Estimation of Peak Discharge Quantiles for Selected Annual Exceedance Probabilities in Northeastern Illinois

By Thomas M. Over, Riki J. Saito, Andrea G. Veilleux, Padraic S. O’Shea, Jennifer B. Sharpe, David T. Soong, and Audrey L. Ishii

Prepared in cooperation with the Illinois Center for Transportation, the Illinois Department of Transportation, and the Federal Highway Administration

Scientific Investigations Report 2016–5050
Version 3.0, June 2021

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

The guidance of the Technical Review Panel for this project, chaired by Matt O’Connor of the Illinois Department of Transportation (IDOT) and including as members Steve Ferguson (IDOT), Dan Ghere (Federal Highway Administration), Rick Gosch (Illinois Department of Natural Resources), Perry Masouridis (IDOT), Tzuoh-Ying Su (U.S. Army Corps of Engineers), Neil VanBebber (IDOT), and Rick Wojcik (IDOT) was critical to the success of this project.

Marvin Harris and John Latour of the U.S. Geological Survey (USGS) Illinois Water Science Center, John Walker and Robert Waschbusch of the USGS Wisconsin Water Science Center, and Donald Arvin of the USGS Indiana Water Science Center provided key input on the quality of streamgage records used in this study.

Allan Curtis (Midwestern Regional Climate Center) assisted with downloading and understanding the National Weather Service cooperative daily precipitation data used in this study.

This publication is based on the results of ICT-R27-144, Development and Implementation of Updated Urban Regional Flood Frequency Equations for Illinois. ICT-R27-144 was conducted in cooperation with the Illinois Center for Transportation; the Illinois Department of Transportation, Division of Highways; and the U.S. Department of Transportation, Federal Highway Administration.
Contents

Acknowledgments ........................................................................................................................................... iii
Abstract .......................................................................................................................................................... 1
Introduction .................................................................................................................................................. 1
  Purpose and Scope ..................................................................................................................................... 2
  Description of Study Area ......................................................................................................................... 2
  Previous Studies ...................................................................................................................................... 7
Data Development ....................................................................................................................................... 7
  Peak Discharge Quantiles at Streamgages ............................................................................................ 8
    Streamgage Selection ............................................................................................................................. 8
    Redundancy Analysis ............................................................................................................................ 9
    Regional Skew Analysis ....................................................................................................................... 12
    Frequency Analysis ............................................................................................................................. 12
  Basin Characteristics ............................................................................................................................ 13
    Basin Delineation .................................................................................................................................. 13
    Spatially Averaged Basin Characteristics ........................................................................................... 13
Regional Temporal Regression Analysis and Adjustment ......................................................................... 13
  Data Used in This Analysis .................................................................................................................... 14
  Temporal Analysis of Urbanization Effect ........................................................................................... 16
Regional Spatial Regression Analyses ....................................................................................................... 22
  Selection of Basin Characteristics ........................................................................................................ 22
  Development of Final Spatial Regression Equations .......................................................................... 27
  Accuracy of Final Spatial Regression Equations .................................................................................. 29
Applications of Regression Equations ....................................................................................................... 30
  Applications of the Spatial Regression Equations .............................................................................. 33
    Ungaged Locations Far from a Streamgage ......................................................................................... 33
    At a Streamgage ................................................................................................................................... 33
    Ungaged Location Near a Streamgage ................................................................................................. 34
  Applications of the Temporal Urbanization Coefficients .................................................................... 34
Example Computations ............................................................................................................................. 35
  Example 1: Ungaged Location in the Study Region, Far from a Streamgage ................................... 35
  Example 2: Ungaged Location in the Study Region, Near a Streamgage ........................................ 37
  Example 3: At a Streamgage in the Study Region ................................................................................ 37
  Example 4: At a Location Outside the Study Region ......................................................................... 37
  Example 5: Adjustment for Effects of Future Urbanization ............................................................... 38
Summary ..................................................................................................................................................... 38
References Cited .......................................................................................................................................... 39
Appendix 1. Northeastern Illinois Regional Skew Analysis .................................................................... 44
  Application of B-WLS/B-GLS Method in Northeastern Illinois ......................................................... 44
  Results of Regional Skew Analysis in Northeastern Illinois .............................................................. 45
References Cited .......................................................................................................................................... 49
Figures

1. Map showing study area and streamgages used in this study..........................3
2. Graph showing population growth in northeastern Illinois, 1850–2010...............5
3. Map showing housing density by decade in northeastern Illinois.....................6
4. Graph showing empirical distribution functions of urbanized fraction of gaged basins in Illinois flood-frequency regions ...........................................8
5. Graphs showing properties of U.S. Geological Survey streamgage records used in this study in northeastern Illinois.........................................................10
6. Graphs showing properties of basins of U.S. Geological Survey streamgages used in this study in northeastern Illinois.........................................................11
7. Map showing locations of precipitation stations used in this study and their Thiessen polygons.................................................................15
8. Graph showing segment intercepts as a function of drainage area for 181 streamgages used in this study in northeastern Illinois.............................18
9. Graph showing urban fraction coefficients from temporal regression analysis of 117 streamgages in northeastern Illinois and adjacent states, as a function of exceedance probability ..................................................19
10. Graph showing segment intercept-subtracted log-transformed annual maximum peak discharge from 181 streamgages in northeastern Illinois and adjacent states as a function of urban fraction with quantile regression and least-squares regression line fits .................................................20
11. Graphs showing changes in the mean, standard deviation, and skewness of the log-transformed annual maximum peak discharge records for 181 streamgages used in this study in northeastern Illinois, resulting from adjustment to 2010 urban fraction values.................................................................21
12. Graphs showing annual exceedance probability of adjusted $R^2$ as a function of the regression coefficient t ratio values for basin characteristic categories considered in this study in northeastern Illinois........................................23
13. Graph showing comparison of square root of fraction of 2011 National Land Cover Dataset for 117 basins used in this study in northeastern Illinois........31
14. Graphs showing ratios of peak discharge quantiles as a function of urban fraction for selected annual exceedance probabilities as implied by the two sets of urbanization coefficients obtained in this study in northeastern Illinois ......32
1–1. Fitting of cross-correlation model of the log-annual maximum peak discharges in this study in northeastern Illinois..................................................46
1–2. Relations between the unbiased at-site skew and urbanization measure for 110 streamgages in northeastern Illinois regional skew study area .............48
Tables

1. U.S. Geological Survey streamgages used in this study in northeastern Illinois and adjacent states (online) ................................................................. 9
2. Estimated peak discharge quantiles for 181 streamgages in northeastern Illinois and adjacent states, at selected annual exceedance probabilities (online) ................ 12
3. Spatially averaged basin characteristics considered for developing spatial regression equations in this study in northeastern Illinois (online) .................. 13
4. Segment information for 181 U.S. Geological Survey streamgages used in this study in northeastern Illinois and adjacent states (online) .................. 16
5. Number of segments per streamgage record used in regression analysis of peak discharge for 117 streamgages, northeastern Illinois and adjacent states .......... 16
6. Observed and urban-adjusted annual maximum peak discharges and associated urbanization and precipitation values at 181 streamgages in northeastern Illinois and adjacent state (online) .................................................................................................................. 16
7. Results of ordinary least-squares linear regression of 117 streamgages in northeastern Illinois and adjacent states to diagnose the temporal effects of urbanization and precipitation on annual maximum peak discharges ....................... 17
8. Quantile regression coefficients from temporal analysis of 117 streamgages in northeastern Illinois and adjacent states, as a function of annual exceedance probability (online) ............................................. 18
9. Quantile regression coefficients of urban fraction from temporal analysis of 117 streamgages in northeastern Illinois and adjacent states, at selected annual exceedance probabilities ............................................................... 19
10. Correlation matrix and variance inflation factor values of basin characteristics selected for use in spatial regression analysis of 117 streamgages in northeastern Illinois and adjacent states ........................................................................ 27
11. Coefficients of the selected spatial regression equations in this study in northeastern Illinois .................................................................................... 28
12. Measures of the average accuracy of the selected spatial regression equations in this study in northeastern Illinois ................................................. 29
13. Components of variance of prediction for the selected spatial regression equations in this study in northeastern Illinois (online) .................. 30
14. Ranges of basin characteristic values used to fit selected spatial regression equations in this study in northeastern Illinois .................................. 33
15. Example peak discharge quantile computations in this study in northeastern Illinois .............................................................................................. 35
1–1. Skew statistics at U.S. Geological Survey streamgages used in the development of the regional skew model in this study in northeastern Illinois (online) ................ 45
1–2. Regional skewness models and corresponding metrics in this study in northeastern Illinois ......................................................................................... 48
1–3. Average regional skew, variance of prediction, and equivalent record length for URBAN regional skew model for various values of NLCD_22_23_24, in this study in northeastern Illinois ....................................................... 48
1–4. Pseudo analysis of variance (ANOVA) statistics for the northeastern Illinois URBAN regional skew model ............................................ 49
Conversion Factors

[U.S. customary units to International System of Units]

<table>
<thead>
<tr>
<th>Multiply</th>
<th>By</th>
<th>To obtain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td></td>
<td></td>
</tr>
<tr>
<td>inch (in.)</td>
<td>2.54</td>
<td>centimeter (cm)</td>
</tr>
<tr>
<td>inch (in.)</td>
<td>25.4</td>
<td>millimeter (mm)</td>
</tr>
<tr>
<td>foot (ft)</td>
<td>0.3048</td>
<td>meter (m)</td>
</tr>
<tr>
<td>mile (mi)</td>
<td>1.609</td>
<td>kilometer (km)</td>
</tr>
<tr>
<td>Area</td>
<td></td>
<td></td>
</tr>
<tr>
<td>acre</td>
<td>4,047</td>
<td>square meter (m²)</td>
</tr>
<tr>
<td>acre</td>
<td>0.4047</td>
<td>hectare (ha)</td>
</tr>
<tr>
<td>acre</td>
<td>0.4047</td>
<td>square hectometer (hm²)</td>
</tr>
<tr>
<td>acre</td>
<td>0.004047</td>
<td>square kilometer (km²)</td>
</tr>
<tr>
<td>square mile (mi²)</td>
<td>259.0</td>
<td>hectare (ha)</td>
</tr>
<tr>
<td>square mile (mi²)</td>
<td>2.590</td>
<td>square kilometer (km²)</td>
</tr>
<tr>
<td>Volume</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cubic foot (ft³)</td>
<td>28.32</td>
<td>cubic decimeter (dm³)</td>
</tr>
<tr>
<td>cubic foot (ft³)</td>
<td>0.02832</td>
<td>cubic meter (m³)</td>
</tr>
<tr>
<td>acre-foot (acre-ft)</td>
<td>1,233</td>
<td>cubic meter (m³)</td>
</tr>
<tr>
<td>acre-foot (acre-ft)</td>
<td>0.001233</td>
<td>cubic hectometer (hm³)</td>
</tr>
<tr>
<td>Flow rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cubic foot per second (ft³/s)</td>
<td>0.02832</td>
<td>cubic meter per second (m³/s)</td>
</tr>
</tbody>
</table>

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows:

\[°F = (1.8 \times °C) + 32.\]

Temperature in degrees Fahrenheit (°F) may be converted to degrees Celsius (°C) as follows:

\[°C = (°F - 32) / 1.8.\]

Datum

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

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

Elevation, as used in this report, refers to distance above the vertical datum.


**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AEP</td>
<td>annual exceedance probability</td>
</tr>
<tr>
<td>AFPD</td>
<td>annual flood-probability discharge</td>
</tr>
<tr>
<td>AMS</td>
<td>annual maximum series</td>
</tr>
<tr>
<td>B-GLS</td>
<td>Bayesian generalized least squares</td>
</tr>
<tr>
<td>B-WLS</td>
<td>Bayesian weighted least squares</td>
</tr>
<tr>
<td>BDF</td>
<td>basin development factor</td>
</tr>
<tr>
<td>CCPN</td>
<td>Cook County Precipitation Network</td>
</tr>
<tr>
<td>COOP</td>
<td>cooperative network</td>
</tr>
<tr>
<td>CSG</td>
<td>crest-stage gage</td>
</tr>
<tr>
<td>DEM</td>
<td>digital elevation model</td>
</tr>
<tr>
<td>EMA</td>
<td>Expected Moments Algorithm</td>
</tr>
<tr>
<td>Esri</td>
<td>Environmental Systems Research Institute, Inc.</td>
</tr>
<tr>
<td>GIS</td>
<td>geographic information system</td>
</tr>
<tr>
<td>GLS</td>
<td>generalized least squares</td>
</tr>
<tr>
<td>HDB</td>
<td>hand-drawn boundary</td>
</tr>
<tr>
<td>ICT</td>
<td>Illinois Center for Transportation</td>
</tr>
<tr>
<td>IDOT</td>
<td>Illinois Department of Transportation</td>
</tr>
<tr>
<td>MACS</td>
<td>MRCC Applied Climate System</td>
</tr>
<tr>
<td>MRCC</td>
<td>Midwestern Regional Climate Center</td>
</tr>
<tr>
<td>MSE</td>
<td>mean squared error</td>
</tr>
<tr>
<td>MWRDGC</td>
<td>Metropolitan Water Reclamation District of Greater Chicago</td>
</tr>
<tr>
<td>NED</td>
<td>National Elevation Dataset</td>
</tr>
<tr>
<td>NHD</td>
<td>National Hydrography Dataset</td>
</tr>
<tr>
<td>NID</td>
<td>National Inventory of Dams</td>
</tr>
<tr>
<td>NLCD</td>
<td>National Land Cover Dataset</td>
</tr>
<tr>
<td>NWI</td>
<td>National Wetlands Inventory</td>
</tr>
<tr>
<td>NWIS</td>
<td>National Water Information System</td>
</tr>
<tr>
<td>NWS</td>
<td>National Weather Service</td>
</tr>
<tr>
<td>OLS</td>
<td>ordinary least squares</td>
</tr>
<tr>
<td>OVB</td>
<td>omitted variable bias</td>
</tr>
<tr>
<td>PDS</td>
<td>partial duration series</td>
</tr>
<tr>
<td>PeakFQ</td>
<td>Peak flow FreQuency program</td>
</tr>
<tr>
<td>PILF</td>
<td>potentially influential low flood</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>PRISM</td>
<td>Parameter-elevation Relationships on Independent Slopes Model</td>
</tr>
<tr>
<td>SE</td>
<td>standard error</td>
</tr>
<tr>
<td>SSURGO</td>
<td>State Soil Survey Geographic database</td>
</tr>
<tr>
<td>USGS</td>
<td>U.S. Geological Survey</td>
</tr>
<tr>
<td>VIF</td>
<td>variance inflation factor</td>
</tr>
<tr>
<td>WLS</td>
<td>weighted least squares</td>
</tr>
<tr>
<td>WREG</td>
<td>Weighted-Multiple-Linear Regression program</td>
</tr>
<tr>
<td>WY</td>
<td>water year; a water year is the 12-month period October 1 through September 30 designated by the calendar year in which it ends.</td>
</tr>
</tbody>
</table>
Estimation of Peak Discharge Quantiles for Selected Annual Exceedance Probabilities in Northeastern Illinois

By Thomas M. Over, Riki J. Saito, Andrea G. Veilleux, Padraic S. O’Shea, Jennifer B. Sharpe, David T. Soong, and Audrey L. Ishii

Abstract

This report provides two sets of equations for estimating peak discharge quantiles at annual exceedance probabilities (AEPs) of 0.50, 0.20, 0.10, 0.04, 0.02, 0.01, 0.005, and 0.002 (recurrence intervals of 2, 5, 10, 25, 50, 100, 200, and 500 years, respectively) for watersheds in Illinois based on annual maximum peak discharge data from 117 watersheds in and near northeastern Illinois. One set of equations was developed through a temporal analysis with a two-step least squares-quantile regression technique that measures the average effect of changes in the urbanization of the watersheds used in the study. The resulting equations can be used to adjust rural peak discharge quantiles for the effect of urbanization, and in this study the equations also were used to adjust the annual maximum peak discharges from the study watersheds to 2010 urbanization conditions.

The other set of equations was developed by a spatial analysis. This analysis used generalized least-squares regression to fit the peak discharge quantiles computed from the urbanization-adjusted annual maximum peak discharges from the study watersheds to drainage-basin characteristics. The peak discharge quantiles were computed by using the Expected Moments Algorithm following the removal of potentially influential low floods defined by a multiple Grubbs-Beck test. To improve the quantile estimates, regional skew coefficients were obtained from a newly developed regional skew model in which the skew increases with the urbanized land use fraction. The skew coefficient values for each streamgage were then computed as the variance-weighted average of at-site and regional skew coefficients. The drainage-basin characteristics used as explanatory variables in the spatial analysis include drainage area, the fraction of developed land, the fraction of land with poorly drained soils or likely water, and the basin slope estimated as the ratio of the basin relief to basin perimeter.

This report also provides the following: (1) examples to illustrate the use of the spatial and urbanization-adjustment equations for estimating peak discharge quantiles at ungaged sites and to improve flood-quantile estimates at and near a gaged site; (2) the urbanization-adjusted annual maximum peak discharges and peak discharge quantile estimates at streamgages from 181 watersheds including the 117 study watersheds and 64 additional watersheds in the study region that were originally considered for use in the study but later deemed to be redundant.

