent exceedance probability quantiles were obtained from the ordered pairs of ranked daily flow data and exceedance probabilities by linear interpolation. The computed FDC quantiles for high flow (0.1 to10 percent) and low flow (20 to 99.9 percent) are provided in tables 4 and 5 (available at https://doi.org/10.3133/sir20145177), along with information on the period of record used in the quantiles.

As a record of the trend test results, quantiles removed from the study because of the trend tests are shown in the tables without quantile values. Overall, among the 151 candidate study streamgages in Indiana remaining after removal of the basins with drainage areas greater than 5,000 mi^2, 109 streamgages, or about 72 percent, passed the low-flow trend test; and 140 streamgages, or about 93 percent, passed the high-flow trend test. The fractions passing the trend test for the Illinois streamgages remaining after removal of the basins with drainage areas greater than 6,500 mi^2 were similar: of 174 such streamgages, 131, or about 75 percent, passed the low-flow trend test; and 159, or about 91 percent, passed the high-flow trend test. A couple notable groups of stations for which both low and high-flow quantiles were removed include a group in the Kankakee River Basin in Indiana and Illinois (streamgages 05515500, 05518000, 05519000, 05520500, 05526000, and 05527500) (figs. 1 and 2) and a group in the Pecatonica River Basin in Wisconsin and Illinois (streamgages 05434500, 05435500, and 05436500) (fig. 2). The significant trends in the annual FDC quantiles of the streamgages in the Kankakee Basin group are all positive; whereas in the Pecatonica Basin group, the low-flow trends are positive, and the high-flow trends are negative. Discussion of possible

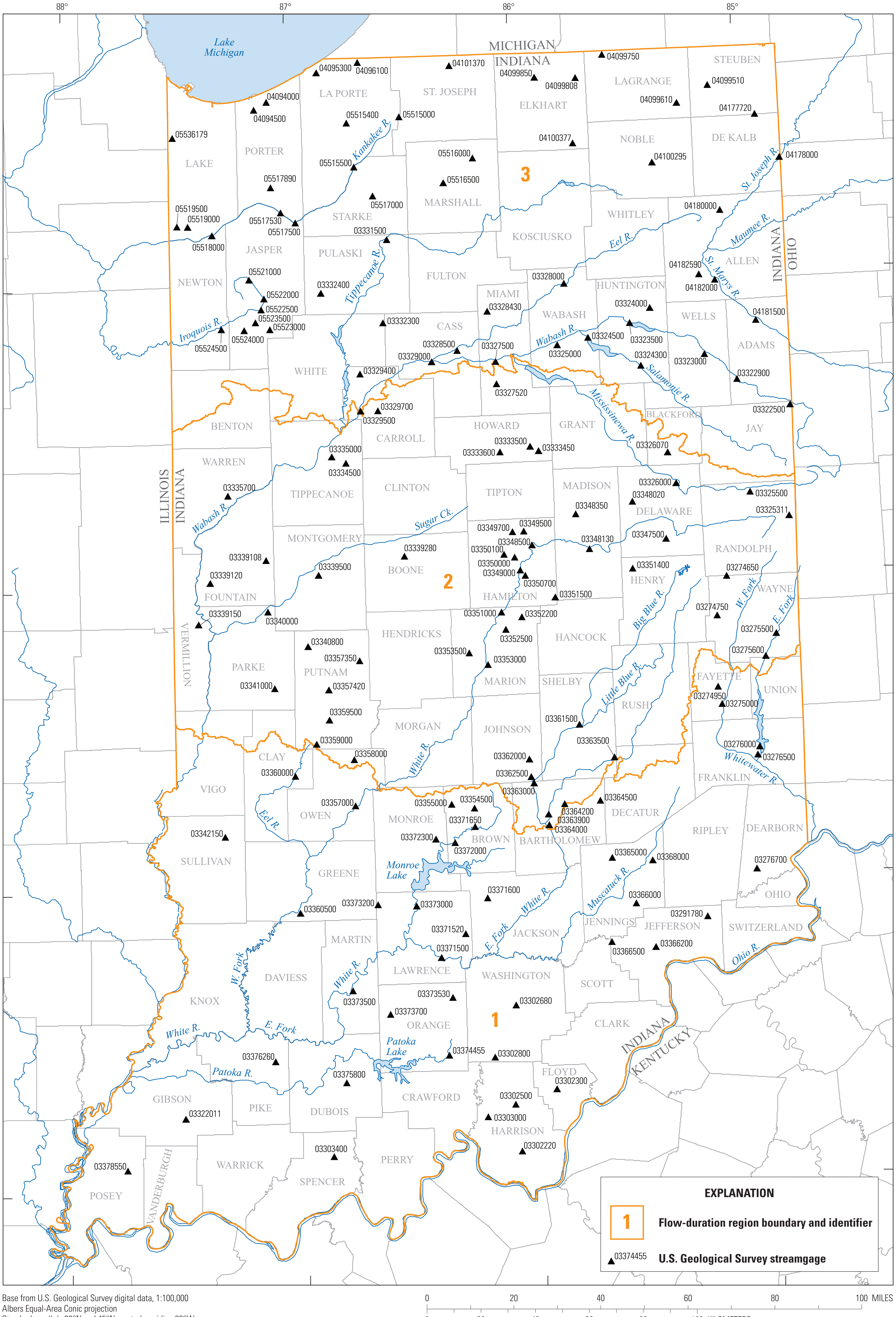
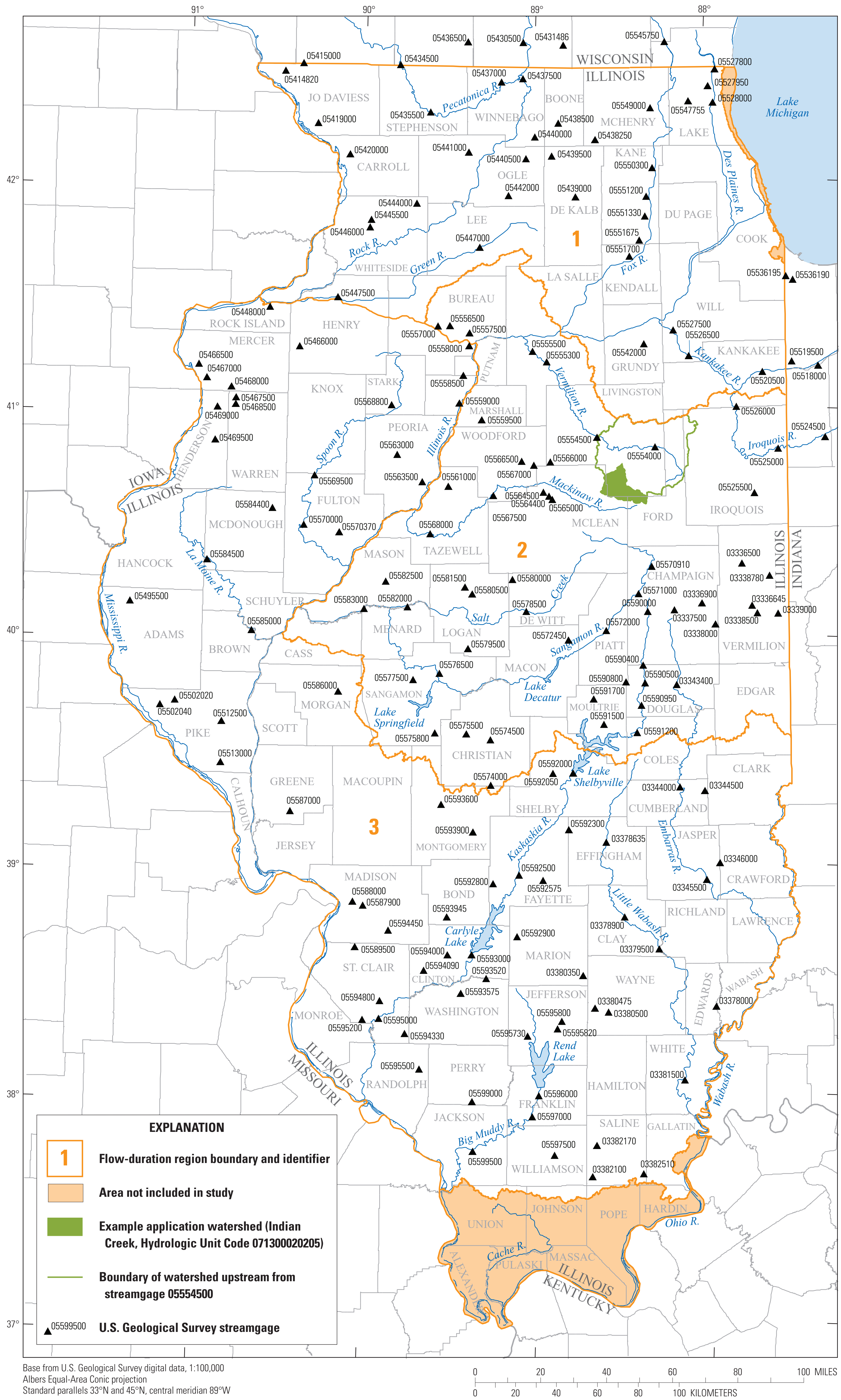


Figure 1. Indiana flow-duration regions.



Base from U.S. Geological Survey digital data, 1:100,000
Albers Equal-Area Conic projection
Standard parallels 33°N and 45°N, central meridian 89°W

Figure 2. Illinois flow-duration regions.

reasons for these trends can be found in Illinois Department of Natural Resources (1998) regarding trends in the Kankakee River Basin and in Markus and others (2013) and references therein regarding trends the Pecatonica River Basin.

Defining Regions

Because FDCs describe the flow properties throughout the full range of conditions from low to high flows, all the various physical factors governing streamflow affect the properties of FDCs, including climate, land use and vegetation, soils, topography, and geology (Searcy, 1959), and may therefore enter into the definition of flow-duration regions. Most of the physiography of Indiana and Illinois is usually divided into three general regions: a northern moraine and lake region, and central till plain region, and a southern hills region, with small areas along the borders assigned to other physiographic regions, though the precise boundaries vary (Leighton and others, 1948; Schneider, 1966; Gray, 2000). The three general regions are the result of glacial activity. The southern region was not glaciated during the Pleistocene or the glacial drift is thin and so the physiography is the result of "normal degradational processes" (Schneider, 1966). The central region consists of mostly uneroded broad plains on deep glacial drift and therefore the terrain is the most flat of the three general regions. The northern region consists of a variety of post-glacial features such as end moraines, outwash plains, and lake plains, resulting in the presence of many lakes, sand dunes, and peat bogs. Notable border regions include part of the Wisconsin Driftless Section in extreme northwestern Illinois and part of the Coastal Plain Province in extreme southern Illinois (Leighton and others, 1948).

