

Prepared in cooperation with the Iowa Department of Transportation and the Iowa Highway Research Board (Project TR-519)

# Methods for Estimating Annual Exceedance-Probability Discharges for Streams in Iowa, Based on Data through Water Year 2010



Scientific Investigations Report 2013–5086

**Cover photograph.** View looking southeast at Oakville, Iowa, June 15, 2008; main channel of the Iowa River is just out of view in the foreground. Photograph by [www.kenpurdy.com](http://www.kenpurdy.com).

# **Methods for Estimating Annual Exceedance-Probability Discharges for Streams in Iowa, Based on Data through Water Year 2010**

By David A. Eash, Kimberlee K. Barnes, and Andrea G. Veilleux

Prepared in cooperation with the Iowa Department of Transportation and the Iowa Highway Research Board (Project TR-519)

Scientific Investigations Report 2013–5086

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

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

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

U.S. Geological Survey, Reston, Virginia: 2013

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://www.usgs.gov/pubprod>

To order this and other USGS information products, visit <http://store.usgs.gov>

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

Eash, D.A., Barnes, K.K., and Veilleux, A.G., 2013, Methods for estimating annual exceedance-probability discharges for streams in Iowa, based on data through water year 2010: U.S. Geological Survey Scientific Investigations Report 2013–5086, 63 p. with appendix.

# Contents

|  |    |
|--|----|
| Abstract.....  | 1  |
| Introduction .....   | 1  |
| Purpose and Scope .....  | 2  |
| Description of Study Area .....  | 2  |
| Previous Studies .....   | 2  |
| Methods for Dataset Development for Streamgages.....   | 6  |
| Basin Characteristics.....   | 7  |
| Geographic Information System Measurements .....   | 7  |
| Annual Peak-Discharge Data .....   | 8  |
| Trend Analyses.....  | 9  |
| Annual Exceedance-Probability Analyses .....   | 10 |
| Bulletin 17B/GB Analyses.....  | 10 |
| Expected Moments Algorithm (EMA/MGB) Analyses.....   | 11 |
| Multiple Grubbs-Beck (MGB) Test for Detecting Low Outliers.....  | 13 |
| Regional Skew Analysis .....   | 14 |
| Regional Regression Analyses to Estimate Annual Exceedance-Probability Discharges for Ungaged Stream Sites.....                  | 16 |
| Definition of Flood Regions .....  | 17 |
| Development of Regional Regression Equations .....   | 20 |
| Multiple-Linear Regression.....  | 20 |
| Ordinary-Least-Squares Regression.....   | 20 |
| Generalized-Least-Squares Regression.....  | 23 |
| Final Regression Equations.....  | 25 |
| Accuracy and Limitations of Regression Equations .....   | 27 |
| Prediction Intervals .....   | 27 |
| Application of Regression Equations .....  | 29 |
| Example 1 .....  | 29 |
| Example 2 .....  | 29 |
| Weighted Method to Estimate Annual Exceedance-Probability Discharges for Streamgages.....  | 30 |
| Example 3 .....  | 31 |
| Weighted Methods to Estimate Annual Exceedance-Probability Discharges for Ungaged Sites on Gaged Streams.....                    | 31 |
| Regression-Weighted Estimates for Ungaged Sites on Gaged Streams .....   | 31 |
| Example 4 .....  | 32 |
| Area-Weighted Estimates for Ungaged Sites on Gaged Streams.....  | 32 |
| Example 5 .....  | 33 |
| Estimates for Ungaged Sites on Gaged Streams Between Two Streamgages.....  | 33 |
| Weighted Method to Estimate Annual Exceedance-Probability Discharges for Ungaged Sites Draining More Than One Flood Region ..... | 33 |
| Region-of-Influence Method to Estimate Annual Exceedance-Probability Discharges for Ungaged Stream Sites.....                    | 34 |
| Comparison of Annual Exceedance-Probability Discharges .....   | 35 |
| Estimates from Annual Exceedance-Probability Analyses .....  | 35 |

|  |    |
|--|----|
| Estimates from Regional Regression Equations ..... | 38 |
| StreamStats .....                                  | 40 |
| Maximum Floods in Iowa.....                        | 40 |
| Summary.....                                       | 42 |
| Acknowledgments.....                               | 44 |
| References Cited.....                              | 44 |
| Appendix.....                                      | 51 |
| Regional Skewness Regression .....                 | 52 |

## Figures

|  |    |
|--|----|
| 1. Map showing location of flood regions and streamgages evaluated for use in the regional skew analysis and annual exceedance-probability regressions for Iowa .....  | 3  |
| 2. Map showing soil regions in Iowa.....   | 4  |
| 3. Map showing landform regions in Iowa .....  | 5  |
| 4. Graph showing annual exceedance-probability curves for Bloody Run Tributary near Sherrill, Iowa (05414605), showing the difference between expected moments algorithm (EMA/MGB) and Bulletin 17B/GB annual exceedance-probability analyses for a crest-stage gage (CSG) with four annual peak discharges below the minimum recording threshold.....                                   | 12 |
| 5. Graph showing annual exceedance-probability curves for West Branch Floyd River near Struble, Iowa (06600300), showing the effects of including or censoring potentially influential low flows identified from the multiple Grubbs-Beck test and of using the updated regional-skew-coefficient constant or the superseded regional skew-coefficient value.....                        | 13 |
| 6. Map showing location of basin centroids for 240 streamgages used for regional skew analysis for Iowa .....  | 15 |
| 7. Graphs showing relation between one-percent annual exceedance-probability discharges and drainage area for <i>A</i> , eight initial; <i>B</i> , three combined; and <i>C</i> , six final flood regions defined for study area .....   | 19 |
| 8. Graphs showing relation between one-percent annual exceedance-probability discharges and drainage area for 127 streamgages in flood region 3 for <i>A</i> , log 10 transformed drainage area and <i>B</i> , power-transformed drainage area.....  | 22 |
| 9. Screen capture of the weighted-multiple-linear regression program (WREG) smoothing function for generalized-least-squares (GLS) correlation of the time series of annual peak discharges as a function of distance between 176 streamgages in flood region 2 with 30 years of concurrent discharge .....  | 23 |
| 10. Graphs showing relation between one-percent annual exceedance-probability discharges computed from observed streamflow and those predicted from regression equations for flood regions in Iowa.....  | 28 |
| 11. Graph showing relative percentage change, by drainage area and type of stream-gage, between expected moments algorithm (EMA/MGB) and standard Bulletin 17B/GB analyses computed using the updated regional-skew-coefficient constant for one-percent annual exceedance-probability discharges for 283 streamgages in Iowa .....  | 37 |
| 12. Graph showing relative percentage change, by drainage area and type of stream-gage, between expected moments algorithm (EMA/MGB) analyses computed using the updated regional-skew-coefficient constant and Bulletin 17B/GB analyses computed using superseded regional skew coefficients for one-percent annual exceedance-probability discharges for 283 streamgages in Iowa ..... | 38 |

13. Graph showing relative percentage change by drainage area between one-percent annual exceedance-probability discharges computed using regional regression equations developed in this study and those developed in the previous study for 185 streamgages in Iowa.....39
14. Graphs showing relation between maximum flood discharge and drainage area for streams in *A*, flood region 1, *B*, flood region 2, and *C*, flood region 3 .....41

## Tables

1. Description of streamgages located in Iowa and in neighboring States within a 50-mile buffer of Iowa that were evaluated for use in the regional skew analysis and annual exceedance-probability regressions for Iowa .....6
2. Basin characteristics tested for significance in developing regression equations.....7
3. Selected basin-characteristic measurements for streamgages evaluated in study.....8
4. Annual exceedance-probability discharges for streamgages evaluated in study based on data through the 2010 water year.....10
5. Annual exceedance probability and equivalent flood recurrence interval for selected probabilities.....10
6. Annual exceedance-probability-discharge comparison data for 283 streamgages in Iowa based on data through the 2010 water year .....12
7. Streamgages removed from regional-regression analyses.....16
8. Significant explanatory variables and predictive accuracies of preliminary statewide regression equations.....17
9. Regression equations for estimating annual exceedance-probability discharges for unregulated streams in flood region 1 of Iowa.....24
10. Regression equations for estimating annual exceedance-probability discharges for unregulated streams in flood region 2 of Iowa.....24
11. Regression equations for estimating annual exceedance-probability discharges for unregulated streams in flood region 3 of Iowa.....24
12. Range of basin-characteristic values used to develop annual exceedance-probability regression equations for unregulated streams in Iowa.....25
13. Values needed to determine the 90-percent prediction intervals for estimates obtained from regional regression equations using covariance matrices in Iowa.....29
14. Variance of prediction values for 394 streamgages included in this study that were weighted using expected moments algorithm (EMA/MGB) and regional-regression-equation estimates of annual exceedance-probability discharges.....30
15. Regional exponents determined from regional regression of log-10 drainage area for area-weighting method to estimate annual exceedance-probability discharges for ungaged sites on gaged streams.....33
16. Significant explanatory variables and predictive accuracies of preliminary region-of-influence equations in Iowa.....34
17. Relative percentage change between annual exceedance-probability discharge estimates based on data through the 2010 water year for 283 streamgages in Iowa using expected moments algorithm (EMA/MGB) and Bulletin 17B/GB analyses and using updated and superseded regional skew coefficient values.....36
18. Relative percentage change from regional-regression-equation estimates computed in the previous study to those computed in this study for 185 streamgages in Iowa .....39