The urbanization-adjustment equations, spatial regression equations, and peak discharge quantile estimates developed in this study will be made available in the web application StreamStats, which provides automated regression-equation solutions for user-selected stream locations. Figures and tables comparing the observed and urbanization-adjusted annual maximum peak discharge records by streamgage are provided at https://doi.org/10.3133/sir20165050 for download.

Introduction

Estimates of the magnitude and frequency of floods, particularly in urban areas where there is an increased density of lives and property at risk, is a critical ingredient to floodplain management, emergency response planning, and infrastructure design tasks such as sizing of bridges and culverts. At the same time, urbanization, particularly the construction of impervious surfaces, changes the response of watersheds to precipitation by reducing infiltration and increasing flow velocities, which increases flood volumes and magnitudes of peaks (Konrad, 2003).

Predictions of the effects of urbanization in ungaged basins or in basins where future urbanization is expected or planned are usually obtained by following one of two general approaches: (1) simulation with rainfall-runoff models driven by historical precipitation (either as design storms or by continuous simulation of long-term records), usually calibrated to observed streamflow data; and (2) regressions of peak discharge quantiles obtained from observed streamflow data at a collection of streamgages in and near the region of interest on relevant basin characteristics such as drainage area and fraction of urbanized area. The two approaches have different advantages and disadvantages (Rosbjerg and others, 2013); generally the simulation model approach is preferred for predicting the response of a basin to changes in the basin properties or climate characteristics, whereas for ungaged basins over a historical period, the regional regression approach may be more accurate (Hodgkins and others, 2007).
Indeed, the regional regression equation approach is a well-developed form of regional calibration of a watershed model, whereas significant challenges remain for regional calibration of simulation models (Vogel, 2006).

Regional regression equations for urbanization-affected areas of northeastern Illinois were first published by Allen and Bejcek (1979) who used data beginning around 1960 and ending in 1976. Since 1976 there have been many years of additional streamflow data, extensive growth of the urbanized area of northeastern Illinois, and several historic floods (Juhl, 2005). As a result, the U.S. Geological Survey (USGS) Illinois Water Science Center in cooperation with the Illinois Department of Transportation (IDOT) and the Federal Highway Administration (FHWA), through the Illinois Center for Transportation (ICT), developed a project to update the spatial regression equations of Allen and Bejcek (1979) for urbanization-affected streams in northeastern Illinois and for adjusting rural peak discharge quantile estimates for the effect of urbanization.

**Purpose and Scope**

This report presents the methods and results of a study that applied a regional spatial regression approach to develop equations for estimating peak discharge quantities with annual exceedance probabilities (AEPs) of 0.5, 0.2, 0.1, 0.04, 0.02, 0.01, 0.005, and 0.002 (often referred to as the 2-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year floods, respectively) for Illinois flood-frequency region 2 (previously defined by Soong and others, 2004). Flood-frequency region 2 consists of northeastern Illinois and parts of north-central Illinois and includes the great Chicago region. This report updates the existing spatial regression equations (Soong and others, 2004; Allen and Bejcek, 1979) for northeastern Illinois in the following ways: (1) it uses streamflow data through water year (WY) 2009, compared to through 1999 for Soong and others (2004) and through 1976 for Allen and Bejcek (1979); (2) it tests and implements wetland-open water and slope variables replacing those of Soong and others (2004); and (3) it uses the updated flood-frequency estimation methods and software, particularly in the estimation of at-site peak discharge quantiles by using multiple Grubbs-Beck potentially influential low flood (PILF) detection, the Expected Moments Algorithm (EMA) for handling censored and historical floods, and regional skew estimation by Bayesian generalized least squares.

The study described in this report also implements the temporal regression approach of Over and others (2016) to the analysis and adjustment of annual maximum peak discharge records for the exceedance probability-dependent effect of urbanization. By this approach the observed peak discharge records are adjusted to current (2010) urbanization conditions before they are used in the spatial regression analysis; and a set of “temporal” urbanization coefficients are obtained that are applicable wherever hydrologic effects of urbanization are similar to those in northeastern Illinois. Because the historical urbanization data used in the temporal regression approach are available beginning 1940, the earliest streamflow data used in this study are from WY 1940.

There are three primary products from the study described in this report: (1) this report, published in ICT and USGS versions, having the same content, and presenting the methods and results, including the two sets of regression equations for estimating peak discharge quantiles; (2) an implementation of the spatial and temporal regression equations in StreamStats (https://streamstats.usgs.gov/); and (3) a collection of figures and tables comparing the observed and urbanization-adjusted peak discharges at https://doi.org/10.3133/sir20165050.

**Description of Study Area**

The study area covers northeastern Illinois and parts of north-central Illinois, northwest Indiana, and southeast Wisconsin (fig. 1). The physiography of the region is determined by recent glacial episodes. The last glacial episode was the Wisconsinan, which ended about 12,000 years ago and during which all except the western lobe of the study area in the valley of the Green River was covered in ice (Hansel and McKay, 2010). The entire study region was covered with ice during the previous glacial episode, the Illinoian, which ended about 125,000 years ago. The glaciers of the Wisconsinan episode left behind various glacial features, the most prevalent of which are a series of moraines (Arnold and others, 1999). The physiographic regions of northeastern Illinois follow these glacial features, with the Chicago Lake Plain and the Wheaton Morainal Country making up the Great Lakes Section, and then, to the west and south, the Till Plains Section (Leighton and others, 1948), with corresponding regions in Indiana (Schneider, 1966). As a result, the surficial geology in the study region consists of unconsolidated glacial drift of varying thickness up to 400 feet (Piskin and Bergstrom, 1975). The character of the drift ranges from end moraines and till plains to outwash fans and lacustrine deposits (Hansel and McKay, 2010).

Subsequent warmer and drier periods added loess to the surface, thinner toward Lake Michigan, so that soil parent materials combine the loess and the underlying drift (Fehrenbacher and others, 1984). Soil orders of the region vary with the original vegetation type. Alfisols, which form under forests, dominate nearer to Lake Michigan, whereas mollisols, which form under grasslands, cover most of the rest of the region (Fehrenbacher and others, 1984). The permeability of the soils varies with the parent material from very low on clayey deposits to high on sandy outwash (Arnold and others, 1999).

The climate of the study region is classified as “Dfa” in the Köppen-Geiger system, which means humid continental with warm summers (Peel and others, 2007; Belda and others, 2014). This classification means winters are cold, summers are warm, and precipitation is common during the whole year. Based on 1971–2000 data, the average temperature at Chicago
Figure 1. Study area and streamgages used in this study: A, entire area; B, detail showing southern Lake, Cook, and DuPage counties in detail.
Estimation of Peak Discharge Quantiles for Selected Annual Exceedance Probabilities in Northeastern Illinois

Figure 1. Study area and streamgages used in this study: A, entire area; B, detail showing southern Lake, Cook, and DuPage counties in detail.—Continued
is about 50 °Fahrenheit (F), with average winter (December–February) daily highs of about 33 °F and lows of 17 °F and average summer (June–August) daily highs of about 82 °F and lows of 62 °F (Changnon and others, 2004). Based on 1981–2010 Parameter-elevation Relationships on Independent Slopes Model (PRISM) data (http://www.prism.oregonstate.edu/normals/, accessed April 15, 2014), precipitation for Chicago is about 35 inches per year, with a minimum monthly amount of about 2 inches in both January and February, increasing to about 4 inches per month during May–August before decreasing through the fall toward the winter minimum. About half the total precipitation is the result of thunderstorms and 3 to 3.5 inches falls as snow, based on 1941–1995 data (Changnon and others, 2004).

In the flood climate classification of Hayden (1988), the study area lies at the border of the “TsuCpSs**” region to the north and the “TsuCpSe*” region to the south, where “Tsu” indicates barotropic (nonfrontal or convective) and unorganized (that is, without tropical cyclones) in the summer, “Cp” indicates that frontal storms are possible throughout the year, and “Ss**” indicates that winter snow cover is seasonal and exceeds 50 centimeters (cm), so that there may be substantial spring snowmelt flooding, whereas “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. The seasonality of precipitation in the study region and its flood classification explain the wide distribution of the timing of the annual maximum peak discharges used in this study (see “Streamgage Selection” section), with the largest monthly fractions in March, April, and June, but also substantial fractions in February and May, with smaller fractions during July to September.

Nonurbanized land cover in the study area includes a substantial fraction of row crop agriculture and also more forest, grassland, and open water and wetlands than the surrounding regions, which are primarily agricultural (Illinois Department of Natural Resources, 1996). Urbanized land use in the study area has been driven by population growth in metropolitan Chicago, which was dominated by that of the City of Chicago until 1940, when its population leveled off and population growth shifted to the suburbs (fig. 2). This suburban population growth was accompanied by even larger growth in developed land use; for example, from 1970 to 1990, the population grew by 4 percent, but developed land use grew by 40 percent, during which time more than 400 square miles (mi²) of farmland that had been previously drained for agricultural use (Juhl, 2005) were developed (Mariner, 2005). This growth in developed land use is shown geographically in figure 3.

The hydrologic effects of this expansion of developed land in northeastern Illinois can be inferred from the series of historic floods that began in the 1950s and have extended to the present (2015) (Juhl, 2005; Changnon, 1999; Changnon and Westcott, 2002a, b; Angel and Changnon, 2008; Changnon, 2010; Changnon, 2011; Fazio and Sharpe, 2012; Villarini and others, 2013), though increases in precipitation
Figure 3. Housing density by decade in northeastern Illinois (modified from Karsten and others, 2013).
have also been identified (Rougé and Cai, 2014; Winters and others, 2015). These floods led to several policy and engineering responses (Juhl, 2005). One such policy response was the introduction of ordinances requiring stormwater detention on new development, the first of which was promulgated in 1972 by the Metropolitan Water Reclamation District of Greater Chicago (MWRDGC). Another policy response was the passage of legislation by the State of Illinois allowing five suburban counties to plan and fund countywide stormwater management programs (Resource Coordination Policy Committee, 1998). A third policy response was the planning and construction of channel improvements and flood-control reservoirs (Resource Coordination Policy Committee, 1998).

Previous Studies

Allen and Bejcek (1979) developed regional spatial regression equations for peak discharge quantiles for the urbanized area of northeastern Illinois using data through WY 1976 as a function of impervious percentage, drainage area, and slope (all log-transformed). They used 103 streamgage records, including 70 crest-stage gauges (CSGs), for many of which they developed stage-discharge ratings as part of the study. Drainage areas of the basins ranged from 0.07 to 630 mi² and impervious fractions from 1 to 38.7 percent. A decreasing effect of urbanization (measured as an estimated fraction of impervious area) was determined with increasing flood magnitude (decreasing AEP), as is generally expected (Konrad, 2003). The authors also suggest that the urbanization effect factor from their equations would be appropriate to use outside their study area elsewhere in Illinois and surrounding regions to adjust existing rural equations for use in urban areas, but that their results would not be appropriate for basins “completely served by underground drainage systems” or where “flood detention or retention reservoirs substantially affect the flood peaks” (Allen and Bejcek, 1979, p. 8).

Sauer and others (1983) completed a nationwide spatial regression study of the effect of urbanization on flood frequency that used the results of Allen and Bejcek (1979). Sauer and others used 199 streamgage records including 18 from Illinois, excluding streamgages with substantial manmade detention effects and excluding those for which the developed fraction had increased by more than 50 percent during the period of record. They developed three sets of spatial regression equations. Two sets include seven explanatory variables and the remaining set has three. The two seven-variable equation sets have the same variables except one uses basin lag time (estimated from observed rainfall-runoff data, where available) in place of a measure of channel efficiency determined following the methods of Espey and Winslow (1974). All three sets of equations include as explanatory variables drainage area, the rural peak discharge quantile, and the basin development factor (BDF), which is a numerical index of the presence of channel improvements, curbs and gutters, and storm sewers, and is not easily computable from standard GIS databases. These three common explanatory variables constitute the variables of the three-variable equation set. Impervious area fraction is included in the two seven-variable sets of equations, but it is not as significant as BDF. The coefficients of their equations also indicate a decreasing effect of urbanization for larger floods.

More recently, Moglen and Shivers (2006) developed an iterative method of adjusting rural flood frequency equations for the effects of time-varying urbanization, measured by one of two GIS-computable quantities, estimated impervious fraction and population density. Among their 78 study streamgages are 22 from Illinois, of which just one (Boneyard Creek at Urbana, USGS streamgage 03337000) is outside northeastern Illinois. One of the conclusions of the study listed as “Model Weaknesses” is that the apparent effects of urbanization are “relatively mild.” In the use of GIS-computable quantities as the urbanization variable and directly incorporating time-varying urbanization, Moglen and Shivers’ study is similar to this study. Their method does not, however, produce an adjusted time series of annual maximum peak discharges for each streamgage; rather, it produces just a set of urbanization-adjusted peak discharge quantiles.

Other studies that have characterized the effect of urbanization on flooding in northeastern Illinois include Changnon and Demissie (1996), Changnon and others (1996), Hejazi and Markus (2009), Villarini and others (2013), and Rougé and Cai (2014).

Soong and others (2004) divided the State of Illinois into seven hydrologic regions and developed regional regression equations for both annual maximum series (AMS) and partial duration series (PDS) peak discharges for rural basins in each region, using data through WY 1999. The basin characteristics used in the regression equations are drainage area, main channel slope, basin length, soil permeability, and percent open water and herbaceous wetland. For region 2, which is the focus area of this study, they used drainage area, main channel slope, and percent open water and herbaceous wetland. The AMS regression equations from the study were later implemented in StreamStats (Ishii and others, 2010).

Data Development

Data development for this study consisted of the computation of two sets of data values: (1) observed peak discharge quantiles, that is, the statistical relation between flood magnitudes and frequencies based on observed floods at selected streamgages and (2) basin characteristics, that is, quantitative descriptors of characteristics of the basins draining to the locations of the streamgages. Regression equations for estimating the peak discharge quantiles at ungauged locations were then developed using the basin characteristics.
Peak Discharge Quantiles at Streamgages

To develop equations for estimating peak discharge quantiles for ungaged stream locations, such quantiles were first computed at selected streamgages with at least 10 years of annual maximum peak discharges. This computation included four steps: (1) streamgage selection; (2) removal of redundant streamgages, where one streamgage of streamgage pairs whose basins are nested with substantial overlap are removed from the set of streamgages on which the regression analyses are done; (3) regional skew analysis, where a regional regression of at-streamgage skewness values is done to provide a regional estimate that is weighted with the at-streamgage skewness values in the frequency analysis; and (4) frequency analysis, where peak discharge quantiles at the selected AEPs are computed for each selected streamgage by fitting to the log-Pearson type III by a generalized method-of-moments method.

Streamgage Selection

The streamgages were chosen from the urbanized area of northeastern Illinois and nearby urbanized areas in northwestern Indiana and southeastern Wisconsin within hydrologic unit 0404, along with region 2 of Soong and others (2004) (figs. 1 and 3). In Indiana, the nearby urbanized area was selected as Lake and Porter Counties, and in Wisconsin, as consisting of the 8-digit hydrologic unit 04040002 (Pike-Root) (fig. 1). The decision to update only region 2 of Soong and others (2004) as part of this study was in part made on the assumption that only this region includes substantial numbers of streamgages identified as rural drainage basins that contain substantial fractions of urbanized land use. This assumption is verified in figure 4, which shows for example that 40 percent of gaged basins in region 2 used by Soong and others (2004) had urbanized fractions greater than 20 percent as of the last year of record used by Soong and others (2004), whereas only about 2 percent of gaged basins outside region 2 used by Soong and others (2004) had such an urbanized fraction.

A minimum record-length criterion of 10 years of record between WYs 1940 and 2009 was applied when selecting stations. Water year 1940 was chosen as the starting point of the study period because that is the earliest year available in the housing density data of Theobald (2005) that is used to estimate historical urbanization in this study. Water year 2009 was chosen as the ending point of the study period for consistency with the analysis of Over and others (2016).

Figure 4. Empirical distribution functions of urbanized fraction of gaged basins in Illinois flood-frequency regions of Soong and others (2004).
The list of streamgages with at least 10 years of record in the study area was screened to eliminate those with potential measurement accuracy issues using several criteria. One was to refer to the list of streamgages used in the most recent published rural flood-frequency studies (Soong and others [2004] in Illinois, Rao [2006] in Indiana, and Walker and Krug [2003] in Wisconsin). John Walker (USGS, written commun., 2013) also provided a list of streamgages being used in an on-going flood-frequency study in Wisconsin. If a streamgage had been used in one of those studies, it was retained for this study unless specific problems were identified by data section staff at the Illinois, Indiana, or Wisconsin USGS Water Science Centers.

The 181 selected streamgages and some properties of their records and drainage basins are listed in table 1 (available at https://doi.org/10.3133/sir20165050). Histograms of the annual maximum peak discharge record properties at the selected streamgages are shown in figure 5. These histograms show a maximum in streamgage records beginning around 1960 and ending around 1980, which were primarily CSGs on small (less than about 10 mi²) drainage basins (table 1). Histograms of the urban fractions of the streamgage drainage basins in figure 6 show substantial urban fraction increases throughout the streamgage periods of record and further urban fraction increases between the ends of the streamgage records and 2010. These increases in urbanization underscore the nonstationary nature of the land use in the study basins and the need for adjustment to current (2010) conditions. Overall, drainage areas range from less than 0.01 mi² to more than 1,000 mi², but the most are between 1 and 300 mi² (fig. 6).

