USGS
science for a changing world

Adjusting Annual Maximum Peak Discharges at Selected Stations in Northeastern Illinois for Changes in Land-Use Conditions

Scientific Investigations Report 2016–5049

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

Adjusting Annual Maximum Peak Discharges at Selected Stations in Northeastern Illinois for Changes in Land-Use Conditions

By Thomas M. Over, Riki J. Saito, and David T. Soong

Scientific Investigations Report 2016–5049

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

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

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

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit http://www.usgs.gov or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit http://store.usgs.gov.

Acknowledgments

Contents

Figures

Tables

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
Length		
inch (in.)	2.54	centimeter (cm)
inch (in.)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
Area		
square mile (mi^2)	259.0	hectare (ha)
square mile (mi^2)	2.590	square kilometer (km^2)
Volume		
cubic foot (ft^3)	28.32	cubic decimeter (dm^3)
cubic foot (ft^3)	0.02832	cubic meter (m^3)
acre-foot (acre-ft)	1,233	cubic meter (m^3)
acre-foot (acre-ft)	0.001233	cubic hectometer (hm^3)
Flow rate		
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)

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 [insert datum name (and abbreviation) here, for example, 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.

Abbreviations

AEP	annual exceedance probability
C/I/T	Commercial/Industrial/Transportation
CSG	crest-stage gage
EROS	Earth Resources Observation and Science
LULC	land use/land cover
MWRDGC	Metropolitan Water Reclamation District of Greater Chicago
NEXRAD	NEXt Generation RADar
NID	National Inventory of Dams
NIPC	Northeastern Illinois Planning Commission
NLCD	National Land Cover Dataset
NOAA	National Oceanic and Atmospheric Administration
NRCS	Natural Resources Conservation Service
SE	standard error
SERGoM	Spatially Explicit Regional Growth Model
USACE	U.S. Army Corps of Engineers
USGS	U.S. Geological Survey
WY	water year

Adjusting Annual Maximum Peak Discharges at Selected Stations in Northeastern Illinois for Changes in Land-Use Conditions

By Thomas M. Over, Riki J. Saito, and David T. Soong

Abstract

The effects of urbanization on annual maximum peak discharges in northeastern Illinois and nearby areas from 1945 to 2009 were analyzed with a two-step longitudinal-quantile linear regression approach. The peak discharges were then adjusted to 2010 land-use conditions. The explanatory variables used were daily precipitation at the time of the peak discharge event and a housing density-based measure of developed land use. The effect of the implementation of stormwater detention was assessed indirectly. Peak discharge records affected by the construction of large reservoirs that affect channel routing were identified and were split into segments at the time of completion of the reservoir. Longitudinal regressions of the peak discharge records on linear and logarithmic transformations of the selected measures of urbanization and precipitation were tested, and the best fitting model was selected for quantile regression and adjustment of the peak discharges.

Because the uncertainties of streamgage-by-streamgage regressions of peak discharges as a function of urbanization are so large, a regional urbanization response was computed. Streamgages used in this study fit the following two criteria: (1) drainage area is at most 200 square miles and, (2) at least 10 consecutive years of peak discharge record are available. In the first step of the regression analysis, linear longitudinal regression models with fixed intercepts estimated for each segment of the peak discharge records were computed. The segment intercepts were then subtracted from the discharge records to homogenize the discharge dataset across the segments in preparation for the quantile regression analysis. From the quantile regression analysis, the effect of urbanization on peak discharge varies strongly with the exceedance probability of the peak discharge event; coefficients monotonically increase from 0.340 to 0.969 over exceedance probabilities from 0.002 to 0.99. The regression analyses yield estimates of the population-wide effect of the explanatory variables on the dependent variables as a function of exceedance probability. These estimates are similar to the coefficients of the regional regression relations in USGS regional flood-frequency studies such as those implemented in the Web application Stream-Stats; although in the longitudinal analysis used in this study, it is the temporal not the spatial (between-streamgage) variations that are taken into account.

The observed and adjusted values for each streamgage are tabulated. To illustrate the overall effect of the adjustments, differences in the mean, standard deviation, and skewness of the log-transformed observed and urbanization-adjusted peak discharge series by streamgage are computed. For almost every streamgage where an adjustment was applied (no increase in urbanization was reported for a few streamgages), the mean increased and the standard deviation decreased; the effect on skewness values was more variable but usually they increased. Significant positive peak discharge trends were common in the observed values, occurring at 27.3 percent of streamgages at a p-value of 0.05 according to a Kendall's tau correlation test; in the adjusted values, the incidence of such trends was reduced to 7.0 percent.

Introduction

The U.S. Geological Survey (USGS), in cooperation with the U.S. Army Corps of Engineers (USACE)—Chicago District, the Illinois Center for Transportation, and the Illinois Department of Transportation, has begun a study to investigate the effects of urban development on peak-flood discharges in the northeastern Illinois (Chicago) region. The objective of this first phase of the study is to develop and implement a method to adjust peak discharge data at USGS streamgages in urbanized northeastern Illinois to values consistent with 2010 land-use conditions. The adjusted peak discharge data can then be applied at the streamgage to problems such as the evaluation of synthesized peak discharges from deterministic rainfall-runoff modeling. The adjusted peak discharge data also are the first step toward the development of regional equations for estimation of peak discharge quantiles in urban areas in the study area. The resulting flood frequencies have broad applications to flood protection, mitigation planning, and risk analysis.

The most recent urban regional flood-frequency study in the Chicago region was done by Allen and Bejcek (1979). As a result of the completion of a flood inundation mapping study (Shaeffer and others, 1970), many streamgage records in the Chicago area ended around that time, but many others also continued in operation, and many were started since then, so considerable additional data is available for a more complete analysis. The use of long-term streamflow data in an urbanizing region, however, requires confronting the fact that the watershed conditions are nonstationary (that is, changing with time), and, as a result, the properties of the discharge data are likely to be nonstationary as well. One approach to this problem is to limit the portion of the data used to periods for which the watershed conditions are approximately stationary and known. This approach has the advantage that the nonstationarity does not have to be accounted for in the analysis, but it ignores potentially valuable data and still requires sufficient knowledge of the historical land-use conditions to determine stationary periods. For this study, an approach was sought that allows peak discharge series from watersheds with nonstationary land-use conditions to be adjusted to land-use conditions at a common year (2010).

Annual maximum peak discharge records from 143 USGS streamgages with drainage areas of at most 200 square miles (mi^2) with record lengths of at least 10 years from the metropolitan Chicago region, and parts of southeast Wisconsin and northwest Indiana that drain into Illinois, were used in this study. Analysis made with Kendall's tau rank-based correlation test showed that 43 of the selected streamgage records had significant trends ($p < 0.05$), of which 39 of the significant trends were positive. Linear regression with respect to time also indicated the possibility of slope changes in addition to the trends. Stormwater detention basins built in response to ordinances that began in 1972 in northeastern Illinois could offset some potential increases from increased imperviousness in a gradual manner. On the other hand, the construction of flood-control reservoirs could affect peak discharges in a more temporally abrupt manner.

Because individual records are too short to allow accurate estimates to be obtained streamgage-by-streamgage, the response of the selected peak discharge records to urbanization was analyzed with a two-step longitudinal-quantile linear regression technique. This technique allows the estimation of coefficient values of the explanatory variables as a function of exceedance probability that apply to all the streamgage records in the study; among the causative factors of changes in peak discharges, precipitation and land use were used as the explanatory variables in the regressions. The effects of the construction of reservoirs are incorporated into the analysis by splitting the records into segments at the year of completion of the reservoir. The results are used to develop and implement a regional regression model for adjusting the peak discharge records to 2010 land-use conditions.

This report describes the methodology and data used in the study, the results of the study, and compares the results to the results of Allen and Bejcek (1979). Adjusted annual maximum peak discharges for each streamgage are tabulated in this report.

Methodology

The methodology used in this study consists of two parts:

1. Apply a two-step regression analysis to obtain regional urbanization coefficients as a function of annual exceedance probability; and

2. Adjust the flood discharge series to 2010 urbanization conditions by means of the urbanization coefficients obtained from the regression analysis.

Regression Analysis

The regression analysis in this study follows the two-step longitudinal-quantile linear regression method suggested by Canay (2011), which was applied to segments of peak discharge data into which the streamgage records were divided at years when major changes occurred in the drainage basin as discussed in the section "Accounting for Effects of Large-Scale Flood-Control Facilities." In the first step of Canay's method, longitudinal least-squares linear regression analysis (Frees, 2004) of the peak discharge data is applied (with each segment having its own fixed intercept value), and in the second step, quantile regression (Koenker and Bassett, 1978) is applied to the peak-flood data with the intercept values subtracted. In both regressions, urban fraction and precipitation are used as independent variables. The least-squares analysis in the first step is used primarily to obtain values of the segment intercepts. The intercept values are then subtracted from the observations to obtain adjusted peak discharge values that are, by hypothesis, homogeneous (that is, segment-independent), to which quantile regression can be applied in the second step. Quantile regression is used in the second step because it provides an estimate of the coefficients of the regression model for different quantiles (that is, different 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; Sauer and others, 1983; Konrad, 2003, and references therein; Moglen and Shivers, 2006) can be estimated, and the adjustment of the peak discharges to the common (2010) urban fraction value can take this disparate effect into account.

Longitudinal Regression Analysis

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

$$y_i(t) = \hat{y}_i(t) + \varepsilon_i(t) = a_i + b_P P_i(t) + b_U U_i(t) + \varepsilon_i(t), i = 1, 2, \ldots, N, \tag{1}$$

where

$y_i(t) = \log_{10} Q_i(t)$	is the base-10 logarithm of the annual maximum peak discharge at the ith 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 estimated intercept for the ith segment,
$P_i(t)$ and $U_i(t)$	are the precipitation and urban fraction, respectively, for the ith segment during year t,
b_P and b_U	are the regression coefficients for precipitation and urban fraction, respectively, and
$\varepsilon_i(t) = y_i(t) - \hat{y}_i(t)$	is the regression error for the ith segment during year t.

There is one intercept for each segment, whereas there is only one value of each of the coefficients. The coefficients, therefore, are estimated from the effect of precipitation and urbanization on all the segments combined, and these values are equivalent to the values that would be obtained from an uncertainty-weighted mean of the coefficients of segment-by-segment least-squares regression analysis (Frees, 2004, p. 32). The intercepts, therefore, contain all the information in the model that distinguishes one segment from another; the effects of urbanization and precipitation, captured in their coefficients in equation 1, are assumed to be the same for all segments. The computation of the longitudinal regression model was done in R (R Core Team, 2014) using the plm function from the plm package (Croissant and Millo, 2008).

Quantile Regression Analysis

To test for the disparate effect of urbanization on peak discharges of different exceedance probabilities and to adjust for the disparate effect if present, a quantile regression analysis (Koenker and Bassett, 1978) of the log-transformed peak discharges after subtraction of the segment intercept values was completed as the second step of the regression analysis. The segment intercept subtraction equation follows:

$$y_i'(t) = y_i(t) + \bar{a} - a_i, \tag{2}$$

where

$y_i'(t)$	is the intercept-subtracted log-transformed peak discharge value associated with the observation $y_i(t)$, and
$\bar{a} = \sum_i T_i a_i \Big/ \sum_i T_i,$	where T_i is the number of observations in the ith segment, is the weighted mean of the intercept values obtained from the longitudinal least-squares regression, and
a_i	is the intercept of the ith segment.

Following the intercept subtraction (homogenization) step, quantile regression was applied to obtain the exceedance probability-dependent response of the peak discharges to urbanization and precipitation. The quantile regression model is as follows:

$$y_i'(t) = \hat{y}_i'(t;q) + \varepsilon_k'(q) = a'(q) + b_P'(q) P_k + b_U'(q) U_k + \varepsilon_k'(q), \tag{3}$$

where

q	is a sequence of annual exceedance probabilities (AEPs),
$\hat{y}_i'(t;q)$	is the fitted value of the quantile with $AEP = q$ of the kth log-transformed peak discharge value $y_i'(t)$,
$a'(q)$	is the intercept of the fitted linear relation between $\hat{y}_i'(t;q)$ and the independent variables,
P_k and U_k	are the precipitation and urban fraction, respectively, of y_k',
$b_P'(q)$ and $b_U'(q)$	are the quantile regression coefficients as a function of q (AEP) for precipitation and urban fraction, respectively, and
$\varepsilon_k'(q) = y_i'(t) - \hat{y}_i'(t;q)$	is the regression residual for $y_i'(t)$ at $AEP = q$.

The quantile regression model was solved with the rq function of the quantreg package (Koenker, 2013) of the R language. Further background information on quantile regression is given in appendix 1.

Two post-processing steps were applied to obtain the final quantile regression coefficients and their standard errors. First, a bootstrap resampling approach suggested by Canay (2011) was applied to obtain standard errors and unbiased estimates of the mean coefficient values at the selected discrete exceedance probabilities, denoted $\bar{a}''(q)$, $\bar{b}_P''(q)$ and $\bar{b}_U''(q)$, respectively.

Second, a seventh-order polynomial was fitted to the bootstrap mean urbanization coefficients to obtain a continuous and monotonic urbanization coefficient function, denoted $b_U^*(q)$, for the adjustment of the peaks to 2010 conditions.

In the bootstrap resampling, N streamgage record segments were sampled with replacement from the complete set of N streamgage record segments, and a set of quantile regression coefficients, denoted $a''(q)$, $b_P''(q)$, and $b_U''(q)$, were obtained from each resampling. One thousand resamples were obtained and analyzed. The mean of the bootstrapped coefficients, $\bar{a}''(q)$, $\bar{b}_P''(q)$, and $\bar{b}_U''(q)$, respectively, were taken as unbiased estimates of the true coefficient values, as argued by the theory of the bootstrap (Hesterberg and others, 2006).

Method of Adjusting Peak Discharges to 2010 Land-Use Conditions

The second part of the methodology is the adjustment of the peaks to 2010 urbanization conditions, making use of the continuous urbanization coefficient function $b_U^*(q)$ obtained in the first part of the methodology. To complete this adjustment, the following equation is applied to each streamgage record segment i:

$$y_i''(t) = a_i^* - a_i + y_i(t) + b_U^*(q)\left[U_i(t^*) - U_i(t)\right],$$ (4)

where

$y_i''(t)$ and $y_i(t)$ are the base-10 logarithm of the adjusted and unadjusted, respectively, peak-flood magnitude value for year t and segment i,

a_i^* and a_i are the intercept values for the last and ith segments, respectively, of the streamgage record containing segment i,

$b_U^*(q)$ is the urban fraction coefficient value corresponding to the exceedance probability q assigned to the peak-flood value $y_i(t)$, and

$U_i(t^*)$ and $U_i(t)$ are the urban fraction values for the basin corresponding to segment i during the year t^* 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.

Because the urbanization coefficient function $b_U^*(q)$ depends on the AEP value q, to apply the adjustment, exceedance probabilities for each peak are needed. These exceedance probabilities were determined by linear interpolation among the log-transformed peak values as predicted by the regression analysis as follows:

$$q_i(t) = q^- + \left(y_i'(t) - y^-\right)\left(\frac{q^+ - q^-}{y^+ - y^-}\right),$$ (5)

where

$q_i(t)$ is the estimated exceedance probability for the intercept-subtracted peak discharge $y_i'(t)$ (and, equivalently, the observed peak discharge $y_i(t)$),

y^+ and y^- are predicted values bounding $y_i'(t)$ above and below, respectively, and

q^+ and q^- are the exceedance probabilities corresponding to y^+ and y^-.

The bounding values y^+ and y^- are selected from a vector $y'(q)$ of values computed as

$$y'(q) = \bar{a}''(q) + \bar{b}_P''(q)P_i'(t) + \bar{b}_U''(q)U_i(t),$$ (6)

where

$\bar{a}''(q)$, $\bar{b}_P''(q)$, and $\bar{b}_U''(q)$ are the bootstrap mean intercepts, urbanization coefficients, and precipitation coefficients from the quantile regressions at the exceedance probability values q,

$U_i(t)$ is the urban fraction corresponding to $y_i'(t)$, and

$P_i'(t)$ is a fitted precipitation value corresponding to $y_i'(t)$.