## Conversion Factors

Inch/Pound to SI

| Multiply                                   | By                     | To obtain  |
|--|------------------------|--|
|  | Length                 |  |
| inch (in.)                                 | 2.54                   | centimeter (cm)                                      |
| foot (ft)                                  | 0.3048                 | meter (m)  |
| mile (mi)                                  | 1.609                  | kilometer (km)                                       |
|  | Area                   |  |
| square mile (mi <sup>2</sup> )             | 2.590                  | square kilometer (km <sup>2</sup> )                  |
| square mile per mile (mi <sup>2</sup> /mi) | 1.609                  | square kilometer per kilometer (km <sup>2</sup> /km) |
|  | Flow rate              |  |
| cubic foot per second (ft <sup>3</sup> /s) | 0.02832                | cubic meter per second (m <sup>3</sup> /s)           |
|  | Hydraulic conductivity |  |
| inch per second (in/s)                     | 25,400                 | micrometers per second (μm/s)                        |

Water year is the 12-month period from October 1 through September 30. The water year is designated by the calendar year in which the water year ends and that includes 9 of the 12 months. Thus, the water year ending September 30, 2010, is called the “2010 water year.”

## Abbreviations and Acronyms

|                    |  |
|--------------------|--|
| $\Delta A$         | Absolute value of the difference between the drainage areas of a streamgage and an ungaged site              |
| Adj-R <sup>2</sup> | Adjusted coefficient of determination  |
| AEPD               | Annual exceedance-probability discharge  |
| $A_{(g)}$          | Drainage area for a streamgage   |
| $A_{(u)}$          | Drainage area for an ungaged site  |
| AVP                | Average variance of prediction   |
| $b$                | Exponent of drainage area from table 15  |
| B17B/GB            | Bulletin 17B annual exceedance-probability analysis with standard Grubbs-Beck low outlier test               |
| BSHAPE             | Shape factor measure of basin shape  |
| Bulletin 17B/GB    | Bulletin 17B annual exceedance-probability analysis with standard Grubbs-Beck low outlier test               |
| B-GLS              | Bayesian generalized-least-squares regression  |
| B-WLS              | Bayesian weighted-least-squares regression   |
| CCM                | Constant of channel maintenance  |
| CSG                | Crest-stage gage   |
| DEM                | Digital elevation model  |
| DESMOIN            | Percent of basin within Des Moines Lobe landform region  |
| DRNAREA            | GIS-determined drainage area   |
| EMA/MGB            | Expected moments algorithm annual exceedance-probability analysis with multiple Grubbs-Beck low-outlier test |

|               |   |
|---------------|---|
| FEMA          | Federal Emergency Management Agency   |
| GB            | Grubbs-Beck low-outlier test  |
| GIS           | Geographic Information System   |
| GLS           | Generalized-least-squares regression  |
| HUC           | Hydrologic unit code  |
| I24H10Y       | Maximum 24-hour precipitation that happens on average once in 10 years                                |
| IDOT          | Iowa Department of Transportation   |
| KSATSSUR      | Average saturated hydraulic conductivity of soil  |
| LP3           | log-Pearson Type III  |
| MEV           | Model error variance  |
| MGB           | Multiple Grubbs-Beck low-outlier test   |
| MSE           | Mean-square error   |
| $MSE_G$       | EMA/MGB analysis estimates of the MSE of skew   |
| NHD           | National hydrography dataset  |
| NRCS          | Natural Resources Conservation Service  |
| NWIS          | National Water Information System   |
| OLS           | Ordinary-least-squares regression   |
| PILF          | Potentially influential low flow  |
| PRESS         | Predicted residual sums of squares  |
| PRISM         | Parameter-elevation Regressions on Independent Slopes Model   |
| Pseudo- $R^2$ | Pseudo coefficient of determination   |
| $Q_{50\%}$    | Annual exceedance-probability discharge of 50 percent (2-year recurrence-interval flood discharge)    |
| $Q_{20\%}$    | Annual exceedance-probability discharge of 20 percent (5-year recurrence-interval flood discharge)    |
| $Q_{10\%}$    | Annual exceedance-probability discharge of 10 percent (10-year recurrence-interval flood discharge)   |
| $Q_{4\%}$     | Annual exceedance-probability discharge of 4 percent (25-year recurrence-interval flood discharge)    |
| $Q_{2\%}$     | Annual exceedance-probability discharge of 2 percent (50-year recurrence-interval flood discharge)    |
| $Q_{1\%}$     | Annual exceedance-probability discharge of 1 percent (100-year recurrence-interval flood discharge)   |
| $Q_{0.5\%}$   | Annual exceedance-probability discharge of 0.5 percent (200-year recurrence-interval flood discharge) |
| $Q_{0.2\%}$   | Annual exceedance-probability discharge of 0.2 percent (500-year recurrence-interval flood discharge) |
| $Q_{hist}$    | Historical flood discharge  |
| $Q_{P(g)r}$   | RRE estimate of flood discharge for AEPD for a streamgage   |
| $Q_{P(g)w}$   | WIE estimate of flood discharge for AEPD for a streamgage   |
| $Q_{P(u)aw}$  | Area-weighted estimate of flood discharge for AEPD for an ungaged site                                |
| $Q_{P(u)r}$   | RRE estimate of flood discharge for AEPD for an ungaged site  |
| $Q_{P(u)rw}$  | Regression-weighted estimate of flood discharge for AEPD for an ungaged site                          |

|             |   |
|-------------|---|
| RMSE        | Root mean square error, also referred to as SEE   |
| RoI         | Region of influence   |
| RRE         | Regional regression equation  |
| SEE         | Average standard error of estimate, also referred to as RMSE  |
| SEM         | Standard error of model   |
| SEP         | Average standard error of prediction  |
| SSURGO      | NRCS Soil Survey Geographic database  |
| StreamStats | USGS Web-based GIS tool ( <a href="http://water.usgs.gov/osw/streamstats/index.html">http://water.usgs.gov/osw/streamstats/index.html</a> ) |
| U           | Covariance matrix   |
| USACE       | U.S. Army Corps of Engineers  |
| USDA        | U.S. Department of Agriculture  |
| USGS        | U.S. Geological Survey  |
| VIF         | Variance inflation factor   |
| WBD         | Watershed boundary dataset  |
| WIE         | Weighted independent estimates  |
| WREG        | Weighted-multiple-linear regression program   |

# Methods for Estimating Annual Exceedance-Probability Discharges for Streams in Iowa, Based on Data through Water Year 2010

By David A. Eash, Kimberlee K. Barnes, and Andrea G. Veilleux

## Abstract

A statewide study was performed to develop regional regression equations for estimating selected annual exceedance-probability statistics for ungaged stream sites in Iowa. The study area comprises streamgages located within Iowa and 50 miles beyond the State's borders. Annual exceedance-probability estimates were computed for 518 streamgages by using the expected moments algorithm to fit a Pearson Type III distribution to the logarithms of annual peak discharges for each streamgage using annual peak-discharge data through 2010. The estimation of the selected statistics included a Bayesian weighted least-squares/generalized least-squares regression analysis to update regional skew coefficients for the 518 streamgages. Low-outlier and historic information were incorporated into the annual exceedance-probability analyses, and a generalized Grubbs-Beck test was used to detect multiple potentially influential low flows. Also, geographic information system software was used to measure 59 selected basin characteristics for each streamgage.

Regional regression analysis, using generalized least-squares regression, was used to develop a set of equations for each flood region in Iowa for estimating discharges for ungaged stream sites with 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent annual exceedance probabilities, which are equivalent to annual flood-frequency recurrence intervals of 2, 5, 10, 25, 50, 100, 200, and 500 years, respectively. A total of 394 streamgages were included in the development of regional regression equations for three flood regions (regions 1, 2, and 3) that were defined for Iowa based on landform regions and soil regions.

Average standard errors of prediction range from 31.8 to 45.2 percent for flood region 1, 19.4 to 46.8 percent for flood region 2, and 26.5 to 43.1 percent for flood region 3. The pseudo coefficients of determination for the generalized least-squares equations range from 90.8 to 96.2 percent for flood region 1, 91.5 to 97.9 percent for flood region 2, and 92.4 to 96.0 percent for flood region 3. The regression equations are applicable only to stream sites in Iowa with flows not significantly affected by regulation, diversion, channelization, backwater, or urbanization and with basin characteristics within the range of those used to develop the equations.