Annual maximum peak discharge data for the selected stations were retrieved from the National Water Information System (NWIS) (U.S. Geological Survey, 2013).

Redundancy Analysis

The generalized least-squares analysis that is used to fit the regression models for the peak discharge quantiles that form the basis of the final results of this study allows for cross-correlation among the observed quantities that is induced by temporal factors such as the same storm hitting two nearby basins and causing the annual maximum peak discharge for both. Generalized least-squares regression, however, does not account for correlation in the model error that arises when two nested basins have an appreciable overlap in drainage area (Veilleux, 2009). When the overlap is large enough, one of the basins is considered redundant for purposes of regional regression modeling (Veilleux, 2009). To address this issue, a redundancy analysis was carried out to determine an optimal nonredundant subset of the original 181 stations.

Because any one gaged basin may be nested within multiple other gaged basins or have multiple other gaged basins nested within it, a way to decide in what order basins should be removed was needed. For this study, the decision was made by assigning a score to each gaged basin based on its record length and drainage-basin properties and its amount of overlap with other gaged basins, removing the basin with the worst (highest) score, followed by re-computation of the scores, removing the remaining basin with the worst recomputed score, and proceeding in this way until no basin remained with an overlap fraction larger than 0.205, where the overlap fraction \( \text{DAR}_y \) for two nested basins \( i \) and \( j \) is defined as

\[
\text{DAR}_y = \min \left[ \frac{\text{DA}_i}{	ext{DA}_j}, \frac{\text{DA}_j}{\text{DA}_i} \right].
\]

where

\( \text{DA}_i \) and \( \text{DA}_j \) are the drainage areas of basins \( i \) and \( j \), respectively.

Originally a maximum overlap fraction of 0.2 as suggested by Veilleux (2009) was chosen, but it was determined that a few basins could be retained by increasing the value slightly to 0.205. A script was written in R (R Core Team, 2014) to implement this algorithm.

The details of computation of the score are as follows:

The score \( S_y(t) \) for basin \( i \) at iteration step \( t \) is computed as the sum, \( \sum s_y(t) \), over all remaining basins \( j \) having basins overlap with basin \( i \), where \( s_y(t) \), the subscore of basin \( i \) with overlapping basin \( j \) at step \( t \), is computed as

\[
s_y(t) = \text{DAR}_y \left[ 2 - \text{RL}_j / \max(\text{RL}) \right] [d_i(t) + 1],
\]

where

\( \text{DAR}_y \) is the overlap fraction between basins \( i \) and \( j \) (eq. 1),

\( \text{RL}_j \) is the record length of the streamgage at basin \( j \) in years,

\( \max(\text{RL}) \) is the maximum record length in the dataset (70 years), and

\( d_i(t) \) is the probability density value of the drainage area of station \( i \) at step \( t \) scaled to lie between 0 and 1.

Based on their definitions, \( \text{DAR}_y \) lies between 0 to 1, \( 2 - \text{RL}_j / \max(\text{RL}) \) between 1 and 2, and \( d_i(t) + 1 \) between 1 and 2. The effect of each of the factors making up the subscore \( s_y(t) \) on its value is as follows: (a) with more overlap of basin areas, \( \text{DAR}_y \) is larger, and so is \( s_y(t) \); (b) with shorter record length \( \text{RL}_j \), \( 2 - \text{RL}_j / \max(\text{RL}) \) is larger and so is \( s_y(t) \); and (c) when the drainage area of basin \( i \) is at a larger density value (where there are many basins of similar drainage areas), \( d_i(t) \) is larger and so is \( s_y(t) \); consequently, larger basin overlaps, shorter records, and larger drainage area overlaps raise the subscore and therefore the score, making the associated streamgage more likely to be removed.

The result of applying this algorithm was to remove 64 of the original 181 to leave 117 nonredundant streamgages, of which 57 were CSGs throughout their peak discharge record, 13 were CSGs during part of their peak discharge record, and 47 had continuous record throughout their peak discharge records. These 117 streamgages were used for all subsequent analyses. Streamgages that were used and streamgages that
Figure 5. Properties of U.S. Geological Survey streamgage records used in this study in northeastern Illinois. A, number of streamgages with annual maximum peak discharges each water year; B, number of streamgages with annual maximum peak discharge records beginning each water year; C, number of streamgages with peak discharge records ending each water year; D, number of annual maximum peak discharges in each month of the year.
A. Drainage area

B. Urban fraction at first year of record

C. Urban fraction at last year of record

D. Urban fraction in 2010

Figure 6. Properties of basins of U.S. Geological Survey streamgages used in this study in northeastern Illinois. A, basin drainage areas, B, urbanized basin fractions at first year of streamgage records, C, urbanized basin fractions at last year of streamgage record, and D, urbanized basin fractions in 2010, where urbanized is defined as Theobald (2005) housing density fewer than 10 acres per house plus commercial/industrial/transportation land use.
were considered redundant are indicated in the study area maps (fig. 1), the histograms describing the basic properties of the streamgage records and their basins (figs. 5 and 6), and in the table listing the basic properties of the streamgage records and their basins (table 1).

Regional Skew Analysis

Although the Expected Moments Algorithm (EMA) (Cohn and others, 1997 and 2001) used in this study to obtain estimates of the peak discharge quantiles generalizes the method-of-moments method of Bulletin 17B (U.S. Interagency Advisory Committee on Water Data, 1982), it still relies on estimates of the mean, variance, and skewness of the logarithms of the annual maximum peak discharge data at each streamgage. Of these moments, the skewness, because it involves raising the data values to the third power, is very sensitive to extreme values and thus has a large uncertainty when estimated from the data at a single site. Therefore, weighting at-site and regional skew estimates in inverse proportions to their uncertainties to obtain a weighted skew estimate was recommended in Bulletin 17B and continues to be recommended. Because the map in Bulletin 17B showing a grid of regional skew values has become outdated, newer and more refined maps have been created, such as the one in Soong and others (2004) for Illinois. These maps, however, do not address the effects of urbanization. Furthermore, a new approach to the estimation of regional skew through Bayesian generalized least-squares (GLS) analysis has been developed (Reis and others, 2005; Gruber and others, 2007; Gruber and Stedinger, 2008; Veilleux, 2011) that has shown the ability to reduce the uncertainty in regional skew estimates and allows the regional skew estimate to take into account basin characteristics where applicable.

Details of the new approach and its application in this study are given in appendix 1. The data analyzed were the annual maximum peak discharges at the nonredundant streamgages adjusted to 2010 urbanization conditions. The selected regional skew model is dependent on the urban fraction as measured by the NLCD_22_23_24 variable, that is, the sum of fractions of National Land Cover Database (NLCD) 2011 classes 22 (developed, low intensity), 23 (developed, medium intensity), and 24 (developed, high intensity), according to the linear relation:

$$\gamma_0 = -0.39 + 0.97 \times NLCD_{22\_23\_24}^{1/2}, \quad \text{(3)}$$

where

$$\gamma_0$$ is the regional skewness.

Skewness increases substantially with urbanization from a value of -0.39 for a basin with no urbanization (NLCD_22_23_24 = 0) to -0.39 + 0.97 = 0.58 for a basin where the NLCD_22_23_24 fraction is 100 percent. The regional skewness estimate from equation 3 was combined with the at-site skewness according to their associated uncertainties to obtain a weighted skew estimate for use in computing the at-site peak discharge quantile estimates as described in the “Frequency Analysis” section.

Frequency Analysis

The peak discharge quantiles with AEPs of 0.50, 0.20, 0.10, 0.04, 0.02, 0.01, 0.005, and 0.002 for redundant and nonredundant streamgages (table 2, available at https://doi.org/10.3133/sir20165050) were computed by applying the EMA approach (Cohn and others, 1997 and 2001), along with the Multiple Grubbs-Beck test (Cohn and others, 2013) to detect and censor PILFs, to the annual maximum peak discharges adjusted to 2010 urbanization conditions as described in the “Regional Temporal Regression Analysis and Adjustment” section below. The EMA methodology generally follows guidelines provided in Bulletin 17B of the U.S. Interagency Advisory Committee on Water Data (1982) by using the log-Pearson Type III distribution fitted by the method-of-moments for estimating discharge frequency; however, the EMA methodology provides updated procedures for incorporating historical and censored peak discharges. Use of the EMA approach and the Multiple Grubbs-Beck test are as recommended by the Hydrologic Frequency Analysis Work Group (https://acwi.gov/hydrology/Frequency/HFAWG-B71B-Recommended-Rev-SOH-12June2013-revised-final.pdf). Software developed by the USGS to analyze peak discharge data (PeakFQ version 7.3, https://water.usgs.gov/software/PeakFQ/) was used for these computations (Veilleux and others, 2013).

As part of the frequency analysis, the skewness values for each streamgage were computed as a weighted average $\gamma_w$ of the station and regional skew values following the method suggested in Bulletin 17B (U.S. Interagency Advisory Committee on Water Data, 1982):

$$\gamma_w = \frac{AVP_{new}\gamma_S + MSE[\gamma_S]\gamma_R}{AVP_{new} + MSE[\gamma_S]}, \quad \text{(4)}$$

where

$\gamma_w$ is the weighted skewness,

$\gamma_S$ is the station skewness computed as described in the “Regional Skew Analysis” section and in appendix 1,

$\gamma_R$ is the regional skewness computed as implemented in PeakFQ for each streamgage,

$AVP_{new} = 0.19$ is the average variance of prediction at a new site computed as described in appendix 1, and

$MSE[\gamma_S]$ is the mean square error of the at-site skewness, which also is computed by using the EMA approach as implemented in PeakFQ.
Basin Characteristics

The starting point in defining basin characteristics for regional statistical analyses is to delineate the basin boundaries. Following delineation, basin characteristics for this study were computed following two main approaches. The first was the traditional approach in projects of this type, using spatially averaged or basin-wide properties such as drainage area, fractions of various land uses and soil types, average soil properties, and physiographic properties such as basin slope. The second approach was to develop and analyze spatially distributed basin characteristics based on properties of estimated instantaneous unit hydrographs (for example, Grimaldi and others, 2010). The spatially distributed approach was proposed as a way to see if the effects of basin shape and the location, not just the amount, of different types of hydrologically relevant characteristics such as urbanized areas and open water and wetland on flood-frequency characteristics were significant and could improve flood-frequency prediction. Preliminary results with the spatially distributed approach were not promising and as a result it was dropped from further consideration. As this approach was not used in the final results, it is not discussed further in this report.

Basin Delineation

Most basins were delineated by using the online Illinois StreamStats application (https://streamstats.usgs.gov/), which is based on Environmental Systems Research Institute’s (Esri’s) Arc Hydro Tools software (Maidment, 2002). Illinois StreamStats uses the 1:100,000-scale National Hydrography Dataset (NHD) for hydrography and the 30-meter National Elevation Dataset (NED) resampled to a 10-meter resolution for elevation information for Illinois, Wisconsin, and Indiana. Seven Indiana basins were delineated using the Indiana StreamStats application, a similar application to Illinois StreamStats, because the results were slightly better in delineating the Kankakee River than the Illinois StreamStats results. Eight Wisconsin basins in the hydrologic unit 0404 (Lake Michigan) were provided by the USGS Wisconsin Water Science Center because the Illinois StreamStats does not cover that area and Wisconsin StreamStats did not exist at the time of the analysis. Resulting drainage areas from the Illinois StreamStats application were compared to the drainage areas of hand-drawn boundary (HDB) basin delineations, which are considered to be the most accurate delineation type. For those basins with discrepancies larger than about 10 percent, it was decided that the HDB delineations were preferable. These basins are usually small urban basins with records that ended around 1980, where it is difficult to determine basin boundaries because of urban development and the related potential for later modifications to the hydrography and topography.

Spatially Averaged Basin Characteristics

A common challenge in projects of this type is to select among the many basin characteristics that are available in GIS databases. As this project is in part an update of the regional regression equations for region 2 from the rural flood-frequency study by Soong and others (2004), which used spatially averaged basin characteristics, it was decided to simplify the process of the selection of basin characteristics by considering only the categories of characteristics used by Soong and others (2004) in that region, which were drainage area, slope, and wetland-open water fraction, plus measures of urbanized area (table 3).

The slope variable used in Soong and others (2004) is main channel slope. The determination of the main channel requires sometimes lengthy computations in StreamStats and is in any case ambiguous, so alternative slope variables were considered for this study. These alternatives included the basin average of digital elevation model (DEM) slopes, the basin average of the State Soil Survey Geographic Database (SSURGO) soil slopes, and basin slope computed as maximum–minimum basin elevations divided by basin perimeter (table 3).

The wetland-open water variable used in Soong and others (2004) was percent water and herbaceous wetland from the 2001 NLCD (Soong and others, 2004, p. 16). Additional open water and wetland variables derived from the 2011 NLCD, National Wetlands Inventory (NWI), and SSURGO were considered in this study (table 3).

The urbanized area variable used for temporal analysis and adjustment of the annual maximum peak discharge records (Over and others, 2016) was derived from a year 2000 U.S. Census housing density database developed by David Theobald (Theobald, 2005). For the spatial modeling of the peak discharge quantiles, it was preferred that the urbanization variable be something that is regularly updated; therefore developed area fractions and impervious percentage from the 2011 NLCD were chosen, with the Theobald housing density urbanization estimates retained for comparison (table 3).

Regional Temporal Regression Analysis and Adjustment

The purpose of the temporal analysis of the annual maximum peak discharges is to provide a means of adjusting the peak discharge records from the streamgages used in the study to their levels of urbanization at a common year (2010). This adjustment is needed to provide approximately stationary and contemporary streamgage records. The method of the analysis is to use regression to determine values of coefficients
that characterize the regional average effect of urbanization on peak discharge quantiles as a function of exceedance probability in Illinois flood-frequency region 2. The coefficients also are used to develop factors for adjusting estimates of peak discharge quantiles in other regions of Illinois for the effect of urbanization. The adjusted annual maximum discharges and associated peak discharge quantiles also are made available herein for general use (see table 2 for quantiles and the “Adjustment of Annual Maximum Peak Discharges to Current (2010) Urban Fractions” section for discharges).

The effects of changes in urbanization in a streamgage record are potentially problematic in a regression analysis of peak discharge quantiles on basin characteristics for at least two reasons. One is that peak discharge records on basins with substantial changes in the fraction of urbanization over the streamgage period-of-record are difficult to use because no single value of urbanization fraction characterizes the entire record. The second is that historical urbanization information is needed to characterize the urban fraction of records that ended some time ago and urbanization of their basins continued after the records ended. Because both of these situations are fairly common in the study region (fig. 6), a method for adjustment of peak discharge records to their levels of urbanization at a common year was needed if such stations were going to be included in the study.

The methods used here for temporal analysis and adjustment were described by Over and others (2016); a summary of the application of their method in this study follows. Over and others use a two-step regression method applied to annual maximum peak discharge records as a function of contemporaneous urbanization and precipitation. In the first step, after breaking the peak discharge records into segments at the time of major abrupt changes in the basins, if any, ordinary least-squares (OLS) regression is used to estimate an intercept for each segment. In the second step, after subtracting the segment intercepts to homogenize the data across all the segments, the homogenized data are analyzed with quantile regression, which provides regional urbanization and precipitation coefficients for a sequence of AEPs and a means of assigning AEPs to each flood-discharge observation. The urbanization coefficient for the AEP assigned to each peak discharge observation is then used to adjust the observation to the estimated 2010 urbanization level in the basin where it was observed. Adjustment for changes in precipitation was not implemented, because, unlike changes in urbanization, changes in precipitation are highly uncertain and may reverse themselves in the future.

**Data Used in This Analysis**

In addition to the annual maximum peak discharge data from the nonredundant streamgages, three types of data were used for the temporal analysis and adjustment method applied in this study: (1) historical urbanization, (2) precipitation associated with each peak discharge, and (3) information on reservoirs and other structural flood-control measures such as channelization.

The historical urbanization data used, as in Over and others (2016), were those of Theobald (2005), which are 1940–2010 decadal housing density data based on the 2000 U.S. Census for 10 categories of housing density (including undeveloped) plus nonhousing urbanization (commercial, industrial, and transportation) (fig. 3). Theobald’s nonhousing urbanization estimates, which are constant year 2000 values, were adjusted in proportion to changes in housing density for each basin according to a technique presented in Over and others to obtain 1940–2010 decadal estimates. Following Over and others, the decadal values of housing density classes 7–10 (fig. 3), which encompass areas with housing densities of no more than 10 acres per unit plus adjusted nonhousing urbanization, were linearly interpolated to annual values and used as the adjustment measure for the temporal analysis and adjustment in this study.

To account for effects of trends in precipitation alongside the effects of urbanization on peak discharges, precipitation values were estimated for each observed peak discharge used in this study. Because the density of daily precipitation gages is much higher than that of hourly gages, daily data were used. Three sets of daily precipitation data were used for the estimates: (1) National Weather Service (NWS) cooperative network (COOP) stations in the study area; (2) the Cook County Precipitation Network (CCPN) (Westcott, 2015); and (3) precipitation data collected by the Argonne National Laboratory (http://www.atmos.anl.gov/ANLMET/, accessed March 10, 2011). Data from COOP were downloaded from the Midwestern Regional Climate Center (MRCC) Applied Climate System (MACS) at https://mrcc.illinois.edu/CLIMATE/. Missing data during the study period (WYs 1940–2009) were filled by using a weighted average of values at neighboring gages, where weights were determined by an inverse-distance weighting technique (Over and others, 2016). Precipitation data were distributed spatially by using Thiessen polygons (Thiessen, 1911) (fig. 7), and daily values for the basin upstream from each streamgage in the study were computed as area-weighted average of the values in each Thiessen polygon overlying the basin. An estimate of the precipitation associated with each annual maximum peak discharge was obtained by selecting the maximum daily precipitation during the period from 3 days before to 1 day after the date of the peak discharge. Precipitation data dated after the peak discharge were used because NWS cooperative precipitation measurements typically are made during the early morning of the date assigned to the data, and therefore include precipitation from the previous day.