When $y_i'(t)$ is not bounded by the $y'(q)$ values, $q_i(t)$ is taken to be 0 or 1, depending on whether $y_i'(t)$ exceeds all $y'(q)$ or is less than all $y'(q)$ values, respectively.

A fitted precipitation value $P_i'(t)$ is used in equation 6 rather than the observed value $P_i(t)$ associated with the peak $y_i'(t)$. It was determined that if the observed precipitation value corresponding to the peak was used, then the exceedance probabilities of larger peaks were sometimes larger than those of smaller peaks, and vice versa. At the same time, as discussed, it was not intended to adjust for precipitation but merely to account for any overall trends in precipitation as the effect of changes

in urbanization were estimated. Therefore, a linear relation between the $y_i'(t)$ and $P_i(t)$ values was computed to capture the general dependence between y' and P and to smooth out the fluctuations that cause the anomalous exceedance probability estimates. This $y' - P$ relation was computed by a robust linear model fit to P as a function of y', that is,

$$P = a + by', \qquad (7)$$

by using the rlm function of the MASS R library (Venables and Ripley, 2002). This fitted relation was used to compute the value of $P_i'(t)$ for use in equation 6 given $y_i'(t)$, that is,

$$P_i'(t) = a + by_i'(t). \qquad (8)$$

With the $P_i(t)$ values computed in this way, equation 6 can be used to compute the bounding values for the interpolation of the exceedance probabilities according to equation 5 so that the adjustment indicated in equation 4 can be implemented.

Explanatory Variables

The types of explanatory variables considered in the study are historical land-use/land-cover (LULC) and climate data. In this section the options available and choices made regarding these two variable types are described.

Historical Land-Use/Land-Cover Data

The National Land Cover Dataset (NLCD; Homer and others, 2015) is a standard source of LULC data. The NLCD, however, is only available for years 1992, 2001, 2006, and 2011, whereas streamflow data in northeast Illinois are available at multiple streamgages since about 1945. Population density can be used as a proxy for urban area: Allen and Bejcek (1979) estimated imperviousness with population density and Moglen and Shivers (2006) used population as a proxy to examine trends in urban peak discharge records. High resolution decadal Census population density data from 1970 to 2000 are available for purchase (see, for example: http://www.geolytics.com). Because the available period of historical population data is still somewhat short for this study and because population data are still only a proxy for land cover, this option was also rejected. At the time this study began (2010), a nationwide analysis of land-cover change based on Landsat data from 1973 to 2000 was being carried out by the USGS Earth Resources Observation and Science (EROS) Center Land Cover Trends team (Trends; Loveland and others, 2002). A combination of Landsat data, which goes back to 1973, and before that, historical aerial photos, is probably the best potential data for assessing land use over multiple decades. However the basis of the estimates being provided by the USGS Land Cover Trends project is land-cover change within 10 by 10 kilometer (km) sample blocks and considers

developed land as a single category, not a complete geographic coverage as would be needed for hydrologic applications in particular basins. Lacking the budget and expertise to work directly from Landsat and aerial photographs ourselves, this best potential data option was not feasible for this study.

Ries and Dillow (2006) used Theobald's (2005) decadal Census-based housing density data to examine trends in peak discharges in their study in Delaware. These data extend from 1940 to 2030 and are available on a 100-meter grid covering the conterminous United States (David M. Theobald, Colorado State University, written commun., 2009). The basis of the data is the 2000 Census block and block-group data; the historical change in housing is estimated by using responses to the Census long-form question "Approximately what year was your house built?". Water areas and lands protected from development are masked out, and provision was made for housing being close to major roads within the Census blocks. Future housing densities, which include the 2010 data, were forecast using a method that Theobald (2005) called the Spatially Explicit Regional Growth Model (SERGoM) in which future growth depends on past growth trends and travel times to the nearest urban core. The resulting housing densities, computed at decadal intervals, are grouped into the 11 classes given in the Explanation of figure 5. Because the completeness and convenience of this dataset, it is used in this study to estimate historic land-use changes. Further details on the processing of this data are given in the section "Census-Based Housing Density Data."

Climate

Because most USGS peak discharge data include the date of the peak discharge event, for this study it was decided to attempt to characterize the precipitation associated with each event. In this characterization process, ideally one would account for the differences in basin lag times, but that would require hourly or higher resolution data of intensive spatial coverage. Although hourly precipitation data are available in the study area, the number of precipitation gages having such data is not large; there are many more daily precipitation gages, and high quality (NEXRAD) radar data are not available before the mid-1990s. Because of these constraints on the availability of precipitation data, the approach adopted in this study is to designate the maximum daily precipitation during the period from three days before to one day after the date of the peak discharge event. It was tested as part of the regression analysis for this study whether the following choices in processing the precipitation data gave a better fit to the peak discharge data: (1) assigning the values from the precipitation gage closest to the basin centroid or from a basin-average constructed by using Thiessen polygons as the precipitation series, (2) filling zeroes in the precipitation data or leaving them as recorded, and (3) applying a logarithmic transformation or no transformation. Filling of zeroes in the precipitation data is considered because it is assumed that zeroes in the precipitation data are either false zeroes, that is, missing

data mistakenly filled with zeroes, or they indicate peak discharges that were the result of snowmelt. In either case, a zero value does not correctly characterize the forcing of the peak discharge.

Accounting for Effects of Stormwater Detention Basins

The coverage and strictness of stormwater detention ordinances have grown continuously in the Chicago region since the first ordinance promulgated by the Metropolitan Water Reclamation District of Greater Chicago (MWRDGC) in 1972. Since the first ordinance, detention of stormwater runoff has been a requirement of the sewer permits within the service area of the MWRDGC in Cook County (Resource Coordination Policy Committee, 1998). By 1986, the State of Illinois passed legislation that authorized northeastern Illinois counties to develop their own regional stormwater management programs (Resource Coordination Policy Committee, 1998). Since 1986 a variety of county-wide ordinances were passed and updated by northeastern Illinois counties and neighboring counties in the study area in Indiana and Wisconsin and by communities within them. Therefore, there is a very complex history of stormwater ordinances in the study area, and the timing of county-wide ordinances can only provide a general timeframe for correlating to observed trends in the peak discharge records.

It was beyond the scope of this study to determine the timing of construction of the individual detention basins in the study watersheds. As an alternative, a global approximate approach was implemented within the longitudinal regression model framework where a detention initiation year was assumed, after which all subsequent development is assumed to be subject to detention, thereby splitting the urbanized fraction into two separate explanatory variables. Mathematically, if $U(t)$ is the total urbanized fraction in year t, $U'(t)$ the urbanized fraction not subject to detention, $U''(t)$ the urbanized fraction subject to detention, and t' the year the detention began, then for $t \leq t'$, $U'(t) = U(t)$ and $U''(t) = 0$, whereas for $t > t'$, $U''(t) = U(t) - U'(t')$ and $U'(t) = U'(t')$. Different detention initiation years t' were tried to see if there was a year that might apply approximately globally throughout the study domain.

Accounting for Effects of Large-Scale Flood-Control Facilities

Whereas the effects of stormwater detention facilities are treated as a type of land use and modeled by means of the fraction of basin area having detention, this approach was expected to be too approximate for large-scale flood-control facilities. Instead, the peak discharge record for a given streamgage was broken into segments at water years (WYs) when new flood-control facilities were completed, and each

segment was treated as a separate record for the purposes of the regression analysis (equation 1), with the year of completion belonging to neither segment.

In the adjustment procedure (equation 4), the effect of reservoir construction appears through the term $a_i^* - a_i$, which indicates the difference in intercept values for the last segment in time of a given streamgage record and the ith such segment. Note that this term has the value zero during the last segment of a streamgage record, including for streamgages having no reservoirs constructed during their period of record. To see the effect on the adjustment when there is a reservoir constructed during the period of record, consider the reservoir whose construction defines the beginning of the last segment a_i^*. If the segment intercept decreases when this occurs, as would be expected, that is, $a_i^* < a_i$, then $a_i^* - a_i$ is negative, and the earlier segments are adjusted downward as if the reservoir had already been built, with all discharges being adjusted by the same proportion, because the adjustment (equation 4) is in terms of the log-discharge. In this way, the adjusted record includes the effects of all the reservoirs built by the end of the observed record. This technique, being completely empirical, has no way to adjust for reservoirs built after the observational record has ended.

To define the segments, information about the year of completion, year of modification if any, the maximum capacity, and location for flood-control facilities in the study watersheds was compiled. A segment break was made at any year when a reservoir with a drainage area of at least 10 percent of the gaged watershed area and a maximum storage capacity per unit area of the gaged watershed of at least 0.40 inches (1 centimeter) was completed. Additionally each segment was required to be at least 5 years long; if it is was not, those years also were eliminated from the regression analysis. When such periods lay between segments, segment intercept values were computed by means of the linear interpolation between the segment intercepts of the neighboring segments, and adjusted values were computed using the interpolated segment intercept values. When such periods were at the beginning or end of the gaged record, no adjusted peak discharges were computed.

Data Used in this Study

The retrieval and preparation of the annual maximum peak discharge data, sources and processes of data for characterizing the variables describing LULC changes and precipitation in the northeast Illinois study region are described in this section. As discussed, implicit in the approach is that the relevant changes to the hydrologic system resulting from urbanization can be divided into two types: (1) land-use (small-scale) changes which primarily affect runoff generation, including construction of impervious surfaces, storm sewers, and detention basins, and (2) large-scale constructed facilities which affect the routing of streamflow, including in-line and off-line reservoirs.

Peak Discharge Records

Streamgages with annual maximum peak discharge data in the northeastern Illinois region (Cook, DeKalb, DuPage, Grundy, Kane, Kankakee, Kendall, Lake, LaSalle, McHenry, and Will Counties in Illinois and counties in neighboring states whose streams flow into Illinois comprising Kenosha and Walworth County of southeastern Wisconsin and Jasper, Lake, and Porter Counties in northwestern Indiana) were considered for inclusion in the study (fig. 1). Those streamgages in the region of interest that fit the following two criteria were retained for use in the study: (1) drainage area is at most 200 square miles and, (2) at least 10 consecutive years of record are available. The upper bound in drainage area is arbitrary but was chosen for the purpose that urban effects remain identifiable; the 10 consecutive years of record was adopted according to the recommendation of the U.S. Interagency Advisory Committee on Water Data (1982, p. 2). A total of 143 streamgages in the region of interest had annual maximum peak discharge records fitting the criteria and are used in this study. The geographical distribution of the streamgages in the study is shown in figures 1 and 2, and basic information on each of the streamgages is given in table 1 (available at http://dx.doi.org/10.3133/sir20165049) and the distribution by county in table 2.

Peak discharge data for the selected 143 streamgages for the study period 1945–2009 were downloaded from the USGS National Water Information System (http://nwis.waterdata.usgs.gov/usa/nwis/peak). Peak discharge values with accompanying codes indicating the value was an historic peak, was censored, or was a maximum daily average, or where the date of the peak was unknown, were deleted from the dataset. A total of 3,976 annual maximum peak discharge values remained after these deletions.

The distribution of the data by year is illustrated in figure 3; panels *A*–*C* (panel *D* will be discussed in the section "Flood-Control Reservoirs"). For 85 of the 143 streamgage records used in the study, data ended by 1980 (table 1). Among these 85 streamgages, all but four of them were crest-stage gages (CSGs); the others had continuous daily discharge data. For the 58 streamgages for which there are data after 1980, four were CSGs during their complete periods of record and 12 were CSGs during part of their period of record. The reason for the large number of observations by CSG in the 1960s and 70s is because a flood inundation study in northeastern Illinois was done which involved a large network of installed CSGs in cooperation with the Northeastern Illinois Planning Commission (NIPC) during that period (Shaeffer and others, 1970). These CSGs generally started just before the WY 1962 and most of them were discontinued by the end of WY 1980.

Watershed Boundaries

Digital watershed boundaries of the 143 streamgages were determined by means of the ArcGIS-based methods used in Illinois StreamStats as described in Ishii and others (2010).

Precipitation Records

Daily precipitation data from National Oceanic and Atmospheric Administration (NOAA) cooperative weather observer sites, the Cook County Precipitation Network (Westcott, 2012) and from Argonne National Laboratory (Murphy and Ishii, 2006) were used. NOAA daily precipitation data were retrieved from the National Climatic Data Center (www.ncdc.noaa.gov). A total of 49 precipitation gages were selected for use in this analysis, and missing data were filled with data from the nearest gage for the period of study (1945 to 2009). The Thiessen polygons created by these gages for use in computing spatial average precipitation for the study watersheds are shown in figure 4.

Census-Based Housing Density Data

As explained in the section "Historical Land-Use / Land-Cover Data," Census-based housing density data created by Theobald (2005) was used to estimate historic urbanization in this study. The data file used, bhcs_fote20080612.zip, was downloaded from http://www.nrel.colostate.edu/ftp/theobald/ on 5/12/2009, and are now downloadable from http://dx.doi.org/10.3133/sir20165049. To illustrate the properties of the data and the growth of urbanization in the northeastern Illinois region, maps of the northeastern Illinois region using this data were created for each decade from 1940–2010 and are shown in figure 5.

The completeness of this housing density dataset makes it convenient for use in this study; however, confidence in its validity would be strengthened by comparison with a more direct measurement of developed land use. For this purpose, the Trends team (Krista Karstensen, 2011, written commun.) provided developed land-cover estimates for the years 1939, 1973, 1980, 1992, and 2000 for the 40 sample blocks in Ecoregion 54 (Central Corn Belt), which covers most of our region of interest (Karstensen and others, 2013). Comparisons between the Trends developed areas and the Theobald (2005) exurban/urban + commercial/industrial/transportation (C/I/T) (classes 7–10; see Explanation of figure 5 for definitions) land-cover areas within the sample blocks are provided in figure 6 for the five time periods. The comparison shows that the two quantities match rather well. Based on this comparison and the other advantages of the Theobald housing density dataset, the Theobald dataset was chosen for use in this study to estimate the fraction of developed land.

The following additional processing steps were applied to the Theobald (2005) data for use in this study:

1. Adjustment of class 10 (C/I/T) fractions, which are given in the dataset as year 2000 values, to estimated values for other decades by the method described in appendix 2.