These physiographic regions also correspond to differences in soils, land use, and vegetation. The central till plains have fertile grassland soils (mollisols) and are mostly used for row crop agriculture (corn and soybeans), and in the flatter areas, extensive agricultural drainage has been implemented, including the construction of ditches and the installation of drainage tiles. The conditions in the northern region are more varied and therefore so are the soils and land use, though outside of urban areas, row-crop agriculture still dominates. The southern region retains the most forest land cover in study region, and the part of this region in Illinois has less permeable soils and thus less infiltration and recharge and subsequent base flow (Singh, 1971).

Based on maps developed from 1981–2010 PRISM data (http://www.prism.oregonstate.edu/normals/, accessed April 15, 2014), climatic variation in Indiana and Illinois follows mainly north-south gradients, independent of topography because of the low relief. The region has mean annual precipitation of about 36 inches in the northern extreme to almost 50 inches at the southern extreme, mean temperatures of about 47 to about 57 degrees Fahrenheit, a wider range of mean January temperatures of about 20 to about 35 degrees, and a narrower range of mean July temperatures of about 72 to about 79 degrees. Hayden (1988) places most of both States in his "flood climate" region "TsuCpSe*", where "Tsu" indicates storm systems that are barotropic (nonfrontal or convective) and "unorganized" (without tropical cyclones) in the summer, "Cp" indicates that frontal storms are possible throughout the year, and "Se*" indicates seasonal, ephemeral snow cover for 10–50 days per year that may contribute to flooding during winter when rain falls on existing snow. Hayden places the part of the region north of a roughly east-west line crossing at about the southern tip of Lake Michigan (not shown) in his "TsuCpSs**" flood climate region, which contrasts with the "TsuCpSe*" region in that winter snow cover is seasonal rather than ephemeral and exceeds 50 centimeters (cm), so that there may be substantial spring snowmelt flooding.

The initial regions within each State into which the stations were grouped for analysis were those used in the regional flood-frequency studies carried out recently in Illinois (Soong and others, 2004) and Indiana (Knipe and Rao, 2005). The regions determined by Soong and others (2004) for Illinois regions combine major river basins based on physiographic and hydrologic characteristics. The regions determined by Knipe and Rao (2005) for Indiana are based on statistical analyses of physiographic and hydrologic similarity. Because many gages that were used in the previous flood-frequency studies were crest-stage gages that measure only peak stages rather than providing a continuous record and because a number of stations were removed based on the stationarity and imperviousness tests described above, the number of stations available in many of the flood-frequency regions were too few to obtain a meaningful set of regional FDC equations. As a result, various combinations of the flood-frequency regions were tested to find combinations that were advantageous in having a sufficient number of stations and range of drainage areas and could reasonably be expected to be homogeneous considering the physiography of the regions. The latter criterion was tested by comparing the error resulting from regressions on the combined regions to the regression error obtained by keeping the regions separate. The minimum number of stations in a region was constrained by the rule of thumb that 10–15 stations are needed per basin characteristic used in the regression equations (USGS training course SW1523, Regionalization of Surface Water Statistics, February 23–27, 2009, written commun., 2009).

The testing of the various options for flow-duration regions resulted in the proposed regions for Indiana and Illinois presented in figures 1 and 2, respectively. These regions correspond to the flood-frequency regions of Knipe and Rao (2005) and Soong and others (2004) as given in table 6. A small area at the extreme southern tip of Illinois (Soong and others [2004], region 7) was excluded from the results because only two of four stations in this physiographically distinct region passed the stationarity test, and these appeared as outliers in the regional regressions when combined with other nearby regions. For the purpose of the analyses completed for this study and presented in tables 2–5, the

Table 6. Flood-frequency regions approximately corresponding to flow-duration regions in this study.

Flow-duration region	Corresponding flood-frequency regions
Indiana region 1	Knipe and Rao (2005) regions 2 and 3 and part of region 1
Indiana region 2	Knipe and Rao (2005) region 4 and part of region 1
Indiana region 3	Knipe and Rao (2005) regions 5, 6, 7, and 8
Illinois region 1	Soong and others (2004) regions 1 and 2
Illinois region 2	Soong and others (2004) region 3
Illinois region 3	Soong and others (2004) regions 4, 5, and 6

region of a streamgage was defined as the region that contained most of the streamgage drainage basin, which is not necessarily the location of the streamgage itself. The regional boundaries shown in figures 1 and 2 were drawn to follow major basin divides to minimize the number of streamgages with basins in more than one region but eliminating all multiregion basins was not possible for the larger drainage basins.

Regression

The smallest positive daily flow published by the USGS in Indiana and Illinois is 0.01 ft^3/s; on a day when the flow averages less than 0.005 ft^3/s, the published value is zero (Jon Hortness and Donald Arvin, oral commun., 2009). For the analyses in this study, these zero values, which appear in most regions for at least some quantiles and in one region for quantiles as common as the 70-percent exceedance probability, are considered to be "censored" in the sense of being below the detection limit of the measurement system. Because log-transformation of discharge quantiles is usually required to obtain linear and homoscedastic (constant error variance) regression fits and because linear least-squares regression requires continuously varying predictor and predictand variables regardless of log-transformation, zero quantiles violate these conditions on linear least-squares regression. The final equations were therefore obtained by censored regression, a generalization of least-squares regression applicable to censored data, which is solved by maximum likelihood estimation and simultaneously provides a linear fit to the noncensored data and a prediction of the data that should be censored (Helsel, 2005). The censored regression computations were carried out using the *survReg* function provided as part of the survival R package (Therneau, 2021).

The basic form of the equations used in the multiple censored regression analysis is

$$\log_{10}Q_p = i + a_1\log_{10}A_1 + a_2\log_{10}A_2 + a_3\log_{10}A_3, \quad (4)$$

which after exponentiation becomes,

$$Q_p = 10^i A_1^{a_1} A_2^{a_2} A_3^{a_3}, \quad (5)$$

where $\log_{10}x$ indicates the base-10 logarithm of x, i is the regression intercept, Q_p is the estimated daily discharge in ft^3/s having exceedance probability p, A_i, i=1,2,3, are the basin characteristics used as explanatory variables, and a_i, i=1,2,3, are the regression coefficients. Note that not all regression equations have three explanatory variables.

The goodness-of-fit of the censored regressions was computed by using a mean-square residual statistic (MSR) developed to account for the presence of censored values, which was computed as

$$MSR = \frac{1}{N}\sum_{i=1}^{n} w_i e_i^2, \quad (6)$$

where N is the number of stations included in the regression, w_i is the weight of the ith station, and e_i is its residual or error. The weight w_i was computed as

$$w_i = Nn_i / \sum_{j=1}^{N} n_j, \quad (7)$$

where n_i is the number of years of record at the ith station, and the residual e_i was computed as

$$e_i = \begin{cases} y_i - \hat{y}_i \text{ when both } y_i \text{ and } \hat{y}_i \text{ are uncensored} \\ y_i - y_c \text{ when only } \hat{y}_i \text{ is censored} \\ y_c - \hat{y}_i \text{ when only } y_i \text{ is censored} \\ 0 \text{ when both } y_i \text{ and } \hat{y}_i \text{ are censored,} \end{cases} \quad (8)$$

where y_i is base-10 logarithm of the observed discharge value at the ith station, $\hat{y}_i$ is the regression estimate of the value of the base-10 logarithm of the observed discharge value at the ith station, and y_c is the base-10 logarithm of the discharge censoring level, so here $y_c=\log_{10}(0.10)=-2$. Notice that in the case of all uncensored observed and estimated values and unit weights, the MSR reduces to the mean square error (MSE).

Because the residuals are defined in terms of the logarithms of the flows, the numerical value of MSR may be difficult to interpret. To aid in interpretation, in the results tables and figures, the MSR is presented as a percent error computed as

$$MSR\% = 100\left\{e^{\left[(\ln 10)^2 MSR\right]} - 1\right\}^{1/2}, \quad (9)$$

according to equations (33) and (34) of Eng and others (2009). This percent error value is the coefficient of variation (that is, the standard deviation divided by the mean) of a lognormal random variable whose variance is given by the MSR value (compare Benjamin and Cornell, 1970, p. 266, equation 3.3.33). When working directly with the base-10 logarithms, it also is convenient to use the root mean square residual (RMSR) = $MSR^{1/2}$, which, like a standard deviation, has the same units as the quantities from which it was computed, and which reduces to the root mean square error (RMSE) in the absence of censoring.

Before the censored regression, basin characteristics were transformed as needed by centering (subtraction of the mean), addition of a constant, or exponentiation to make them positive and of a wide range of variation before being used in the regressions. Then optimal combinations of basin characteristics were sought by enumeration using least-squares regression followed by ranking by goodness-of-fit and excluding regressions with a high degree of correlation between the basin characteristics proposed as predictor variables as measured by the variance inflation factor (Helsel and Hirsch, 2002, p. 305). All regressions, least-squares or censored, were performed with FDC quantiles weighted by the number of days in the record normalized to an average value of 1 as shown in equation (7) above, so that stations with long records had more effect on the result than those with shorter records, because the FDC quantiles from the longer records are subject to less sampling variability caused by climate variation for the period of record.

Drainage-Area Ratio Method

When regional FDC regression equations are unavailable and rainfall-runoff modeling is deemed to be infeasible, the usual option for predicting daily flows is to use the drainage-area ratio (DAR) method (U.S. Environmental Protection Agency, 2007a, b; Mohamoud, 2008; Stedinger and others, 1993, p. 18.54–18.55; Parajka and others, 2013, p. 238–239). In the DAR method, daily flows are computed as follows:

$$Q(t)=(A/A_{index})Q_{index}(t), \quad (10)$$

where $Q(t)$ is the daily flow being estimated, A is the area of its drainage basin, $Q_{index}(t)$ is the daily flow at the gaged watershed being used as the predictor (the "index" station), and A_{index} is its drainage area.

To test the accuracy of the FDCs resulting from application of the DAR method, the DAR method was applied in each region following two different approaches. In both approaches, the DAR method was used to estimate one or more daily flow records at each station and estimated FDC quantiles were computed from this estimated record by means of the same method as was used to compute the quantiles from the observed record. The squared residuals between the estimated and observed FDC quantiles were averaged for all stations in each region to compute an MSR value for each quantile. The approaches differ in how index stations were selected for use in estimating daily flow records.

In the first approach, for each streamgage, a collection of index stations was used to compute the estimated daily flow records. Results are presented for "strict" and "broad" criteria for selection of the index stations. The strict criterion requires an interbasin centroid distance of 25 miles or less and a DAR between 0.5 and 2.0. The broad criterion is more relaxed; it requires an interbasin centroid distance of 100 miles or less and a DAR between 0.1 and 10. Only a small percentage of streamgage pairs in a region usually satisfies the strict criterion (usually around 4 percent; see table 7), whereas most satisfy the broad criterion. Therefore, the results from the first approach include a range of the possible errors, ranging from the case when an applicable streamgage is quite near and of similar drainage area to the case where a streamgage is picked almost at random from the region.