Information on reservoirs and other structural flood-control measures were obtained by accessing the 2013 National Inventory of Dams (NID; U.S. Army Corps of Engineers, 2013), reviewing relevant reports, inquiring of various Federal, State, county, and local agencies involved in flood control in northeastern Illinois, and searching historical imagery on Google Earth. By reviewing this information, reservoirs and
Figure 7. Locations of precipitation stations used in this study and their Thiessen polygons.
other structural flood-control measures (for example, channelization) within the basins of the streamgages used in this study were determined. For the identified reservoirs, three particular quantities were obtained: (1) the year of construction, (2) the volumetric capacity, and (3) the basin area upstream from the reservoir or other flood-control measure.

The volumetric capacity and upstream area were used to estimate the magnitude of the hydrologic effect of the reservoir. When a reservoir controlling at least 10 percent of the streamgage drainage area and with a volumetric storage capacity divided by the streamgage drainage area of 0.40 inches or other substantial change in the effect of a basin’s structural flood-control measures was deemed to have changed during the period-of-record of a streamgage used in the study, the year of the change was dropped from the temporal analysis and the remaining record was broken into segments at that year. After the segment y-intercept values were obtained by a preliminary run of the first, OLS, step of the temporal analysis procedure, if the intrastreamgage y-intercept values were anomalous (for example, showing a significant increase following the construction of a flood-control reservoir), further research was completed on the basin history. If no reason to explain the anomaly was found, the segment break where the anomaly occurred was dissolved and the record was reconnected. The final set of segments used is presented in table 4 (available at https://doi.org/10.3133/sir20165050). According to table 5, all but 18 of the 117 nonredundant streamgages have only 1 segment, and only 3 streamgage records were split into more than 2 segments. Peak discharges associated with segments shorter than 5 years, historic peak discharges, censored peak discharges, and peak discharges for which the date of the event is not known also were dropped from the temporal analysis.

### Temporal Analysis of Urbanization Effect

The annual maximum peak discharges at the nonredundant streamgages during the study period, separated into segments as necessary based on the construction of structural flood-control measures with certain years of data dropped as described in the “Data Used in This Analysis” section, along with associated precipitation and urban fraction values, constituted the database for the temporal analysis of the effect of urbanization (table 6, available at https://doi.org/10.3133/sir20165050). The results of this analysis were used to adjust the annual maximum peak discharges to their urban fractions at a common year (2010) based on an analysis to determine the regional effect of changes in urban fraction on peak discharge magnitude.

In addition to the peak discharge values dropped from the temporal analysis because of segment breaks, certain coded peak discharge values were also dropped. Peak codes as they appear in the peak discharge data downloaded from NWIS are given in table 6, column titled “NWIS peak codes,” and whether or not a particular peak discharge value was used in the temporal regression analysis to perform the urbanization adjustment is indicated in table 6, column titled “Discharge value used in adjustment regression.” As documented by those two columns of table 6, peak discharge values with codes of 1, 4, 7, 8, or B were dropped from the temporal regression analysis. These codes indicate that the published values are subject to censoring (codes 1, 4, 8), are historic peaks (code 7), or the date of the peak discharge event is not known (code B). These peak discharge values were nevertheless adjusted, and the adjusted values appear in table 6. These values also were used in the spatial regression analysis, because the spatial analysis does not require knowing the date of the event and the EMA methodology is designed to handle censored and historic values, as discussed in the “Frequency Analysis” section.

The temporal analysis was carried out by the method described in Over and others (2016), which follows a two-step linear regression technique suggested by Canay (2011), where the first step uses OLS regression to estimate intercepts for each segment of peak discharge data, and the second step applies quantile regression (Koenker, 2005) to the peak discharge data with the segment intercept values subtracted. The segment intercepts are subtracted from the peak discharge observations to obtain segment-independent observations that are approximately homogeneous (that is, segment-independent), to which quantile regression can be applied. Quantile regression is used in the second step because it provides an estimate of the coefficients of the regression model for different quantiles (exceedance probabilities). Because of this property, the disparate effect of urbanization at different exceedance probabilities (Espey and Winslow, 1974, and references therein; Allen and Bejcek, 1979; Konrad, 2003, and references therein; Over and others, 2016, and references therein) can be estimated, and the adjustment of the peak discharges to their urban fractions at a common year (2010) can take this disparate effect into account.

The independent variables in the two-step regression analysis are urban fraction and precipitation. Precipitation

<table>
<thead>
<tr>
<th>Number of segments per streamgage record</th>
<th>&gt;0</th>
<th>&gt;1</th>
<th>&gt;2</th>
<th>&gt;3</th>
<th>&gt;4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of records</td>
<td>117</td>
<td>18</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>
is used in the regression analysis not so that the peak discharges can be adjusted to a present (2010) precipitation regime, as it was beyond the scope of this study to establish one, but so that by including precipitation, trends that happened to exist in the precipitation causing the observed peak discharges would not be falsely interpreted as the effect of changes in the urban fraction.

The following linear regression model was used for the least-squares regression analysis and was the first step of the method:

\[
y_i(t) = \hat{y}_i(t) + e_i(t) = a_i + b_1 P_i(t) + b_2 U_i(t) + e_i(t), \quad i = 1, 2, \ldots, N,
\]

where

- \( y_i(t) = \log_{10} Q_i(t) \) is the base-10 logarithm of the annual maximum peak discharge at the \( i \)th segment during year \( t \),
- \( \hat{y}_i(t) \) is the fitted value of \( y_i(t) \) according to the regression relation,
- \( a_i \) is the intercept of the \( i \)th segment,
- \( P_i(t) \) and \( U_i(t) \) are the precipitation and urban fraction, respectively, for the \( i \)th segment during year \( t \),
- \( b_1 \) and \( b_2 \) are the regression coefficients for precipitation and urban fraction, respectively, and
- \( e_i(t) = y_i(t) - \hat{y}_i(t) \) is the regression error for the \( i \)th segment during year \( t \).

Notice there is one intercept for each segment, whereas there is only one value of each of the coefficients; therefore they are estimated from the effect of precipitation and urbanization on all the segments combined, and are equivalent in value to an uncertainty-weighted mean of the coefficients of segment-by-segment least-squares regression analysis (Frees, 2004, p. 32). Therefore, the segment intercepts contain all the information that distinguishes one segment from another; the effects of urbanization and precipitation, captured in their coefficients in equation 5, are assumed to be the same for all segments. The computation of the regression model was done in R (R Core Team, 2014) using the plm function from the plm package (Crossett and Millo, 2008).

The coefficients obtained by fitting the regression relation (eq. 5) are given in table 7, and the values of the intercepts for each segment are given with the other segment information in table 4 and are plotted in figure 8. As can be seen from the linear relation with limited scatter in figure 8, the segment intercept values are mostly but not completely explained by drainage area. According to the coefficients in table 7, there is a small, positive effect of precipitation, which is highly significant because the magnitude of the coefficient is much larger than its standard error (SE). The urban fraction coefficient is also positive but larger and also highly significant. Because the regression relation is of the form \( \log_{10} Q = a + b_2 U \), the ratio of peak discharge for two different values of urban fraction \( U_1 \) and \( U_2 \) at the same site is given by \( Q_2/Q_1 = 10^{b_2(U_2-U_1)} \); the value of the urban fraction coefficient \( b_2 \) indicates that for a 10-percent increase in urban fraction, the ratio of peak discharges is \( Q_2/Q_1 = 10^{0.507 \times 0.10} = 1.124 \), or about a 12-percent increase. Although the urbanization coefficient from this step of the analysis method is not used for adjustment of the peak discharges, the median value from the results of the next step is quite similar, so this coefficient gives an approximation of the average result.

In the second, quantile regression-based, step of the method, the distinction between the segments is removed by subtracting the applicable segment intercept value from the dependent variable values for each segment, that is,

\[
y'_i(t) = y_i(t) + a - a_i,
\]

where

\[
a = \frac{\sum_i T_i a_i}{\sum_i T_i},
\]

where \( T_i \) is the number of observations in the \( i \)th segment, is the weighted mean of the segment intercepts obtained from the least-squares regression.

---

**Table 7.** Results of ordinary least-squares linear regression of 117 streamgages in northeastern Illinois and adjacent states to diagnose the temporal effects of urbanization and precipitation on annual maximum peak discharges.

<table>
<thead>
<tr>
<th>Precipitation coefficient ( b_1 )</th>
<th>Precipitation coefficient SE</th>
<th>Urban fraction coefficient ( b_2 )</th>
<th>Urban fraction coefficient SE</th>
<th>Weighted mean segment intercept (log units)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.110</td>
<td>0.005</td>
<td>0.508</td>
<td>0.042</td>
<td>2.388a</td>
</tr>
</tbody>
</table>

*See equation 6 for the computation of the weighted mean segment intercept value.*
The independent variables are again the urban fractions and precipitation values associated with each peak discharge.

The quantile regression model is as follows:

\[ y'_k = \hat{y}'_k(p) + \epsilon'_k(p) = a'(p) + b'_1(p)P_k + b'_2(p)U_k + \epsilon'_k(p), \quad (7) \]

where

- \( p \) is an element of a sequence of AEPs (table 8),
- \( \hat{y}'_k(p) \) is the fitted value of the quantile with \( AEP = p \) of the \( k \)th log-transformed peak discharge value \( y'_k \),
- \( a'(p) \) is the intercept of the fitted linear relation between \( y'_k(p) \) and the independent variables,
- \( P_k \) and \( U_k \) are the precipitation and urban fraction, respectively, of \( y'_k \),
- \( b'_1(p) \) and \( b'_2(p) \) are the quantile regression coefficients as a function of \( p \) (AEP) for precipitation and urban fraction, respectively, and
- \( \epsilon'_k(p) = y'_k - \hat{y}'_k(p) \) is the regression error for the \( k \)th observation at \( AEP = p \).

This model was solved with the rq function of the quantreg package (Koenker, 2013) of the R language.

Two post-processing steps were applied to obtain the final quantile regression coefficients and their SEs. First, a bootstrap resampling approach suggested by Canay (2011) was applied to obtain SEs and unbiased mean coefficient values. Second, a continuous and monotonic version of the bootstrap mean urban fraction coefficients as a function of exceedance probability was obtained by fitting a seventh-order polynomial. The original and bootstrap mean and SE values for all the quantile regression coefficients, along with the fitted urban fraction coefficients, are given in table 8 (available at https://doi.org/10.3133/sir20165050), and the three sets of urban fraction coefficients are plotted in figure 9. A subset of the urbanization coefficients at the AEPs of the standard flood-frequency quantiles are given in table 9.

The results in tables 7 through 9 indicate that the median urban fraction and precipitation coefficients are quite similar to the values computed by the OLS regression carried out in the first step of the analysis, with values of about 0.5 and 0.1. Results in these tables indicate that the precipitation coefficient varies only slightly with exceedance probability, whereas the urban fraction coefficient has a strong, mostly monotonic trend with exceedance probability, from a maximum value of 0.929 at an exceedance probability of 0.99 down to 0.294 at an exceedance probability of 0.002 (table 8). The dependence of the urban fraction coefficient on exceedance probability gives direct empirical evidence that urbanization in the study area...
**EXPLANATION**

- Quantile regression (QR) coefficient
- Mean bootstrapped QR coefficient
- Seventh-order polynomial fit to mean QR coefficient
- Standard error of mean QR coefficient

**Figure 9.** Urban fraction coefficients from temporal regression analysis of 117 streamgages in northeastern Illinois and adjacent states, as a function of exceedance probability.

**Table 9.** Quantile regression coefficients of urban fraction from temporal analysis of 117 streamgages in northeastern Illinois and adjacent states, at selected annual exceedance probabilities.

<table>
<thead>
<tr>
<th>AEP, annual exceedance probability</th>
<th>0.5</th>
<th>0.2</th>
<th>0.1</th>
<th>0.04</th>
<th>0.02</th>
<th>0.01</th>
<th>0.005</th>
<th>0.002</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>0.487</td>
<td>0.405</td>
<td>0.391</td>
<td>0.359</td>
<td>0.331</td>
<td>0.312</td>
<td>0.301</td>
<td>0.294</td>
</tr>
<tr>
<td>Standard error</td>
<td>0.075</td>
<td>0.080</td>
<td>0.078</td>
<td>0.085</td>
<td>0.086</td>
<td>0.134</td>
<td>0.112</td>
<td>0.115</td>
</tr>
</tbody>
</table>
affects the magnitudes of the more common, large exceedance probability peak discharges more than it does the rarer, small exceedance probability peak discharges.

**Adjustment of Annual Maximum Peak Discharges to Current (2010) Urban Fractions**

Following the application of the two-step method to obtain exceedance probability-dependent urbanization and precipitation coefficients, each peak discharge is assigned an exceedance probability by a procedure described in Over and others (2016). This procedure was developed to adjust only for urbanization and therefore reduces the three-dimensional log$_10Q$, $U$, and $P$ space in which the least-squares and quantile regressions were carried out to a two-dimensional log$_10Q$ and $U$ space in which the quantile regression results are a set of lines in that space (fig. 10), where the diminishing effect of urbanization on the larger, rare floods can be seen graphically.

Using the exceedance probability assigned to each peak discharge observation and the continuous urban fraction coefficient function $K_i(p)$, the annual maximum peak discharges were adjusted to the 2010 urban fraction obtained from the Theobald (2005) data by using the following equation applied to each streamgage record segment $i$:

$$y_i^*(t) = a_i - a_i + y_i(t) + b_i(p)[U_i(t) - U_i(t)]$$

where $y_i^*(t)$ and $y_i(t)$ are the base-10 logarithm of the adjusted and unadjusted, respectively, peak discharge value for year $t$ and segment $i$.

**Figure 10.** Segment intercept-subtracted log-transformed annual maximum peak discharge from 181 streamgages in northeastern Illinois and adjacent states as a function of urban fraction with quantile regression and least-squares regression line fits.
\(a_i^*\) and \(a_i\) are the segment intercept values for the most recent and current segments, respectively, of the streamgage record containing segment \(i\).

\(\beta_2(p)\) is the urban fraction coefficient value corresponding to the exceedance probability \(q\) assigned to the peak discharge value \(y_i(t)\), and \(U_i(i^*)\) and \(U_i(t)\) are the urban fraction values for the basin corresponding to segment \(i\) during the year \(i^*\) to which the peak discharges are being adjusted (here, 2010) and the year \(t\) of the observation of the peak discharge \(y_i(t)\), respectively.

Notice that if a streamgage record has only one segment, then \(a_i^* = a_i\), and the quantity \(a_i^* - a_i\) drops out.

For the peak discharges and censoring levels that were dropped from the temporal analysis, where the correct segment could be determined, the adjustment method allows those values also to be adjusted so that these data can be used for at-site frequency analysis. For peak discharge values between two segments, the adjustment was interpolated. As a result, only peak discharge values associated with segments shorter than 5 years were not adjusted and thus were dropped from the at-site frequency analysis.

The set of adjusted values \(Q_i(t) = 10^{\beta_2(p)}\) and the assigned exceedance probability values, \(p\), on which the adjustment depends, are given in table 6. The effects of the adjustment on mean, standard deviation, and skewness of the log-transformed peak discharges at each streamgage in the study are shown in figure 11. The change in mean, which results from adjusting upward peaks with urban fractions less than the 2010 value, is almost always positive, and the change in standard deviation, which results from removal of urban fraction-induced trends in peak discharge in the observed data by the adjustment process and from the adjustment of smaller peak discharges more than larger ones, is almost always negative. The skewness increases more often than it decreases, also showing the effect of adjust-

**Figure 11.** Changes in the mean, standard deviation, and skewness of the log-transformed annual maximum peak discharge records for 181 streamgages used in this study in northeastern Illinois, resulting from adjustment to 2010 urban fraction values.
ing the smaller peaks more than the larger ones. The overall increase in skewness as peaks are adjusted to more urbanized conditions is also consistent with the increase in skewness with urban fraction determined in the spatial analysis, as discussed in the “Regional Skew Analysis” section and appendix 1.

**Regional Spatial Regression Analyses**

After the annual maximum peak discharges were adjusted to their levels of urbanization at a common year (2010) as described in the “Regional Temporal Regression Analysis and Adjustment” section, peak discharge quantiles for the selected AEPs were computed as described in the “Frequency Analysis” section, including weighting with the regional skew model. The computed quantiles constitute the dependent variable for the spatial regression analyses; the spatially averaged basin characteristics developed as described in the “Spatially Averaged Basin Characteristics” section constitute the independent variables.