2. Linear interpolation between decadal values to obtain annual values.
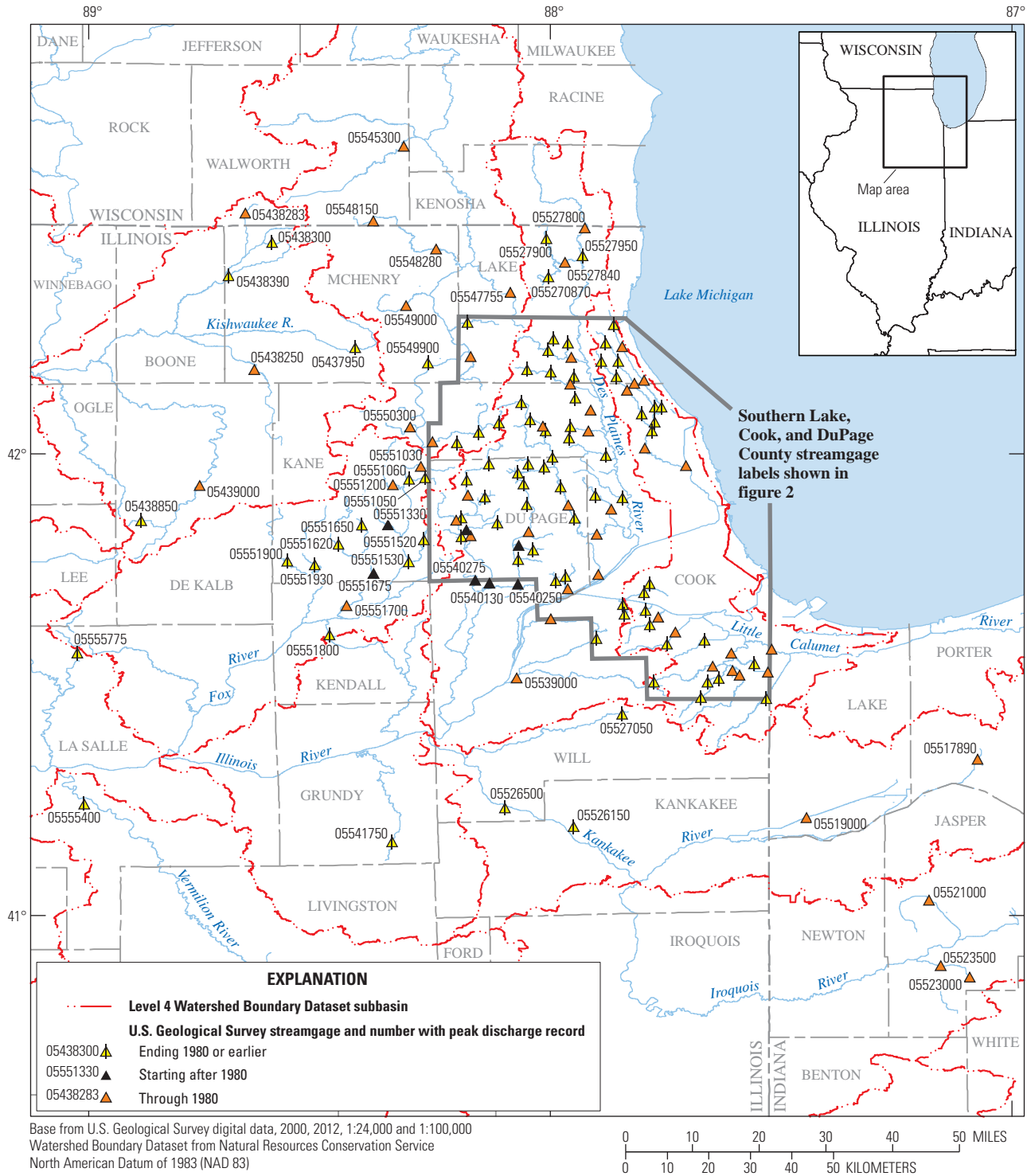
Figure 1.   Locations of the U.S. Geological Survey streamgages used in this study.
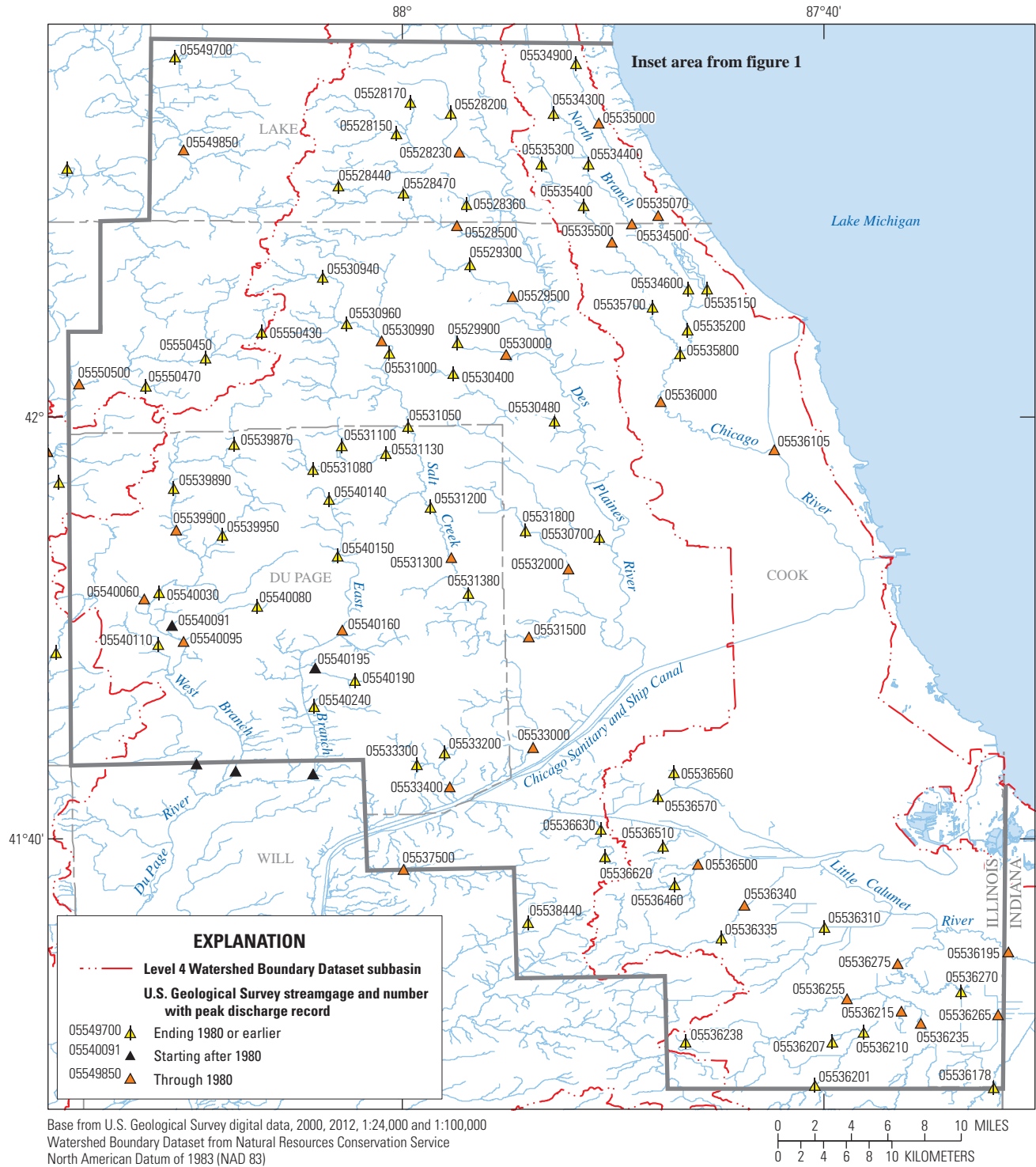
Figure 2. Locations of the U.S. Geological Survey streamgages used in this study (detail).

Table 2.   Inventory of U.S. Geological Survey streamgages used in this study in northeastern Illinois.

[mi², square miles; --, not applicable]

County	Streamgages with systematic record after 1980				Streamgages with systematic record ending by 1980			
	Number of stations	Minimum drainage area (mi²)	Median drainage area (mi²)	Maximum drainage area (mi²)	Number	Minimum drainage area (mi²)	Median drainage area (mi²)	Maximum drainage area (mi²)
Illinois counties								
Cook	19	7.93	20.9	115	32	1.6	9.97	92.2
DeKalb	1	77.7	77.7	77.7	1	1.67	1.67	1.67
DuPage	9	6.83	18.1	91.5	18	2.1	8.82	101
Grundy	0	--	--	--	1	4.52	4.52	4.52
Kane	5	14	38.9	55	8	2.09	14.1	32.6
Kankakee	0	--	--	--	1	0.19	0.19	0.19
Kendall	1	70.2	70.2	70.2	1	0.45	0.45	0.45
Lake	8	13	28.6	123	15	1.03	8.17	145
LaSalle	0	--	--	--	2	0.14	0.25	0.36
McHenry	3	15.5	85.1	192	4	0.07	7.62	88.1
Will	3	75.8	107.5	123	2	0.8	6.45	12.1
Indiana counties								
Jasper	3	21.8	35.6	83.7	0	--	--	--
Lake	2	90	106.5	123	0	--	--	--
Porter	1	30.3	30.3	30.3	0	--	--	--
Wisconsin counties								
Kenosha	0	--	--	--	0	--	--	--
Walworth	3	9.58	13.6	110	0	--	--	--
Total	58	6.83	28.05	192	85	0.07	8.81	145

3. Testing of weighted sums of different fractions of classes Theobald data in the longitudinal regression model fitting and selection of an unweighted sum of the fractions of classes 7–10 as adjusted for the regression analysis. The sum of these fractions was used for comparison with the Trends data in figure 6; histograms of the distribution of this sum are shown in figure 7. The other fractions and weighting schemes tested did not improve the longitudinal regression model fits.

Flood-Control Reservoirs

An initial list of dams in the study area was obtained using the National Inventory of Dams (NID, URL: http://crunch.tec.army.mil/nidpublic/webpages/nid.cfm). Important local references were "Our Community and Flooding" (Resource Coordination Policy Committee, 1998) and final reports of detailed watershed plans by the MWRDGC for various watersheds in Cook County (MWRDGC, 2009 to

2011), and the USACE study report, "Upper Des Plaines River and Tributaries" (U.S. Army Corps of Engineers, 2011). The information collected regarding the dams was verified with the assistance of county and Natural Resources Conservation Service (NRCS) offices.

Using the collected reservoir information as compiled in table 3 (available at http://dx.doi.org/10.3133/sir20165049) and the segmentation procedure described in the section "Accounting for the Effects of Large-Scale Flood-Control Facilities," 29 of the 143 streamgage records were broken into two or more (as many as four) segments, yielding a total of 172 segments (table 4, available at http://dx.doi.org/10.3133/sir20165049). Of the 3,976 peak discharges retained in the downloaded dataset, 3,874 (97.4 percent) were used in the regression analysis and 3,927 (98.8 percent) have adjusted peak discharge values, of which 53 were adjusted with the use of an interpolated segment intercept value (fig. 3D). There were 49 peak discharges in the dataset not adjusted because they could not be assigned a segment number and did not fall between segments.
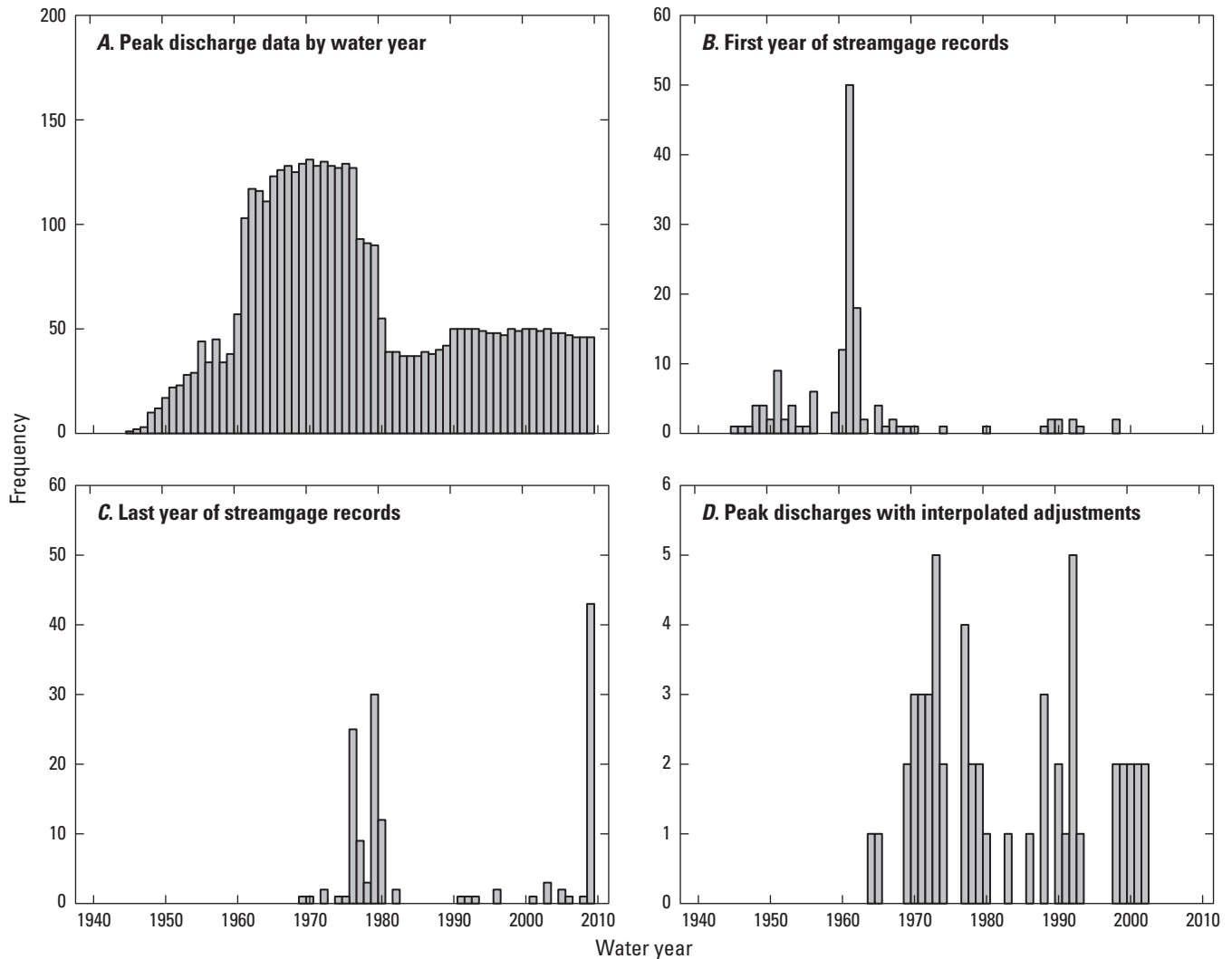
Figure 3. Streamgage data used in this study in northeastern Illinois for each water year with *A*, number of streamgages having peak discharge data each water year; *B*, number of streamgages whose data began each water year; *C*, number of streamgages whose data ended each water year; and *D*, number of streamgages each water year with adjusted discharges computed by interpolation because they fall during years when major dams were completed or in a short period between two such dam completions. In panel *C*, many streamgage records are shown as ending in 2009 because that is the last year of the study, not because the record did not continue into 2010.

Results

The segments of the peak discharge records at the selected streamgages were analyzed according to the two-step longitudinal-quantile linear regression analysis described in the "Methodology" section. The results of the regression analysis were then used to adjust the peak discharge records to 2010 urbanization conditions. The details of the results of these procedures are presented and discussed in this section.