In the second approach, termed the "nearest-neighbor centroid approach", for each streamgage whose FDC quantiles are to be estimated, the streamgage record from just the one basin whose centroid was nearest to the centroid of the basin whose FDC quantiles are being estimated was used to

Table 7. Fraction of gage pairs satisfying interbasin centroid distance and drainage-area ratio criteria in each flow-duration region.

[The strict criterion requires an interbasin centroid distance of 25 miles or less and a drainage area ratio (DAR) between 0.5 and 2.0. The broad criterion requires an interbasin centroid distance of 100 miles or less and a DAR between 0.1 and 10]

Range of exceedance probabilities (percent)	Number of stations	Number of gage pairs	Fraction of gage pairs satisfying strict criterion	Fraction of gage pairs satisfying broad criterion
		Indiana region 1		
20–99.9	23	253	0.047	0.700
0.1–10	30	435	0.041	0.630
		Indiana region 2		
20–99.9	49	1,176	0.045	0.639
0.1–10	60	1,770	0.038	0.656
		Indiana region 3		
20–99.9	37	666	0.039	0.476
0.1–10	50	1,225	0.037	0.518
		Illinois region 1		
20–99.9	25	300	0.090	0.710
0.1–10	35	595	0.057	0.578
		Illinois region 2		
20–99.9	48	1,128	0.043	0.658
0.1–10	55	1,485	0.044	0.653
		Illinois region 3		
20–99.9	58	1,653	0.015	0.343
0.1–10	69	2,346	0.016	0.353

compute an estimated daily flow record by means of the DAR method. A wider range of drainage-area ratios is possible with this approach than when the choice of an index station is constrained according to some criterion as in the first approach.

Results and Discussion

The values of the regression coefficients computed by the censored regression technique for the selected basin characteristics and for drainage area alone for each FDC quantile in each study region are presented in tables 8–13 (available at https://doi.org/10.3133/sir20145177). The minimum and maximum values of each of the basin characteristics used for each flow regime (low- or high-flow quantiles) in each region are presented in table 14. The accuracy of the results will decrease, perhaps substantially, if the regression equations are used to estimate FDC quantiles for basins whose characteristics are outside the bounds of the minimum and maximum values. The centroid latitude (CLat) and centroid longitude (CLon) characteristics are included in table 14 only for reference, because the geographic extent of each region is already defined in figures 1 and 2.

Table 14. Minimum and maximum values of basin characteristics used in the regressions for low and high-flow quantiles in each flow-duration region.

[mi, miles; PRISM, Parameter-elevation Regressions on Independent Slopes Model; in, inches; QSS_PermB, index of permeability of Quaternary surface sediments; cm, centimeters; STATSGO, State Soil Geographic Database; in/hr, inches per hour; PermBXThick, index of permeability of Quaternary surface sediments multiplied by their thickness; Low flow, 99.9–20-percent exceedance probability quantiles; High flow, 10–0.01-percent exceedance probability quantiles; "–", basin characteristic not used for this region and flow regime]

Region	Range of exceedance probabilities (percent)	Minimum ("min") or maximum ("max") values	Drainage area (DA), mi²		Centroid latitude (CLat), degrees North		Centroid longitude (CLon), degrees West		Sep-Oct-Nov monthly average PRISM precipitation (SON.Precip), in.	
			Low flow	High flow	Low flow	High flow	Low flow	High flow	Low flow	High flow
IN 1	99.9–70	min	6.80	6.80	--	--	--	--	--	--
IN 1	99.9–70	max	4,930	4,930	--	--	--	--	--	--
IN 1	60–0.1	min	6.80	6.80	--	--	--	--	--	--
IN 1	60–0.1	max	4,930	4,930	--	--	--	--	--	--
IN 2	99.9–20	min	2.99	2.99	--	--	--	--	--	--
IN 2	99.9–20	max	4,680	4,680	--	--	--	--	--	--
IN 2	10–0.1	min	2.99	2.99	39.32637	39.32637	--	--	--	--
IN 2	10–0.1	max	4,680	4,680	40.57001	40.57001	--	--	--	--
IN 3	99.9–20	min	2.99	2.99	--	--	--	--	2.795	2.779
IN 3	99.9–20	max	4,070	4,070	--	--	--	--	3.564	3.564
IN 3	10–0.1	min	2.99	2.99	--	--	--	--	--	--
IN 3	10–0.1	max	4,070	4,070	--	--	--	--	--	--
IL 1	99.9–0.1	min	12.1	12.1	--	--	--	--	--	--
IL 1	99.9–0.1	max	2,549	6,366	--	--	--	--	--	--
IL 2	99.9–95	min	5.70	6.91	--	--	--	--	--	--
IL 2	99.9–95	max	2,618	5,094	--	--	--	--	--	--
IL 2	90–50	min	5.70	5.70	--	--	--	--	--	--
IL 2	90–50	max	2,618	5,094	--	--	--	--	--	--
IL 2	40–0.1	min	5.70	6.91	--	--	87.69787	87.57228	--	--
IL 2	40–0.1	max	2,618	5,094	--	--	89.85162	89.85162	--	--
IL 3	99.9–0.1	min	5.63	5.63	--	--	--	--	2.860	2.860
IL 3	99.9–0.1	max	5,190	5,190	--	--	--	--	3.568	3.631

Basin Characteristics and their Coefficient Values

Drainage area is used in all the multiple regression equations not only because it is typically the basin characteristic with the most explanatory power but also to provide a comparison with the drainage area-only results. For both the drainage area-only and multiple regression equations, the drainage-area coefficient shows a consistent pattern of decreasing from a value greater than one (between 1.1 and 2.8) at high exceedance probabilities (low flows), then passing through the value one, usually at about an exceedance probability of 5 or 10 percent, and ending up at a value between 0.7 and 0.9 for the lowest exceedance probability (highest flow quantile) (fig. 3). At low flows, the drainage-area coefficient is highest, on average, for Illinois region 3 (southern and west-central Illinois) and next highest for Indiana region 1 (southern Indiana) and Illinois region 2 (east-central Illinois), where there also are large fractions of censored discharge values (tables 8, 12, and 13). Similar overall behavior of the dependence of drainage-area coefficients of FDC quantiles on exceedance probability in Illinois, including the low-flow coefficients being larger in southern and central parts of the State, was reported by Singh (1971). Indiana region 1 and Illinois region 3 are also the parts of the study area where the glacial till is thinner or bedrock is exposed. In the southern part of Illinois region the soils are also low in permeability relative to the rest of the state, as noted by Singh (1971). In Illinois region 2 the glacial till is thick but there is little topographic relief and reduced natural channel development, though this is somewhat compensated by artificial drainage ditches. The observed high-flow behavior also is consistent with that of peak flow drainage-area coefficients in this region (Soong and others, 2004; Knipe and Rao, 2005) and more generally throughout the United States, depending on the flood-generation mechanism (Gupta and Dawdy, 1995). The appearance of this pattern of drainage-area coefficient behavior for FDCs in Indiana as well as Illinois with this updated data and a wider range of exceedance probabilities used in this study than was used by Singh (1971) suggests it may be a robust behavior for this type of physiographic and climatic region.

Dec-Jan-Feb monthly average PRISM precipitation (DJF.Precip), in.		QSS_PermB		Available water content of soil, 0–100cm (AWC), cm		Soil drainage index (Drain.Index)		STATSGO soil permeability (STAT.Perm), in/hr		PermBXThick	
Low flow	High flow	Low flow	High flow	Low flow	High flow	Low flow	High flow	Low flow	High flow	Low flow	High flow
--	--	--	--	11.12	11.12	--	--	--	--	--	--
--	--	--	--	19.52	19.52	--	--	--	--	--	--
2.666	2.666	--	--	--	--	--	--	--	--	--	--
3.660	3.660	--	--	--	--	--	--	--	--	--	--
2.033	2.033	1.00	1.00	--	--	--	--	--	--	--	--
2.773	2.773	25.48	25.48	--	--	--	--	--	--	--	--
--	--	--	--	--	--	--	--	0.579	0.576	--	--
--	--	--	--	--	--	--	--	1.627	3.020	--	--
--	--	--	--	--	--	--	--	--	--	--	--
--	--	--	--	--	--	--	--	--	--	--	--
--	--	--	--	--	--	--	--	--	--	46.6	46.6
--	--	--	--	--	--	--	--	--	--	30,040.6	30,040.6
--	--	--	--	--	--	3.36	3.36	--	--	--	--
--	--	--	--	--	--	16.00	15.90	--	--	--	--
--	--	--	--	16.19	16.19	--	--	--	--	26.0	26.0
--	--	--	--	22.16	22.16	--	--	--	--	18,491.6	18,491.6
--	--	--	--	--	--	--	--	0.511	0.511	26.0	26.0
--	--	--	--	--	--	--	--	2.380	2.380	18,491.6	18,491.6
--	--	--	--	--	--	--	--	0.511	0.511	--	--
--	--	--	--	--	--	--	--	2.380	2.380	--	--
--	--	--	--	--	--	--	--	--	--	--	--
--	--	--	--	--	--	--	--	--	--	--	--