**Selection of Basin Characteristics**

The predictive capability of the basin characteristics in each category (table 3) were investigated by fitting weighted least-squares (WLS) linear regression models with regression weights in proportion to record length for each AEP for all possible combinations of basin characteristics where at most one variable is selected from each category with log_{10}-transformed drainage area always included. For these exploratory regressions, a preliminary set of peak-flow quantiles developed using regional skew were used. An R script (R Core Team, 2014) was written to do the regression computations. The general form of the regression models fitted in this investigation was as follows in equation 9:

\[
\log_{10} Q_p = (\beta_0)_p + (\beta_1)_p \log_{10} A + (\beta_2)_p (U + c_2)^{d_2} + (\beta_3)_p (W + c_3)^{d_3} + (\beta_4)_p (S + c_4)^{d_4},
\]

\[p = 0.50, 0.20, 0.10, 0.05, 0.02, 0.01, 0.005, 0.002,\]

where

- \(Q_p\) is the estimated peak discharge quantile with regional skew, in cubic feet per second, with annual exceedance probability \(p\),
- \((\beta_0)_p\), \((\beta_1)_p\), \((\beta_2)_p\), \((\beta_3)_p\), and \((\beta_4)_p\) are coefficients estimated by the WLS procedure,
- \(A\) is the drainage area, in square miles,
- \(U\) is an urbanization measure (a decimal fraction),
- \(W\) is a basin water and wetness measure (a decimal fraction),
- \(S\) is a basin slope measure, in feet per mile,
- \(d_2\), \(d_3\), and \(d_4\) are exponents used to transform the distribution of urbanization, water and wetness, and slope measures, respectively, to an approximately Gaussian shape (\(d_i = 1, l/2, 0\), where 1 implies no transformation, \(l/2\) a square-root transformation, and 0 a log_{10}-transformation were tested), and
- \(c_2\), \(c_3\), and \(c_4\) are constants added to the urbanization, water and wetness, and slope measures when a log_{10}-transformation was used and the basin characteristic had zero values).

The selected transformations are listed in table 3.

Various plots of the regression results were made to facilitate examination of the results; the most useful plot was determined to be the adjusted coefficient of determination \((R^2)\) (Helsel and Hirsch, 2002, p. 313) as a function of the \(t\) ratio of the coefficients of variables of the urbanization, slope, and wetness categories (fig. 12), where the \(t\) ratio is the coefficient value divided by its standard error (Helsel and Hirsch, 2002, p. 238). Such a plot shows which models provide the best fit (highest adjusted \(R^2\)) and the most significant coefficient values (largest magnitude \(t\) statistics), and, by retaining the sign of the coefficient values, such plots show if the sign of the coefficient is physically reasonable. The variables whose corresponding regression model fits have the largest magnitude \(t\) ratio values and highest adjusted \(R^2\), along with the sign of the coefficient being physically reasonable, were selected as candidate variables for the development of the final spatial regression equations.

Plots of this type for each of the categories (urbanization, slope, and water and wetland) are shown in figure 12. The urbanization variable with the largest \(t\) ratio values and highest adjusted \(R^2\) values is usually the square root of NLCD_22_23_24, except for AEPs 0.5 and 0.2 where Theobald_Urban is of similar quality, as shown in figure 12_A. Both NLCD_22_23_24 and
Figure 12. Annual exceedance probability (AEP) of adjusted coefficient of determination ($R^2$) as a function of the regression coefficient $t$ ratio values for basin characteristic categories considered in this study in northeastern Illinois: A, urbanization variables; B and C, water and wetland variables; and D, slope variables. When the absolute value of the $t$ ratio is greater than 2, then the coefficient is significantly different from zero at about the $\alpha = 0.05$ significance level (Helsel and Hirsch, 2002, p. 238). The variable names in the figure explanations are defined in Table 3; the strings preceding the variable names in the figure explanations indicate the transformation applied to the variable, where “lin” means “linear” (that is, no transformation), “sqrt” means a square root transformation, and “log10” means a log10 transformation.
Figure 12. Annual exceedance probability (AEP) of adjusted coefficient of determination ($R^2$) as a function of the regression coefficient $t$ ratio values for basin characteristic categories considered in this study in northeastern Illinois: $A$, urbanization variables; $B$ and $C$, water and wetland variables; and $D$, slope variables. When the absolute value of the $t$ ratio is greater than 2, then the coefficient is significantly different from zero at about the $\alpha = 0.05$ significance level (Helsel and Hirsch, 2002, p. 238). The variable names in the figure explanations are defined in table 3; the strings preceding the variable names in the figure explanations indicate the transformation applied to the variable, where "lin" means "linear" (that is, no transformation), "sqrt" means a square root transformation, and "log10" means a log$_{10}$ transformation—Continued.
Figure 12. Annual exceedance probability (AEP) of adjusted coefficient of determination ($R^2$) as a function of the regression coefficient $t$ ratio values for basin characteristic categories considered in this study in northeastern Illinois: $A$, urbanization variables; $B$ and $C$, water and wetland variables; and $D$, slope variables. When the absolute value of the $t$ ratio is greater than 2, then the coefficient is significantly different from zero at about the $\alpha = 0.05$ significance level (Helsel and Hirsch, 2002, p. 238). The variable names in the figure explanations are defined in table 3; the strings preceding the variable names in the figure explanations indicate the transformation applied to the variable, where “lin” means “linear” (that is, no transformation), “sqrt” means a square root transformation, and “log10” means a log10 transformation—Continued.
Figure 12. Annual exceedance probability (AEP) of adjusted coefficient of determination ($R^2$) as a function of the regression coefficient $t$ ratio values for basin characteristic categories considered in this study in northeastern Illinois: A, urbanization variables; B and C, water and wetland variables; and D, slope variables. When the absolute value of the $t$ ratio is greater than 2, then the coefficient is significantly different from zero at about the $\alpha = 0.05$ significance level (Helsel and Hirsch, 2002, p. 238). The variable names in the figure explanations are defined in table 3; the strings preceding the variable names in the figure explanations indicate the transformation applied to the variable, where "lin" means "linear" (that is, no transformation), "sqrt" means a square root transformation, and "log10" means a log10 transformation—Continued.
Theobald_Urban have positive coefficients, as is physically expected. Because only the NLCD-based urbanization variables are expected to be updated in the future, Theobald_Urban was eliminated from consideration for the spatial regression equations and NLCD_22_23_24 was chosen for all AEPs as the candidate urbanization variable for the final equations. Because there are too many water and wetland variables to distinguish on a single plot, the results for these are plotted on two sets of plots, figure 12B and 12C. The variable with the highest adjusted $R^2$ values and largest magnitude $t$ ratio values varies from smaller to larger peak discharge quantiles, being NWI_emergent for the 0.5 annual exceedance probability (AEP) flood, NWI_total for 0.2 to 0.04 AEP floods, and DrainageClass1a for the larger floods (0.02 to 0.002 AEP), and all have physically expected negative coefficient values. Because NWI_total is almost as good as NWI_emergent for the 0.5 AEP flood, it was selected as the candidate variable for the smaller floods (AEPs of 0.5 to 0.04), and DrainageClass1a was selected for the larger floods (AEPs of 0.02 to 0.002). It can be seen in figure 12D that the slope variable with the highest adjusted $R^2$ values and largest magnitude $t$ statistic values for all AEPs (and physically expected positive coefficients) is $\log_{10}(DEM_1_0_0_P)$, so $\log_{10}(DEM_1_0_0_P)$ was chosen for all AEPs as the candidate urbanization variable for the final equations.

### Development of Final Spatial Regression Equations

Following the selection of the most promising basin characteristics in each category by the WLS regression investigation, the final equations were computed with generalized leastsquares (GLS) regression, which generalizes WLS regression by taking into account cross-correlations that arise among nearby streamgages with concurrent records. Stedinger and Tasker (1985) showed that GLS regression for regional spatial flood-frequency analysis provides more accurate coefficients and error estimates than OLS, with more modest improvement compared to WLS regression depending on the amount of cross-correlation among the streamgage records.

The GLS regression models were fitted with the function WREG.GLS from the package WREG (Farmer, 2017) in R version 3.6.3 (R Core Team, 2020). WREG.GLS implements a GLS approach to the estimation of peak discharge quantiles fit to a log-Pearson type III distribution using techniques described in Eng and others (2009) as developed by Stedinger and Tasker (1985), Tasker and Stedinger (1989), Griffis and Stedinger (2007), and Griffis and Stedinger (2009).

The cross-correlations $\hat{\rho}_i$ among streamgage records are estimated by selecting streamgage record pairs with a minimum number of concurrent years and fitting a function of the form

$$\hat{\rho}_i = \theta^{d_i/(\alpha \cdot d_i + 1)}$$

where $d_i$ is the distance between streamgages $i$ and $j$, and $\alpha$ and $\theta$ are dimensionless parameters fitted by a graphical method.

For this study the minimum number of concurrent years was taken as 45, and $\alpha$ and $\theta$ were taken as 0.005 and 0.975, respectively.

When the selected water and wetland variables were used in the GLS regression analysis, it was determined that use of DrainageClass1a gave a better fit than NWI_total from AEP 0.002 to AEP 0.20 and had only a very small difference at AEP 0.50, in the final equations, so DrainageClass1a is used for all AEPs in the final spatial regression equations. Values of the selected variables for each streamgage are given in table 1 and their cross-correlations and variance inflation factors (VIFs) in table 10.

### Table 10. Correlation matrix and variance inflation factor (VIF) values of basin characteristics selected for use in spatial regression analysis of 117 streamgages in northeastern Illinois and adjacent states.

<table>
<thead>
<tr>
<th>$\log_{10}(DA)$</th>
<th>NLCD_22_23_24</th>
<th>DrainageClass1a</th>
<th>$\log_{10}(DEM_1_0_0_P)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\log_{10}(DA)$</td>
<td>1</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>NLCD_22_23_24</td>
<td>-0.156</td>
<td>1</td>
<td>--</td>
</tr>
<tr>
<td>DrainageClass1a</td>
<td>0.288</td>
<td>0.221</td>
<td>1</td>
</tr>
<tr>
<td>$\log_{10}(DEM_1_0_0_P)$</td>
<td>-0.816</td>
<td>0.077</td>
<td>-0.215</td>
</tr>
</tbody>
</table>

Variance inflation factor

| 3.27 | 1.12 | 1.19 | 3.02 |
The general form of selected equations follows in equation 11:

\[
\log_{10} Q_p = (b_0)_p + (b_A)_p \log_{10} A + (b_U)_p U^{1/2} + (b_W)_p W^{1/2} + (b_S)_p \log_{10} S,
\]

\[p = 0.50, 0.20, 0.10, 0.05, 0.02, 0.01, 0.005, 0.002,\]

where

- \(Q_p\) is the peak discharge quantile with weighted skew, in cubic feet per second, with annual exceedance probability \(p\);
- \(A\) is the drainage area, in square miles,
- \(U\) is the selected urbanization measure for the spatial equations NLCD\_22\_23\_24 (a decimal fraction),
- \(W\) is the basin soil wetness measure DrainageClass1a (a decimal fraction),
- \(S\) is the basin slope measure DEM\_1\_0\_P (basin elevation range divided by basin perimeter), in feet per mile, and

\((b_0)_p, (b_A)_p, (b_U)_p, (b_W)_p,\) and \((b_S)_p\) are coefficients estimated by using WREG (table 11).

After inverting the logarithmic transformation, the \(Q_p\) prediction equation 12 is the result:

\[
Q_p = 10^{(b_0)_p} 10^{(b_A)_p} A^{(b_U)_p} 10^{(b_W)_p} W^{1/2} S^{(b_S)_p},
\]

\[p = 0.50, 0.20, 0.10, 0.05, 0.02, 0.01, 0.005, 0.002.\]
Accuracy of Final Spatial Regression Equations

Several measures of the average accuracy of the final GLS spatial regression equations are presented in Table 12, along with comparative values of the standard model error in percent from the corresponding WLS equations and the previous studies by Allen and Bejcek (1979) and Soong and others (2004). All these measures are averages for all the streamgages in the analysis. The pseudo $R^2$ measures the fraction of variability in the dependent variable that is explained by the regression model after removing the effect of the time-sampling error (Eng and others, 2009) and is analogous to the standard $R^2$. The average variance of prediction includes both the model error and time-sampling error components and indicates the expected accuracy of a prediction at an ungauged location with basin characteristics that are near the centroid of the basin characteristics used in the GLS analysis. The standard model error indicates the accuracy of the model fit without the component in the average variance of prediction that indicates the additional error of making a prediction at an ungauged location; as a result it is always somewhat smaller than the average variance of prediction. All the accuracy measures decrease as the AEP decreases, as expected, because of the fewer data available for less frequent events.

Techniques for computing accuracy estimates at individual locations (streamgages or ungauged locations) are also available, including the variance and standard error of prediction and confidence interval values. According to Hodge and Tasker (1995), the variance of prediction $V_i$ at an individual location based on a GLS regression analysis is

$$V_i = \gamma^2 + x_i \left( X' \Lambda^{-1} X \right)^{-1} x'_i,$$

where

- $\gamma^2$ is the model error variance (Table 12),
- $x_i = \left[ 1, \log_{10} A_i, U_{spat,i}, W_i^{1/2}, \log_{10} S_i \right]$ is a row vector specifying the basin characteristics 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 ($n = 117$ is the number of streamgages and $p = 5$ is the number of basin characteristics plus 1)
- $\Lambda^{-1}$ is the matrix inverse of $\Lambda$, the $(n \times n)$ covariance matrix used in the GLS regression analysis, and
- $X'$ is the matrix transpose of $X$; and $x'_i$ is the matrix transpose of $x_i$.

### Table 12. Measures of the average accuracy of the selected spatial regression equations in this study in northeastern Illinois.

[R^2, coefficient of determination; WLS, weighted least squares; --, no data]

<table>
<thead>
<tr>
<th>Annual exceedance probability</th>
<th>Pseudo $R^2$ (percent)</th>
<th>Average variance of prediction (log units)</th>
<th>Average standard error of prediction (percent)</th>
<th>Model error variance $\gamma^2$ (log units)</th>
<th>Standard model error (percent)</th>
<th>Standard error of estimate (percent) (from Allen and Bejcek, 1979; table 4)</th>
<th>Standard model error (from Soong and others, 2004; table 4, applicable to regions 2, 6, and 7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>87.5</td>
<td>0.0362</td>
<td>46.0</td>
<td>0.0341</td>
<td>44.5</td>
<td>45.2</td>
<td>36</td>
</tr>
<tr>
<td>0.2</td>
<td>85.6</td>
<td>0.0384</td>
<td>47.5</td>
<td>0.0360</td>
<td>45.9</td>
<td>46.3</td>
<td>38</td>
</tr>
<tr>
<td>0.1</td>
<td>83.9</td>
<td>0.0414</td>
<td>49.6</td>
<td>0.0387</td>
<td>47.7</td>
<td>48.6</td>
<td>40</td>
</tr>
<tr>
<td>0.04</td>
<td>81.0</td>
<td>0.0473</td>
<td>53.4</td>
<td>0.0439</td>
<td>51.2</td>
<td>52.4</td>
<td>43</td>
</tr>
<tr>
<td>0.02</td>
<td>78.6</td>
<td>0.0523</td>
<td>56.6</td>
<td>0.0485</td>
<td>54.2</td>
<td>55.8</td>
<td>45</td>
</tr>
<tr>
<td>0.01</td>
<td>76.5</td>
<td>0.0571</td>
<td>59.5</td>
<td>0.0528</td>
<td>56.8</td>
<td>59.4</td>
<td>48</td>
</tr>
<tr>
<td>0.005</td>
<td>74.2</td>
<td>0.0622</td>
<td>62.5</td>
<td>0.0573</td>
<td>59.6</td>
<td>63.2</td>
<td>--</td>
</tr>
<tr>
<td>0.002</td>
<td>71.7</td>
<td>0.0681</td>
<td>66.0</td>
<td>0.0624</td>
<td>62.6</td>
<td>68.5</td>
<td>52</td>
</tr>
</tbody>
</table>
The $(X'X)^{-1}$ matrices for the selected AEPs are given in table 13 (available at https://doi.org/10.3133/sir20165050).

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, $S_p = V_i^{1/2}$ and in percent units by using the following formula:

$$S_p = 100 \left\{ \exp \left[ \ln(10) V_i \right] \right\}^{1/2}$$

(Eng and others, 2009).

The confidence intervals of the predicted discharge quantile $Q_i$ are computed as

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

where $t_{\alpha/2,n-p}$ is the critical value of the $t$ distribution at the alpha level $\alpha$ (for example, $\alpha = 0.05$ for 90-percent confidence intervals) and $n - p$ degrees of freedom, where $n = 117$ is the number of streamgages used in the spatial regressions and $p = 5$ is the number of basin characteristics plus 1. After inverting the logarithmic transformation, the interval is

$$Q_i = 10^{\log_{10} Q_i - t_{\alpha/2,n-p} S_p} < Q_i < 10^{\log_{10} Q_i + t_{\alpha/2,n-p} S_p}.$$  \hspace{1cm}(16)

In equation 16, $Q_i$ is the median prediction, and $Q_i 10^{\log_{10} Q_i - t_{\alpha/2,n-p} S_p}$ and $Q_i 10^{\log_{10} Q_i + t_{\alpha/2,n-p} S_p}$ are the upper and lower 1 - $2\alpha$ confidence limits, respectively. For example, for $\alpha = 0.05$ (therefore, 90-percent 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 $Q_i 10^{\log_{10} Q_i + t_{\alpha/2,n-p} S_p}$ is $\alpha = 0.05$; likewise the probability that the true value of the discharge quantile is less than the lower limit $Q_i 10^{\log_{10} Q_i - t_{\alpha/2,n-p} S_p}$ is $\alpha = 0.05$, so that the probability that the true value lies between the limits is 90 percent.