These regression equations will be implemented within the U.S. Geological Survey StreamStats Web-based geographic information system tool. StreamStats allows users to click on any ungaged site on a river and compute estimates of the eight selected statistics; in addition, 90-percent prediction intervals and the measured basin characteristics for the ungaged sites also are provided by the Web-based tool. StreamStats also allows users to click on any streamgage in Iowa and estimates computed for these eight selected statistics are provided for the streamgage.

## Introduction

Reliable estimates of annual exceedance-probability discharges (AEPDs) are essential for the economic planning and safe design of bridges, dams, levees, and other structures located along rivers and streams, and for the effective management of flood plains. Knowledge of AEPDs allows engineers and planners to standardize risk factors. For example, 1- and 0.2-percent AEPDs are used in the design for, and estimate of, scour at bridges (Arneson and others, 2012; Fischer, 1995) and to manage development on flood plains through the National Flood Insurance Program, administered by the Federal Emergency Management Agency (FEMA; Federal Emergency Management Agency, 2002). Methods that are as accurate as possible, yet easy to apply, are needed to estimate AEPDs at ungaged stream sites in Iowa because long-term annual peak-discharge data are available at few gaged sites.

Streamgages operated by the U.S. Geological Survey (USGS) are the primary source of long-term annual peak-discharge data in Iowa. Regression analyses performed on AEPDs computed from annual peak-discharge data collected at streamgages are used to develop equations to estimate AEPDs at ungaged sites. The equations are developed by statistically relating AEPDs to significant basin characteristics for selected streamgages. AEPDs computed for streamgages are statistics that can change as more annual peak-discharge data become available. Statistics become more reliable as longer-term data are collected and used in the computations.

In response to the need to update and improve the predictive accuracy of estimates of AEPDs for ungaged stream sites

## 2 Methods for Estimating Annual Exceedance-Probability Discharges for Streams in Iowa

in Iowa, the USGS, in cooperation with the Iowa Department of Transportation (IDOT) and the Iowa Highway Research Board, initiated a statewide study in 2006. This study updates AEPD estimation equations for ungaged stream sites in Iowa, and AEPDs for streamgages in Iowa, with data collected through September 30, 2010. Major accomplishments of the study included (1) performing a Bayesian weighted least-squares/generalized least-squares regression (B-WLS/GLS) analysis to update regional skew coefficients for Iowa; (2) computing eight selected AEPDs using a new annual exceedance-probability analysis method, named expected moments algorithm (EMA), at 518 streamgages within Iowa and adjacent States with at least 10 years of annual peak-discharge record, based on data through September 30, 2010; (3) measuring 59 basin characteristics for each streamgage; (4) defining three flood regions for the State and developing 24 regional regression equations based on basin characteristics to estimate the eight selected AEPDs at ungaged stream sites; (5) calculating weighted AEPDs at 394 streamgages using the weighted independent estimates (WIE) method; and (6) calculating AEPD relative percentage change for streamgages in Iowa between estimates from different annual exceedance-probability analyses based on data through the 2010 water year (October 1, 2009 through September 30, 2010) and between regional regression equations developed in this study and a previous study (Eash, 2001). A water year is the period October 1 through September 30 and is designated by the year in which it ends.

### Purpose and Scope

Regression equations for estimating AEPDs were developed for use in Iowa and are described in this report. The regression equations relate AEPDs to physical and climatic characteristics of drainage basins. In addition, the regression equations developed from this study also will be included in the USGS StreamStats Web-based geographic information system (GIS) tool (<http://water.usgs.gov/osw/streamstats/index.html>). StreamStats allows users to obtain selected streamflow-statistic estimates, upstream drainage-basin characteristics, and other information for user-selected stream sites. Using a GIS-based interactive map of Iowa, the user can “point and click” on a stream site and StreamStats will delineate the basin boundary upstream from the selected site. The user also can “point and click” on USGS streamgages and receive selected streamflow statistics and other streamgage information.

This report presents 24 regional regression equations (RREs) that can be used to estimate eight selected annual exceedance-probability statistics for ungaged sites on unregulated streams in Iowa. The eight selected annual exceedance-probability statistics are flood discharges that have probabilities of 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent, which are equivalent to annual flood-frequency recurrence intervals of 2, 5, 10, 25, 50, 100, 200, and 500 years, respectively; hereafter, these statistics are referred to as  $Q_{50\text{-percent}(\%)}$ ,  $Q_{20\%}$ ,

$Q_{10\%}$ ,  $Q_{4\%}$ ,  $Q_{2\%}$ ,  $Q_{1\%}$ ,  $Q_{0.5\%}$ , and  $Q_{0.2\%}$ , respectively. This report also presents the results of a regional skew analysis performed for Iowa to develop updated regional skew coefficients for all streamgages in the State.

The regional regression equations were developed using AEPDs computed for streamgages unaffected by regulation, diversion, channelization, backwater, or urbanization that are located in Iowa and in adjacent States within a 50-mile (mi) buffer of Iowa (all gaged drainage basins are within the buffer). AEPDs computed for 518 streamgages using the new EMA annual exceedance-probability analysis are presented in this report. AEPDs for these 518 streamgages were computed using annual peak-discharge data collected through September 30, 2010, and were computed using 10 or more years of record. The accuracy and limitations of the regression equations and the methodology used to develop the equations are described in the report.

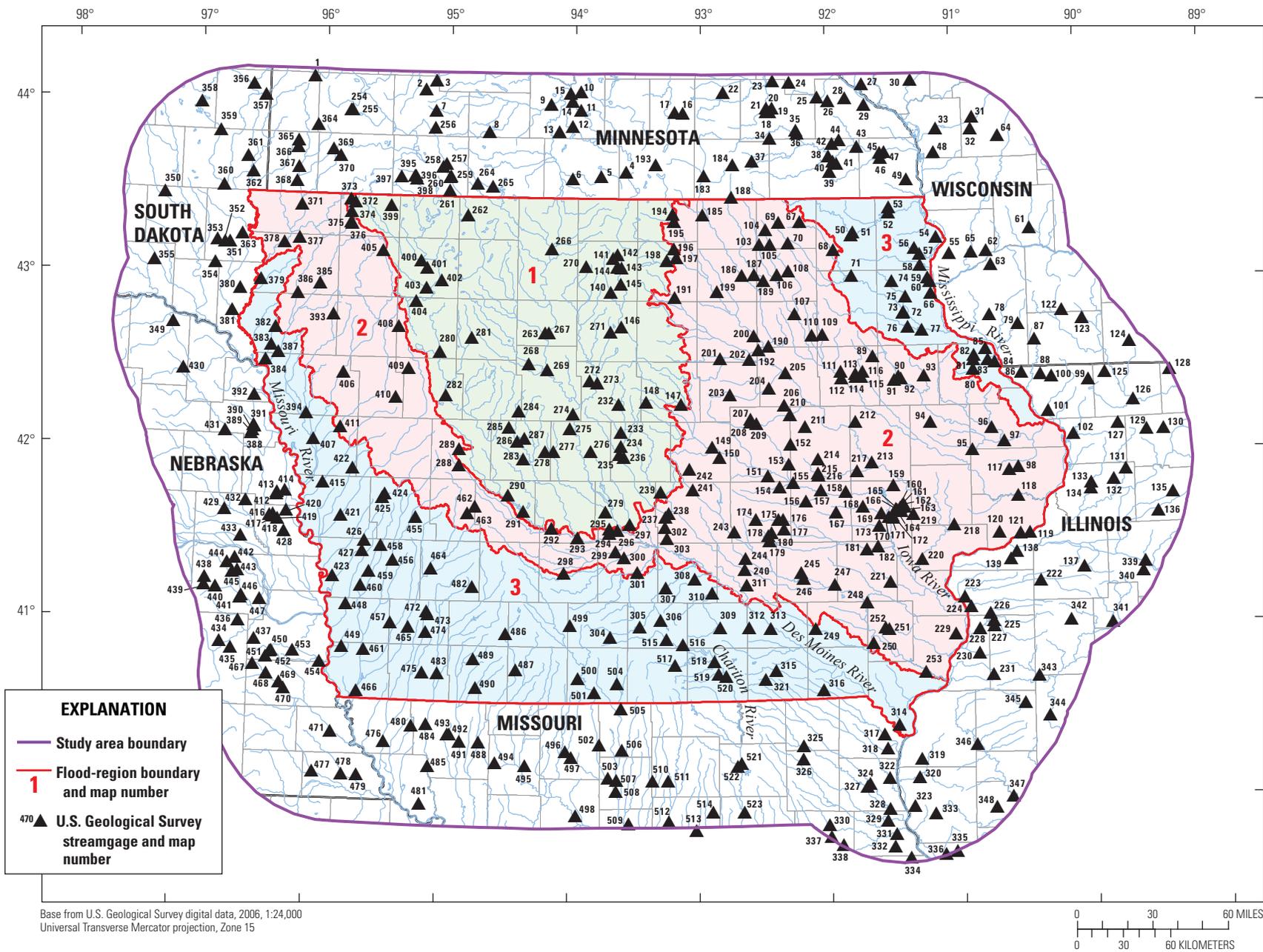
### Description of Study Area

The study area (fig. 1) includes the entire State of Iowa and adjacent areas within a 50-mi buffer of Iowa in the neighboring states of Illinois, Minnesota, Missouri, Nebraska, South Dakota, and Wisconsin. A map of Iowa soil regions created by the National Cooperative Soil Survey and the Natural Resources Conservation Service (NRCS) is shown in figure 2 (<ftp://ftp-fc.sc.egov.usda.gov/IA/technical/IowaSoilRegionsMap.html>). Oschwald and others (1965) present a detailed description of soils in Iowa. There are 10 landform regions in Iowa, each having distinctive topography and geology (fig. 3). A brief description of the landform regions in Iowa is presented in Eash and Barnes (2012) and a detailed description is presented by Prior (1991). Prior and others (2009) describe updates to landform regions in Iowa.