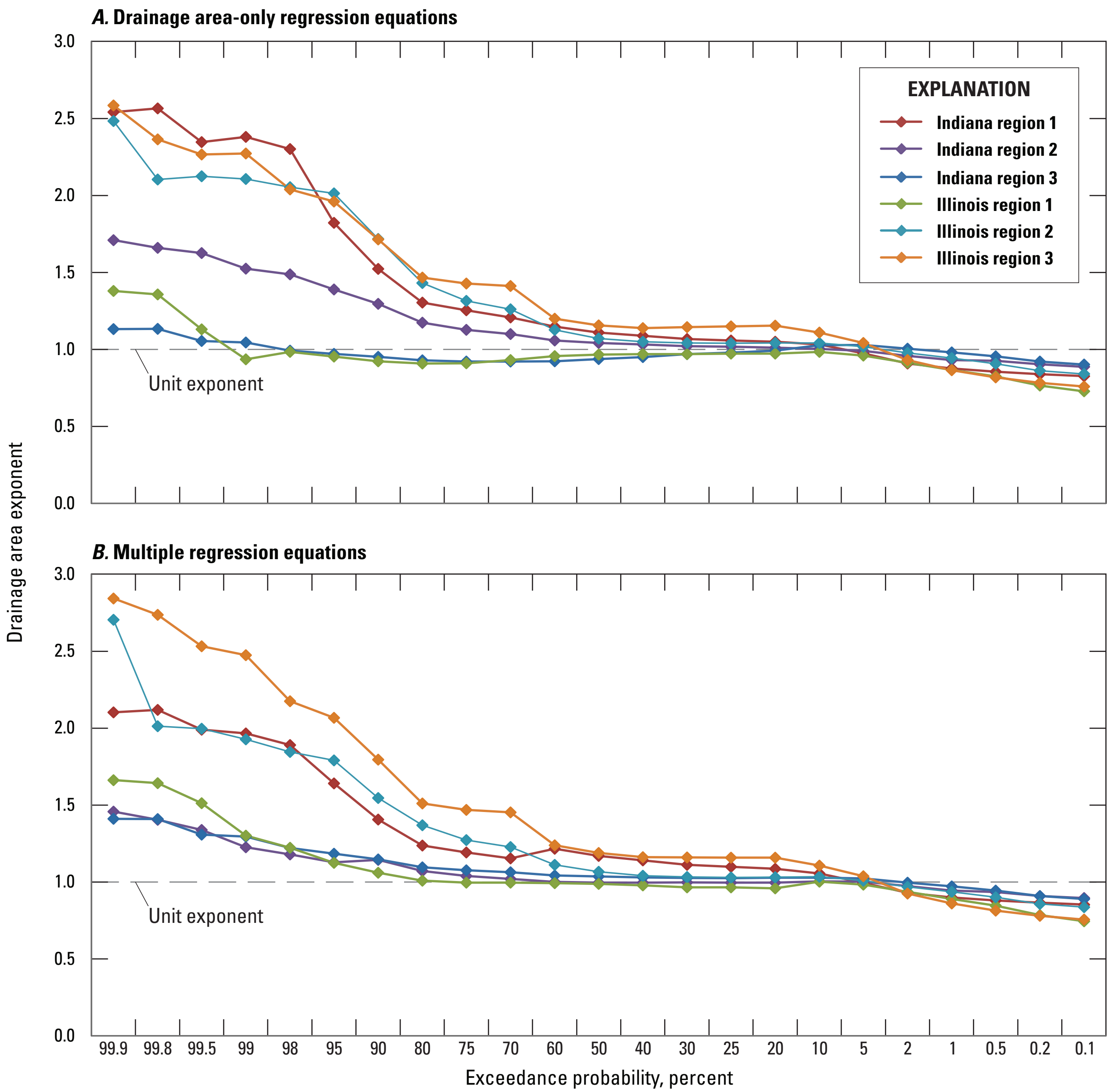


Figure 3. Drainage-area exponents from *A*, drainage area-only and *B*, multiple regression equations.

An important implication of this pattern of dependence of the drainage area coefficient on exceedance probability is that it indicates that smaller streams have, on average, greater variability or flashiness as compared to larger streams. Variability of daily flows is given by the overall slope of the FDC: the larger the slope, the more variable the streamflow (Searcy, 1959). This variability often is quantified using the streamflow variability index V, usually defined as the standard deviation of the base-10 logs of the 5 to 95-percent FDC quantiles at 10-percent intervals, that is,

$$V = \sqrt{\sum_i \left(\log_{10} Q_i - \langle \log_{10} Q_i \rangle\right)/9}, \; i = 5, 15, 25, ..., 95 \text{ percent,} \quad (11)$$

where Q_i is the i-percent daily flow FDC quantile and $\langle X \rangle$ indicates the mean of X (Lane and Lei, 1950; Mitchell, 1957; Searcy, 1959). To see the effect of the drainage area coefficient on the FDC slope, consider two basins, one with a basin area of 10 mi^2 and another with a drainage area of 1,000 mi^2 that are otherwise identical. If the 0.1-percent exceedance probability drainage area coefficient is 0.5, then using equation (5), the 0.1-percent (high-flow) quantile $Q_{0.1}^{(2)}$ for the larger basin will be $Q_{0.1}^{(2)}/Q_{0.1}^{(1)}$=$1{,}000^{0.5}/10^{0.5}$=10 times as large, whereas if for the 10-percent quantile the coefficient is 1.0, then the ratio of quantiles will be 100, and if for the 99.9-percent (low-flow) quantile the coefficient is 2.0, then the ratio of quantiles will be 100^2=10,000. So the effect could be quite dramatic at the extremes. Lane and Lei (1950) and Searcy (1959) commented that the variability index could be expected to decrease with increasing drainage area, and Mitchell (1957) includes a graph showing this behavior in the dependence of V on drainage area for streamflow records in Illinois.

In the multiple regression equations, there are various additional basin characteristics that appear in the chosen equations, including geographic indicators (centroid latitude, CLat, and centroid longitude, CLon), measures of seasonal precipitation (September-October-November and December-January-February average precipitation from PRISM), three soil property indicators (available water content, AWC, soil drainage index, Drain.Index, and soil permeability, STAT.Perm), an index of Quaternary surface sediment permeability (QSS_PermB), and an index of the thickness and character of the surficial geology (PermBXThick). The centroid latitude CLat appears only in the high-flow quantile equations in Indiana region 2 along with drainage area and STAT.Perm, and has a negative coefficient, indicating that greater high flows are found as latitude decreases, that is, to the south, which agrees with the trend in precipitation. The STAT.Perm coefficient value in these equations is negative, indicating that higher soil permeability suppresses high flows in this region, presumably by increasing infiltration. The centroid longitude CLon appears only in the moderate to high-flow equations for Illinois region 2 along with drainage area and, like CLat in Indiana region 2, STAT.Perm. The centroid longitude CLon in this region has a negative coefficient, indicating that the corresponding flows decrease as west longitude increases, that is, to the west. It is not clear what physical property is behind this result; one possibility is that although precipitation gradients are mainly north-south, there is some decrease in the westward direction as well. The STAT.Perm coefficient at moderate to high flows in Illinois region 2 switches from positive to negative as the flows increase, which agrees with the Indiana region 2 coefficient at high flows and indicates that the higher infiltration indicated by higher permeability tends to increase the magnitude of lower flows.

The equations in Illinois region 2 that were not discussed previously, when the use of CLon was being discussed, are those corresponding to moderate to low flows. These equations use PermBXThick and drainage area, along with available water content (AWC) for the lowest flows and STAT.Perm for the moderately low flows. PermBXThick for these equations has a positive coefficient of increasing magnitude as the exceedance probability increases, indicating that deeper and more permeable surficial geologic sediments are conducive to higher low flows. Available water content (AWC) likewise has positive coefficients whose magnitudes increase as the exceedance probability increases. This behavior of AWC coefficients is repeated in the other region where it is used, that is, Indiana region 1 (southern Indiana), where for moderate to low flows, it is the only basin characteristic other than drainage area. The other place PermBXThick is used is for high flows in the Indiana region 3 equations, where it is the only characteristic other than drainage area and its coefficient is negative and increasing in magnitude as exceedance probabilities increase, indicating that deeper and more permeable sediments tend to decrease high flows. The moderate to low-flow equations in Indiana region 3 use September-October-November average precipitation (SON.Precip) along with drainage area. The coefficient of SON.Precip is positive and of increasing magnitude as the exceedance probability increases. Since September-October-November includes the period of lowest flows, the sign and trend in magnitude of this coefficient seem reasonable. The equations for all exceedance probabilities in Illinois region 3 also use SON.Precip, in addition to drainage area. The SON.Precip coefficients have a trend that is opposite of that of Indiana region 3, decreasing from small positive values to larger negative values with increasing exceedance probability. Based on the geographic distribution of fall precipitation in Illinois, which increases to the south (not shown), this behavior indicates that SON.Precip is acting in this equation as a proxy for another variable that depends on geographic location. Because southern Illinois has generally less permeable soils and therefore less infiltration, recharge, and base flow (Singh, 1971) but larger SON.Precip values, the coefficient of SON.Precip is increasingly negative with higher exceedance probability. The remainder of the equations in Indiana region 1 (those for moderate to high flows) use drainage area and December-January-February average precipitation (DJF.Precip). DJF.Precip has a positive coefficient of increasing magnitude as exceedance probability increases, indicating, as seems reasonable, that winter precipitation is less important for higher flows. The low- to moderate-flow quantiles for Indiana region 2 also use DJF.Precip, with the coefficient decreasing with increasing exceedance probability, indicating that the magnitude of lower flows is decreased by more winter precipitation. This apparently anomalous behavior, similar to that of SON.Precip in Illinois region 3, indicates that winter precipitation is acting as a proxy for some other basin property. Also appearing in the low- to moderate-flow quantiles for Indiana region 2 is the index of Quaternary sediment permeability (QSS_PermB), the coefficients of which increase with increasing exceedance probability, indicating that more permeable sediments increase the magnitude of low flows.

The remaining region to be discussed is Illinois region 1. All the multiple regression equations in this region use the drainage index (Drain.Index) and drainage area. The coefficient of Drain.Index is large and positive for low flows and decreases with decreasing exceedance probability, becoming near zero for moderate flows and negative for high flows. The behavior of the Drain.Index coefficient indicates that better soil drainage increases the magnitude of low flows but decreases high flows, presumably because higher values of the Drain.Index implies more infiltration and recharge and less surface runoff.

Accuracy of the Estimation Equations

A comparison of the accuracy of the different FDC quantile estimation methods as measured by the percent error (MSR%) is shown in figures 4 and 5. The most striking feature of these results is that for all methods the percent error varies across a wide range as a function of exceedance probability, with the minimum of 10 to 20 percent at moderately high flows of around a 5- to 10-percent exceedance probability, where, according to figure 1 and tables 8–13, the drainage area coefficient is approximately one for the regression methods. The error increases slightly to 17 to 38 percent from this minimum for the highest flows, and increases moderately toward lower flows until around the 90- to 99-percent exceedance probabilities, depending on the region and method, where the MSR usually exceeds 100 percent, even for the selected multiple regression equations. The maximum MSR for the multiple regression method ranges from 100 to 745 percent.

Overall, as expected, the selected multiple regression equations give the best fit and the DAR method using broad selection criteria the worst fit. The DAR method with strict selection criteria (which, as discussed in the Drainage-Area Ratio Method section, is only occasionally possible to satisfy at a given location) and regression with basin area alone are usually in the middle, though occasionally the strict selection DAR method fits better than even the selected multiple regression equations (for example, Illinois region 3, mid to high flows). Usually the regression equations provide a better fit relative to the DAR method as the difficulty of fitting increases, that is, at the lowest flows. There is not a consistent relation of errors of the basin centroid-based nearest neighbor DAR method to the other DAR methods, though it usually lies between the errors of the strict and broad DAR criteria. An analysis of the median nearest-neighbor centroid distances and median DARs in each region indicates that the average nearest-neighbor DAR errors increase as the nearest neighbor centroid distances and median DARs increase, especially the latter. These results, along with the strong dependence of the DAR method errors on the strict criteria as compared to the broad criteria, indicate that if the DAR method is to be used, great care needs to be taken in selection of the index station. A gaged basin close to the site of interest is helpful, but it does not guarantee good performance; the use of the regional regression equations will usually be better, often much better, especially at higher exceedance probabilities (lower flows). The appearance of usually smaller errors for the regional regression equations with drainage-area alone as compared to the various DAR method and the wide range of drainage area exponents obtained in this study (fig. 3) also indicates that the accuracy of the DAR method could be improved by considering the use of non-unit exponents, as was reported by Emerson and others (2005) and Asquith and others (2006).