### Applications of Regression Equations

There are two sets of regression equations presented in this report, the temporal (eq. 7 and table 9) and the spatial (eq. 12 and table 11). In both sets, the values of the urbanization coefficients are positive, as expected, and the urbanization coefficients from the temporal analysis decrease substantially as the peak discharge quantiles increase, that is, with decreasing AEP. The changes in the urbanization coefficients from the spatial regression equations as a function of AEP are more modest and have a minimum at AEP $= 0.1$, with increases toward both smaller and larger AEP values; in fact the largest urbanization coefficient occurs at AEP $= 0.002$. The increase of the urbanization coefficients from the spatial analysis for decreasing AEPs beyond AEP $= 0.1$ does not agree with the expectation that the effect of urbanization decreases with decreasing AEP and therefore is considered anomalous. Reasons for this anomaly and implications for applications of the spatial equations are discussed below.

Because the urbanization measures to which these coefficients apply are different but approximately commensurate (fig. 13), a direct comparison of their implications regarding the effect of urbanization can be considered as meaningful. Plots of the ratios of peak discharge quantiles for the different AEPs considered here are shown in figure 14. These plots show that except for very small amounts of urbanization beginning with a completely rural basin, the larger coefficients from the temporal analysis imply larger effects of urbanization, especially for the larger AEPs (smaller floods). They also show the effect of the minimum value of the spatial coefficient occurring at AEP $= 0.1$; for example, the largest effect of urbanization in the spatial curves is seen for AEP $= 0.002$.

The urbanization coefficients in the temporal equations are estimated by a temporal analysis, that is, directly from the changes in the annual peak discharges as the basins urbanize. In the spatial equations, the urbanization coefficients are not computed from observed changes in the gaged basins, but indirectly, from the differences in flood-discharge quantiles among basins of different levels of urbanization. When applying flood-discharge regression equations, one is typically interested in the behavior at an ungaged location, and this spatial transfer of information is entirely appropriate. When it comes to urbanization, however, the primary interest is the temporal question, that is, how did or how will the flood-discharge quantiles change when the basin is urbanized. Because they are based on a direct analysis of the results of this process, the temporal urbanization coefficients are more likely to reflect the effects of urbanization on a given basin.

Two general statistical issues, at least, that affect the values of regression coefficients in a multiple regression framework also need to be considered when the value of the coefficients is of interest, as in this study with respect to the urbanization coefficients. These issues are (1) omitted variable bias (OVB) and (2) measurement errors in the explanatory variables. The issue of OVB arises when an omitted variable is correlated with both the variable of interest and the dependent variable (Greene, 1997). In this case, an omitted variable will cause a negative or positive bias in the coefficient of the variable of interest, though it does not bias the predictions of the equations. In the context of the present analysis, if there is some variable that is important in predicting the flood-discharge properties that is left out of the spatial or temporal analyses but is correlated with urbanization, the urbanization coefficient will not have its correct value. Measurement errors in an explanatory variable cause attenuation (reduction of the absolute value) of the variable, again without biasing the prediction of the equations (Fuller, 1987).
It is possible that OVB is affecting either or both sets of coefficients, but because the urbanization coefficients from the spatial analysis are significantly smaller than those from the temporal analysis and measurement error always causes a reduction in positive coefficients, whereas OVB may cause an error in either direction, the first hypothesis that the results suggest is that the spatial urbanization coefficients are subject to a measurement error. In particular, if the relation between urbanization and its effects on peak discharges varies among basins because of different development practices related to stormwater and flood control, this variation would induce a measurement error relative to the hydrologically effective urbanization value. Such an error would, in addition, affect the spatial analyses more than the temporal ones, because the temporal analyses depend only on the changes in time of the peak discharges with interbasin differences captured in the segment intercepts, whereas the spatial analyses depend directly on the values of the peak discharges.

Given this concern regarding the accuracy of the urbanization coefficients from the spatial models (which nevertheless does not bias the predictions) and the previous assertion that the basic question when it comes to the effect of urbanization is a temporal one, it seems clearly preferable to use temporal equations to predict the effect of urbanization where possible. It is not possible to use the temporal equations for adjustment of rural estimates in ungaged basins in region 2 because there are not enough truly rural basins to develop rural-only equations unaffected by urbanization (fig. 4). Therefore to obtain flood-quantile estimates for ungaged basins in region 2, the spatial regression equations should be used. But for situations where peak discharge quantile estimates are being adjusted for the effects of urbanization, the temporal equations should be used. Such situations include ungaged basins outside region 2, assuming urbanization has proceeded similarly as for the gaged basins analyzed in this study. Such situations may also include the prediction of future effects of urbanization for basins in region 2, again assuming urbanization will proceed as it did for the basins analyzed in this study. These situations are discussed in more detail in the “Applications of the Temporal Urbanization Coefficients” section.
Figure 14. Ratios of peak discharge quantiles $Q_p(U)/Q_p(U_0)$ as a function of urban fraction $U$ for selected annual exceedance probabilities (AEPs) $p$ as implied by the two sets of urbanization coefficients obtained in this study in northeastern Illinois: A, basin initially having no urbanization ($U_0 = 0$); B, basin initially 50-percent urbanized ($U_0 = 0.5$).
Applications of the Spatial Regression Equations

The details of application of the updated Illinois region 2 spatial regression equations depend on the nearness of the location of interest to a streamgage. If the location of interest is far from a streamgage, the spatial regression equations are used. If the location is at a streamgage, a weighted average of the quantiles from regional equations and the quantiles from the streamgage record itself are used. If the location is near a streamgage, a weighted average of the quantiles from regional equations and an adjusted version of the quantiles from the streamgage record itself are used. The details of the different applications are described in this section.

Ungaged Locations Far from a Streamgage

For a general ungaged location in Illinois flood-frequency region 2, the final, real-space regression equation 12 should be used. The USGS StreamStats application (https://streamstats.usgs.gov) could be used to select the location of interest, delineate the basin, and compute the basin characteristics and the peak discharge quantiles.

It is important to realize that the regression equations represented by equation 12 only apply within the ranges of the basin characteristics used to fit the equations. These ranges are given in table 14. StreamStats will not enforce this limitation, though it does provide a warning.

At a Streamgage

Because of the finite length of streamgage records, the regional equations can improve the accuracy of the peak discharge quantile estimates at streamgages by incorporating regional information. The procedure recommended by Cohn and others (2012; see also Tasker, [1975]) is to compute this peak discharge quantile from the weighted average of the regression equation estimate and the result of the frequency analysis of the streamgage record, where the weights are the inverses of the variance of each of the discharge estimates. The weighted discharges are computed with the following equation:

\[
\log_{10}\left(Q_{p,g,w}\right) = \left(V_{p,g,r} + V_{p,g,s} + V_{p,g,gr} + V_{p,g,gs}\right) \left(V_{p,g,r} + V_{p,g,s} + V_{p,g,gr} + V_{p,g,gs}\right)
\]

where

- \(Q_{p,g,w}\) is the weighted peak discharge quantile at the streamgage for an AEP of \(p\),
- \(Q_{p,g,s}\) is the peak discharge quantile for an AEP of \(p\) computed from the streamgage record by using EMA with weighted skew as described in the “Frequency Analysis” section,
- \(Q_{p,g,gr}\) is the peak discharge quantile for the selected AEP obtained from regression equation 12 applied at the streamgage,
- \(V_{p,g,r}\) is the variance of prediction of \(Q_{p,g,r}\) for the an AEP of \(p\) computed by using equation 13, and
- \(V_{p,g,s}\) is the variance of prediction of \(Q_{p,g,s}\) for the selected AEP computed by PeakFQ as part of the frequency analysis.

The values of \(Q_{p,g,s}\), \(Q_{p,g,gr}\), and \(Q_{p,g,gs}\) for both the redundant and nonredundant stations in this study are tabulated in table 2, in the third, fourth, and fifth rows, respectively, for each streamgage.

The variance of prediction for the weighted discharge \(Q_{p,g,w}\) is given by

\[
\left(V_{p,g,w}\right) = \left(V_{p,g,s}\right) + \left(V_{p,g,gr}\right) + \left(V_{p,g,gs}\right)
\]

(Tasker, 1975). With this value of \(V_{p,g,w}\), the standard error of prediction \(S_{p,g,w}\) can be computed as \(S_{p,g,w} = \left(V_{p,g,w}\right)^{1/2}\), and then equations 14–16 can be used to compute the standard error of prediction in percent units and confidence intervals.

<table>
<thead>
<tr>
<th>Basin characteristic name</th>
<th>StreamStats name</th>
<th>Units</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drainage area</td>
<td>DRNAREA</td>
<td>square mile</td>
<td>0.078</td>
<td>1.351</td>
<td>13.6</td>
</tr>
<tr>
<td>NLCD_22_23_24</td>
<td>FLC11DVLMH</td>
<td>decimal fraction</td>
<td>0.0022</td>
<td>0.979</td>
<td>0.391</td>
</tr>
<tr>
<td>Theobald_Urban</td>
<td>URBTHE2010</td>
<td>decimal fraction</td>
<td>0.000</td>
<td>1.00</td>
<td>0.580</td>
</tr>
<tr>
<td>DrainageClass1a</td>
<td>FSSURGDC78</td>
<td>decimal fraction</td>
<td>0.000</td>
<td>0.256</td>
<td>0.0511</td>
</tr>
<tr>
<td>DEM_1_0_P</td>
<td>RELRELF</td>
<td>foot per mile</td>
<td>0.821</td>
<td>37.2</td>
<td>4.79</td>
</tr>
</tbody>
</table>
Ungaged Location Near a Streamgage

If the ungaged location of interest is near to and on the same stream as a streamgage, the accuracy of the flood-discharge 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, 2007). There are a few different methods in the literature for this adjustment; the method used here follows that of Soong and others (2004) (as corrected August 10, 2010), which is the same as the method presented in the IDOT Drainage Manual (Drainage Manual Committee, 2011). According to this method, the near-gage adjustment has an effect only if the ratio $A_i/A_g$ of the drainage area of the ungaged basin of interest $A_i$ to that of a gaged basin $A_g$ is between 0.5 and 1.5 (see eq. 19); 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

$$w_a = \begin{cases} 2 & \text{if } 0.5 < (A_i/A_g) < 1.5 \\ 1 & \text{otherwise} \end{cases}$$  \hspace{1cm} (19)

where

- $A_i$ 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:

$$(Q_p)_{u,w} = w_a (Q_p)_{u,r} + (1-w_a) (Q_p)_{g,w} (A_i/A_g)$$  \hspace{1cm} (20)

where

- $(Q_p)_{u,w}$ is the gage-adjusted flood-discharge quantile estimate for an AEP of $p$ at the ungaged location of interest,
- $(Q_p)_{u,r}$ is the flood-discharge quantile estimate for an AEP of $p$ at the ungaged location of interest from the spatial regression equations (eq. 12),
- $(Q_p)_{g,w}$ is the flood-discharge quantile estimate for an AEP of $p$ at the gage, weighted with the spatial regression equations (eq. 17, table 2), and
- $w_a$, $A_i$, and $A_g$ are as defined for equation 19.

According to equations 19 and 20, the near-gage adjustment has no effect when $A_i/A_g \leq 0.5$ or $A_i/A_g \geq 1.5$, because $w_a = 1$ (eq. 19) and therefore (eq. 20) the near-gage adjusted value $(Q_p)_{u,w}$ is identical to the value $(Q_p)_{u,r}$ from the regional regression equations. At a streamgage, $A_g = A_g$ and the near-gage adjusted value reduces to the weighted flood-discharge at the gage $(Q_p)_{g,w}$ because $w_a = 0$. When $0.5 < A_i/A_g < 1.5$, then both $(Q_p)_{u,w}$ and $(Q_p)_{g,w}$ contribute to the value of $(Q_p)_{u,w}$ according to the values of the weights $w_a$ and $1-w_a$.

Applications of the Temporal Urbanization Coefficients

As discussed, there are two situations where the temporal urbanization coefficients may be appropriate for estimating the effects of urbanization, if it is the judgment of the analyst that the hydrologic effects of urbanization are or will be similar to those of the basins analyzed in this study: (1) to adjust the prediction of rural regression equations outside of Illinois flood-frequency region 2 for urbanization, and (2) to predict the future effects of urbanization for any basin in Illinois. For either situation the urbanization-adjusted flood-discharge quantile $Q_p(U)$ is given by:

$$Q_p(U) = Q_p(U_0) 10^{(b_u)_{p} (U - U_0)}$$  \hspace{1cm} (21)

or, in log-transformed form,

$$\log_{10} Q_p(U) = \log_{10} Q_p(U_0) + (b_u)_{p} (U - U_0)$$  \hspace{1cm} (22)

where

- $Q_p(U)$ is the peak discharge quantile at the location of interest with AEP $p$ adjusted to correspond to Theobald (2005) urbanization fraction $U$,
- $Q_p(U_0)$ is the peak discharge quantile with AEP $p$ at the initial urbanization fraction $U_0$,
- $(b_u)_{p}$ is the temporal urbanization coefficient for AEP $p$ from table 9.

The urbanization-adjustment factors $10^{(b_u)_{p} (U - U_0)}$ in equation 21 are the same ratios $Q_p(U)/Q_p(U_0)$ as are plotted with solid lines in figure 14A and 14B for particular values of $U_0$ (0.0 and 0.5, respectively). Because of the functional form, the urbanization effect as indicated by this relation depends only on the coefficient $(b_u)_{p}$ and the urban fraction increase $U - U_0$; the effect is estimated to be the same for equivalent increases of $U - U_0$ regardless of the value of the initial urban fraction $U_0$.

For an application outside region 2 using StreamStats, the initial urbanization fraction $U_0$ would be zero, and the unadjusted flood-discharge quantile $Q_p(U_0)$ would come from the rural regional flood-frequency equation where the basin of interest is found. For an application inside region 2 using StreamStats, the initial urbanization fraction $U_0$ might be positive, and the unadjusted flood-discharge quantile $Q_p(U_0)$ would come from the spatial regression equations for region 2, such as the methods described in the section “Applications of the Spatial Regression Equations.” Whether inside or outside of region 2, the unadjusted flood-discharge quantile $Q_p(U_0)$ also could be a quantile estimate at or near
a gage. The Theobald (2005) data for 2010 as adjusted has been installed in StreamStats so that fractions of urbanization appropriate for use with equations 21 and 22 can be computed within StreamStats.

The variance of prediction $V_U$ of the urbanization-adjusted flood-discharge quantile in log units, $\log_{10} Q_p(U)$, can be derived from equation 22 by computing the variances of both sides as

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

(23)

where

- $V_{U_0}$ is the variance of prediction of the unadjusted flood-discharge in log units, $\log_{10} Q(U_0)$,
- $V_{b_U}$ is the variance of the urbanization coefficient $(b_U)_p$, which is the square of the standard error given in table 9.

The standard error of prediction of $\log_{10} Q_p(U)$, $S_U$, can be computed from $V_U$ as $S_U = \sqrt{V_U}$, and then $S_U$ can be used in equations 14–16 to obtain a standard error in percent and confidence intervals.

Near-gage adjustment of peak discharge quantiles outside region 2 would be applicable only if the peak discharge quantiles at the streamgage arose from the same urbanization conditions as for the basin upstream from the site of interest.

**Example Computations**

To provide further clarification on the use of the results presented in this report, five example computations are provided in this section. Basic information about the examples is given in table 15.

**Example 1: Ungaged Location in the Study Region, Far from a Streamgage**

Consider a hypothetical ungaged basin with a drainage area ($A$) of 50 mi$^2$, urbanized land use fractions ($U$) NLCD_22_23_24 of 20 percent and Theobald_Urban of 40 percent, water and wetland fraction ($W$) DrainageClass1a of 15 percent, 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, it can be seen that the basin characteristic values are well within their corresponding ranges (table 14). If the basin is within region 2, the peak discharge quantiles should be computed by using the spatial regression equations, defined by equations 11 or 12, with coefficient values given in table 11. Taking the peak discharge quantile with AEP = 0.01 as an example, using equation 12, this quantile is computed as follows:

$$Q_{0.01} = 10^{(b_U)_0 A^{0.53} U^{1.2} W^{0.2} S^{0.53}}$$

(24)

$$Q_{0.01} = 10^{1.968 \times 50^{0.53} \times 0.2^{1.2} \times 0.15^{0.2} \times 3.67^{0.53}}$$

$$Q_{0.01} = 1,790 \text{ ft}^3/\text{s},$$

where an NLCD_22_23_24 value of 20 percent is used for the urbanization measure $U$. 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 equations 13–16. The variance $V_i$ of the estimate is computed from equation 13:

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

where

- $\gamma^2 = 0.0528$ is the model error variance (table 12),
- $x_i = \left[ 1, \log_{10} A, U^{1/2}, W^{1/2}, \log_{10} S \right]$,
- $x_i = \left[ 1, \log_{10} 50, 0.20^{1/2}, 0.15^{1/2}, \log_{10} 3.67 \right]$,
- $x_i = \left[ 1, 1.699, 0.4472, 0.3873, 0.5647 \right]$.