Most precipitation in the study area results from storms moving inland primarily from the Gulf of Mexico and secondarily from the Pacific Ocean (Soenksen and Eash, 1991). Annual precipitation, which is mostly rain, ranges from 26 inches (in.) in the extreme northwest to as much as 38 in. in the southeast; the statewide average is around 34 in. (National Climatic Data Center, 2012). About 75 percent of the annual precipitation is received during April through September. Typically, during August through February, streamflow in most unregulated streams in the study area is base flow; during March through July, streamflow is significantly greater, primarily as a result of snowmelt during late February through early April and rainfall during May through July. Annual maximum streamflows are typically during April through July.

### Previous Studies

This is the seventh in a series of reports that describe flood characteristics for Iowa streams. The first report (Schwob, 1953) contained information on AEPDs for 55 continuous-record streamgages in Iowa using annual



**Figure 1.** Location of flood regions and streamgages evaluated for use in the regional skew analysis and annual exceedance-probability regressions for Iowa.

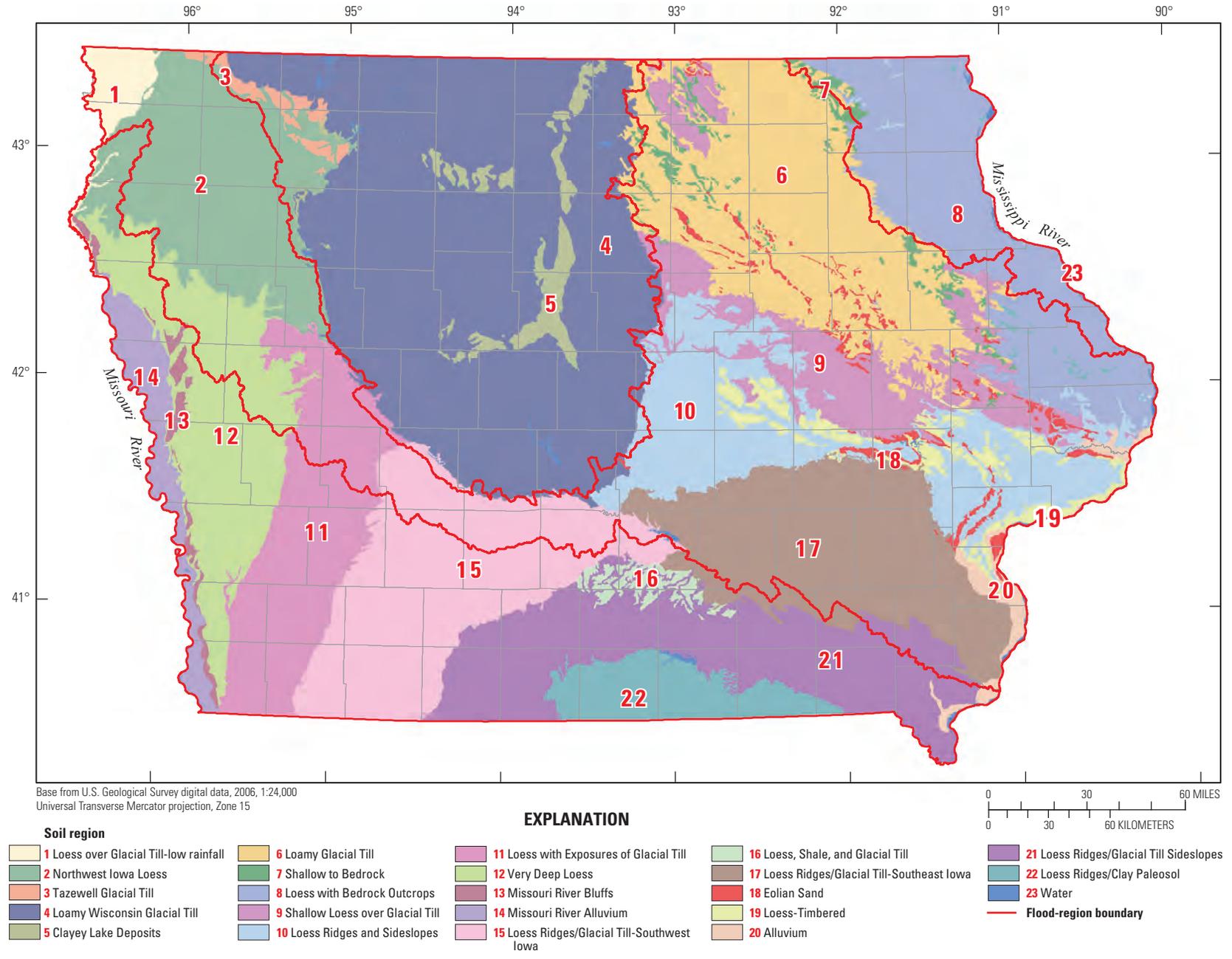
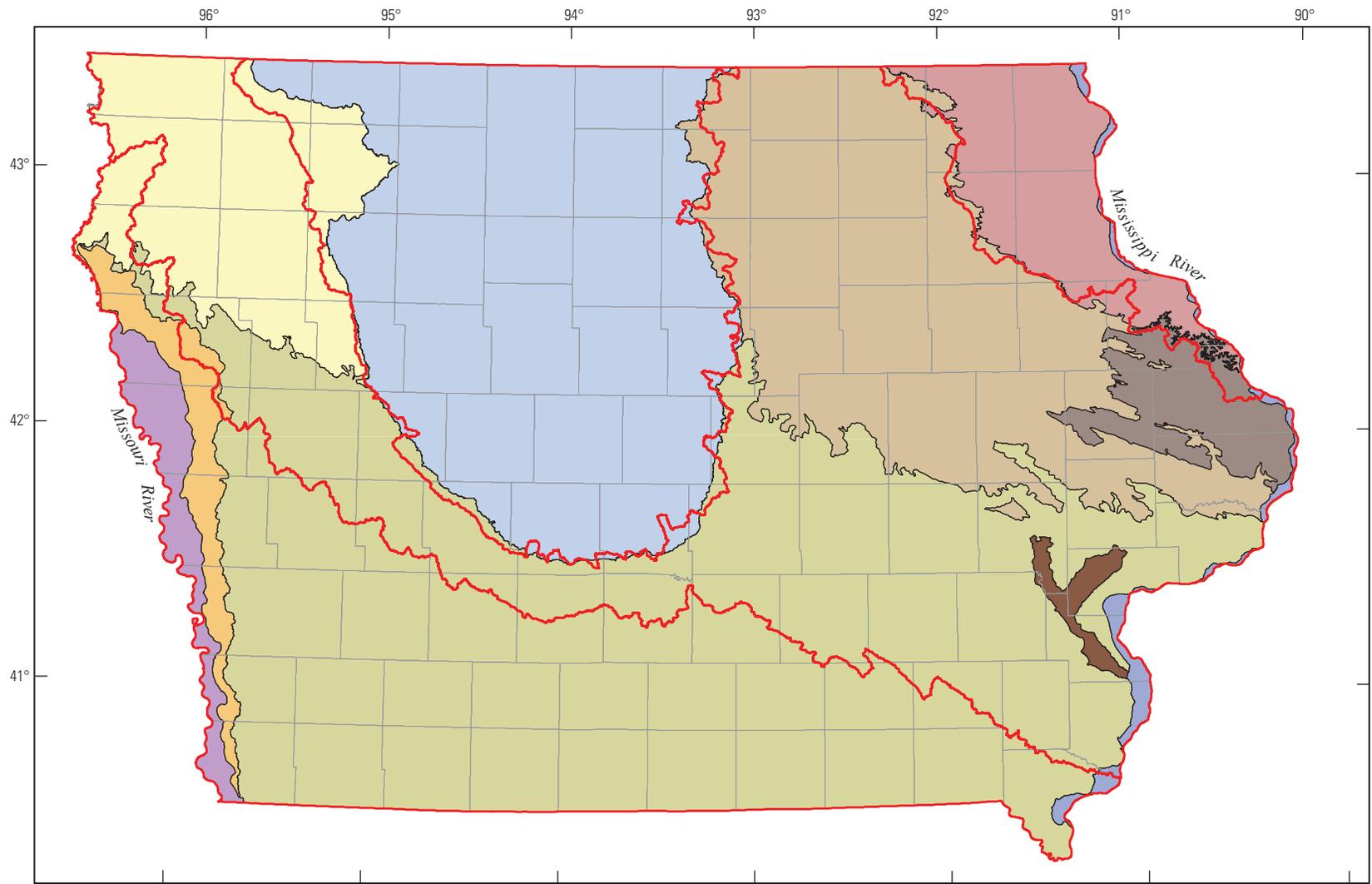
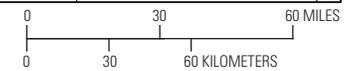


Figure 2. Soil regions in Iowa.