Example Application

The Illinois Environmental Protection Agency (IEPA) began sampling for water quality constituents, including nitrate, nitrite, and phosphorus, total suspended solids, dissolved oxygen, temperature and pH, at five stations in the watershed of Indian Creek (fig. 6), a tributary of the Vermilion River in Ford, Livingston, and McLean Counties, Illinois (fig. 2), in May, 2010, as part of a program to monitor the implementation of agricultural practices aimed at reducing nitrogen loadings (Trevor Sample, IEPA, written commun., 2014). The monitoring program also includes the establishment and operation of a USGS streamgage near Fairbury, Illinois (streamgage number 05554300; fig. 6), at which there is continuous flow data beginning July 2011 and continuous nitrate data beginning September 2011.

Daily streamflow is available at the streamgage near Fairbury (fig. 6), but estimates of daily streamflow at the upstream monitoring stations also are needed to convert the sampled nutrient concentrations to loads. Estimating daily FDC quantiles is the first step in the so-called QPPQ method of estimating streamflow (Fennessey, 1994; Smakhtin, 1999; Mohamoud, 2008; Archfield and others, 2010; Straub and Over, 2010, appendix A; Linhart and others, 2012; Stuckey and others, 2012). Although estimating daily streamflow is beyond the scope of this study, estimating the FDCs that are needed to implement the QPPQ method serves as a useful example of the application of the results given in this report.

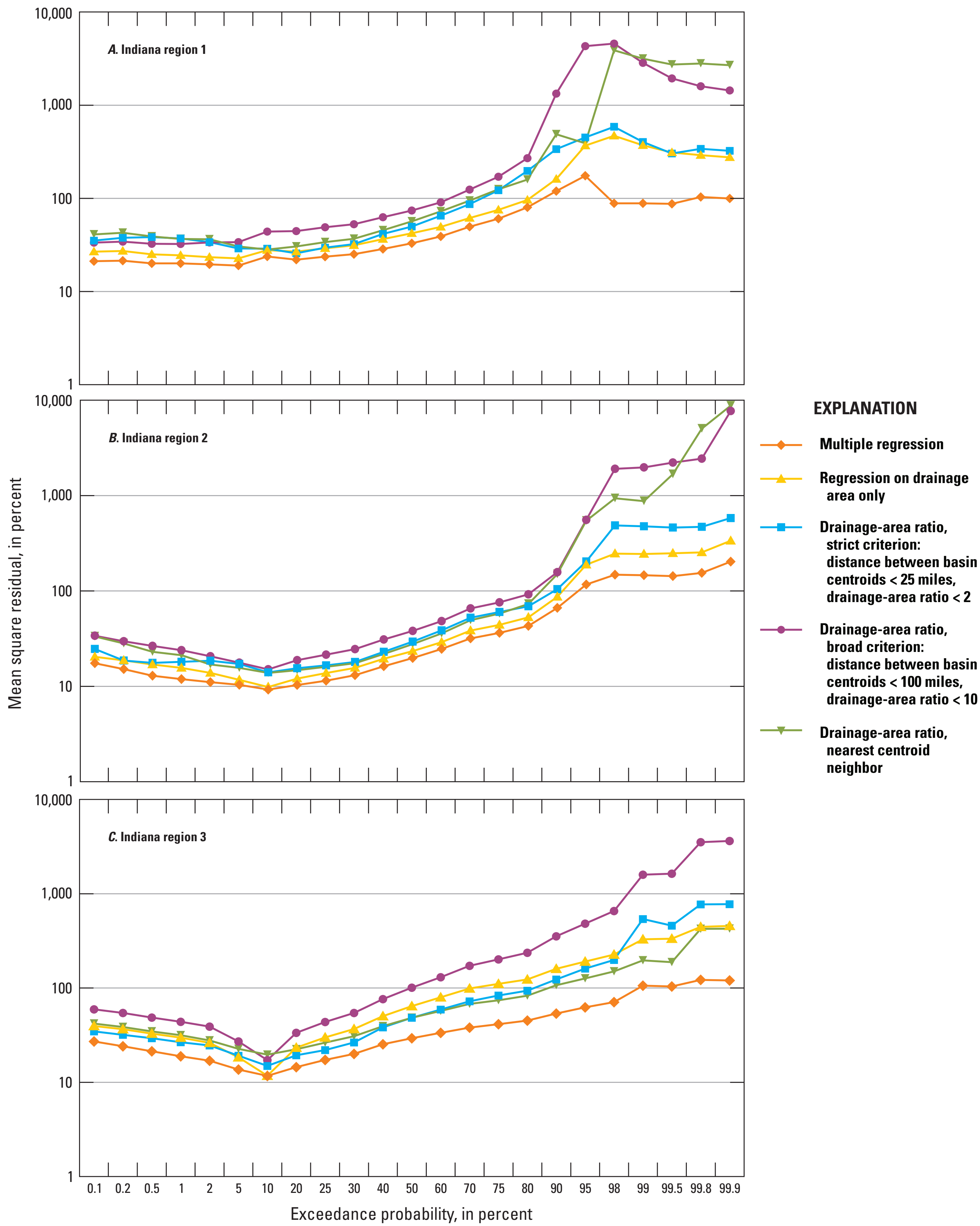


Figure 4. Comparison of goodness-of-fit of flow-duration quantiles estimated by different methods as measured by mean square residual (MSR) for Indiana flow-duration regions *A*, 1; *B*, 2; and *C*, 3.

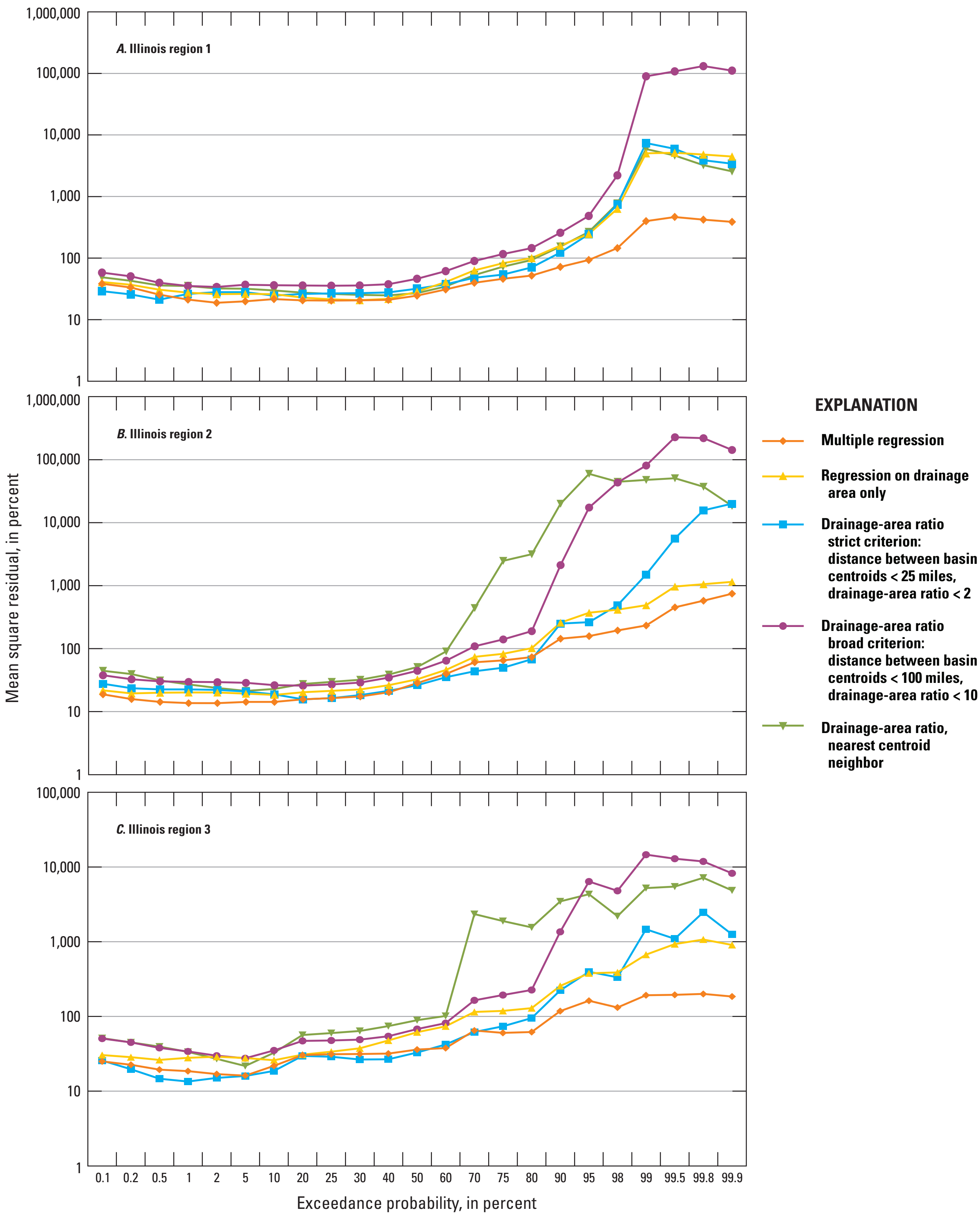


Figure 5. Comparison of goodness-of-fit of flow-duration quantiles estimated by different methods as measured by mean square residual (MSR) for Illinois flow-duration regions *A*, 1; *B*, 2; and *C*, 3.