**Table 15.** Example peak discharge quantile computations in this study in northeastern Illinois.

<table>
<thead>
<tr>
<th>Number</th>
<th>Description</th>
<th>Equation used to compute discharge values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ungaged location in study region, far from a streamgage</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>Ungaged location in study region, near a streamgage</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>At a streamgage in study region</td>
<td>17</td>
</tr>
<tr>
<td>4</td>
<td>Locations outside study region</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>Adjusting for future effects of urbanization</td>
<td>21</td>
</tr>
</tbody>
</table>

1 In this study, an ungaged location is considered to be near a streamgage if ratio $A_s/A_d$ of the drainage area at the ungaged location $A_s$ to the drainage area at the gage $A_d$ is between 0.5 and 1.5 and they are on the same stream.

2 Values were computed with equation 17 and are tabulated in table 2.
and the matrix

\[
\left( X' \Lambda^{-1} X \right)^{-1} = 0.001
\]

\[
\begin{bmatrix}
-7.6988 & 1.1394 & 10.542 & -1.2200 & 2.4035 \\
-4.5311 & -2.6149 & -1.2200 & 40.324 & -1.2044 \\
-20.598 & 5.9206 & 2.4035 & -1.2044 & 18.346
\end{bmatrix}
\]

(matrix data from Table 13). With these values, the second term of \( V_i \), \( x_i \left( X' \Lambda^{-1} X \right)^{-1} x'_i \), is computed as

\[
x_i \left( X' \Lambda^{-1} X \right)^{-1} x'_i =
\]

\[
\begin{bmatrix}
-7.6988 & 1.1394 & 10.542 & -1.2200 & 2.4035 \\
-4.5311 & -2.6149 & -1.2200 & 40.324 & -1.2044 \\
-20.598 & 5.9206 & 2.4035 & -1.2044 & 18.346
\end{bmatrix}
\]

\[
0.001[1, 1.699, 0.4472, 0.3873, 0.5647]
\]

\[
\begin{bmatrix}
-7.6988 & 1.1394 & 10.542 & -1.2200 & 2.4035 \\
-4.5311 & -2.6149 & -1.2200 & 40.324 & -1.2044 \\
-20.598 & 5.9206 & 2.4035 & -1.2044 & 18.346
\end{bmatrix}
\]

\[
\begin{bmatrix}
1.699 \\
0.4472 \\
0.3873 \\
0.5647
\end{bmatrix}
\]

\[
= 0.00214.
\]

Therefore \( V_i = 0.0528 + 0.00214 = 0.0549 \).

Given the \( V_i \) value, the standard error of prediction in log units is \( S_i = V_i^{1/2} = 0.234 \), and from equation 14, the standard error of prediction in percent is

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

\[
S_{pi} = 100 \left\{ \exp \left[ \left( \ln 10 \right)^2 0.0549 \right] - 1 \right\}^{1/2}
\]

\[
S_{pi} = 58.1 \text{ percent.}
\]

The confidence intervals are \( \log_{10} Q_i \pm t_{n-2, \alpha \cdot \cdot \cdot}, S_i \) (eq. 15), where \( t_{n-2, \alpha \cdot \cdot \cdot} \) is the critical value of the t distribution at the alpha level \( \alpha \) and \( n - p \) degrees of freedom, where \( n = 117 \) is the number streamages used in the spatial regressions and \( p = 5 \) is the number of basin characteristics plus 1. Here, for the 90-percent confidence intervals \( t_{n-2, \alpha \cdot \cdot \cdot} = t_{0.95, 112} = 1.659 \), so the confidence intervals in log units are

\[
\log_{10} Q_i \pm t_{n-2, \alpha \cdot \cdot \cdot} S_i = \log_{10} \left( 1790 \right) \pm 1.659 * 0.234
\]

\[
\log_{10} Q_i \pm t_{n-2, \alpha \cdot \cdot \cdot} S_i = \left[ 2.864, 3.641 \right].
\]

After inverting the logarithmic transformation, the 90-percent confidence intervals are (eq. 16)

\[
\left[ Q_i, 10^{-t_{n-2, \alpha \cdot \cdot \cdot} S_i}, Q_i 10^{t_{n-2, \alpha \cdot \cdot \cdot} S_i} \right] = \left[ 1790 \cdot 10^{-1.659 \cdot 0.234}, 1790 \cdot 10^{1.659 \cdot 0.234} \right]
\]

(27)
Applications of Regression Equations

\[ Q_{1.00} = 10^{(2.864 \pm 3.641)} \]

\[ Q_{1.00} = [730, 4,380] \text{ ft}^3/\text{s}. \]

Summing up, the estimated 1-percent AEP peak discharge quantile \( Q_{0.01} \) for this hypothetical ungaged basin on a stream in region 2 far from a streamgage is 1,790 ft\(^3\)/s with a standard error of prediction of 58.1 percent and a 90-percent confidence interval of [730, 4,380] ft\(^3\)/s.

**Example 2: Ungaged Location in the Study Region, Near a Streamgage**

Next assume that this hypothetical ungaged basin is located, still in region 2, upstream from USGS streamgage 05527950, Mill Creek at Old Mill Creek, Illinois, which has a drainage area of 59.87 mi\(^2\) (table 1). In this case the value of the ratio \( A_g/A_s \) is 50/59.87 = 0.835, and the adjustment weighting factor \( w_a \) takes the value \( 2[50/59.87 - 1] = 0.330 \) (eq. 19), so a near-gage adjustment (eq. 20) is applicable (see discussion immediately following eq. 20). From table 2, the weighted AEP 0.01 at-gage flood-discharge quantile \( Q_{0.01} \) \(_{g,w} \) at streamgage 05527950 is 2,050 ft\(^3\)/s, and as previously computed the regional regression estimate at the ungaged site of interest \( Q_{0.01} \) \(_{r} \) is 1,790 ft\(^3\)/s. Therefore, the weighted estimate at the ungaged site of interest is

\[ Q_{0.01} = w_a (Q_{0.01})_{r} + (1 - w_a) (Q_{0.01})_{g,w} (A_g/A_s) \]  

(28)

\[ (Q_{0.01})_{u,w} = (0.330 * 1,790) + (0.670 * 2,050 * 0.835) \]

\[ (Q_{0.01})_{u,w} = 1,740 \text{ ft}^3/\text{s}. \]

Summing up, the estimated 1-percent AEP peak discharge quantile \( Q_{0.01} \) for this hypothetical ungaged basin on a stream in region 2 near to streamgage 05527950 \( (A_g/A_s = 0.835) \) is 1,740 ft\(^3\)/s, compared to 1,790 ft\(^3\)/s without the near-gage adjustment.

**Example 3: At a Streamgage in the Study Region**

To obtain an estimate at an applicable streamgage in the study region, no computations are needed. Although the method of computation is presented in the “At a Streamgage” section (eq. 17), the results of the computations have been tabulated (table 2). The particular value from table 2 to be used is the same weighted at-gage flood-discharge quantile \( Q_{0.01} \) \(_{g,w} \) used in the previous near-gage adjustment computation. Following that example, for USGS streamgage 05527950, Mill Creek at Old Mill Creek, Illinois, the appropriate AEP 0.01 at-gage flood-discharge quantile \( Q_{0.01} \) \(_{g,w} \) is 2,050 ft\(^3\)/s.

**Example 4: At a Location Outside the Study Region**

If instead the hypothetical ungaged basin were outside region 2, then the adjustment of a rural estimate for urbanization described in the “Applications of the Temporal Urbanization Coefficients” section would be applicable (eq. 21). For example, if the AEP = 0.01 quantile estimate from the rural regional regression equations from Soong and others (2004) is 2,000 ft\(^3\)/s, then the urbanization-adjusted AEP = 0.01 quantile value at the location of interest is

\[ Q_{u} = Q_{0.01} (U_o) 10^{(h_o (U - U_o))} \]  

(29)

where the urban fraction of interest, \( U \), is, by assumption, 40 percent or 0.4, and the current urban fraction, \( U_o \), is 0, because the discharge quantile value being adjusted, \( Q_{0.01} (U_o) \), is being estimated by using a rural regression equation. Therefore,

\[ Q_{0.01} (U) = 2,000 * 10^{0.312 * (0.4 - 0.0)} \]

\[ Q_{0.01} (U) = 2,000 * 1.333 \]

\[ Q_{0.01} (U) = 2,670 \text{ ft}^3/\text{s}, \]

where \( (h_o)_{0.01} = 0.312 \) is taken from table 9.

The variance of prediction of this estimate in log units can be computed from equation 23 as follows:

\[ V_U = V_{U_o} + V_{h_o} (U - U_o)^2 \]  

(30)

\[ V_U = 0.0549 + 0.134^2 (0.4 - 0.0)^2 \]

\[ V_U = 0.05777, \]

where \( V_{U_o} \), the variance of prediction of the unadjusted flood-discharge in log units, \( log_{10} Q(U_o) \), is assumed, just for this example, to have the same value as \( V_U \) which was computed previously for region 2 (in an actual computation it would be
necessary to compute this value by the relevant method, such as the method given in Soong and others (2004), which is currently (2016) implemented in StreamStats, and \( V_{bi} \) is the square of the standard error of \( \hat{b}_i \), the value of which is given in Table 9.

Given this value of \( V_{bi} \), the standard error is
\[
S_{\text{std}} = \left( V_{bi} \right)^{1/2} = 0.240,
\]
and the standard error in percent units is
\[
S_{\%} = 100 \left\{ \exp \left[ \left( \ln 10 \right)^2 V_{bi} \right] - 1 \right\}^{1/2}.
\]

Confidence intervals for this estimate can be computed by using equation 16 as already shown; the resulting 90-percent confidence interval in log10 units is \([3.0269, 3.8507] \) in log10 units and \([1,070, 7,090] \) in cubic feet per second, \([1,060, 6,680] \).

Summing up, the estimated 1-percent AEP peak discharge quantile \( Q_{0.01} \) for this hypothetical ungaged basin on a stream outside region 2 increases from 2,000 to 2,670 ft³/s as a result of adjusting for the effects of urbanization increasing from 0 to 40 percent. The estimated adjusted quantile has a standard error of prediction of 59.9 percent and a 90-percent confidence interval of \([1,060, 6,680] \) ft³/s.

Example 5: Adjustment for Effects of Future Urbanization

Finally, assume the hypothetical basin is as originally assumed, with the same basin characteristics, but an estimate of the peak discharge quantile at "build-out" (100 percent urbanization) is desired. If the future urbanization is expected to have similar hydrologic effects as the basins used in this study, the temporal urbanization coefficients would again be applicable, and the estimate of the AEP = 0.01 peak discharge quantile would be obtained by equation 21 as follows:

\[
Q_{\text{out}}(U) = Q_{\text{out}}(U_0) 10^{(b \lambda_{\text{un}}(U-U_0)),}
\]
where
\[
Q_{0.01}(U) = 1,790 * 1.539
\]
\[
Q_{0.01}(U) = 2,750 \text{ ft}^3/\text{s},
\]
where
\[
Q_{0.01}(U_0) = 1,790 \text{ ft}^3/\text{s},
\]
\[
(b \lambda_{\text{un}})_{0.01} = 0.312
\]
\[
\text{is taken from the first example calculation (eq. 24), and}
\]
\[
\text{is taken from Table 9.}
\]
The uncertainty of this estimate can be obtained from equation 23 as in the previous example; the resulting uncertainty \( S_{\%} \) is 62.0 percent, and the 90-percent confidence interval is \([3.0286, 3.8507] \) in log10 units and \([1,070, 7,090] \) in cubic feet per second.

Summing up, the estimated 1-percent AEP peak discharge quantile \( Q_{0.01} \) for this hypothetical ungaged basin increases from 1,790 ft³/s to 2,750 ft³/s as a result of adjusting for the effects of urbanization increasing from 40 to 100 percent. The estimated adjusted quantile has a standard error of prediction of 62.0 percent and a 90-percent confidence interval of \([1,070, 7,090] \) ft³/s.

A future increase in urbanization does not need to take the urban fraction to completed build-out, that is, \( U = 1.0 \). Any value of \( U \) greater than the current urban fraction (\( U_0 = 0.40 \), in the example) could be used to estimate the effects of future increases in urbanization.

**Summary**

This report provides two sets of equations for estimating peak discharge quantiles at annual exceedance probabilities (AEPs) of 0.50, 0.20, 0.10, 0.04, 0.02, 0.01, 0.005, and 0.002 (recurrence intervals of 2, 5, 10, 25, 50, 100, 200, and 500 years, respectively) for urbanized and urbanizing watersheds in Illinois based on analysis of 117 streamgage records in and near Illinois flood-frequency region 2. One set of equations was developed through a temporal analysis by using a two-step least squares-quantile regression technique that measures the average effect of changes in the urbanization of the gaged watersheds used in the study and were used to adjust the annual maximum peak discharge records to 2010 urbanization conditions. The other set of equations was developed through a spatial analysis using generalized least-squares (GLS) regression applied to quantiles estimated from the urbanization-adjusted records.

The effect of urbanization indicated by the two sets of equations is substantially different, with the temporal analysis indicating a larger effect, especially for larger AEPs (smaller recurrence intervals). This difference in urbanization effect was attributed to coefficient biases in the spatial analysis, and as a result, the urbanization coefficients from spatial equations are recommended for use only at ungaged locations in...
the study region. Other applications discussed in the report, which are estimating the effect of urbanization in regions of Illinois outside flood-frequency region 2 in combination with rural regression equations and estimating the effect of future urbanization anywhere in Illinois, use the results of the temporal analysis.

The peak discharge quantiles for the spatial analysis were computed by using the Expected Moments Algorithm following the censoring of potentially influential low floods identified by a multiple Grubbs-Beck test. To improve the skew estimates used for the peak discharge quantile estimation, weighted skew coefficients were computed as the variance-weighted average of at-site and regional skew coefficients. The regional skew coefficients were obtained from a new regional skew model, which estimates skewness as an increasing linear function of the urbanized land use fraction, with values ranging from -0.39 to +0.58.

Several urbanization, water and wetland, and slope variables, in addition to drainage area, were considered for use as basin characteristics in the spatial equations in a preliminary weighted least-squares analysis. The combination of variables giving the best results—drainage area, the fraction of developed land, the fraction of land with poorly drained soils or likely water, and the basin slope estimated as the ratio of the basin relief to basin perimeter—were used in the GLS analysis to develop the final spatial regression equations.

In addition to this report, which details the development of the two sets of equations and provides guidance on their application, including numerical examples, the products of the study include the urbanization-adjusted annual maximum peak discharge records and peak discharge quantile estimates at 181 streamgages: the 117 streamgages used to develop the spatial regression equations and 64 additional streamgages in the region that originally were considered for use in the study but later deemed to be redundant. The equations and quantile estimates are available in the web application StreamStats, and the urbanization-adjusted peak discharge records are provided in a table as part of this report and as a collection of tables and plots by streamgage at https://doi.org/10.3133/sir20165050.

References Cited


Cohn, T.A., England, J.F., Berenbrock, C.E., Mason, R.R.,
Grubbs-Beck test statistic for detecting multiple potentially
influential low outliers in flood series: Water Resources
Research, v. 49, p. 5047–5058. [Also available at
https://doi.org/10.1002/wrcr.20392.]

Cohn, T.A., Lane, W.L., and Baier, W.G., 1997, An algorithm
for computing moments-based flood quantile estimates
when historical flood information is available: Water Resources
Research, v. 33, no. 9, p. 2089–2096.

Cohn, T.A., Lane, W.L., and Stedinger, J.R., 2001, Confidence
intervals for expected moments algorithm flood quantile
estimates: Water Resources Research, v. 37, no. 6, p. 1695–
1706.

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
Documents/Classification-of-Wetlands-and-Deepwater-
Habitats-of-the-United-States.pdf.]

Croissant, Y., and Millo, G., 2008, Panel data econometrics in
R—The plm package: Journal of statistical software, v. 27,
no. 2, 51 p., accessed December 29, 2010, at
http://www.jstatsoft.org/v27/i02/.

Drainage Manual Committee for IDOT Division of Highways,
2011, IDOT drainage manual: Springfield, Ill., Illinois
Department of Transportation [variously paged].

Eng, Ken, Chen, Yin-Yu, 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 Meth-

Espey, W.H., Jr., and Winslow, D.E., 1974, Urban flood fre-
quency characteristics: Journal of the Hydraulics Division,
Proceedings of the American Society of Civil Engineers,
v. 100, no. HY2, p. 279–293.

Farmer, W.H., 2017, WREG: USGS WREG v. 2.02 R package
version 2.02: U.S. Geological Survey WREG software web
page, accessed November 15, 2019, at https://rdrr.io/github/
USGS-R/WREG/.

Fazio, D.J., and Sharpe, J.B., 2012, Flood of September
13–16, 2008, in northeastern Illinois: U.S. Geological Sur-
vey Data Series 726, 40 p. [Also available at

Fehrenbacher, J.B., Alexander, J.D., Jansen, I.J., Darmody,
R.G., Pope, R.A., Flock, M.A., Voss, E.E., Scott, J.W.,
University of Illinois, College of Agriculture, Agricultural
Experiment Station, and U.S. Department of Agriculture,

Frees, E.W., 2004, Longitudinal and panel data: Analysis and
applications in the social sciences: Cambridge, Cambridge
University Press, 467 p.

Fuller, W.A., 1987, Measurement error models: New York,
John Wiley and Sons, 440 p.