Longitudinal Regression Model Fits

The urban land-use fraction computed as the sum of classes 7–10 of Theobald's (2005) housing density data and maximum daily precipitation occurring from 3 days before through 1 day after the date of the recorded annual peak event were determined to be highly significant ($p < 0.001$) variables in the longitudinal regression analysis. The form of the precipitation variable giving the best fits was determined to be the basin average computed by means of Thiessen polygons, with zeroes filled with the median of the nonzero values. As shown in the first four rows of table 5, between linear and log-transformations, the best regression model fit was obtained when both urban fraction and precipitation were untransformed
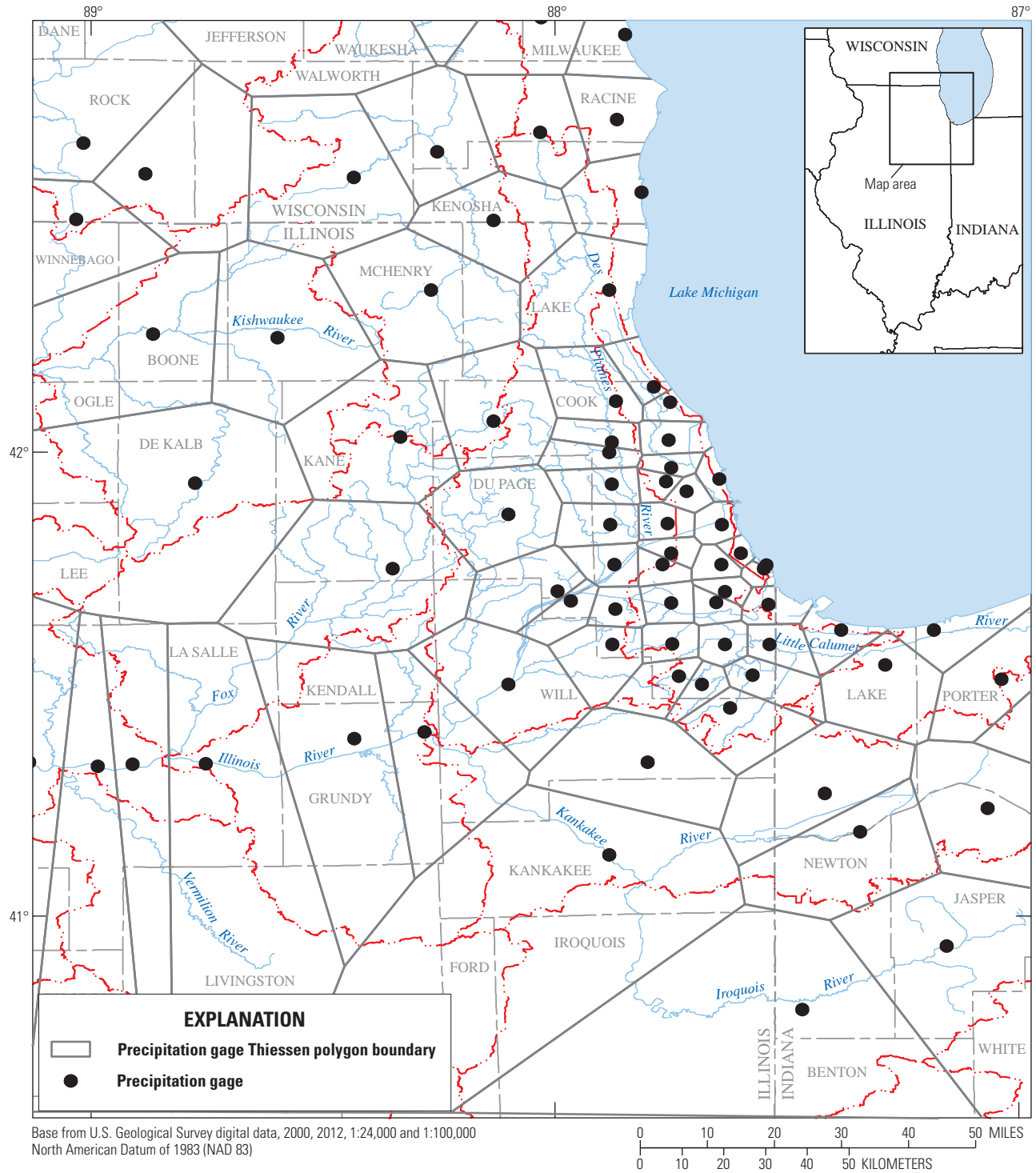
Figure 4.   Locations of the precipitation gages used in this study and their Thiessen polygons.
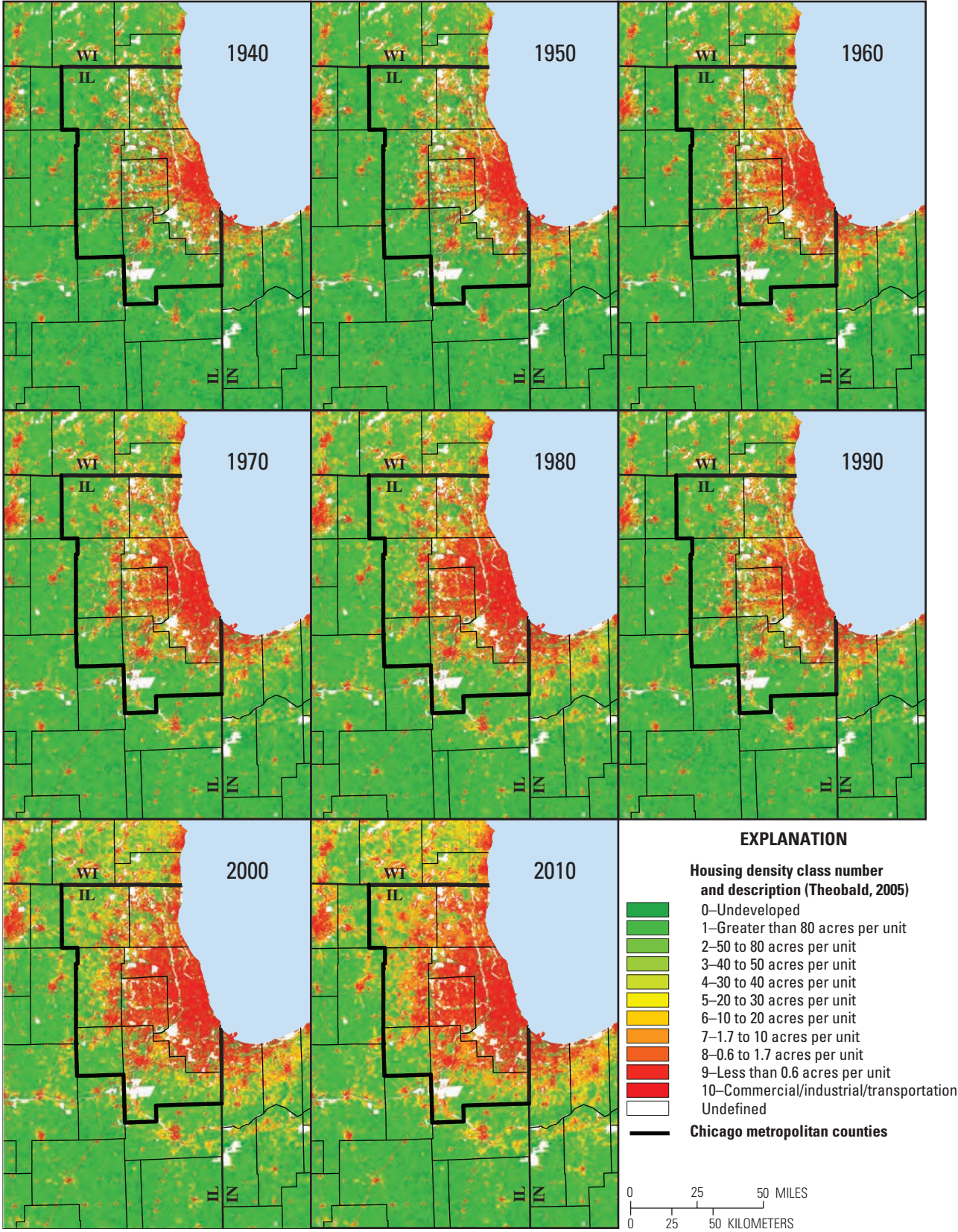
Figure 5. Housing density by decade in northeastern Illinois, 1940–2010 (modified from Karstensen and others, 2013).
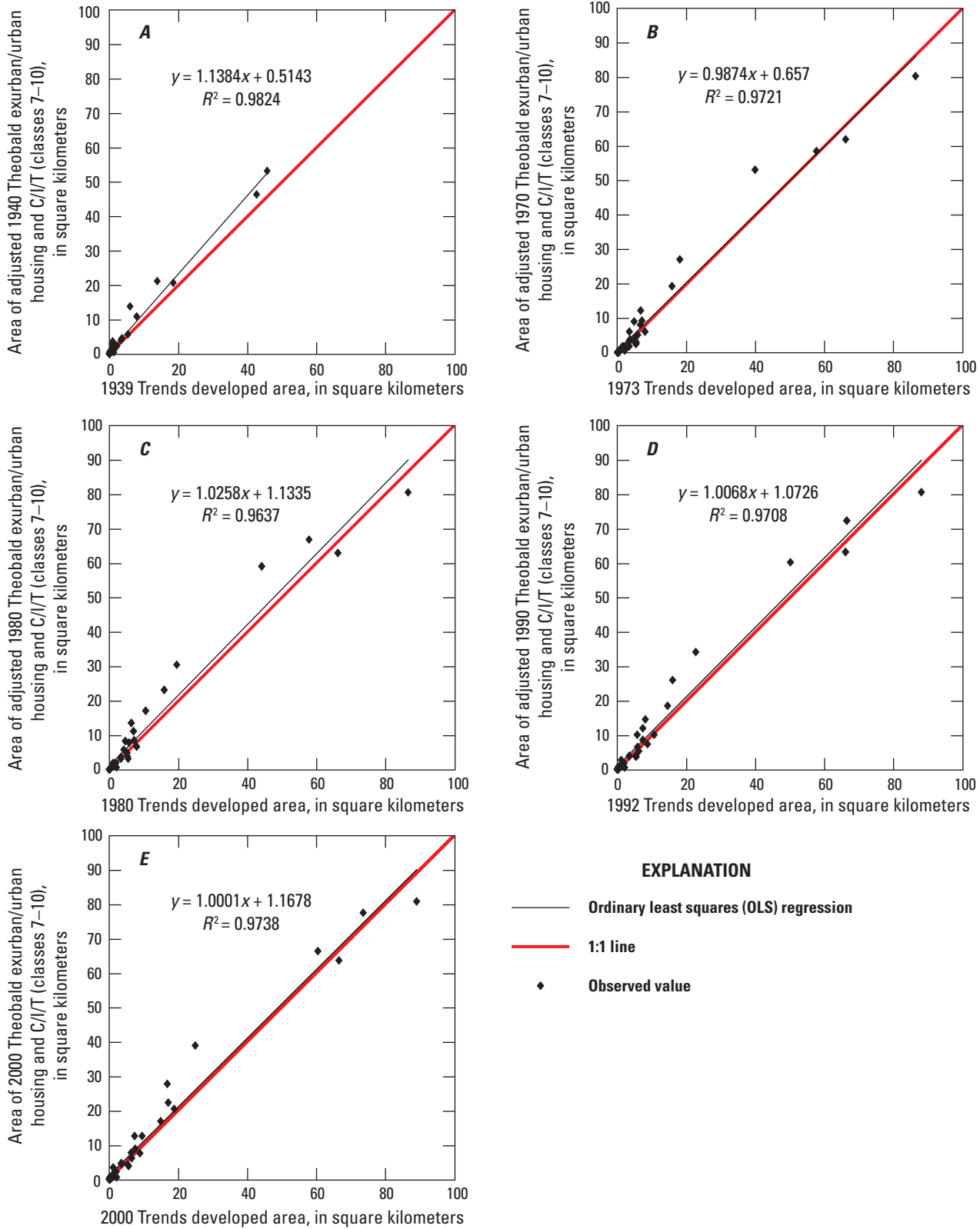
Figure 6.   Comparisons between the Theobald (2005) exurban/urban housing plus adjusted commercial/industrial/transportation (C/I/T) (classes 7–10) land-cover areas and USGS Land Cover Trends project developed land-cover areas within the 40 100-square kilometer sample blocks in Ecoregion 54 (Central Corn Belt) for *A*, 1940–1939; *B*, 1970–1973; *C*, 1980–1980; *D*, 1990–1992; and *E*, 2000–2000. The C/I/T land areas in the Theobald (2005) data are from the year 2000, and so for other years, these values were adjusted proportionally to the changes in housing density by the method explained in appendix 2.
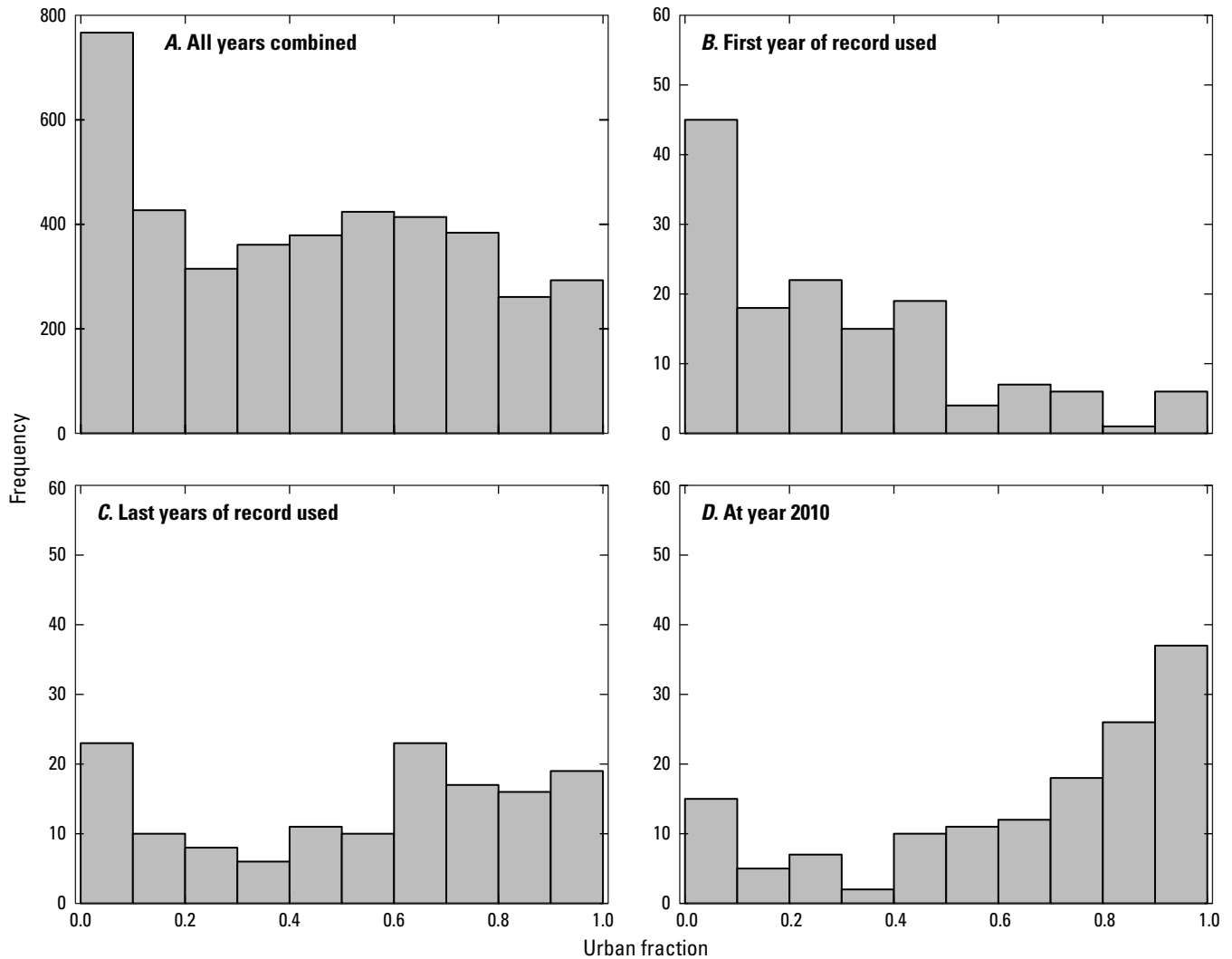
Figure 7. Urban fraction computed as sum of the frequencies of Theobald (2005) housing density classes 7–10 for the 143 U.S. Geological Survey streamgages in this study in northeastern Illinois at *A*, all years of record; *B*, first year of record; *C*, last year of record; and *D*, year 2010.

(linear), so this transformation was selected for both variables for the analysis in this study.

The final five rows in table 5 give the longitudinal regression results for different assumptions on the timing of the initiation of detention. Based on these results, there is some evidence that a significantly higher coefficient of the urbanized fraction may apply before an assumed detention initiation year, particularly for the assumed detention initiation year of 1975. For the other detention initiation years tested, the coefficients are also higher before the assumed detention initiation years than after, but the confidence intervals defined by their standard errors overlap. Considering the large number of streamgages whose records ceased between 1975 and 1979 (see figure 3C), the earliness of this date (given the first stormwater detention ordinance in the region was promulgated by MWRDGC in 1972), the crudely approximate nature of assuming all watersheds initiated detention at the time, and the

increase in the uncertainty of value of the urbanized area coefficients when estimating two such coefficients rather than one, for purposes of adjusting the records in this study, the model in which there is only a single urbanized fraction was chosen. Further investigation of the effect of detention in selected watersheds for which the timing of the initiation of detention can be estimated with some confidence might be a useful direction for further analysis.

Based upon the selection of the longitudinal least-squares regression model of the untransformed urbanization and precipitation and a single urbanization coefficient (table 5, row 1), the segment intercepts a_i are obtained (table 4). The column in table 4 with the header "Intra-streamgage record segment intercept differences" shows, for the streamgages with multiple segments, whether the intercept decreased (a negative value) or increased (a positive value). Because the breaks between segments were defined based on substantial increases

Table 5.   Longitudinal regression model coefficients and standard errors from the analysis of 172 segments derived from 143 U.S. Geological Survey streamgage records in the study area in northeastern Illinois for different transformations of independent variables and different assumptions on timing of initiation of stormwater detention.