Base from U.S. Geological Survey digital data, 2006, 1:24,000  
 Universal Transverse Mercator projection, Zone 15



**EXPLANATION**

|                               |                                  |                               |                              |
|-------------------------------|----------------------------------|-------------------------------|------------------------------|
| <b>Landform regions</b>       |                                  |                               |                              |
| Des Moines Lobe               | Iowan Surface                    | Missouri River Alluvial Plain | Southern Iowa Drift Plain    |
| East-Central Iowa Drift Plain | Loess Hills                      | Northwest Iowa Plains         | <b>Flood-region Boundary</b> |
| Iowa-Cedar Lowland            | Mississippi River Alluvial Plain | Paleozoic Plateau             |                              |

**Figure 3.** Landform regions in Iowa.

peak-discharge data collected through the 1950 water year. Schwob (1953) defined eight flood regions for Iowa using 34 of the 55 streamgages with AEPD information and presented a method for estimating AEPDs for ungaged sites with drainage areas greater than about 100 square miles (mi<sup>2</sup>). The method used graphs developed for each region to relate drainage area to mean annual floods to estimate a mean annual flood-discharge statistic for an ungaged site, and then used a graph relating recurrence interval to the ratio of the mean annual flood to determine a ratio value for a recurrence interval. AEPDs for as large as the  $Q_{2\%}$  flood discharge were then estimated by multiplying the mean annual flood discharge for an ungaged site by the ratio for a recurrence interval. Predictive accuracies were not determined for the AEPD estimates.

The second report (Schwob, 1966) contained information on AEPDs for 147 continuous-record and crest-stage (partial-record) streamgages using annual peak-discharge data collected through the 1965 water year. Schwob (1966) defined two flood regions for Iowa using 147 streamgages and presented a two- or three-variable regression equation for each region for estimating a mean annual flood-discharge statistic for ungaged sites with drainage areas greater than 1 mi<sup>2</sup>. A graph relating recurrence interval to the ratio of the mean annual flood was used to determine a ratio value for a recurrence interval. AEPDs for as large as the  $Q_{2\%}$  flood discharge were then estimated by multiplying the mean annual flood value for an ungaged site by the ratio for a recurrence interval. Average standard errors of estimate for the two regression equations ranged from 30.4 to 37.9 percent. Three basin characteristics were measured manually for each streamgage.

The third report (Lara, 1973) contained information on AEPDs for 136 continuous-record and crest-stage streamgages using annual peak-discharge data collected through the 1972 water year. Lara (1973) defined two flood regions for Iowa using the 136 streamgages and presented one- or two-variable regression equations for each region for estimating six annual exceedance-probability statistics for as large as the  $Q_{1\%}$  flood discharge for ungaged sites with drainage areas greater than 2 mi<sup>2</sup>. Average standard errors of estimate for the regression equations ranged from 26 to 44 percent. Thirteen basin characteristics were manually measured for each streamgage.

The fourth report (Lara, 1987) contained information on AEPDs for 263 continuous-record and crest-stage streamgages using annual peak-discharge data collected through the 1984 water year. Lara (1987) defined five flood regions for Iowa using 251 of the 263 streamgages and presented one-variable regression equations for each region for estimating six annual exceedance-probability statistics for as large as the  $Q_{1\%}$  flood discharge for ungaged sites with drainage areas greater than 0.04 mi<sup>2</sup>. Average standard errors of estimate for the 30 regression equations ranged from 20 to 61 percent. Two basin characteristics were measured manually for each streamgage.

The fifth report (Eash, 1993) contained information on AEPDs for 188 continuous-record and crest-stage streamgages using annual peak-discharge data collected through the 1990

water year. Eash (1993) defined one flood region for Iowa using 164 streamgages and presented three-variable, statewide, drainage-basin-characteristic regression equations for estimating six annual exceedance-probability statistics for as large as the  $Q_{1\%}$  flood discharge for ungaged sites with drainage areas greater than 0.34 mi<sup>2</sup>. Average standard errors of prediction for the drainage-basin regression equations ranged from 38.6 to 50.2 percent. A GIS program developed by the USGS, named Basinsoft (Harvey and Eash, 1996), was used to automate measurements of 26 basin characteristics for each streamgage. Eash (1993) also defined two flood regions for Iowa using 157 streamgages and presented two sets of one- or two-variable channel-geometry-characteristic regression equations for either region for estimating six annual exceedance-probability statistics for as large as the  $Q_{1\%}$  flood discharge for ungaged sites with bankfull widths greater than 9.6 feet (ft). Average standard errors of prediction for the channel-geometry regression equations ranged from 30.3 to 70.0 percent. The channel-geometry-characteristic regression equations required the collection of channel-geometry measurements at ungaged sites.

The sixth report (Eash, 2001) contained information on AEPDs for 291 continuous-record and crest-stage streamgages using annual peak-discharge data collected through the 1997 water year. Eash (2001) defined three flood regions for Iowa using 241 of the 291 streamgages and presented one-, two-, or three-variable regression equations for each region for estimating eight annual exceedance-probability statistics as large as the  $Q_{0.2\%}$  flood discharge for ungaged sites with drainage areas greater than 1.30 mi<sup>2</sup>. Average standard errors of prediction for the regression equations ranged from 30.8 to 42.7 percent. The Basinsoft program was used to measure 38 basin characteristics for each streamgage. A regional skew analysis was performed as part of this study using 239 of the 291 streamgages and an updated regional-skew-coefficient constant was used for computing annual exceedance-probability analyses for streamgages in Iowa.

## Methods for Dataset Development for Streamgages

Data used in this report were collected for 523 active and inactive continuous-record and crest-stage streamgages located in Iowa and within a 50-mi buffer of Iowa in the neighboring States of Illinois, Minnesota, Missouri, Nebraska, South Dakota, and Wisconsin (fig. 1; [table 1](#), [link to Excel file](#)). Streamgages with at least 10 years of annual peak discharges and unaffected by regulation or diversion initially were selected for evaluation in the study, which included 284 streamgages in Iowa and 239 streamgages in neighboring States. Streamgages from neighboring States were used to improve the representativeness of annual exceedance-probability statistics and basin characteristics in Iowa border areas and to provide better estimates of the error of the regression equations for ungaged sites near the State border. Of these

original 523 streamgages, five of them included for evaluation in the regional skew analysis for Iowa subsequently were removed resulting in a total of 518 streamgages evaluated for use in annual exceedance-probability regressions for Iowa. See the sections Regional Regression Analyses to Estimate Annual Exceedance-Probability Discharges for Ungaged Stream Sites and Definition of Flood Regions for more information on the removal of the five streamgages from the original dataset of 523 ([table 1](#), link to Excel file).

## Basin Characteristics

Physical processes controlling floods vary from one region to another region and from one stream site to another stream site, but they generally are related to storm events (precipitation intensity) and drainage area. Flood peaks generated by snowmelt have a different impetus than those generated by rainfall. Peak discharges are a function of many interrelated factors that include runoff response to geology, soils, slope, and land cover; surface storage such as wetlands, lakes, and flood plains; and routing related to drainage density, basin shape, channel length, and slope (C.P. Konrad, U.S. Geological Survey, written commun., 2009). Basin characteristics investigated in this study as potential explanatory variables in the regression analysis were selected on the basis of their theoretical relation to peak discharges, results of previous studies in similar hydrologic areas, and the ability to quantify the basin characteristics using GIS technology and digital datasets. The use of GIS enables the automation of the basin-characteristic measurements and solution of the RREs using StreamStats.

Using GIS technology, 59 basin characteristics were measured for each of the 518 streamgages evaluated for use in the development of the regression equations. [Table 2](#) (link to Excel file) provides a brief description of each basin characteristic and the data source used to measure the characteristic. Basin-characteristic names used in this study were selected to maintain consistency with the names of explanatory variables in the USGS StreamStats Web-based GIS tool (<http://water.usgs.gov/osw/streamstats/bcdefinitions1.html>).