The first step was to delineate the boundaries of the basins upstream from each of the monitoring stations (including the USGS streamgage) by using Illinois StreamStats (https://streamstats.usgs.gov/ss/, accessed 2014) with adjustments obtained from 12-digit hydrologic unit code (HUC) watershed boundary data from the Watershed Boundary Dataset (U.S. Geological Survey and U.S. Department of Agriculture, Natural Resources Conservation Service, 2013). To provide an example of the degree of repeatability to be expected when computing these quantities, basin characteristics for the next downstream streamgage (05554500; fig. 2) used in this study also were computed. Using the delineated basin boundaries, basin characteristics for each basin were computed from the coverages used in this study. The Indian Creek watershed is located in Illinois flow-duration region 2, so values of drainage area (DA), centroid

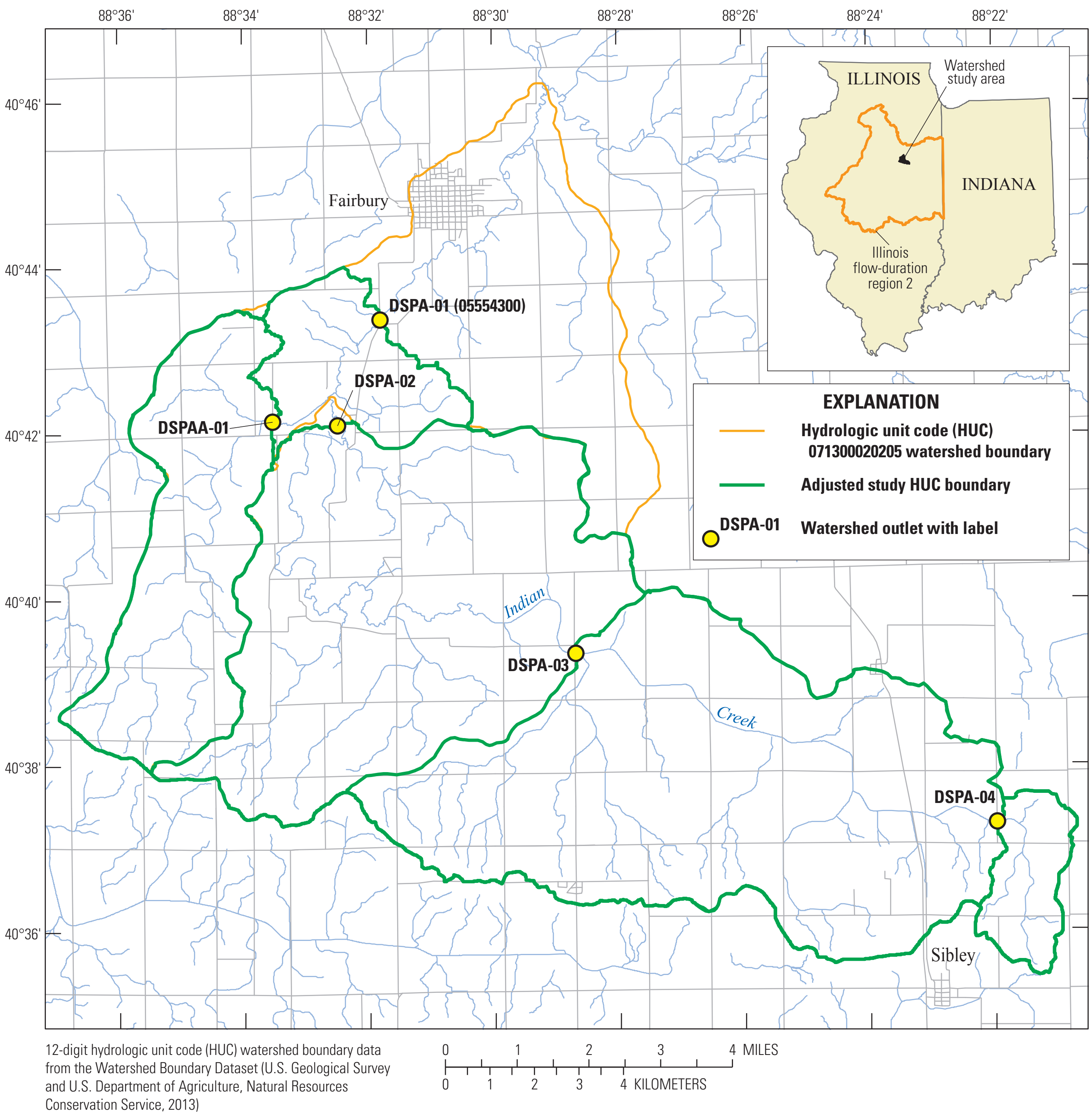


Figure 6. Map of Indian Creek watershed in Ford, Livingston, and McLean Counties, Illinois, showing streamflow and nutrient monitoring stations established by the Illinois Environmental Protection Agency.

longitude (CLon), available water content (AWC), soil permeability (STAT.Perm), and of the index of permeability and thickness of Quaternary sediments (PermBXThick) are needed (table 12). The computation of the AWC and PermBXThick values, which require computation outside of ArcMap, is shown in tables 15 and 16. The computation of the PermBXThick characteristic also is described by equation (2). The complete set of basin characteristics is given in table 17. Both the new and original values are provided for streamgage (05554500). Notice the value of drainage area for station DSPA-04 (2.325 mi^2) is lower than the minimum used in the Illinois region 2 regression equations (6.8 mi^2; see table 14). The estimated quantiles for DSPA-04 should therefore be used only for reference.

With the basin characteristics computed, it is possible to compute the FDC quantiles. These are given in table 18 (available at https://doi.org/10.3133/sir20145177). For example, to use the drainage area-only regression equation to obtain an estimate of the 99-percent exceedance probability FDC quantile at station DSPA-01, the drainage area is read from table 17 as 67.0 mi^2, and the Illinois flow-duration region 2 drainage area-only regression intercept of −5.361 and drainage area coefficient of 2.108 are read from table 12. These are used to obtain an estimate of the base-10 logarithm of the quantile of interest as follows:

$$\log_{10}\left(Q_{0.99}\right)=i+a_1\log_{10}\left(\mathrm{DA}\right)=-5.361+2.106*\log_{10}\left(67.0\right)=-1.782,$$

so that the quantile estimate is $Q_{0.99}=10^{-1.782}=0.0165$ ft^3/s or 0.02 ft^3/s when rounded to hundredths of ft^3/s, as observed daily flows usually are and as is done in table 18.

To use the multiple regression equation to estimate this same quantile for this same basin, the rest of the basin characteristics are obtained from table 17 and the coefficients from table 12. The quantile estimate is then computed as follows:

$$\log_{10}\left(Q_{0.99}\right)=i+a_1\log_{10}\left(\mathrm{DA}\right)+a_2\log_{10}\left(\mathrm{PermBXThick}\right)+a_3\log_{10}\left(10^{(\mathrm{AWC}-18.3287)}\right),$$

so, inserting the numerical values,

$$\log_{10}\left(Q_{0.99}\right)=-10.251+1.926*\log_{10}\left(67.0\right)+1.631*\log_{10}\left(137.9\right)+0.247*\log_{10}\left(3.962\right)=-3.096,$$

so that the multiple regression quantile estimate is $Q_{0.99}=10^{-3.096}=0.0008$, which is rounded to 0.00 in table 18.

The estimated quantiles for this example are plotted in figure 7 for the drainage area-only and the multiple regression equation estimates. As the example computation suggests, the major difference between the two sets of estimates is for the larger exceedance probabilities (lower flows), where multiple regression quantile estimates are as much as two orders of magnitude (a factor of 0.01) smaller than the drainage area-only estimates. Because both sets of equations were fitted on the same set of streamflow data, this difference would not arise on average, but results from the particular set of characteristics of these basins. It is also less surprising to see such a difference at the lower flows, because, as figures 4 and 5 show, the uncertainty for the lower flows can be substantial. In particular, for the 99-percent quantile in Illinois region 2 where the example basins lie, the mean-square residual (MSR) values (equations 6–9) are 491 and 234 percent for the drainage area-only and multiple regression equations, respectively, or, as a root mean square residual (RMSR) = MSR$^{1/2}$ in $\log_{10}$ units, the values are 0.780 and 0.594, respectively (table 12). The RMSR values constitute approximate standard errors on the $\log_{10}$ quantile estimates, which would then be written as -1.782 ± 0.780 and -3.096 ± 0.594. Because $-1.782-0.780=-2.562$ and $-3.096+0.594=-2.502$, these intervals do overlap somewhat and the magnitude of their difference is seen to be in line with the uncertainties as expressed in the MSR values. By way

Table 15. Computation of available water content (AWC) values for selected basins in Illinois flow-duration region 2.

[mi^2, square miles; cm, centimeters]

Station identifer/ AWC class	Areas of available water content of soil, 0–100 cm (AWC), classes (mi^2)						Total Area (mi^2)	Area-weighted AWC
	12 cm	16 cm	18 cm	19 cm	20 cm	21 cm		
DSPA-01 (05554300)	0.00	0.00	35.97	0.00	31.05	0.00	67.0	18.927
DSPA-02	0.00	0.00	24.03	0.00	28.36	0.00	52.4	19.083
DSPA-03	0.00	0.00	15.94	0.00	13.50	0.00	29.4	18.917
DSPA-04	0.00	0.00	2.33	0.00	0.00	0.00	2.33	18.000
DSPAA-01	0.00	0.00	7.54	0.00	2.69	0.00	10.2	18.526
05554500 (newly computed values)	162.87	16.36	259.66	33.14	31.43	77.08	581	16.824

Table 16. Computation of PermBXThick values for selected basins in Illinois flow-duration region 2.

[mi^2, square miles; ft, feet; PermBXThick, index of permeability of Quaternary surface sediments multiplied by their thickness]

Station identifier	Total area (mi^2)	Areas of classes of surficial Quaternary sediments (mi^2)				Index of Quaternary sediment permeability (QSS_PermB)	Areas of classes of total thickness of Quaternary sediments (mi^2)					Area-weighted average thickness (QSS_Thick) (ft)	PermBXThick
		Coarse-grained stratified sediment (code 101)	Fine-grained stratified sediment (code 102)	Till (code 103)	Bedrock (code 105)		0-50 ft (code 201)	50-100 ft (code 202)	100-200 ft (code 203)	200-400 ft (code 204)	400-600 ft (code 205)		
DSPA-01 (05554300)	67.0	0.00	0.00	67.02	0.00	1.000	1.88	36.25	14.60	14.29	0.0	137.9	137.9
DSPAA-01	10.2	0.00	0.00	10.23	0.00	1.000	0.784	5.106	1.967	2.375	0.0	137.8	137.8
DSPA-02	52.4	0.00	0.00	52.39	0.00	1.000	0.383	27.46	12.63	11.92	0.0	143.9	143.9
DSPA-03	29.4	0.00	0.00	29.44	0.00	1.000	0.00	10.53	8.26	10.65	0.0	177.4	177.4
DSPA-04	2.33	0.00	0.00	2.325	0.00	1.000	0.00	0.00	0.00	2.325	0.0	300.0	300.0
05554500 (newly computed values)	581	47.3	110.9	422.2	0.00	9.074	73.6	107.9	217.1	182.0	0.0	167.2	1517.4

Table 17. Basin characteristics for estimation of flow-duration curves for selected basins in Illinois flow-duration region 2.