[SE, standard error; NA, not applicable; NC, not computed. Mathematical notation references equation 1]

Year detention assumed to have been initiated	Precipitation transformation	Urban fraction transformation	Precipitation coefficient, b_P ($\log_{10}$[ft³/s]/ inch)	Standard error of precipitation coefficient, SE(b_P) ($\log_{10}$[ft³/s]/ inch)	Coefficient of urban fraction, b_U ($\log_{10}$[ft³/s]/ inch)	Standard error of coefficient of urban fraction, SE(b_U) ($\log_{10}$[ft³/s]/ inch)	Coefficient of urban fraction before detention, $b_{U'}$ ($\log_{10}$[ft³/s]/ inch)	Standard error of coefficient of urban fraction before detention, SE($b_{U'}$) ($\log_{10}$[ft³/s]/ inch)	Coefficient of urban fraction after detention, $b_{U''}$ ($\log_{10}$[ft³/s]/ inch)	Standard error of coefficient of urban fraction after detention, SE($b_{U''}$) ($\log_{10}$[ft³/s]/ inch)	Standard error of model residuals, SE(ε) ($\log_{10}$[ft³/s])	Weighted average of segment intercepts, $\bar{a}$ ($\log_{10}$[ft³/s])
no detention	none (linear)	none (linear)	0.0827	0.0035	0.598	0.039	NA	NA	NA	NA	0.21693	2.199
no detention	$\log_{10}$	none (linear)	0.201	0.011	0.571	0.041	NA	NA	NA	NA	0.22261	NC
no detention	$\log_{10}$	$\log_{10}$	0.202	0.011	0.377	0.032	NA	NA	NA	NA	0.22427	NC
no detention	none (linear)	$\log_{10}$	0.0824	0.0036	0.394	0.031	NA	NA	NA	NA	0.21885	NC
no detention	NA	none (linear)	NA	NA	0.647	0.042	NA	NA	NA	NA	0.23251	NC
1970	none (linear)	none (linear)	0.0828	0.0035	NA	NA	0.661	0.071	0.542	0.066	0.21690	NC
1975	none (linear)	none (linear)	0.0830	0.0035	NA	NA	0.677	0.056	0.449	0.084	0.21681	NC
1980	none (linear)	none (linear)	0.0828	0.0035	NA	NA	0.628	0.048	0.492	0.106	0.21689	NC
1985	none (linear)	none (linear)	0.0828	0.0035	NA	NA	0.618	0.046	0.503	0.122	0.21691	NC
1990	none (linear)	none (linear)	0.0827	0.0035	NA	NA	0.604	0.044	0.562	0.143	0.21693	NC

in basin storage, the expected behavior is for a decrease in the segment intercept. The history of basin modifications at streamgages with significant increases (relative to the standard errors of the intercept estimates) in segment intercepts was investigated, and if no reason for the increase could be determined, the segments were combined and the least-squares longitudinal regression was recomputed. One significant segment intercept increase remains, at streamgage 05536560, where channelization of the stream channel immediately upstream from the streamgage at about the same time as the construction of a flood-control reservoir upstream from the channelized portion of the stream channel (Jon Grabowy, MWRDGC, written commun., 2012) led, according to the segment intercept values in table 4, to significantly higher peak discharges.

Some physical understanding of the segment intercepts can be obtained by plotting them in relation to drainage area (fig. 8). This plot shows that the segment intercept values, which are proportional to the logarithms of peak discharge when the effects of precipitation and urbanization are removed (equation 1), are mostly explained by a linear relation with the logarithm of drainage area. This behavior is expected because of the often observed linear relation between the logarithms of peak discharge quantiles and drainage area (see, for example, Soong and others, [2004]).

Quantile Regression Model Fits

As described in the "Methodology" section, the input data for the quantile regression analysis (equation 3) are the logarithms of the peak discharges with the segment intercepts
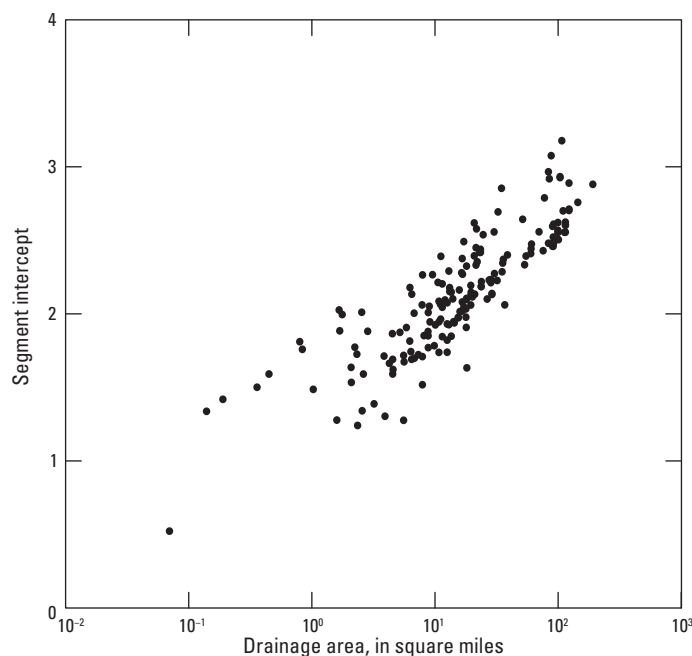
Figure 8. Segment intercepts as a function of drainage area for U.S. Geological Survey streamgage records used in this study in northeastern Illinois.

minus the weighted average of the segment intercepts subtracted (equation 2); see table 5 for the value of the weighted average of segment intercepts. Quantile regression for a sequence of AEP values q was applied to model the peak discharges as a linear function of urban fraction and precipitation (equation 3). The results of this analysis are shown by the more lightly shaded planes in figure 9. The slopes of these planes, being gently upward and approximately parallel along the precipitation axis and more significantly upward with slope varying from the lower (larger AEP) to higher (smaller AEP) planes in the urban fraction axis, indicate the quantile regression coefficients, which are listed in table 6. Notice the median ($q = 0.5$) quantile regression coefficients agree approximately with the least-squares longitudinal regression coefficients, as both are about 0.6 for urban fraction and 0.08 for precipitation (tables 5 and 6), which is expected, as both indicate the central tendency of the dependence of the peak discharges on urban fraction and precipitation. Given the relatively small magnitude of the precipitation coefficients, the AEP-dependence of the precipitation coefficients is fairly small, varying from about 0.07 to 0.13 for the bootstrapped mean coefficients, with their minimum at AEP = 0.9 and maximum at AEP = 0.002, whereas the urban fraction coefficients are much larger and vary almost monotonically with AEP from about 0.97 at AEP = 0.99 (smaller discharges) and about 0.34 at AEP = 0.002 (higher discharges), decreasing by a factor of about 2.8. This almost monotonically decreasing dependence of the urban fraction coefficient on decreasing AEP verifies the expected result that the effect of urbanization is smaller for larger discharges. The dependence of the bootstrap mean urban fraction coefficients (table 6) on AEP was smoothed and made monotonic by fitting a seventh-order polynomial to the urban fraction coefficients as a function of AEP (fig. 10). The polynomial-fitted urban fraction coefficients also define an urban fraction coefficient for any AEP value, as is needed for the adjustment to 2010 urbanization.

Because peak discharges are being adjusted for the effect of urbanization alone (not including precipitation), as described in the "Methodology" section, a robust linear regression model fit of log-transformed segment intercept-subtracted peak discharge to precipitation (equation 7) was used to reduce the three-dimensional space of the quantile regression analysis to a two-dimensional space of peak discharge and urbanization. The fitted equation obtained follows as equation 9:

$$P = -1.1704 + 1.0129 y', \qquad (9)$$

where

P	is precipitation, and
y'	is log-transformed, segment intercept-subtracted peak discharge.

The linear regression model fit of peak discharge to precipitation is represented in figure 9 by the heavier-shaded plane sloping upward and to the left without variation along the urban fraction axis. The intersection of this plane with the quantile regression planes gives a set of lines, portrayed in
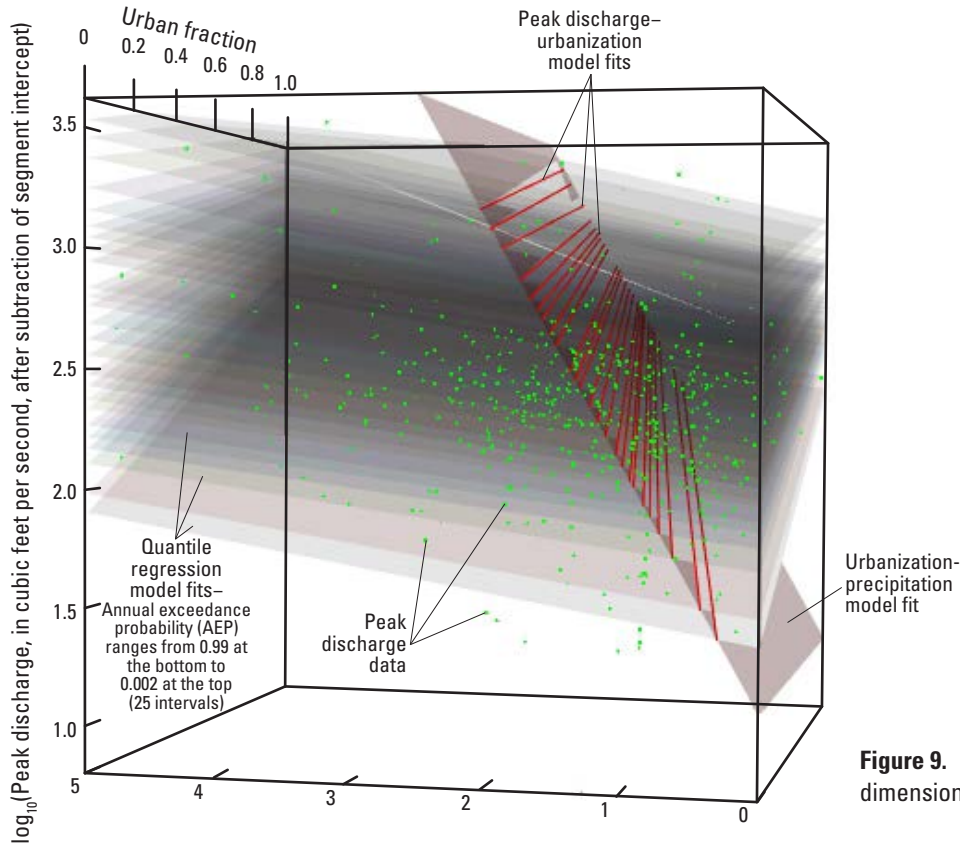
Figure 9.   Quantile regression model fit in three dimensions for this study in northeastern Illinois.

red in figure 9, indicating the dependence of peak discharge on urban fraction alone, given the fitted relation between peak discharge and precipitation. When viewed from the right-hand side of the cube in figure 9 along the precipitation axis, these red lines and the peak discharge data can be portrayed in a two-dimensional plot, as shown in figure 11, which shows the AEP-dependence of the urban fraction coefficients more clearly.

Discussion

The positive signs of the regression coefficients in the longitudinal and quantile regression models are certainly in agreement with physical reasoning, though the magnitude of the precipitation coefficient is small compared to the value one might expect if the causal precipitation (or, more generally, runoff generation forcing) for each event was accurately captured. Consider that the meaning of the precipitation coefficient is the proportional change in discharge to a unit change in precipitation, where the units of precipitation are inches. If a peak discharge is caused by, for example, a two-inch precipitation event rather than a one-inch event, the longitudinal precipitation coefficient of about 0.08 indicates that the peak discharge would increase by only about 8 percent. This

seems quite small, which indicates this value has been attenuated by the difficulty of measuring the causal precipitation. As such, the precipitation coefficient should not be taken as a physically realistic value; it is included in the model to help to correct for the portion of the trends in peak discharge series that are caused by trends in the precipitation rather than in the developed fraction of the watersheds. For a longitudinal linear regression model with only urbanization in the model (no precipitation), the urbanization coefficient increased slightly from 0.598 to 0.647 (table 5), implying that there was a generally increasing trend in precipitation, as was observed and is discussed in the section "Changes in Moments and Trends."

With respect to the urban fraction coefficient, where the change in peak discharge is proportional to differences in the urban fraction, if the urban fraction increases by 0.10, regardless of the starting point (whether 0.0–0.10 or 0.90–1.0 or any such interval), the peak discharge is predicted to increase by a factor of $10^{0.10 b_U^*(q)}$. If the urban fraction increased from 0 to 1, then the peak discharge is predicted to increase by a factor of $10^{b_U^*(q)}$. In either case, the proportional increase depends on the exceedance probability. The resulting values are tabulated in table 6 and vary, for an urban fraction increase of 0.10, from 1.25 at an AEP of 0.99 down to 1.08 at an AEP of 0.002. For an urban fraction increase from 0 to 1, the proportional change

Table 6. Quantile regression model coefficients for this study in northeastern Illinois.

[ft³/s, cubic feet per second. Mathematical notation references equations 3 and 4]

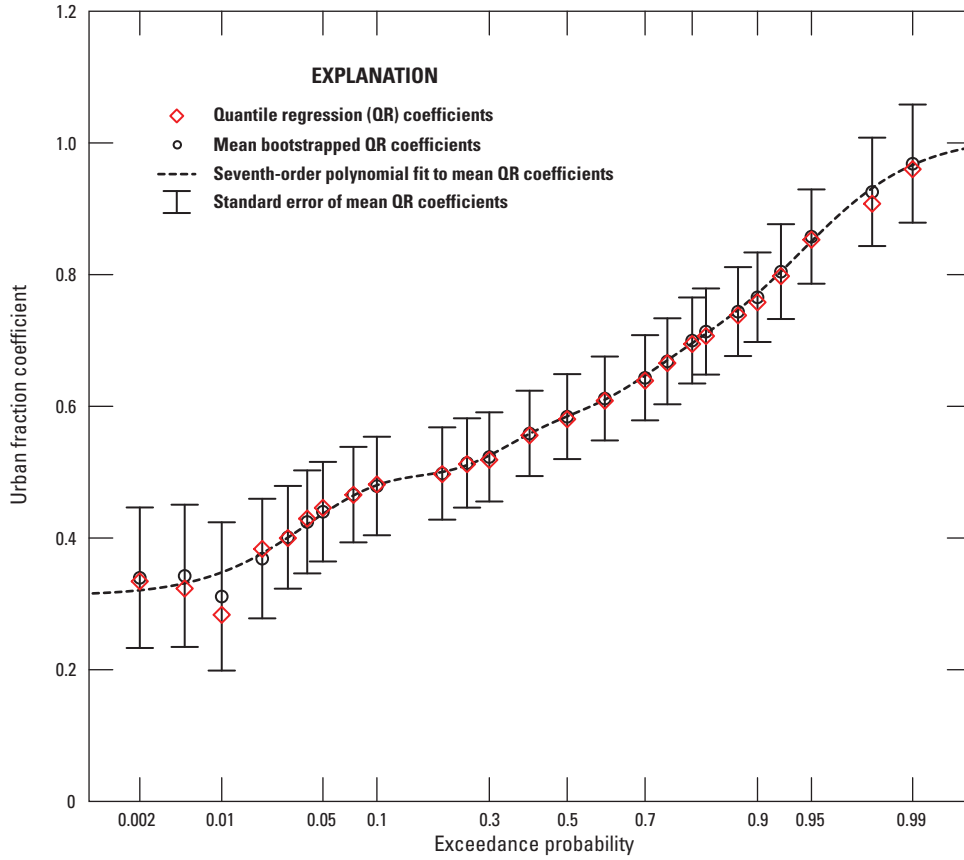| Exceedance probability q	Quantile regression coefficients			Mean bootstrapped quantile regression coefficients			Standard errors of mean bootstrapped coefficients			Urban fraction coefficients fitted by seventh-order polynomial $b_U^*(q)$ (log$_{10}$[ft³/s])	Effect of a 10-percent increase in urban fraction	Effect of a 100-percent increase in urban fraction
	Intercept $a'(q)$ (log$_{10}$[ft³/s])	Urban fraction $b_U'(q)$ (log$_{10}$[ft³/s])	Precipitation $b_P'(q)$ (log$_{10}$[ft³/s])/inch	Intercept $\bar{a}''(q)$ (log$_{10}$[ft³/s])	Urban fraction $\bar{b}_U''(q)$ (log$_{10}$[ft³/s])	Precipitation $\bar{b}_P''(q)$ (log$_{10}$[ft³/s])/inch	Intercept $SE(\bar{a}''(q))$ (log$_{10}$[ft³/s])	Urban fraction $SE(\bar{b}_U''(q))$ (log$_{10}$[ft³/s])	Precipitation $SE(\bar{b}_P''(q))$ (log$_{10}$[ft³/s])/inch			
0.99	1.448	0.960	0.087	1.428	0.969	0.096	0.067	0.090	0.011	0.967	1.25	9.27
0.98	1.538	0.908	0.094	1.537	0.926	0.090	0.058	0.082	0.012	0.931	1.24	8.53
0.95	1.725	0.853	0.078	1.727	0.858	0.076	0.056	0.072	0.008	0.851	1.22	7.10
0.925	1.826	0.798	0.071	1.822	0.805	0.070	0.058	0.072	0.007	0.805	1.20	6.38
0.9	1.888	0.758	0.067	1.885	0.766	0.067	0.052	0.068	0.006	0.771	1.19	5.90
0.875	1.927	0.738	0.068	1.924	0.744	0.069	0.053	0.067	0.007	0.746	1.19	5.57
0.825	1.991	0.707	0.071	1.989	0.714	0.071	0.051	0.065	0.006	0.710	1.18	5.13
0.8	2.014	0.695	0.074	2.012	0.700	0.073	0.052	0.065	0.006	0.696	1.17	4.97
0.75	2.062	0.666	0.075	2.060	0.668	0.075	0.052	0.065	0.005	0.670	1.17	4.68
0.7	2.105	0.639	0.076	2.102	0.643	0.076	0.052	0.065	0.006	0.647	1.16	4.44
0.6	2.168	0.608	0.080	2.168	0.612	0.079	0.051	0.064	0.005	0.610	1.15	4.07
0.5	2.220	0.581	0.082	2.219	0.584	0.082	0.049	0.065	0.004	0.585	1.14	3.85
0.4	2.269	0.556	0.086	2.268	0.559	0.085	0.049	0.065	0.005	0.558	1.14	3.61
0.3	2.330	0.519	0.089	2.329	0.523	0.088	0.050	0.068	0.005	0.525	1.13	3.35
0.25	2.356	0.512	0.090	2.355	0.514	0.090	0.049	0.068	0.005	0.511	1.12	3.24
0.2	2.392	0.497	0.090	2.390	0.498	0.091	0.050	0.070	0.005	0.500	1.12	3.16
0.1	2.473	0.481	0.096	2.476	0.479	0.095	0.051	0.075	0.007	0.480	1.12	3.02
0.075	2.506	0.466	0.101	2.511	0.466	0.098	0.050	0.073	0.007	0.465	1.11	2.92
0.05	2.566	0.446	0.094	2.564	0.440	0.097	0.052	0.076	0.008	0.437	1.11	2.74
0.04	2.587	0.430	0.100	2.589	0.424	0.100	0.053	0.078	0.011	0.421	1.10	2.64
0.03	2.620	0.400	0.110	2.621	0.401	0.109	0.053	0.078	0.013	0.401	1.10	2.52
0.02	2.661	0.383	0.120	2.670	0.369	0.118	0.059	0.091	0.015	0.377	1.09	2.38
0.01	2.767	0.283	0.138	2.776	0.311	0.121	0.070	0.113	0.022	0.348	1.08	2.23
0.005	2.870	0.323	0.114	2.840	0.343	0.122	0.071	0.108	0.021	0.331	1.08	2.14
0.002	2.931	0.334	0.109	2.897	0.340	0.128	0.090	0.107	0.035	0.321	1.08	2.09