The basin characteristics can be separated into four categories: morphometric (physical or shape) characteristics, hydrologic characteristics, pedologic (soils)/geologic/land-use characteristics, or climatic characteristics. Morphometric characteristics were measured from one to three data sources, which are described in the following section Geographic Information System Measurements. Hydrologic characteristics initially were computed as observed values for 208 continuous-record streamgages using daily mean discharge data and subsequently were mapped using a kriging procedure to compute interpolated values for a low-flow estimation study performed for Iowa (Eash and Barnes, 2012). A list of the 208 streamgages included in the low-flow study, descriptions of the hydrologic-characteristic computations and kriging procedure, and isoline maps created from kriged grids for three of the five hydrologic characteristics are presented

in Eash and Barnes (2012). The pedologic/geologic/land-use characteristics were computed from the NRCS Soil Survey Geographic (SSURGO) Database (Soil Survey Staff, 2012) for seven soil characteristics, from the Iowa Geological and Water Survey Des Moines Lobe landform region boundary for the Des Moines Lobe geologic characteristic (Prior and others, 2009), and from the Multi-Resolution Land Characteristics Consortium 2001 National Land Cover Database (Multi-Resolution Land Characteristics Consortium, 2012) for the land-use characteristic that measured percent area of row crops (<http://www.mrlc.gov/index.php>; Homer and others, 2004). The climatic characteristics were computed from Oregon State University Parameter-elevation Regressions on Independent Slopes Model (PRISM) datasets (Parameter-Elevation Regressions on Independent Slopes Model Climate Group, 2008) and from Midwest Climate Center Bulletin 71 (Huff and Angel, 1992).

## Geographic Information System Measurements

Three primary GIS-data layers were processed to produce the Iowa StreamStats data layers. These data layers were needed to delineate accurate stream networks and basin boundaries, and the layers were used to measure 26 morphometric basin characteristics ([table 2](#), link to Excel file). The three primary GIS-data layers include the 1:24,000-scale USGS National Hydrography Dataset (NHD) (<http://nhd.usgs.gov/>; Simley and Carswell, 2009), the 1:24,000-scale U.S. Department of Agriculture (USDA)/NRCS Watershed Boundary Dataset (WBD) (<http://datagateway.nrcs.usda.gov/>; USGS and NRCS, 2009) using 12-digit hydrologic unit codes (HUCs), and the 10-meter (m) (32.81 ft) USGS National Elevation Dataset (<http://ned.usgs.gov/>; Gesch, 2007).

Several preprocessing steps were needed for each of the three data layers to facilitate rapid determination of basin characteristics. Preprocessing of the NHD included removing flow-line paths that represent man-made features (a stream network that only represents natural streams is needed) and selection of the primary flow path in those areas where the NHD indicated split flow (such as might happen when flow splits around an island in a river or with a braided channel). The NHD and WBD had to be verified that the stream from the NHD only crossed the watershed boundary (from the WBD) at the outlet (unless the watershed is downstream from another watershed, in which case the main stem stream will enter the watershed at one place); and watershed outlets should align exactly to the confluences of the streams. For the national elevation dataset, downloaded blocks were mosaicked into one tile, data were extracted for a 4-kilometer (km) (2.5 mi) buffer area around each 8-digit HUC, and projected from decimal degrees to Universal Transverse Mercator Zone 15. A hydro-corrected digital elevation model (DEM) was then developed by filling depressions or sinks, using the basin boundaries from the WBD to conserve known drainage divides, and using the streams from the NHD to create well-defined flow paths through the elevation data.

## 8 Methods for Estimating Annual Exceedance-Probability Discharges for Streams in Iowa

ArcHydro Tools, version 1.3, a set of utilities developed to operate in the ArcGIS, version 9.3, environment (Environmental Systems Research Institute, 2009) was used to process 58 HUCs consisting of eight digits to create StreamStats data layers for the entire State. To calculate basin characteristics to develop the RREs for estimating AEPDs for Iowa, additional data layers were generated. These primary base-grid data layers include catchments, flow accumulation, flow direction, and an artificial flow-path grid used to delineate drainage basins. These additional layers then were used to create layers that control the StreamStats delineation of a watershed, subwatersheds, and stream networks within these watersheds, including the created layers named AdjointCatchment, Catchment, DrainageLine, DrainagePoint, LongestFlowPathCat, and LongestFlowPathAdjCat. Once processing was complete for all 58 processing units, a global geodatabase was created to direct StreamStats as to how all units relate to each other. In addition, the DEM was resampled to 150 m for use in the basin-length calculations. All 59 basin characteristics listed in [table 2](#) (link to Excel file) were measured using ArcHydro Tools or Spatial Analyst tools in ArcGIS, version 9.3 (Environmental Systems Research Institute, 2009).

To measure basin characteristics for streamgages located outside of Iowa, similar preprocessing steps were performed on GIS data layers for 27 additional 8-digit HUCs located in neighboring states. These 27 HUCs are not part of the GIS data layers used by StreamStats for Iowa. Because certified WBD data were not available at the time in adjacent states, the preprocessing of these 27 HUCs did not include the “walling” of basin boundaries using WBD, but did include the “burning” of streams from the NHD into the national elevation dataset; however, a global geodatabase was not created for these 27 HUCs because none of the streamgages within these HUCs accumulated flow from more than one HUC.

GIS measurements of four hydrologic basin characteristics ([table 2](#), link to Excel file) were interpolated by area-weighting values for streamgage watershed boundaries from grids that were created using a kriging procedure (Eash and Barnes, 2012). GIS measurements of seven soil characteristics ([table 2](#), link to Excel file) were made using a three-step process. First, the NRCS Soil Data Viewer tool, built as an extension of ArcMap (Environmental Systems Research Institute, 2009), was used to create four 8-digit HUC data layers for the soil characteristics. Second, a shapefile was created for the hydrosols data layer (includes the four hydrologic soil types A, B, C, and D; ([table 2](#), link to Excel file), and a grid was created for each of the SAND, CLAY, and KSATSSUR data layers. Third, the ArcMap attribute selection tool was used to calculate a percent-area value for each hydrologic soil type, and the Spatial Analyst tool was used to calculate area-weighted values for SAND, CLAY, and KSATSSUR for each streamgage watershed boundary. The geologic characteristic DESMOIN, the land-use characteristic ROWCROP, and 19 of the 20 climatic characteristics ([table 2](#), link to Excel file) were all measured from grids as area-weighted values for each streamgage watershed (PRC5\_7 was calculated for

each watershed as the mean of May, June and July mean precipitation).

[Table 3](#) (link to Excel file) lists two drainage area values for each streamgage included in the study. Each streamgage has a drainage area that is listed in the USGS National Water Information System (NWIS) data base (U.S. Geological Survey, 2012), which is referred to as the “published” drainage area. Published drainage areas were determined primarily from 1:24,000-scale topographic maps by manual planimetry or GIS digitizing methods at the time streamgage operation began. Drainage area values listed in [table 3](#) (link to Excel file) as “GIS” drainage areas, for the basin characteristic DRNAREA, were measured as part of this study using a two-step process within ArcHydro Tools. First, a streamgage location was selected using the point generation tool; second, one of the watershed delineation tools (such as Batch Watershed Delineation) was used to automatically delineate the watershed boundary using hydro-corrected DEM data. The watershed delineation process in the second step delineates the basin boundary from the DEM data proceeding from the streamgage location until an existing basin boundary is reached within the WBD data, and then the delineation follows the WBD boundary for the remainder of the watershed delineation. For some streamgages with small drainage areas that are located completely within a 12-digit HUC, the entire watershed delineation was made from the DEM data.

GIS delineations of watershed boundaries were inspected for streamgages with drainage area differences greater than 5 percent from published values. Basin boundaries of several GIS-delineated watersheds were edited where the delineation did not match well with digital raster graphics elevation contours. Most edits made only a small difference in the drainage area value for the watershed. If the GIS-delineated basin boundary was accurate according to the 8-digit HUC, WBD line work, and digital raster graphics contour lines, then the GIS delineation was accepted even if it exceeded a 5-percent difference from the published drainage area. GIS delineations generally are believed to be more accurate than the published drainage areas. Most of the GIS watershed delineations are using part of the WBD boundaries, which have been certified by NRCS, and use of the WBD data accounts for some of the differences between GIS and published values of drainage areas. GIS measurements of drainage area (DRNAREA) were used to develop the regression equations because StreamStats will use the same GIS data layers and delineation methods for determining watershed boundaries and drainage areas for ungaged stream sites. Drainage areas of the 518 streamgages ranged from 0.05 to 7,783 mi<sup>2</sup> ([table 3](#), link to Excel file).