[mi^2, square miles; in/hr, inches per hour; cm, centimeters]

Station identifier	Drainage area, mi^2	Centroid longitude (CLon), degrees West	$10^{CLon-89.0207}$	Soil permeability (STAT.Perm) (in/hr)	Available water content (AWC), cm	$10^{AWC-18.3287}$	Index of surficial Quaternary sediment permeability and thickness (PermBXThick)
DSPA-01 (05554300)	67.0	88.49427	0.2976	0.9379	18.93	3.962	137.9
DSPAA-01	10.2	88.58125	0.3635	0.7722	18.53	1.573	137.8
DSPA-02	52.4	88.47360	0.2837	0.9626	19.08	5.675	143.9
DSPA-03	29.4	88.43434	0.2592	0.8796	18.92	3.879	177.4
DSPA-04	2.33	88.35797	0.2174	0.5670	18.00	0.469	300.0
05554500 (values used to compute regression equations in this study)	581	88.36744	0.2222	1.1703	16.83	0.0315	1,519.4
05554500 (newly computed values)	581	88.36721	0.2221	1.1612	16.82	0.0313	1,517.4

of comparison, the 1-percent exceedance probability FDC quantile estimates at DSPA-01 are 514 and 593 ft^3/s with regional MSR values of 20.2 and 13.7 percent for the drainage area-only and multiple regression estimates, respectively, or, in base-10 log units, 2.711 ± 0.087 and 2.773 ± 0.059, respectively (table 18). As 593/514=1.154 or a 15.5-percent difference, and $2.711 + 0.087 = 2.799$ and $2.773 - 0.059 = 2.715$, it is clear that these estimates also lie within the uncertainty bounds as estimated by these MSR-based quantities.

At the scale of the graphs in figure 7, the match between the observed and estimated FDCs at streamgage 05554500 appears to be acceptably good; for one thing, their differences are better for the higher flows but worse for the lower flows, agreeing with the uncertainties measured by MSR and the differences between drainage area-only and multiple regression estimates at DSPA-01 discussed above. For a more quantitative analysis of these differences, data are included in table 18 providing, for the quantile values in base-10 log units the differences between the observed and estimated FDC quantiles, and for the values in ft^3/s units the percent differences between the observed and estimated FDC quantiles. Following these data are the applicable RMSR or percent MSR values from table 12. It is evident that the observed and estimated FDC quantile differences generally lie within the expected uncertainties as expressed by the RMSR or percent MSR values, except at the 99-percent exceedance probability quantile, which is the lowest FDC quantile for which the observed FDC is not censored. Unlike the observations, the regional equations do not produce estimates that indicate censoring at this streamgage; a partial explanation of this disagreement is that the MSR values indicate very large uncertainties for these quantiles.

Another characteristic to note in the graph of the estimated FDC quantiles at these example basins is the greater steepness of the FDCs for the smaller basins at the bottom of the graphs as compared to the larger basins at the top. This difference exemplifies the decreases of FDC slope or streamflow variability with increasing drainage area as discussed in the Basin Characteristics and their Coefficient Values section, although among these basins there are differences in other basin characteristics that also affect the shapes of the FDCs.

Summary

Regional regression equations for estimating FDC quantiles at the 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 25, 30, 40, 50, 60, 70, 75, 80, 90, 95, 98, 99, 99.5, 99.8, and 99.9-percent exceedance probabilities for rural, unregulated streams in Indiana and Illinois with temporally stationary records have been computed as a function of GIS-derived basin characteristics by using linear regression techniques and streamflow data through water year 2007. The techniques used account for censored values below 0.01 ft^3/s, which are observed at exceedance probabilities as low as 70 percent (that is, occurring at least 30 percent of the time). The basin characteristics used are suitable for automatic computation by the USGS Web-based application, StreamStats, and are available for all U. S. Environmental Protection Agency (EPA) Region V states and the larger Great Lakes area, with some specific local exceptions. Indiana and Illinois were each divided into three regions, and a different set of equations was computed for each region.

The error of estimation of the FDC quantiles varies somewhat by region but more strongly by exceedance probability, with a minimum error of 10 to 20 percent at an exceedance probability of 5 or 10 percent, but rising to 17 to 38 percent at the high-flow end of the FDCs (the 0.1-percent quantile) and 100 to 745 percent at the low-flow end. For comparison, errors of estimation also were computed for FDC quantiles estimated by linear regression on drainage area alone and by using the drainage-area ratio method. The use of these simpler methods usually resulted in higher errors of estimation, except occasionally in the case of the drainage-area ratio method with a strict index station selection criterion, a criterion that is rarely possible to satisfy.

An example application of the estimated equations to one gaged and a few ungaged locations in a watershed in the study area was included to illustrate the steps required, which are the computation of the basin characteristics and, using those together with the estimated equations, the FDC quantiles and their uncertainties. This application also provides an example of how FDC quantile estimates based on the drainage-area only and multiple regression equations and their uncertainties may differ.

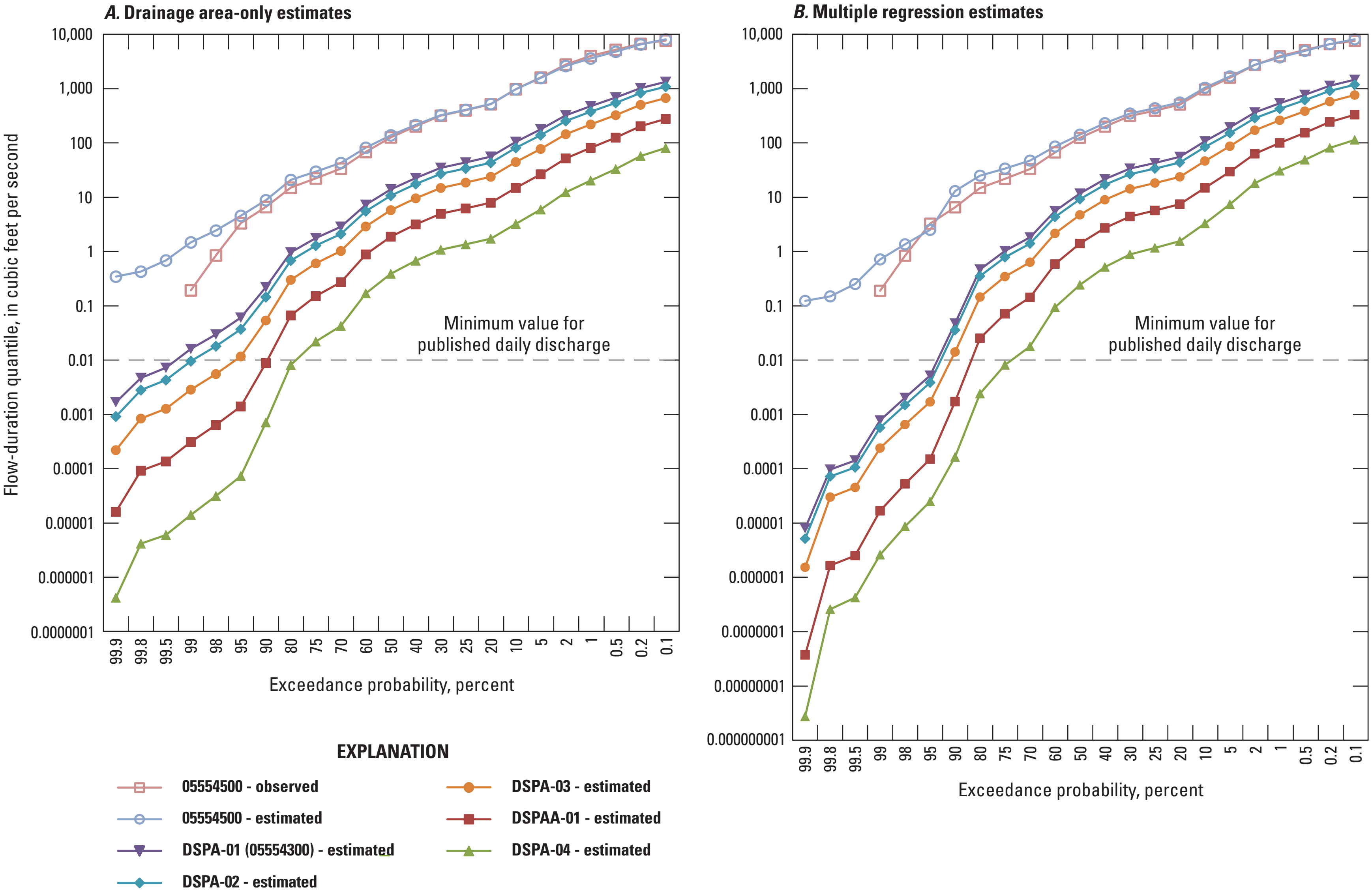


Figure 7. Estimates of flow-duration curves for selected basins in Illinois flow-duration region 2: *A*, drainage-area only estimates; *B*, multiple regression estimates.

References Cited

Alley, W.M., 1988, Using exogenous variables in testing for monotonic trends in hydrologic time series: Water Resources Research, v. 24, p. 1955–1961.

Archfield, S.A., Vogel, R.M., Steeves, P.A., Brandt, S.L., Weiskel, P.K., and Garabedian, S.P., 2010, The Massachusetts Sustainable-Yield Estimator—A decision-support tool to assess water availability at ungaged stream locations in Massachusetts: U.S. Geological Survey Scientific Investigations Report 2009–5227, 41 p. plus CD-ROM.

Arihood, L.D., and Glatfelter, D.R., 1991, Method for estimating low-flow characteristics of ungaged streams in Indiana: U.S. Geological Survey Water-Supply Paper 2372, 18 p.

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

Benjamin, J.R., and Cornell, C.A., 1970, Probability, statistics, and decision for civil engineers: New York, McGraw-Hill, Inc., 684 p.

Bonnin, G.M., Martin, D., Lin, B., Parzybok, T., Yekta, M., and Riley, D., 2006, Precipitation-frequency atlas of the United States: NOAA Atlas 14, v. 2, 71 p.

Bonta, J.V., and Cleland, B., 2003, Incorporating natural variability, uncertainty, and risk into water quality evaluations using duration curves: Journal of the American Water Resources Association, v. 39, no. 6, p. 1481–1496.

Cleland, B.R., 2002, TMDL development from the "bottom up" – Part II—Using duration curves to connect the pieces: National TMDL Science and Policy—WEF Specialty Conference, Phoenix, Arizona, p. 687–697.

Cowardin, L.M., Carter, V., Golet, F.C., and LaRoe, E.T., 1979, Classification of wetlands and deepwater habitats of the United States: Washington, D.C., U.S. Department of the Interior, Fish and Wildlife Service, FWS/OBS–79/31.

Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J., and Pasteris, P.A, 2008, Physiographically-sensitive mapping of temperature and precipitation across the conterminous United States: International Journal of Climatology, v. 28, p. 2031–2064.

Emerson, D.G., Vecchia, A.V., and Dahl, A.L., 2005, Evaluation of drainage-area ratio method used to estimate streamflow for the Red River of the North Basin, North Dakota and Minnesota: U.S. Geological Survey Scientific Investigations Report 2005–5017, 13 p.

Eng, K., Chen, Y.-Y., and Kiang, J.E., 2009, User's guide to the weighted-multiple-linear-regression program (WREG version 1.0): U.S. Geological Survey Techniques and Methods, book 4, chap. A8, 21 p.

Esralew, R., and Smith, S.J., 2010, Methods for estimating flow-duration and annual mean-flow statistics for ungaged streams in Oklahoma: U.S. Geological Survey Scientific Investigations Report 2009–5267, 131 p.