Figure 10. Urban fraction coefficients from quantile regression as a function of exceedance probability for this study in northeastern Illinois.
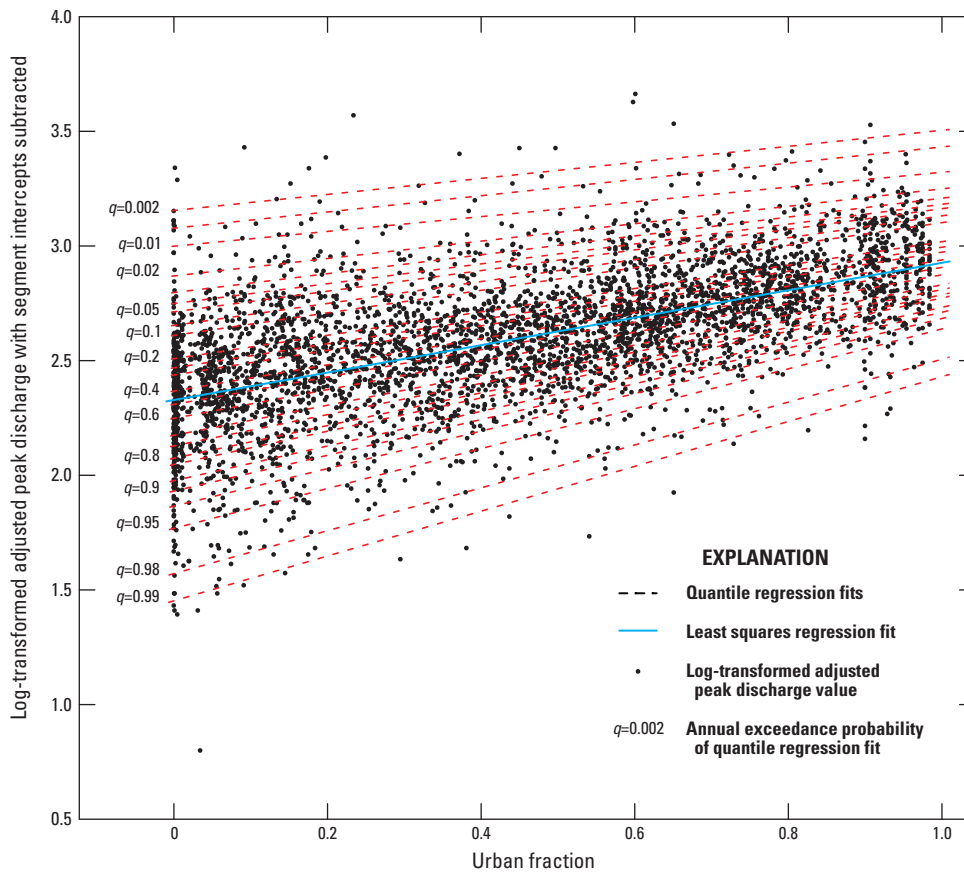

Figure 11. Quantile regression model fit viewed along the precipitation axis for this study in northeastern Illinois.

in peak discharge varies from 9.27 at an AEP of 0.99 down to 2.09 at an AEP of 0.002.

The urban fraction coefficient values obtained in this study are not commensurate with the coefficients in previous studies. These previous studies assumed a log-log relationship between discharge and impervious area (Allen and Bejcek, 1979; Sauer and others, 1983) and are investigating a slightly different problem: the effect of between-streamgage varia-tion in urbanization rather than the present case of temporal variation. Nevertheless a numerical comparison of the implied effect is illustrative. From Allen and Bejcek (1979), looking at the 2-year flood Q_2 (that is, AEP = 0.5), a relation between peak discharge and impervious area was fitted as a power-law with an exponent of 0.313, that is, $Q_2 \propto I^{0.313}$, where I is the frac-tion of impervious area. If it is further assumed that the devel-oped area is 30 percent impervious (see Allen and Bejcek, 1979, figures 3 and 4), then a 10 percent increase in urban fraction corresponds to a 3 percent increase in impervious-ness. The ratio of the original and more urbanized Q_2 values is given as $Q_2(I_2)/Q_2(I_1) = [I_2/I_1]^{0.313}$, which shows that the ratio of more developed to less developed Q_2 values depends on the ratio of larger impervious value I_2 to the smaller I_1. For example, if I goes from 15 to 18 percent then the Q_2 ratio is (18/15)0.313 = 1.059 or about a 6 percent increase, while if I goes from 1 to 4 percent, the Q_2 ratio is 1.543 or about a 54 percent increase. If I goes from 1 to 30 percent (from fully undeveloped to fully urbanized if urban land is 30 percent impervious as assumed here and because Allen and Bejcek (1979) assume that completely undeveloped watersheds have impervious fractions of 1 percent), then the Q_2 ratio is (30/1)0.313 = 2.900, or a 190 percent increase.

For a complete comparison between the results of the present study and those of Allen and Bejcek (1979), their figure 6, which shows predicted increases in peak discharge quantiles compared to such an undeveloped watersheds hav-ing $I = 0.01$, is reproduced here along with the prediction of this study assuming 30 percent imperviousness of urbanized land (fig. 12). According to this figure, though the curvatures are in the opposite direction, the magnitudes of the effects of urbanization on the peak discharge quantiles according to Allen and Bejcek (1979) and this study are fairly similar. Both indicate substantial increases in the effect of urbanization for increasing exceedance probabilities. The difference in curva-ture means that Allen and Bejcek (1979) predict larger initial effects increases in imperviousness with an attenuation of the effects as the increases become large, whereas the pres-ent study predicts smaller initial effects but increasing effects for larger increases. As discussed, in the present study, the log-linear model was selected because it produced a better fit to the data than the log-log (power-law) model in the longitu-dinal regression analysis. Whether Allen and Bejcek (1979) considered a log-linear model is not known.
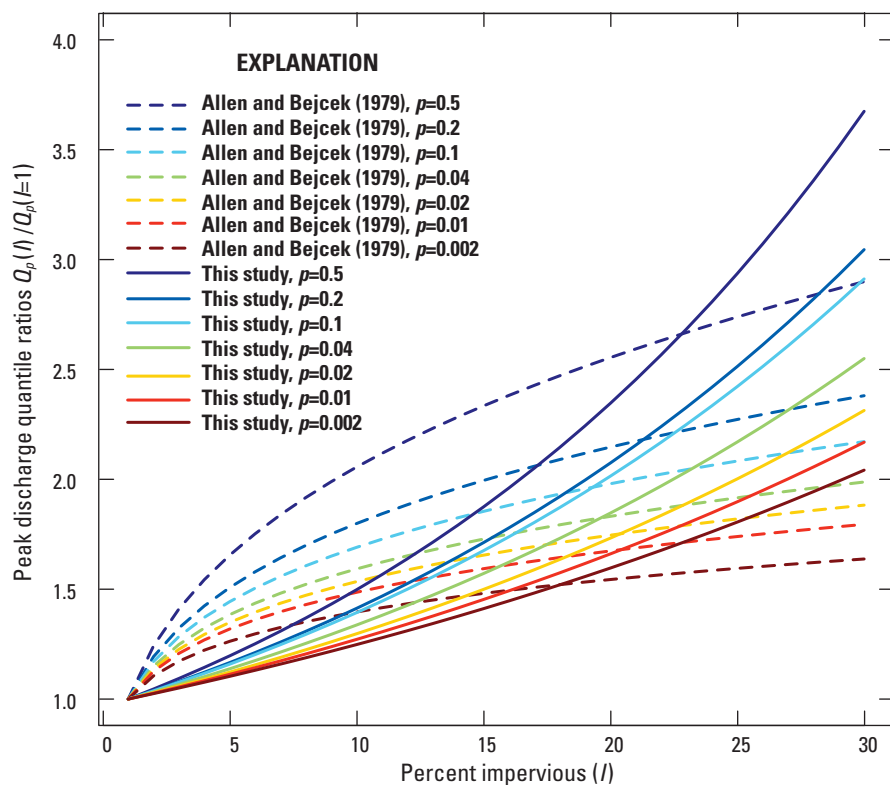

Figure 12. Comparison between Allen and Bejcek (1979) figure 6 and this study in northeastern Illinois, assuming urbanized area is 30-percent impervious.

Adjusted Peak Discharges

Exceedance probabilities were assigned to each peak discharge as described in the "Methodology" section, and the adjustment equation 4 was applied, using the fitted urban fraction coefficient function $b_U^*(q)$ (table 6) and segment intercepts for streamgage records having multiple segments (table 4), to obtain peak discharges adjusted to 2010 urbanization conditions (table 7, available at http://dx.doi.org/10.3133/sir20165049). The effects of the adjustment are illustrated below in two example streamgages and in an analysis of changes in peak discharge moments and trends.

Examples

The first example, USGS streamgage 05437950, Kishwaukee River near Huntley, Illinois (fig. 1), has a small (14.01 mi²) watershed. Its peak discharge record runs from 1965 to 1978, and there were no reservoirs constructed for its period of record, so it has only one segment (table 8 and fig. 13). Its urban fraction increased somewhat over the period of record from a small base but subsequently rose to an estimated 0.502 in 2010 (tables 1 and 8); therefore, the adjustment of the record is relatively large. The magnitude of the discharge adjustment, measured as a ratio to the observed discharge, is larger earlier in the record because the change in

urban fraction to the 2010 value is larger, and is also larger for the smaller peaks because of the increase in the urban fraction coefficients with exceedance probability (table 6 and fig. 10).

The second example, USGS streamgage 05535800, North Branch Chicago River at Morton Grove, Illinois (fig. 2), has a 92.04 mi² watershed. Its peak discharge record runs from 1960 to 1979 (table 9 and fig. 13B). Reservoirs were completed upstream of the gage in 1970, 1973, and 1979 (table 3); the first segment runs from 1960 to 1969 and the second segment from 1974 to 1978. The peak discharge values for the years from 1970 to 1973 and 1979 were not used in regressions, as indicated by the segment number of 0; 1970, 1973, and 1979 were removed because they are years when reservoirs were completed, and the period between 1970 and 1973 is too short to be defined as a separate segment. The peak discharges for 1970 to 1973 are adjusted by means of the interpolation technique (equation 10); the discharge for 1979 was not adjusted because interpolation cannot be applied at an end of the streamgage record. The urban fraction increases modestly throughout the period of record, and for 2010 conditions is estimated as 0.830 (tables 1 and 9). The values of the ratio of adjusted to observed peak discharges reflect increases to account for the changes in urban fraction and variations in the exceedance probability, and, for segment 1, a decrease of 0.127 (in log10 units), equivalent to applying a factor of $10^{-0.127} = 0.746$ to the adjusted discharge values in segment 1, resulting from the difference in the intercepts between the two

Table 8. Observed and urban-adjusted annual maximum peak discharges for U.S. Geological Survey streamgage 05437950, Kishwaukee River near Huntley, Illinois.