### Annual Peak-Discharge Data

A standard, continuous-record streamgage typically records stage (the gage height or water-surface elevation) every 15 minutes throughout the course of a water year. A crest-stage gage (CSG) is a partial-record streamgage that only provides information on the highest stage since the streamgage

was last visited. CSGs are the primary source of annual peak-discharge data for small drainage basins in Iowa (U.S. Geological Survey, Iowa Water Science Center, Flood Information at Selected Bridge and Culvert Sites; <http://ia.water.usgs.gov/projects/ia006.html>). Annual peak discharges are computed for continuous-record streamgages and CSGs by use of a stage-discharge relation (Rantz and others, 1982). The stage-discharge relation, or rating, is used to determine discharges for all recorded stages at streamgages. The largest discharge during a water year is the annual peak discharge, and the annual peak-discharge record is the compilation of all recorded annual peak discharges. Annual peak-discharge records collected through the 2008 water year (through September 30, 2008) initially were retrieved for 330 streamgages for use in computing station skew coefficients and the mean square error (MSE) of station skews for a regional skew analysis performed for Iowa that is described in a following section, Regional Skew Analysis. Annual peak-discharge records collected through the 2010 water year (through September 30, 2010) subsequently were retrieved for 518 streamgages for use in computing annual exceedance-probability analyses that are described in the section Expected Moments Algorithm (EMA/MGB) Analysis. All annual peak-discharge records analyzed in this study were retrieved from the USGS NWIS database (U.S. Geological Survey, 2012; <http://nwis.waterdata.usgs.gov/usa/nwis/peak>). Annual peak-discharge data were reviewed to eliminate data affected by regulations or diversions from biasing the computation of AEPDs. Annual peak-discharge records were reviewed by using the PFRports computer program described by Ryberg (2008).

## Trend Analyses

Annual peak-discharge records retrieved for the 518 streamgages were analyzed for the entire period of record for trends using the Kendall's tau hypothesis test in the PeakFQ program (Flynn and others, 2006). Trends in the annual peak-discharge data could introduce a bias into the annual exceedance-probability analyses because a major assumption of probability analyses is annual peak discharges are independent and stationary with time. The Kendall's tau test computes the monotonic relation between peaks (discharge) and time (water years) (Helsel and Hirsch, 2002). A p-value threshold of 5 percent ( $\alpha=0.05$ ) was used in this study for the Kendall's tau test, and p-values less than or equal to 5 percent were flagged as having statistically significant trends (positive or negative). Results of the Kendall's tau tests ([table 1](#), link to Excel file) indicated statistically significant trends for 58 of the 518 streamgages tested using the entire period of record, for which 22 of the trends were negative and 36 of the trends were positive.

Wahl (1998) describes how Kendall's tau test results may be sensitive to multiyear sequences of larger or smaller discharges if the sequences happen near the beginning or end of the period of record used in the test. Although trend

results are relatively insensitive to individual outliers, multi-year sequences of extremes near either end of the record can have a significant effect on the test results, but may imply no systematic change. Annual peak-discharge records for the 58 streamgages initially indicated to have significant trends were retested using the Kendall's tau test after a few annual peak discharges were removed, consecutively, from either the beginning or the end of the record, or from the beginning and the end of the record. A record-length threshold of 94 percent was used for the retesting of the trend analysis. Initially, a 95-percent threshold was tested on the basis of the assumption that if a significant trend is not identified using 95 percent of the record, then there probably is not a trend in the data; because of rounding for several streamgages, a 94-percent threshold subsequently was selected for the retesting. For example, a streamgage with a record length of 50 years could have no more than 3 years of record removed for the retest. Results of the Kendall's tau retests indicated statistically significant trends for 25 of the 58 streamgages retested using 94 percent of the entire record length ([table 1](#), link to Excel file). A review of the annual peak-discharge records for these 25 sites indicated that 22 of them have either short records of less than 15 years or broken records with sequences of missing years in their records because of (1) intermittent gage operation, or (2) annual peak discharges that were censored because they were lower than the recordable level of the CSG. Because of the short or broken records for these 22 streamgages and because the remaining 3 streamgages are isolated and not supported by trends at other nearby streamgages, there is uncertainty about whether the records of these 25 streamgages represent actual trends or are anomalies. Therefore, the 25 streamgages that recorded significant trends were included in the regression analyses and were removed only if they were indicated to be a significant outlier in the regression model.

The number of uncensored annual peak-discharge record lengths used in the study for the 518 streamgages ranged from 9 to 108 years with a mean of 35.4 years and a median of 28 years ([table 1](#), link to Excel file). Uncensored annual peak discharges include all systematic annual peak discharges collected at continuous-record streamgages and include all systematic annual peak discharges collected at CSGs that were higher than the minimum recordable level of the CSG. Annual peak-water levels that did not reach the bottom of the CSG were recorded with a "less-than discharge value," which is the minimum recordable discharge threshold of the CSG. These "less-than discharge values," or  $Q_{\text{less-than}}$ , are censored annual peak-discharge data; estimates for these censored discharges bounded between 0 and  $Q_{\text{less-than}}$  can be used in the log-Pearson Type III exceedance-probability analysis. For the CSG Drainage Ditch 97 tributary near Britt, Iowa (streamgage 0548065350, map number 270, [fig. 1](#)), the one streamgage in [table 1](#) (link to Excel file) that lists 9 years of uncensored data, the inclusion of seven censored annual peak discharges (7 years for which the discharge did not reach the minimum recordable discharge) with the nine uncensored annual peak discharges (9 years for which the discharge exceeded the

minimum recordable discharge) provides a total of 16 years of annual peak-discharge record for the annual exceedance-probability analysis. This streamgage was included in this study because its annual peak-discharge record included at least 10 years of record.

## Annual Exceedance-Probability Analyses

To estimate AEPDs at continuous-record streamgages and CSGs for this study, such as the  $Q_{1\text{-percent}(\%)}$  flood discharge, EMA exceedance-probability analyses were performed. EMA analyses provide a new alternative method to standard Bulletin 17B (Interagency Advisory Committee on Water Data, 1982) exceedance-probability analyses. To specify the different low-outlier tests used by either exceedance-probability analysis method in this study, hereafter EMA analyses will be referred to as EMA/MGB and Bulletin 17B analyses will be referred to as Bulletin17B/GB or B17B/GB. MGB refers to the multiple Grubbs-Beck test for detecting low outliers and GB refers to the standard Grubbs-Beck test for detecting low outliers. Because the Hydrologic Frequency Analysis Work Group now recommends use of the multiple Grubbs-Beck (MGB) test for detecting low outliers for exceedance-probability analyses ([http://acwi.gov/hydrology/Frequency/minutes/Minutes\\_HFAWG\\_meeting\\_mar19\\_2012\\_040212.pdf](http://acwi.gov/hydrology/Frequency/minutes/Minutes_HFAWG_meeting_mar19_2012_040212.pdf)) and because the EMA analysis method needs a consistent statistical test (MGB) to identify potentially influential low flows in an annual peak-discharge series to properly reduce the effect of low outliers (N.A. Barth, U.S. Geological Survey, written commun., 2012), it is important to specify for this study that all EMA analyses included the MGB test and all standard Bulletin 17B analyses included the standard Grubbs-Beck test. Additional information on both types of Grubbs-Beck tests are presented in the section Multiple Grubbs-Beck (MGB) Test for Detecting Low Outliers.

For this report, annual exceedance probabilities were estimated for the  $Q_{50\%}$ ,  $Q_{20\%}$ ,  $Q_{10\%}$ ,  $Q_{4\%}$ ,  $Q_{2\%}$ ,  $Q_{1\%}$ ,  $Q_{0.5\%}$ , and  $Q_{0.2\%}$  flood discharges using EMA/MGB for 518 streamgages listed in [table 4](#) (link to Excel file). The annual magnitude and probability of flood discharges or AEPDs for a streamgage are computed from an exceedance-probability analysis that relates observed annual peak discharges to annual exceedance probability. Annual exceedance probability is an estimate of the likelihood of a flood of a specific magnitude happening in any 1 year.

Annual exceedance probabilities formerly were reported as flood recurrence intervals expressed in years. For example, a flood magnitude that has a 1-percent chance (annual exceedance probability=0.01) of being exceeded during any particular year is expected to be exceeded on average once during any 100-year period (recurrence interval). Percent probability is the inverse of the recurrence interval multiplied by 100. Because of widespread confusion caused in recent years by two or more “100-year floods” happening in a period of much less than 100 years, the scientific and engineering community

**Table 5.** Annual exceedance probability and equivalent flood recurrence interval for selected probabilities.

| Annual exceedance probability (percent) | Recurrence interval (years) |
|---|-----------------------------|
| 50                                      | 2                           |
| 20                                      | 5                           |
| 10                                      | 10                          |
| 4                                       | 25                          |
| 2                                       | 50                          |
| 1                                       | 100                         |
| 0.5                                     | 200                         |
| 0.2                                     | 500                         |

has begun expressing the annual likelihood of occurrence of flood discharges as a probability. Selected annual exceedance probabilities and equivalent flood recurrence intervals are listed in [table 5](#). Although the annual probability is an estimate of the likelihood of a flood discharge of a specific magnitude happening in any 1 year, more than one flood discharge with a specific magnitude and annual probability could happen in the same year.

Annual exceedance-probability analyses were not computed for regulated annual peak-discharge records of streamgages located downstream from U.S. Army Corps of Engineers (USACE) reservoirs located on the Chariton, Des Moines, Iowa, or Missouri Rivers or downstream from locks and dams located on the Mississippi River. [Table 1](#) (link to Excel file) lists one streamgage located downstream from the USACE reservoir on the Iowa River (streamgage 05454500, map number 170, [fig. 1](#)) and four streamgages located downstream from the USACE reservoir on the Chariton River (streamgages 06903900, 06904000, 06904500, and 06905000; map numbers 518, 520, 521, and 523, [fig. 1](#), respectively) for which annual exceedance-probability analyses were computed for unregulated annual peak-discharge records. Information on regulated flow-frequency studies for the Iowa and Des Moines Rivers are available from U.S. Army Corps of Engineers (2009, 2010).