Gesch, D.B., Oimoen, M.J., Greenlee, S.K., Nelson, C.A.,
Dataset: Photogrammetric Engineering and Remote Sens-
ing, v. 68, no. 1, p. 5–11. [Also available at https://www.
highlight.html.]


Grimaldi, S., Petroselli, A., Alonso, G., and Nardi, F., 2010,
Flow time estimation with spatially variable hillslope velocity
in ungauged basins: Advances in Water Resources, v. 33,
p. 1216–1223.

regression in regional hydrologic analyses: Journal of
Hydrology, v. 344, p. 82–95.

distribution and its application in flood frequency analysis.
III—Sample skew and weighted skew estimators: Journal of

Gruber, A.M., Reis, D.S., Jr., and Stedinger, J.R., 2007, Models
of regional skew based on Bayesian GLS regression,
Paper 40927–3285, in Kabbes, K.C., ed., Restoring our
natural habitat—Proceedings of the 2007 World Envi-
ronmental and Water Resources Congress—May 15–19,
Tampa, Florida: Reston, Va., American Society of Civil
Engineers [variously paged].

regional skew, data selection and Bayesian GLS regression,
Paper 596, in Babcock, R.W., and Walton, Raymond, eds.,
World Environmental and Water Resources Congress 2008–
Ahupua’a—Proceedings of the congress—May 12–16, 2008,
Honolulu, Hawaii: Reston, Va., American Society of Civil
Engineers [variously paged].

Hansel, A.K., and McKay, E.D., III, 2010, Quaternary period,
in Kolata, D.R., and Nimz, C.K., eds., Geology of Illinois:
Champaign, Ill., Illinois State Geological Survey, p. 216–
247.
References Cited


Appendix 1. Northeastern Illinois Regional Skew Analysis

For the log-transformation of annual peak discharges, Bulletin 17B (U.S. Interagency Advisory Committee on Water Data, 1982) recommends using a weighted average of the station skew coefficient and a regional skew coefficient to improve estimates of annual flood-probability discharges (AFPDs). Bulletin 17B supplies a national map but also encourages hydrologists to develop more specific local relations. From the publication of the national map based on data through 1973 through the end of this study in 2009, some 36 years of additional information has accumulated and improved spatial estimation procedures have been developed (Stedinger and Griffis, 2008). Furthermore, prior national and Illinois analyses of regional skew such as that in Soong and others (2004) did not consider urbanization. For these reasons, for this study in northeastern Illinois a regression analysis was done to develop a regional skew model.

Reis and others (2005), Gruber and others (2007), and Gruber and Stedinger (2008) developed a Bayesian generalized least-squares (GLS) regression model for regional skewness analyses. The method provides a more reasonable description of the model error variance than either the generalized least-squares method-of-moments or maximum likelihood point estimates (Veilleux, 2011). However, because of complications introduced by the use of the Expected Moments Algorithm (EMA), with multiple Grubbs-Beck censoring of potentially influential low floods (Cohn and others, 2013) and large cross-correlations between annual peak discharges at pairs of streamgages, an alternate regression procedure was developed to provide stable and defensible results for regional skew regression (Veilleux and others, 2012; Veilleux, 2011; Lamontagne and others, 2012). This alternate procedure is referred to as the Bayesian weighted least-squares/Bayesian generalized least-squares (B–WLS/B–GLS) regression framework (Veilleux and others, 2011; Veilleux, 2011; Veilleux and others, 2012).

The B–WLS/B–GLS regression analysis uses an ordinary least-squares (OLS) analysis to fit an initial regional skewness model; the OLS model is then used to generate a regional skew-coefficient estimate for each streamgage. This regional estimate is the basis for computing the variance of each station skew-coefficient estimator used in the weighted least-squares (WLS) analysis. Then, B–WLS is used to generate estimators of the regional skew-coefficient model parameters. Finally, B–GLS is used to estimate the precision of those WLS parameter estimators, to estimate the model error variance and the precision of that variance estimator, and to compute various diagnostic statistics. The methodology for the regional skewness model is described in detail in Eash and others (2012).

Application of B–WLS/B–GLS Method in Northeastern Illinois

This regional skew study is based on the temporally adjusted annual peak discharge data from the 117 nonredundant streamgages in northeastern Illinois and the surrounding states that were included in the regional regression analyses described in the main report (table 1). As stated in Bulletin 17B, the skew coefficient of the station is sensitive to extreme events, and more accurate estimates can be obtained from longer records. Thus, for regional skew studies it is preferred that each streamgage have a minimum of 30 to 35 years of record. However, because of the nature of the data available in northeastern Illinois, in particular at the urban stations, the minimum was reduced to 15 years to ensure that the entire study area was represented. Applying this minimum reduced the number of streamgages used in the regional skew study to 110 (table 1–1, available at https://doi.org/10.3133/sir20165050).

Because the dataset includes censored data and historic information, the effective record length used to compute the precision of the skewness estimators is no longer simply the number of annual peak discharges at a streamgage. Instead, a more complex calculation is used to take into account the availability of historic information and censored values. Although historic information and censored peaks provide valuable information, they often provide less information than an equal number of years with systematically recorded peaks (Stedinger and Cohn, 1986). The calculations made to compute the pseudo record length (PRL) are described in Eash and others (2013). PRL equals the systematic record length if such a complete record is all that is available for a site.

The station logarithmic skew coefficient, \( \hat{\gamma}_s \) (table 1–1), and its mean square error, \( \text{MSE}[\hat{\gamma}_s] \), were computed by using expected moments analysis (EMA) (Cohn and others, 1997; Griffis and others, 2004). The streamgage skewness estimates are ensured to be unbiased by using the correction factor developed by Tasker and Stedinger (1986) and used in Reis and others (2005). In addition to the skew data, selected basin characteristics for each of the streamgages were available as explanatory variables for the regional skew study.
A model for the cross-correlation of the logarithms of the annual maximum peak discharges between streamgage pairs is needed for the analysis because sample cross-correlations commonly are unavailable and are subject to large uncertainty, usually including unphysical negative values. Such a model for the northeastern Illinois study area was developed by using the 39 streamgages with at least 45 years of concurrent systematic peaks. A logit model, termed the Fisher Z transformation (Kendall and Stuart, 1961), where \(Z = \ln\left[\frac{1+r}{1-r}\right]\), was used to transform the sample correlations \(r_i\) from their \((-1,1)\) range to a \((-\infty, +\infty)\) range that is appropriate for least-squares fitting. Various models relating the transformed cross-correlations to various basin characteristics were considered. The adopted model relates the cross-correlations to the distance \(D_i\) between streamgage basin centroids according to the fitted equation

\[Z_{ij} = 0.32 + \exp\left[-0.40 - 0.042D_i\right]\] (fig. 1–1A).

The fitted relation between \(Z\) and the distance between basin centroids together with sample data from the 567 streamgage pairs of data is shown in figure 1–1A. The equivalent function relation between the cross-correlations and basin centroid distance, which is obtained by back-transforming from \(Z\) to \(r\) as

\[r = \frac{\exp(2Z) - 1}{\exp(2Z) + 1}\]

is shown in figure 1–1B.

### Results of Regional Skew Analysis in Northeastern Illinois

The more promising basin characteristics from a preliminary analysis of the regional spatial regressions were selected for testing as explanatory variables in the regression analysis for regional skew. These selected basin characteristics included drainage area, the urbanization measure \(NLCD_{22\_23\_24}\), the DEM-based slope variable \(DEM_{\text{slope}}\), and several water and wetness variables: \(\text{NLCD}_{\text{total}}, \text{NLCD}_{\text{emergent}},\) DrainageClass1, DrainageClass1a, and \(NLCD_{11\_95}\) (see table 3 for definitions). Of these, only \(NLCD_{22\_23\_24}\) was statistically significant in explaining the site-to-site variability in skewness, with a square-root transformation providing the better result as compared to untransformed values. Thus, the best model, as classified by having the smallest model error variance, \(\sigma^2\), and largest pseudo \(R^2\), is the model using \(NLCD_{22\_23\_24}\) values. Results for the URBAN model, as well as the constant regional model, denoted CONSTANT, for comparison are provided in table 1–2.

Pseudo \(R^2\), included in table 1–2, describes the estimated fraction of the variability in the true skewness from site-to-site explained by each model (Gruber and others, 2007; Parrett and others, 2011). A constant model does not explain any variability, so the pseudo \(R^2\) for the CONSTANT model is equal to 0 percent. The URBAN model has a pseudo \(R^2\) of 14 percent indicating that inclusion of the \(NLCD_{22\_23\_24}\) basin characteristic in the model explains 14 percent of the variability in the true skews. Although the URBAN model accounts for a relatively small portion of the total variation in skews, it is important to note that the focus of the study is on the effects of urbanization. A statistically significant dependence of skewness on the \(NLCD_{22\_23\_24}\) characteristic, a measure of urbanization, suggests that not including it in the skewness model could negatively affect the final spatial regression equations. Also, the posterior mean of the model error variance, \(\sigma^2\), for the URBAN model was 0.16, which is smaller than that of the CONSTANT model (\(\sigma_{\text{CONST}}^2 = 0.19\)). Thus, the URBAN model was chosen as the final regional skew model for the northeastern Illinois study area.

The average variance of prediction at a new site \(\text{AVP}_{\text{new}}\) describes the precision of the regional skew. In table 1–2, the URBAN model has the lower \(\text{AVP}_{\text{new}}\) equal to 0.19. This AVP new is an average value computed by averaging the variance of prediction at a new site \(\text{VP}_{\text{new}}\) for all of the 110 streamgages in the analysis. Just as regional skew varies from streamgage to streamgage, depending on \(NLCD_{22\_23\_24}\), so too do the values of \(\text{VP}_{\text{new}}\). Values of the variance of prediction for the regional skew, \(\text{VP}_{\text{new}}\) and effective record length (ERL) for the URBAN model for values of \(NLCD_{22\_23\_24}\) between 0 and 1 are given in table 1–3. Thus, the URBAN skew model for northeastern Illinois has \(\text{VP}_{\text{new}}\) values ranging from 0.18 to 0.22 and effective record lengths ranging from 39 to 43 years, depending on \(NLCD_{22\_23\_24}\). It is important to note that for the purposes of this study, the AVP new of 0.19 was used when weighting the at-site skew with the regional skew.

The URBAN model provides a reasonable fit for the northeastern Illinois regional skew data (fig. 1–2). As compared to the CONSTANT model, the URBAN model provides smaller values of regional skew for basins with less development (smaller \(NLCD_{22\_23\_24}\) values) and larger values of regional skew for basins with more development (larger \(NLCD_{22\_23\_24}\) values). The URBAN model indicates that skewness increases substantially with urbanization from a value of -0.39 for a basin with \(NLCD_{22\_23\_24}\) equal to 0 to +0.58 for a basin where the \(NLCD_{22\_23\_24}\) fraction is equal to one.

Pseudo analysis of variance (Pseudo ANOVA) statistics for the northeastern Illinois regional skew analysis were determined as additional diagnostics for the selected model (table 1–4). Explanations of how the statistics were computed can be found in Eash and others (2013).
Estimation of Peak Discharge Quantiles for Selected Annual Exceedance Probabilities in Northeastern Illinois

\[ r = \frac{\exp(2Z) - 1}{\exp(2Z) + 1} \]

Streamgage pairs

\[ Z = 0.32 + \exp[-0.40 - 0.42D] \]

Fisher Z cross-correlation of concurrent annual-maximum peak discharges with greater than 45 years of concurrent records

Fisher Z cross-correlation of concurrent annual-maximum peak discharges with greater than 45 years of concurrent records

Figure 1–1. Fitting of cross-correlation model of the log-annual maximum peak discharges in this study in northeastern Illinois: A, relation between Fisher Z transform of cross-correlation and distance; B, relation between cross-correlation and distance.
Table 1–2. Regional skewness models and corresponding metrics in this study in northeastern Illinois.

\[ \beta_1, \text{skew model intercept}; \text{SE}, \text{standard error}; \beta_2, \text{skew model slope}; \sigma^2, \text{model error variance}; \text{ASEV, average sampling error variance}; \text{AVP}_{\text{new}}, \text{average variance of prediction for a new site}; \text{pseudo } R^2, \text{fraction of the variability in the true skews explained by each model (Gruber and others, 2007)}; \hat{\gamma}, \text{estimated regional skewness}; \text{--}, \text{not applicable}; \text{NLCD}_{22\_23\_24}, \text{fraction of the basin that has 2011 National Land Cover Database land cover classes 22, 23, and 24 (low, medium, and high intensity development, respectively).} \]

<table>
<thead>
<tr>
<th>Skew model</th>
<th>Intercept ( \beta_1 )</th>
<th>Intercept SE</th>
<th>Slope ( \beta_2 )</th>
<th>Slope SE</th>
<th>Bayesian plausibility, in percent</th>
<th>Error variance ( \sigma^2 )</th>
<th>Error variance SE</th>
<th>ASEV</th>
<th>AVP_{\text{new}}</th>
<th>Pseudo ( R^2 ), in percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSTANT model</td>
<td>( \hat{\gamma} = \beta_1 )</td>
<td>0.16</td>
<td>0.15</td>
<td>--</td>
<td>--</td>
<td>0.19</td>
<td>0.05</td>
<td>0.022</td>
<td>0.21</td>
<td>0</td>
</tr>
<tr>
<td>URBAN model</td>
<td>( \hat{\gamma} = \beta_1 + \beta_2 \left( \text{NLCD}_{22_23_24} \right)^{0.5} )</td>
<td>-0.39</td>
<td>0.19</td>
<td>0.97</td>
<td>0.34</td>
<td>0.16</td>
<td>0.05</td>
<td>0.032</td>
<td>0.19</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 1–3. Average regional skew, variance of prediction, and equivalent record length for URBAN regional skew model for various values of \text{NLCD}_{22\_23\_24}, in this study in northeastern Illinois.

\[ \text{VP}_{\text{new}}, \text{variance of prediction}; \text{ERL}, \text{equivalent record length}; \text{NLCD}_{22\_23\_24}, \text{the fraction of the basin that has 2011 National Land Cover Database land cover classes 22, 23, and 24 (low, medium, and high intensity development, respectively).} \]

<table>
<thead>
<tr>
<th>\text{NLCD}_{22_23_24} (fraction)</th>
<th>Average regional skew</th>
<th>\text{VP}_{\text{new}}</th>
<th>ERL (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>-0.39</td>
<td>0.196</td>
<td>42</td>
</tr>
<tr>
<td>0.1</td>
<td>-0.29</td>
<td>0.180</td>
<td>43</td>
</tr>
<tr>
<td>0.2</td>
<td>-0.20</td>
<td>0.180</td>
<td>42</td>
</tr>
<tr>
<td>0.3</td>
<td>-0.10</td>
<td>0.182</td>
<td>41</td>
</tr>
<tr>
<td>0.4</td>
<td>0.00</td>
<td>0.186</td>
<td>40</td>
</tr>
<tr>
<td>0.5</td>
<td>0.10</td>
<td>0.191</td>
<td>39</td>
</tr>
<tr>
<td>0.6</td>
<td>0.19</td>
<td>0.196</td>
<td>39</td>
</tr>
<tr>
<td>0.7</td>
<td>0.29</td>
<td>0.203</td>
<td>39</td>
</tr>
<tr>
<td>0.8</td>
<td>0.39</td>
<td>0.208</td>
<td>40</td>
</tr>
<tr>
<td>0.9</td>
<td>0.48</td>
<td>0.215</td>
<td>40</td>
</tr>
<tr>
<td>1.0</td>
<td>0.58</td>
<td>0.222</td>
<td>41</td>
</tr>
</tbody>
</table>
**Figure 1–2.** Relations between the unbiased at-site skew and urbanization measure for 110 streamgages in northeastern Illinois regional skew study area. The lines represent a model based on a constant skew (CONSTANT) and a model with a linear relation between skew and the square root of fraction of National Land Cover Dataset classes 22, 23, and 24 (NLCD_22_23_241/2) urbanization measure (URBAN). The models were developed from Bayesian weighted least-squares/Bayesian generalized least-squares analyses.

**Table 1–4.** Pseudo analysis of variance (ANOVA) statistics for the northeastern Illinois URBAN regional skew model.

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees-of-freedom Equations</th>
<th>URBAN model</th>
<th>Sum-of-squares Equations</th>
<th>URBAN model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>(k)</td>
<td>1</td>
<td>(n\left[\sigma^2_0 - \sigma^2_k\right])</td>
<td>2.9</td>
</tr>
<tr>
<td>Model error</td>
<td>(n-k-1)</td>
<td>108</td>
<td>(n\left[\sigma^2_k\right])</td>
<td>18</td>
</tr>
<tr>
<td>Sampling error</td>
<td>(n)</td>
<td>110</td>
<td>(\sum_{i=1}^{n} Var(\hat{\gamma}_i))</td>
<td>37</td>
</tr>
<tr>
<td>Total</td>
<td>(2n-1)</td>
<td>219</td>
<td>(n\left[\sigma^2_k\right] + \sum_{i=1}^{n} Var(\hat{\gamma}_i))</td>
<td>58</td>
</tr>
<tr>
<td>EVR</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>2.1</td>
</tr>
<tr>
<td>MBV*</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>5.7</td>
</tr>
<tr>
<td>Pseudo (R^2) (in percent)</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>14</td>
</tr>
</tbody>
</table>
References Cited