[ft³/s; cubic foot per second; NA, not applicable; NC, not computed]

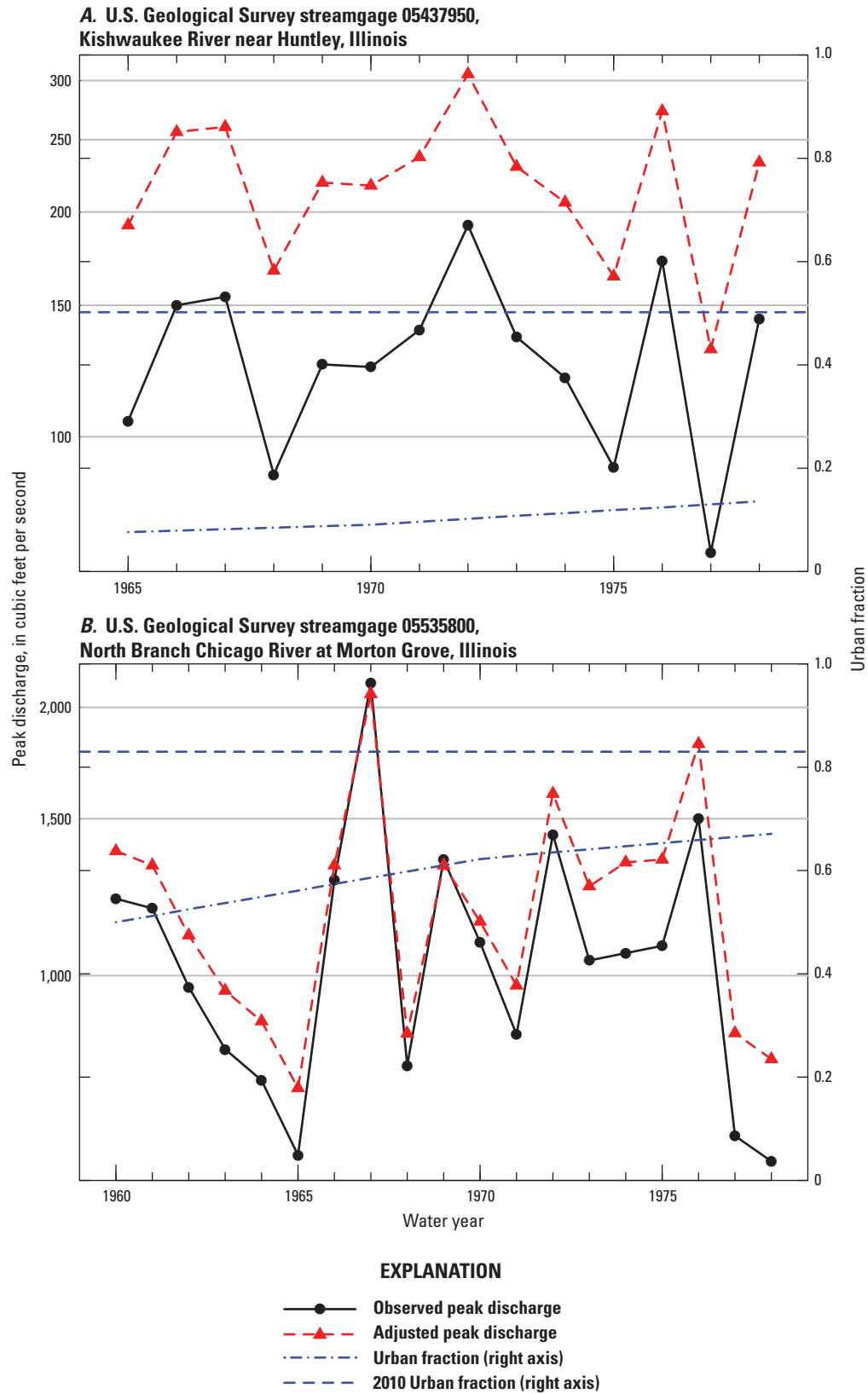| Water year	Segment	Cumulative maximum storage (acre-feet)	Urban fraction	Precipitation (inches)	Observed peak discharge (ft³/s)	Urbanization-adjusted peak discharge (ft³/s)	Ratio of urbanization-adjusted to observed peak discharges	Exceedance probability of peak discharges
1965	1	0	0.076	0.040	105	192	1.829	0.624
1966	1	0	0.079	0.616	150	256	1.707	0.367
1967	1	0	0.082	0.855	154	260	1.688	0.352
1968	1	0	0.085	1.770	89	167	1.876	0.725
1969	1	0	0.088	1.771	125	219	1.752	0.510
1970	1	0	0.091	2.110	124	217	1.750	0.519
1971	1	0	0.096	0.468	139	237	1.705	0.436
1972	1	0	0.102	1.045	192	306	1.594	0.224
1973	1	0	0.108	1.504	136	230	1.691	0.465
1974	1	0	0.113	0.979	120	206	1.717	0.568
1975	1	0	0.119	0.010	91	164	1.802	0.739
1976	1	0	0.124	2.461	172	273	1.587	0.314
1977	1	0	0.130	0.873	70	131	1.871	0.852
1978	1	0	0.136	2.641	144	233	1.618	0.449
2010	NA	NC	0.502	NA	NA	NA	NA	NA

Figure 13. Observed and urbanization-adjusted annual maximum peak discharges for example U.S. Geological Survey streamgages. *A*, 05437950, Kishwaukee River near Huntley, Illinois, and *B*, 05535800, North Branch Chicago River at Morton Grove, Illinois.

Table 9.   Observed and urbanization-adjusted peak discharges for U.S. Geological Survey streamgage 05535800, North Branch Chicago River at Morton Grove, Illinois.

[ft^3/s; cubic foot per second; NA, not applicable; NC, not computed]

Water year	Segment	Cumulative maximum storage (acre-feet)	Urban fraction	Precipitation (inches)	Observed peak discharge (ft^3/s)	Urbanization-adjusted peak discharge (ft^3/s)	Ratio of urbanization-adjusted to observed peak discharges	Exceedance probability of peak discharges
1960	1	0	0.500	0.546	1,220	1,380	1.131	0.352
1961	1	0	0.512	1.310	1,190	1,330	1.118	0.386
1962	1	0	0.525	0.251	971	1,110	1.143	0.595
1963	1	0	0.537	1.284	827	962	1.163	0.726
1964	1	0	0.549	0.811	763	889	1.165	0.787
1965	1	0	0.561	0.143	629	748	1.189	0.882
1966	1	0	0.574	1.251	1,280	1,330	1.039	0.390
1967	1	0	0.586	1.413	2,130	2,070	0.972	0.084
1968	1	0	0.598	1.178	793	861	1.086	0.811
1969	1	0	0.610	1.713	1,350	1,330	0.985	0.384
1970	0	961	0.622	1.708	1,090	1,150	1.055	0.558
1971	0	961	0.629	0.386	859	975	1.135	0.714
1972	0	961	0.635	1.172	1,440	1,600	1.111	0.207
1973	0	1,561	0.641	1.201	1,040	1,260	1.212	0.451
1974	2	1,561	0.647	1.082	1,060	1,340	1.264	0.381
1975	2	1,561	0.653	0.950	1,080	1,350	1.250	0.371
1976	2	1,561	0.659	1.441	1,500	1,820	1.213	0.142
1977	2	1,561	0.665	0.806	661	862	1.304	0.810
1978	2	1,561	0.671	1.448	619	806	1.302	0.846
1979	0	2,961	0.677	0.904	1,000	NC	NC	NC
2010	NA	NC	0.830	NA	NA	NA	NA	NA

segments (table 4), as indicated in the adjustment equation 4. As a result of the decrease of 0.127 applied to segment 1 peak discharge values, the adjustment of these discharge values to 2010 conditions accounts for the effects of the reservoirs built of the period of the streamgage and the increases in urbanization through 2010.

Changes in Moments and Trends

To illustrate the overall effect of the adjustments of the peak discharges, differences in the mean, standard deviation, and skewness of the original and adjusted peak discharge series by streamgage are given in figure 14. From this figure it is evident that the urbanization adjustment, where any change occurred, almost always increased the mean log-peak discharge and decreased the standard deviation, and more often than not increased the skewness. The cause of the increase in mean is obvious, because the peak discharges are being increased in the adjustment process to reflect the effect

of increases in urbanization, with occasional offsets from decreases in segment intercepts as seen in the example of streamgage 05535800. The decrease in the standard deviation is interpreted to come from two effects: one is that, for the most part, the peak discharge records are having their time trends reduced (fig. 15); the other is that because of the exceedance probability-dependence of the adjustment coefficients (fig. 10), smaller peaks are increased more than are larger ones. The overall decrease in skewness also results from the exceedance probability-dependence of the adjustment coefficients, because the larger increase in smaller values as compared to larger values will tend to shorten the left tail of the probability distributions of the peak discharges at a given streamgage more than the increases in the higher values stretches out the right tail.

The effect of the adjustments on trends is shown in figure 15. In these figures, the trends are measured by the Kendall's tau correlation between the data and time. Kendall's tau is a nonparametric rank-based measure of correlation which measures the strength of a monotonic, not necessarily linear,
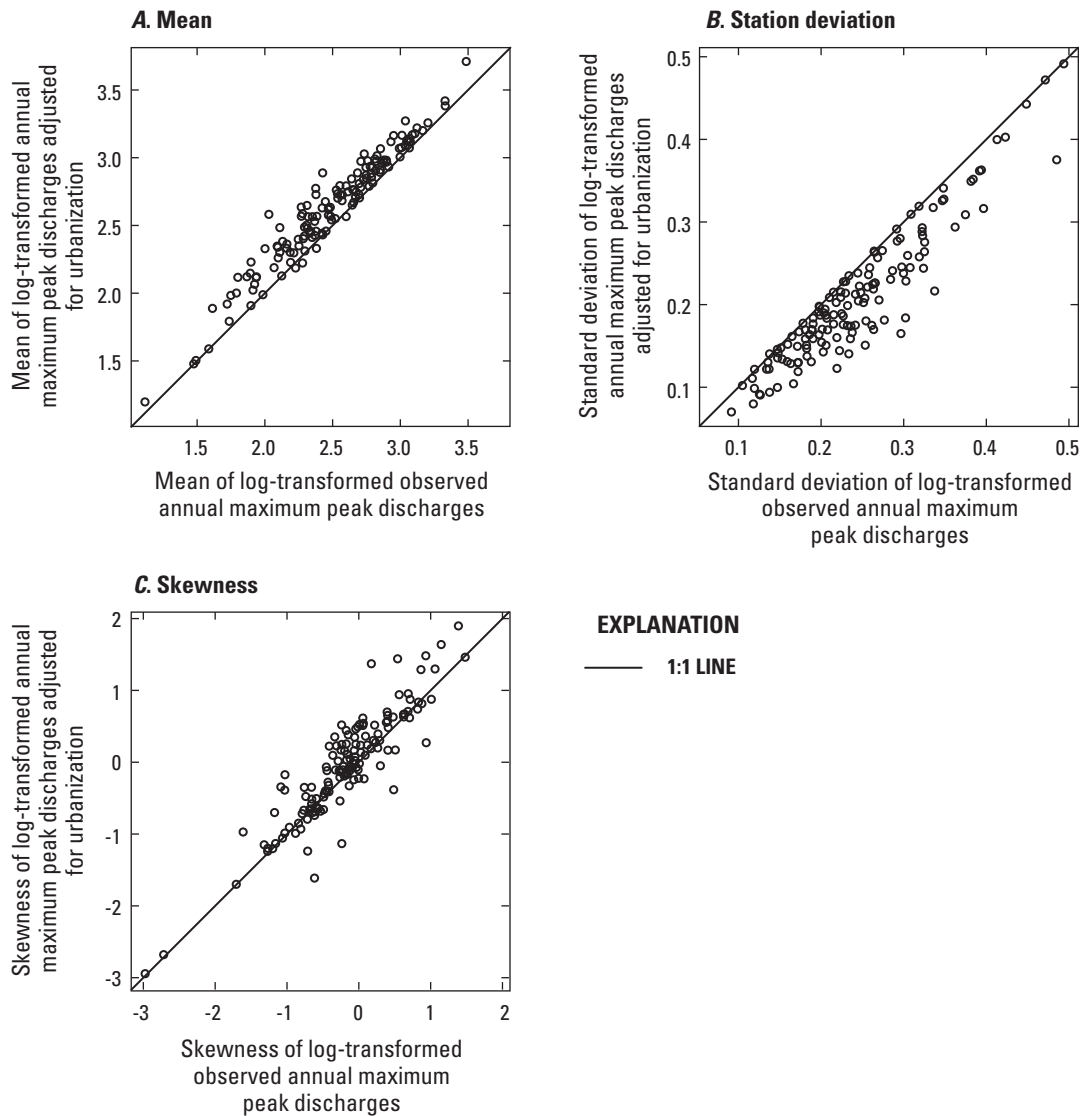
Figure 14. Relation between moments urbanization-adjusted and observed log-transformed peak discharge values for the 143 U.S. Geological Survey streamgages in this study in northeastern Illinois, with *A*, mean; *B*, standard deviation; and *C*, skewness.

relationship, and Kendall's tau was chosen because of its resistance to nonlinearity and skewness (Helsel and Hirsch, 2002). Figure 15*A, B* shows that the number, size, and significance of positive trends were reduced as a result of the urbanization adjustment of the peak discharges, but positive trends still predominate slightly. For example, 76.9 percent of observed peak discharge records have positive Kendall's tau correlation values, and 27.3 percent are positive and significant at

a *p*-value of 0.05 or less. For the adjusted peak discharge records, 59.4 percent have positive Kendall's tau correlations, and 7.0 percent are positive and significant at a p-value of 0.05 or less. At least partly the continued predominance of positive time trends in the adjusted peak discharge records occurs because of more positive than negative precipitation trends (fig. 15*C*), which the adjustment procedure was not designed to remove.
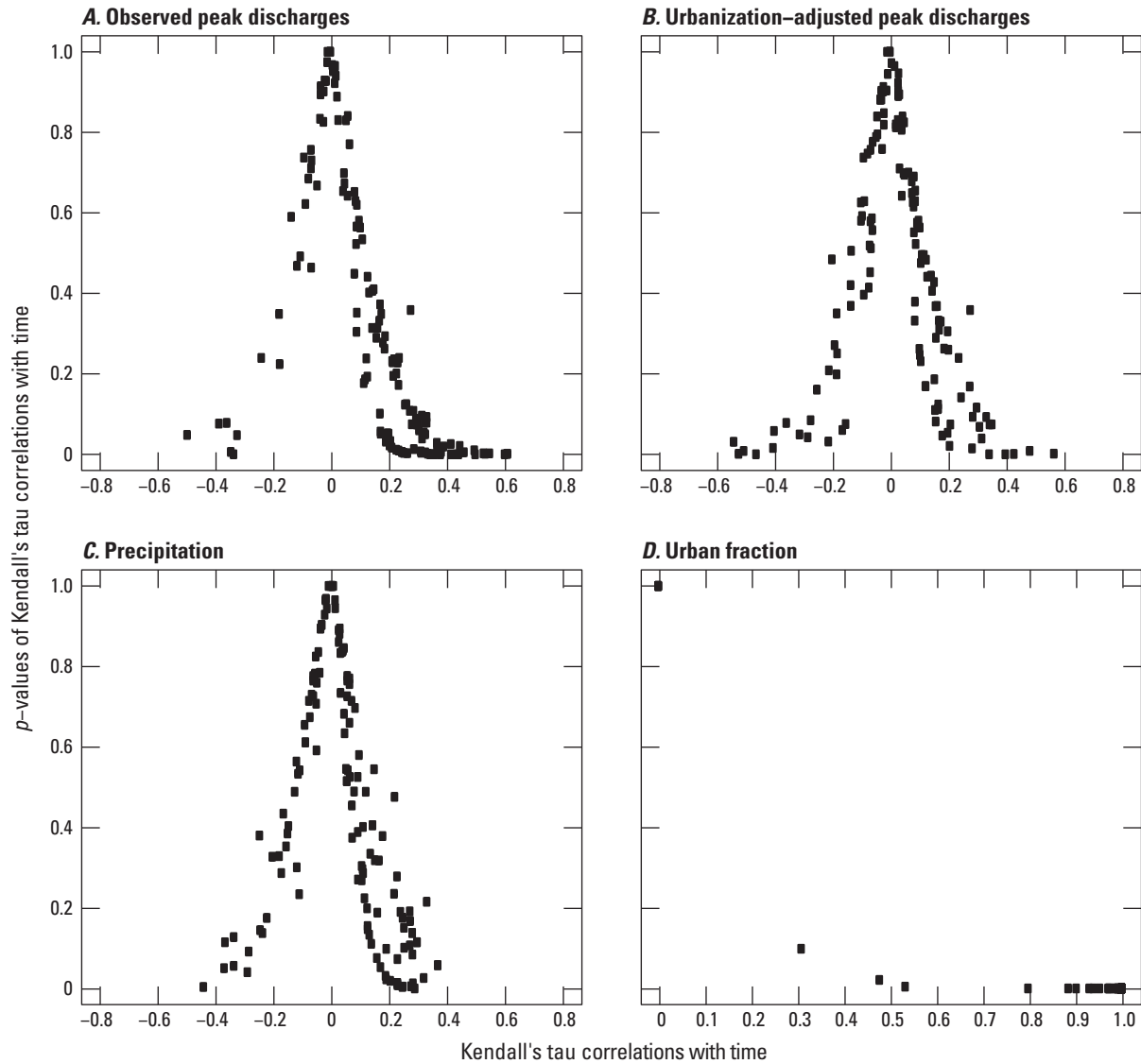
Figure 15. Relations between Kendall's tau *p*-values and correlations with time of peak discharge records at 143 U.S. Geological Survey streamgages in this study in northeastern Illinois for *A*, observed peaks; *B*, urbanization-adjusted peaks; *C*, precipitation; and *D*, urban fraction.