Fennessey, N.M., 1994, A hydro-climatological model of daily streamflow for the northeast United States: Medford, Massachusetts, Tufts University, Ph.D. dissertation, variously paged.

Fennessey, N.M. and Vogel, R.M., 1990, Regional flow duration curves for ungaged sites in Massachusetts: Journal of Water Resources Planning and Management, v. 116, no. 4, p. 530–549.

Flynn, R.H., 2003, Development of regression equations to estimate flow durations and low-flow frequency statistics in New Hampshire streams: U.S. Geological Survey Water-Resources Investigations Report 02–4298, 66 p.

Fowler, K.K., and Wilson, J.T., 1996, Low-flow characteristics of Indiana streams: U.S. Geological Survey Water-Resources Investigations Report 96–4128, 313 p.

Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M., and Tyler, D., 2002, The National Elevation Dataset: Photogrammetric Engineering and Remote Sensing, v. 68, no. 1, p. 5–32.

Gray, H.H., 2000, Physiographic divisions of Indiana: Indiana Geological Survey Special Report 61, 15 p.

Gupta, V.K., and Dawdy, D.R., 1995, Physical interpretations of regional variations in the scaling exponents of flood quantiles: Hydrological Processes, v. 9, p. 347–361.

Hayden, B.P., 1988, Flood climates, *in* Baker, V.R., Kochel, R.C., and Patton, P.C., eds., Flood geomorphology: New York, John Wiley and Sons, p. 13–26.

Helsel, D.R., 2005, Nondetects and data analysis: Statistics for censored environmental data: New York, John Wiley and Sons, 250 p.

Helsel, D.R. and Hirsch, R.M., 2002, Statistical methods in water resources, Techniques of Water-Resources Investigations of the U.S. Geological Survey, book 4, chap. A3, 510 p.

Hirsch, R.M., 1979, An evaluation of some record reconstruction techniques, Water Resources Research, v. 15, p. 1781–1790.

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A., VanDriel, J.N., and Wickham, J., 2007, Completion of the 2001 National Land Cover Database for the conterminous United States: Photogrammetric Engineering and Remote Sensing, v. 73, no. 4, p. 337–341.

Illinois Department of Natural Resources, 1998, Kankakee River Area Assessment, volume 2—Water resources: Champaign, Illinois, Illinois Department of Natural Resources, Office of Scientific Research and Analysis, Illinois State Water Survey, 124 p.

Johnson, S.L., Whiteaker, T., Maidment, D.R., 2009, A tool for automated load duration curve creation: Journal of the American Water Resources Association, v. 45, no. 3, p. 654–663.

Kim, J.G., Engel, B.A., Park, Y.S., Theller, L., Chaubey, I., Kong, D.S., Lim, K.J., 2012, Development of Web-based load duration curve system for analysis of total maximum daily load and water quality characteristics in a waterbody: Journal of Environmental Management, v. 97, p. 46–55.

Knapp, H.V., Terstriep, M.L., Singh, K.P., and D.C. Noel, D.C., 1985, Sangamon River Basin streamflow assessment model—Hydrologic analysis: Illinois State Water Survey Contract Report 368., 74 p.

Knipe, D. and Rao, A.R., 2005, Estimation of peak discharges of Indiana streams by using log Pearson (III) Distribution: Report FHWA/IN/JTRP-2005/1, 194 p.

Koltun, G.F., and Whitehead, M.T., 2002, Techniques for estimating selected streamflow characteristics of rural, unregulated streams in Ohio: U.S. Geological Survey Water-Resources Investigations Report 02–4068, 50 p.

Lane, E.W., and Lei, K., 1950, Streamflow variability: American Society of Civil Engineers, Transactions, v. 115, p. 1084–1134.

Leighton, M.M., Ekblaw, G.E., and Horberg, L., 1948, Physiographic divisions of Illinois: The Journal of Geology, v. 56, no. 1, p. 16–33.

Linhart, S.M., Nania, J.F., Sanders, C.L., Jr., and Archfield, S.A., 2012, Computing daily mean streamflow at ungaged locations in Iowa by using the Flow Anywhere and Flow Duration Curve Transfer statistical methods: U.S. Geological Survey Scientific Investigations Report 2012–5232, 50 p.

Maidment, D.R., ed., 2002, Arc Hydro—GIS for water resources: Redlands, Calif., ESRI Press, 203 p.

Markus, M., Knapp, H.V., Fleger, A., McConkey, S., and Thomas, W.O., 2013, Episodic change analysis of the annual flood peak time series for a flood insurance study: Journal of Hydrologic Engineering, v. 18, no. 1, p. 85–91.

Miller, D.A. and White, R.A., 1998, A conterminous United States multi-layer soil characteristics data set for regional climate and hydrology modeling: Earth Interactions, v. 2, no. 2, p. 1–26.

Mitchell, W.D., 1957, Flow duration of Illinois streams: Illinois Department of Public Works and Buildings, Division of Waterways, 189 p.

Mohamoud, Y.M., 2008, Prediction of daily flow duration curves and streamflow for ungauged catchments using regional flow duration curves: Hydrological Sciences Journal, v. 53, no. 4, p. 706–724.

Parajka, J., and others, 2013, Prediction of runoff hydrographs in ungauged basins, *in* Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H., eds., 2013, Runoff prediction in ungauged basins—Synthesis across processes, places, and scales: Cambridge, Cambridge University Press, p. 227–269.

Perry, C.A., Wolock, D.M., and Artman, J.C., 2004, Estimates of flow duration, mean flow, and peak-discharge frequency values of Kansas stream locations: U.S. Geological Survey Scientific Investigations Report 2004–5033, 651 p.

Ries, K.G., III, and Friesz, P.J., 2000, Methods for estimating low-flow statistics for Massachusetts streams: U.S. Geological Survey Water-Resources Investigations Report 00–4135, 81 p.

Ries, K.G., III, Guthrie, J.G., Rea, A.H., Steeves, P.A., and Stewart, D.W., 2008, StreamStats—A water resources web application: U.S. Geological Survey Fact Sheet 2008–3067, 6 p.

Risley, J., Stonewall, A., and Haluska, T., 2008, Estimating flow-duration and low-flow frequency statistics for unregulated streams in Oregon: U.S. Geological Survey Scientific Investigations Report 2008–5126, 22 p.

Schwarz, G.E., and Alexander, R.B., 1995, State Soil Geographic (STATSGO) Data Base for the conterminous United States: U.S. Geological Survey Open-File Report 95–449, http://water.usgs.gov/GIS/metadata/usgswrd/XML/ussoils.xml. [Also available at http://pubs.usgs.gov/dds/dds38/.]

Searcy, J.K., 1959, Flow-duration curves, manual of hydrology—Part 2, Low-flow techniques: U.S. Geological Survey Water-Supply Paper 1542–A, 33 p.

Schneider, A.F., 1966, Physiography, *in* Lindsey, A.A., ed., Natural features of Indiana: Indianapolis, Indiana Academy of Science, p. 40–56.

Simley, J.D., and Carswell Jr., W.J., 2009, The National Map—Hydrography: U.S. Geological Survey Fact Sheet 2009–3054, 4 p.

Singh, K.P., 1971, Model flow duration and streamflow variability: Water Resources Research, v. 7, no. 4, p. 1031–1036.

Smakhtin, V.Y., 1999, Generation of natural daily flow time-series in regulated rivers using a non-linear spatial interpolation technique: Regulated Rivers: Research and Management, v. 15, p. 311–323.

Soller, D.R., and Berg R.C., 1992, Using regional geologic information to assess relative aquifer contamination potential–an example from the central United States: U.S. Geological Survey Open-File Report 92–694, Map, scale 1:1,000,000.

Soller, D.R., and Packard, P.H., 1998, Digital representation of a map showing thickness and character of Quaternary sediments in the glaciated United States east of the Rocky Mountains: U.S. Geological Survey Digital Data Series DDS-38. [Also available at http://pubs.usgs.gov/dds/dds38/.]

Soong, D.T., Ishii, A.L., Sharpe, J.B., and Avery, C.F., 2004, Estimating flood-peak discharge magnitudes and frequencies for rural streams in Illinois: U.S. Geological Survey Scientific Investigations Report 2004–5103, 158 p.

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

Stiles, T.C., 2001, A simple method to define bacteria TMDLs in Kansas: ASIWPCA / ACWF / WEF TDML Science Issues Conference, On-site Program, St. Louis, MO, p. 375–378.

Straub, T.D., and Over, T.M., 2010, Pier and contraction scour prediction in cohesive soils at selected bridges in Illinois: Illinois Center for Transportation Research Report FHWA-ICT-10-074, 130 p.

Stuckey, M.H., Koerkle, E.H., and Ulrich, J.E., 2012, Estimation of baseline daily mean streamflows for ungaged locations on Pennsylvania streams, water years 1960-2008: U.S. Geological Survey Scientific Investigations Report 2012–5142, 61 p.

Sullivan, J.A., 2002, Use of load duration curves for the development of nonpoint source bacteria TMDLs in Texas: Proceedings, ASAE Watershed Management to Meet Emerging TMDL Regulations Conference, Fort Worth, TX, p. 355–360.

Therneau, T., 2021, A package for survival analysis in R: R package version 3.1–12, accessed August 21, 2020, at https://CRAN.R-project.org/package=survival.

U.S. Environmental Protection Agency, Office of Wetlands, Oceans, and Watersheds, Watershed Branch, 2007a, An approach for using load duration curves in the development of TMDLs: EPA 841-B-07-006, 68 p.

U.S. Environmental Protection Agency, Office of Wetlands, Oceans, and Watersheds, Watershed Branch, 2007b, Options for expressing daily loads in TMDLs, June 22, 2007 draft, 50 p., accessed March 27, 2014, at http://water.epa.gov/lawsregs/lawsguidance/cwa/tmdl/techsupp.cfm.

U.S. Geological Survey and U.S. Department of Agriculture, Natural Resources Conservation Service, 2013, Federal standards and procedures for the national Watershed Boundary Dataset (WBD) (4 ed.): U.S. Geological Survey Techniques and Methods, book 11, chap. A3, 63 p. [Also available at http://pubs.usgs.gov/tm/11/a3/.]

Vogel, R.M., and Fennessey, N.M., 1995, Flow duration curves II – A review of applications in water resources planning: Water Resources Bulletin, v. 31, no. 6, p. 1029–1039.

Wolock, D.M., 1997, STATSGO soil characteristics for the conterminous United States: U.S. Geological Survey Open-File Report 97–656, accessed September 15, 2020, at https://doi.org/10.3133/ofr97656.

USGS

Over and others—**Estimation of Regional Flow-Duration Curves for Indiana and Illinois**—Scientific Investigations Report 2014–5177, ver. 2.0

ISSN 2328-0328 (online)
https://doi.org/10.3133/sir20145177