Summary

In this study, a two-step longitudinal-quantile linear model regression technique was developed to determine the effect of urbanization on annual maximum peak discharges, to provide a means to adjust the peak discharge series to urbanization conditions at a selected year, and to adjust for the effect of reservoirs completed during the period of record for each streamgage. The technique was applied, using urban fraction and precipitation as explanatory variables, to peak discharges from 143 streamgage records in northeastern Illinois and surrounding areas having drainage areas of at most 200 mi² and at least 10 years of record. In the first step of the regression analysis, linear longitudinal regression models with fixed intercepts estimated for each segment of the peak discharge records were computed. The segment intercepts were then subtracted from the discharge records to homogenize the discharge dataset across the segments in preparation for the quantile regression analysis. In the second step of the regression analysis, quantile regression was applied to estimate the regional effect of the explanatory variables as a function of exceedance probability. The effect of urbanization as estimated by the urban fraction coefficient function obtained from the quantile regression analysis was determined to decrease monotonically from 0.969 to 0.340 with exceedance probabilities decreasing from 0.99 to 0.002 (small to large peak discharges), quantifying the degree to which urbanization affects smaller peak discharges more strongly than large peak discharges in this study.

Based on the regression modeling results, the peak discharges were adjusted to 2010 urbanization conditions. The resulting adjusted peak discharge records were seen to have, on average, larger means, smaller standard deviations, and larger skews, and had a reduced incidence of significant trends as measured by Kendall's tau correlation. In particular, for almost every streamgage where the urbanization increased and thus an adjustment was applied, the mean increased and the standard deviations decreased, whereas the effect on skews was more variable, though increasing on average. Significant Kendall's tau correlation ($p < 0.05$) was observed in 27.3 percent of streamgage records with the observed peak discharges but in only 7.0 percent of streamgage records after adjustment.

The adjusted discharges and associated explanatory variables also can provide a basis for a regionalization study to provide equations to estimate peak discharge quantiles at ungaged sites in northeastern Illinois, as an update of Allen and Bejcek (1979) and as a complement to the regionalization study of Soong and others (2004) for rural streams in Illinois. Implementation of the results of such a study in StreamStats (http://streamstats.usgs.gov) could further assist the application and distribution of the peak discharge statistics for scientific and engineering purposes. The techniques described here also could be applied to daily discharge statistics to help extend our understanding of urban effects on a wider range of streamflow statistics, such as low flows, flow duration curves, and seasonal flows.

References Cited

Allen, H.E., Jr. and Bejcek, R.M., 1979, Effects of urbanization on the magnitude and frequency of floods in northeastern Illinois: U.S. Geological Survey Water Resources Investigations 79–36; 48 p.

Canay, I.A., 2011, A simple approach to quantile regression for panel data: The Econometrics Journal, v. 14, p. 368–386.

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/.

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

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

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

Hesterberg, T., Moore, D.S., Monaghan, S., Clipson, A., and Epstein, R., 2006, Bootstrap methods and permutation tests, in Moore, D.S., and McCabe, G.P., (2005), Introduction to the practice of statistics (5th ed.): New York, W.H. Freeman and Co. [variously paged].

Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., and Megown, K., 2015, Completion of the 2011 National Land Cover Database for the conterminous United States—Representing a decade of land cover change information: Photogrammetric Engineering and Remote Sensing, v. 81, no. 5, p. 345–354.

Ishii, A.L., Soong, D.T., and Sharpe, J.B., 2010, Implementation and evaluation of the Streamflow Statistics (Stream-Stats) Web application for computing basin characteristics and flood peaks in Illinois: U.S. Geological Survey Scientific Investigations Report 2009–5197, 37 p.

Karstensen, K.A., Shaver, D.K., Alexander, R.L, Over, T.M., and Soong, D., 2013, Land change in the Central Corn Belt Plains Ecoregion and hydrologic consequences in developed areas—1939–2000: U.S. Geological Survey Open-File Report 2013–1157, 21 p. [Also available at http://pubs.usgs.gov/of/2013/1157/.]

Koenker, R., 2013, quantreg—Quantile Regression: R package version 5.05, accessed November 6, 2014, at http://CRAN.R-project.org/package=quantreg.

Koenker, R., and Bassett, G., 1978, Regression quantiles: Econometrica, v. 46, no. 1, p. 33–50.

Konrad, C.P., 2003, Effects of urban development on floods: U.S. Geological Survey Fact Sheet FS-076–03, 4 p.

Loveland, T.R., Sohl, T.L., Stehman, S.V., Gallant, A.L., Sayler, K.L., and Napton, D.E., 2002, A strategy for estimating the rates of recent United States land cover changes: Photogrammetric Engineering and Remote Sensing, v. 68, no. 10, p. 1091–1099.

Metropolitan Water Reclamation District of Greater Chicago, 2009, Detailed watershed plan for the Calumet-Sag Channel Watershed—Volume 1: [variously paged].

Metropolitan Water Reclamation District of Greater Chicago, 2009, Detailed watershed plan for the Upper Salt Creek Watershed—Volume 1: [variously paged].

Metropolitan Water Reclamation District of Greater Chicago, 2010, Detailed watershed plan for the Poplar Creek Watershed—Volume 1: [variously paged].

Metropolitan Water Reclamation District of Greater Chicago, 2010, Little Calumet River detailed watershed plan Phase B report—Volume 1 of 2: Prepared by CDM, in association with AECOM, FluidClarity, and Molly O'Toole and Associates [variously paged].

Metropolitan Water Reclamation District of Greater Chicago, 2011, Detailed watershed plan for the North Branch of the Chicago River and Lake Michigan Watershed—Volume 1, [variously paged].

Moglen, G.E., and Shivers, D.E., 2006, Methods for adjusting USGS rural regression peak discharges in an urban setting: U.S. Geological Survey Scientific Investigation Report 2006–5270, 65 p.

Murphy, E.A. and Ishii, A.L., 2006, Watershed Data Management (WDM) database for Salt Creek streamflow simulation, DuPage County, Illinois: U.S. Geological Survey Open-File Report 2006–1248, 34 p.

R Core Team, 2014, R—A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing, accessed November 6, 2014, at http://www.R-project.org.

Resource Coordination Policy Committee, 1998, Our community and flooding: A report of the status of floodwater management in the Chicago Metropolitan Area, 72 p.

Ries, G.K., and Dillow, J.A., 2006, Magnitude and frequency of floods on nontidal streams in Delaware: U.S. Geological Survey Scientific Investigation Report 2006–5146, 38 p.

Sauer, V.B., Thomas, W.O., Jr., Stricker, V.A., and Wilson, K.V., 1983, Flood characteristics of urban watersheds in the United States: U.S. Geological Survey Water-Supply Paper 2207, 63 p.

Shaeffer, J.R., Ellis, D.W., and Spieker, A.M., 1970, Flood-hazard mapping in metropolitan Chicago: U.S. Geological Survey Circular 601–C, 14 p.

Soong, D.T., Ishii, A.L., Sharpe, J.B., and Avery, C.F., 2004, Estimating flood-peak discharge magnitudes and frequencies for rural streams in Illinois: U.S. Geological Survey Scientific Investigations Report 2004–5013, 147 p.

Theobald, D., 2005, Landscape patterns of exurban growth in the USA from 1980 to 2020: Ecology and Society, v. 10, no. 1, article 32, 29 p. [Also available at http://www.ecologyandsociety.org/vol10/iss1/art32/.]

U.S. Army Corps of Engineers, Chicago District, 2011, Upper Des Plaines River and tributaries—Integrated feasibility report and environmental assessment—Volume 1, Study Overview: U.S. Army Corps of Engineers, 52 p.

U.S. Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood flow frequency—Bulletin #17B of the Hydrology Subcommittee (revised and corrected): U.S. Geological Survey, Office of Water Data Coordination, 28 p. [plus appendixes].

Venables, W.N., and Ripley, B.D., 2002, Modern applied statistics with S (4th ed.): New York, Springer, 498 p.

Westcott, N., 2012, Continued operation of a 25-raingage network for collection, reduction, and analysis of precipitation data for Lake Michigan diversion accounting—Water year 2011: Illinois State Water Survey, Champaign, Illinois, 80 p.

Appendixes 1–2

Appendix 1.   Quantile Regression

Quantile regression (Koenker and Bassett, 1978; Cade and Noon, 2003; Koenker, 2005; Hao and Naiman, 2007; Davino and others, 2014) provides estimates of conditional quantiles rather than conditional expectations that are obtained by applying least-squares regression. Quantile regression estimates are computed by using a quantile-dependent weighting function, whereas linear least-squares regression finds the coefficient estimates $\hat{\beta}$ of the conditional mean expectation of a linear model by minimizing the squares of the residuals, that is,

$$\hat{\beta} = \min_{b} \sum_{i=1}^{n} \left(\left(y_i - Xb \right)^2 \right), \tag{10}$$

where

y	is the vector of observations,
X	is the matrix of explanatory variables, and
b	is the vector of regression coefficients.

Linear quantile regression obtains quantile-dependent coefficients $\hat{\beta}_\tau$ of the conditional quantile function by minimizing

$$\hat{\beta}_\tau = \min_{b} \sum_{i=1}^{n} \left(\rho_\tau \left(y_i - Xb \right) \right), \tag{11}$$

where
τ indicates a probability, that is, $\Pr\left(Y \le y_\tau \right) = \tau$, and

$$\rho_\tau(u) = I(u > 0) \tau |u| + I(u \le 0)(1 - \tau)|u|, \tag{12}$$

where

$I(x)$ is the indicator function, that is, $I(x) = 1$ if x is true else $I(x) = 0$.
When $\tau = 0.5$, which indicates the conditional median function is being computed, $\rho_\tau(u)$ reduces to

$$\rho_{0.5}(u) = (1/2)|u|\left[I(u > 0) + I(u \le 0) \right] = (1/2)|u|, \tag{13}$$

and $\hat{\beta}_\tau$ to

$$\hat{\beta}_{0.5} = \min_{b} \sum_{i=1}^{n} \left((1/2)|y_i - Xb| \right) = \min_{b} \sum_{i=1}^{n} \left(|y_i - Xb| \right), \tag{14}$$

so quantile regression reduces to the computation of least absolute deviations, which is well-known as the method of computation of the conditional median function (Gilchrist, 2008).

The quantile regression minimization problem turns out to be a linear programming problem. The software used in this study to solve the linear quantile regression problem was the rq function from the R package quantreg (Koenker, 2013).

References Cited

Cade, B.S., and Noon, B.R., 2003, A gentle introduction to quantile regression for ecologists: Frontiers in Ecology and the Environment, v. 1, no. 8, p. 412–420.

Davino, C., Furno, M., and Vistocco, D., 2014, Quantile regression—Theory and applications: Chichester, John Wiley and Sons, 276 p.

Gilchrist, W., 2008, Regression revisited: International Statistical Review, v. 78, no. 3, p. 401–418. [Also available at http://dx.doi.org/10.1111/j.1751-5823.2008.00053.x.]

Hao, L., and Naiman, D.Q., 2007, Quantile regression: Thousand Oaks, Calif., SAGE Publications, 126 p.

Koenker, R., 2005, Quantile regression: Cambridge, England, Cambridge University Press, 349 p.

Koenker, R., 2013, quantreg—Quantile Regression: R package version 5.05, accessed November 6, 2014, at http://CRAN.R-project.org/package=quantreg.

Koenker, R., and Bassett, G., 1978, Regression quantiles: Econometrica, v. 46, no. 1, p. 33–50.

Appendix 2. Adjustment of Commercial/Industrial/ Transportation Land Use Values in Census-Based Housing Density Data

One problem with the housing density dataset (Theobald, 2005) used in this study is that the commercial/industrial/transportation (C/I/T) land-cover class (class 10) is fixed at its year 2000 coverage, when in reality it varies in time just as the housing density values do. To develop a remedy to this problem, the dependence of this class on the other classes in the year 2000 dataset was explored, so that the class 10 fraction of each study watershed could be adjusted decadally following the variation of another class or classes. The best relation determined was between class 10 and class 9 (areas with less than 0.6 acres per housing unit). This relation is shown in figure 2–1, where each data point represents the land use for a basin in the study. A quadratic equation was fit to the relation, constrained to go through (0,0) with the assumption that when there is no housing of the highest density there is likely no commercial/industrial/transportation (C/I/T) land use either, and through (1,0) based on the more certain fact that when the class 9 fraction is 1.0, there is no other land use. The scatter around the fitted line is accounted for in the procedure used to estimate the C/I/T fraction for decades before and after 2000, which is as follows:

1. The fitted equation from figure 2–1 is used to obtain an initial estimate of the C/I/T fraction $f_{C/I/T-est}(t)$ for that year t based on the concurrent class 9 fraction $f_{C/I/T-est}(t)$:

    $$f_{C/I/T-est}(t) = 0.13774 - 0.55096 \left(f_{Class9}(t) - 0.5 \right)^2 \cdot$$

2. This initial estimate is adjusted by multiplying by the ratio of the predicted and observed C/I/T fractions for the year 2000 to obtain an adjusted estimate $f'_{C/I/T-est}(t)$:

    $$f'_{C/I/T-est}(t) = \left(\frac{f_{C/I/T}(2000)}{f_{C/I/T-est}(2000)} \right) f_{C/I/T-est}(t) \text{ if } f_{C/I/T-est}(2000) > 0 \text{ ; otherwise}$$
    $$f'_{C/I/T-est}(t) = 0 \cdot$$

3. Notice that if $f_{Class9}(2000) = 0$, the steps above imply $f_{C/I/T-est}(2000) = 0$ and thus $f'_{C/I/T-est}(t) = 0$. However, sometimes it happens that $f_{C/I/T}(2000) > 0$ when $f_{Class9}(2000) = 0$. In that case, $f'_{C/I/T-est}(t)$ was set to $f_{C/I/T}(2000)$ rather than zero.

4. Finally, if $f'_{C/I/T-est}(t) > f_{C/I/T}(2000)$ then set $f'_{C/I/T-est}(t) = f_{C/I/T}(2000)$. This condition keeps the total urbanization from exceeding 1 as occasionally occurred for $t = 2010$. As one check on the method, notice that for the year 2000 the final estimate $f'_{C/I/T-est}(t)$ is equal to the observed value $f_{C/I/T}(2000)$.
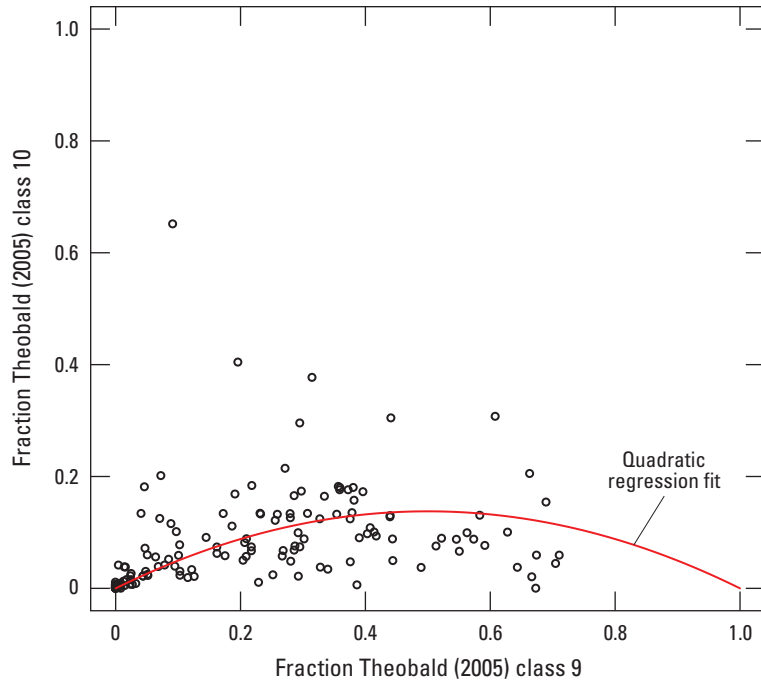
Figure 2–1.   Relation between class 10 (commercial/industrial/
transportation) fraction and year 2000 housing density class 9 (lot size <0.6
acres) fraction in Theobald's (2005) housing density dataset for the 143
watersheds used in this study.

Reference Cited

Theobald, D., 2005, Landscape patterns of exurban growth in the USA from 1980 to 2020:
Ecology and Society, v. 10, no. 1, article 32, 29 p. [Also available at http://www.ecologyand-
society.org/vol10/iss1/art32/.]