## Bulletin 17B/GB Analyses

The Interagency Advisory Committee on Water Data has established standard methods for estimating annual exceedance probabilities for streamgages by fitting a log-Pearson Type III (LP3) distribution to the logarithms (base 10) of the annual peak discharges as described in Bulletin 17B (Interagency Advisory Committee on Water Data, 1982). Before this study, exceedance-probability analyses for streamgages in Iowa were computed using standard Bulletin 17B/GB analyses (Flynn and others, 2006). Standard Bulletin 17B/GB and EMA/MGB analyses use a LP3 distribution to compute AEPDs. Fitting the LP3 distribution requires calculating the mean, standard deviation, and skew coefficient of the logarithms of the annual peak-discharge record, which describe

the midpoint, slope, and curvature of the annual exceedance-probability curve, respectively (Gotvald and others, 2012). Estimates of AEPDs are computed using these three statistics from the LP3 distribution in the following equation:

$$\log Q_p = \bar{X} + K_p S \quad (1)$$

where

- $Q_p$  is the P-percent AEPD, in cubic feet per second (ft<sup>3</sup>/s);
- $\bar{X}$  is the mean of the logarithms of the annual peak discharges;
- $K_p$  is a factor based on the skew coefficient and the given percent annual exceedance probability and is obtained from appendix 3 in Bulletin 17B (Interagency Advisory Committee on Water Data, 1982); and
- $S$  is the standard deviation of the logarithms of the annual peak discharges, which is a measure of the degree of variation of the annual values about the mean value.

The mean, standard deviation, and skew coefficient can be estimated from the available sample data (observed annual peak discharges). The station skew coefficient measures the asymmetry of the probability distribution of a set of annual peak discharges (Gotvald and others, 2009; Feaster and others, 2009; Weaver and others, 2009). Large positive station skew coefficients can result from steep basin and channel slopes, low infiltration rates, fast conveyance through systems, and (or) one or more extremely large peak discharges (high outliers). Conversely, large negative station skew coefficients can result from low mean basin slopes, high infiltration rates, high channel losses, a substantial percentage of a basin controlled by lakes or swamps, and (or) one or more extremely low peak discharges (low outliers) (Ahearn, 2003). The station skew coefficient is sensitive to extreme events; therefore, the station skew coefficient for short records may not provide an accurate estimate of the population or true skew coefficient (Gotvald and others, 2009; Feaster and others, 2009; Weaver and others, 2009). Thus, guidelines in Bulletin 17B (Interagency Advisory Committee on Water Data, 1982) recommend that the skew coefficient calculated from streamgage data (station skew) needs to be weighted with a generalized, or regional, skew determined from an analysis of selected long-term streamgages in the study area (Gotvald and others, 2012). The weighted skew is determined by weighting the station skew and the regional skew inversely proportional to their respective mean square errors, as shown in the following equation (Interagency Advisory Committee on Water Data, 1982):

$$G_w = \frac{MSE_R(G_s) + MSE_s(G_R)}{MSE_R + MSE_s}, \quad (2)$$

where

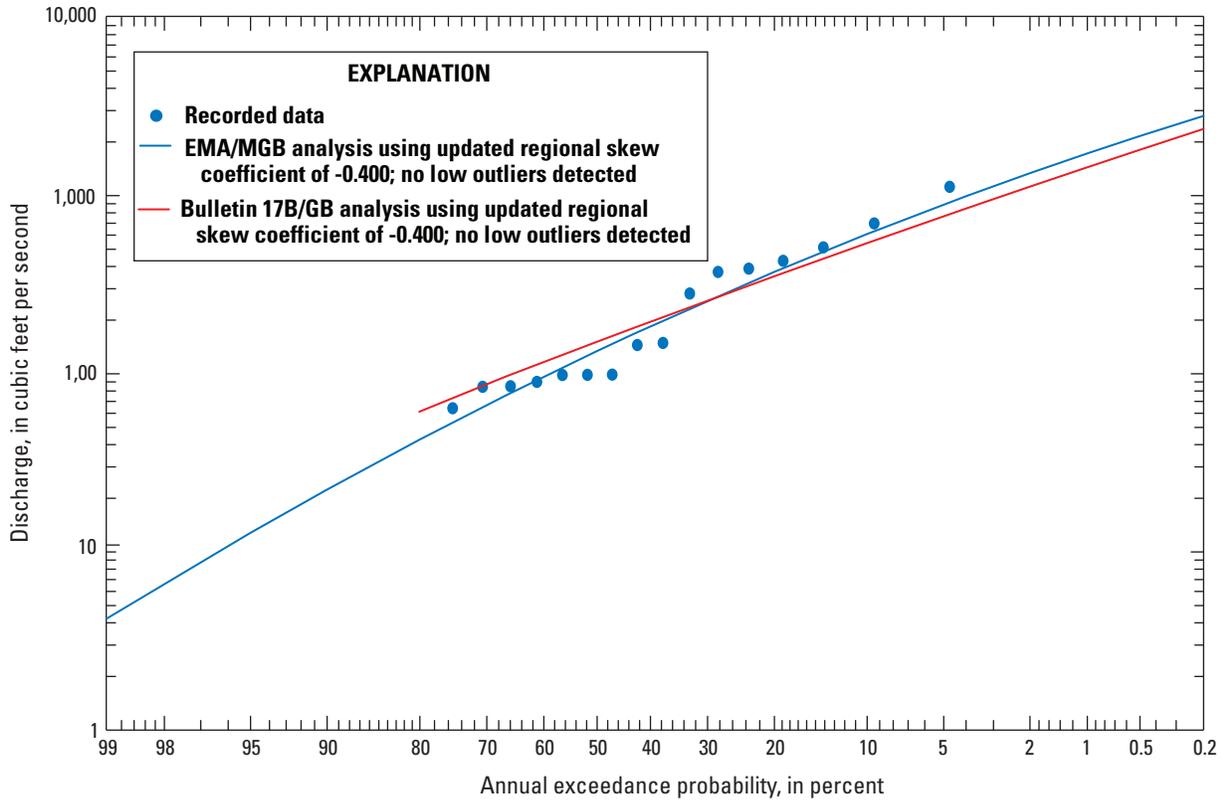
- $G_w$  is the weighted skew,
- $G_s$  is the station skew,
- $G_R$  is the regional skew, and

$MSE_R$  and  $MSE_s$  are the mean square error of the regional and station skew, respectively.

An annual peak-discharge record at a streamgage may include extremely small or extremely large discharge values that are statistically determined to be low or high outliers in the record. The peak-discharge record also may include information about historic floods that happened outside of the period of streamgage operation, or systematic record. Historic floods typically are considered to be the largest peak discharges during an extended period of time that is longer than the systematic record. Bulletin 17B (Interagency Advisory Committee on Water Data, 1982) provides guidelines for detecting outliers and interpreting historical floods, and provides computational methods for appropriate adjustments to the LP3 distribution to account for outliers and historical flood information (Gotvald and others, 2012). Although these adjustments generally improve estimates of AEPDs at a streamgage, the EMA/MGB method integrates censored discharges (low and high outliers) and historical flood discharges more efficiently (Cohn and others, 1997) than the guidelines provided in Bulletin 17B. The station estimates of mean, standard deviation, and skew for streamgages included in this study were computed using EMA/MGB.

### Expected Moments Algorithm (EMA/MGB) Analyses

In this study, the EMA/MGB method was used to compute LP3 exceedance-probability analyses for all 518 streamgages evaluated for use in the development of regression equations and for all 330 streamgages evaluated for use in the regional skew analysis for Iowa. PeakfqSA versions 0.960 and 0.974, an EMA/MGB program developed by Cohn (2011), were used to compute all EMA/MGB annual exceedance-probability analyses for this study. Identical results were obtained using either version of PeakfqSA for the streamgages tested. For streamgages that have systematic annual peak-discharge records for complete periods, no low outliers, and no historical flood information, the EMA/MGB method provides identical log estimates of the three LP3 statistics (mean, standard deviation, and skew coefficient) as the standard LP3 method described in Bulletin 17B (Gotvald and others, 2012). The EMA/MGB method allows for the integration of censored and interval peak-discharge data in the analysis. For historical periods that extend outside of the period of systematic record or between periods of systematic record, censored data may be expressed in terms of discharge perception thresholds. Thus, there are two types of censored data: (1) annual peak discharges collected at CSGs for which the discharge is only known to be less than the minimum recordable discharge threshold, or (2) in the case of historical periods, annual peak discharges that are only known not to have exceeded a recorded historical flood discharge. For example, the Thompson River at Davis City (streamgage 06898000, map number 501, fig. 1) has historic information that indicates a recorded

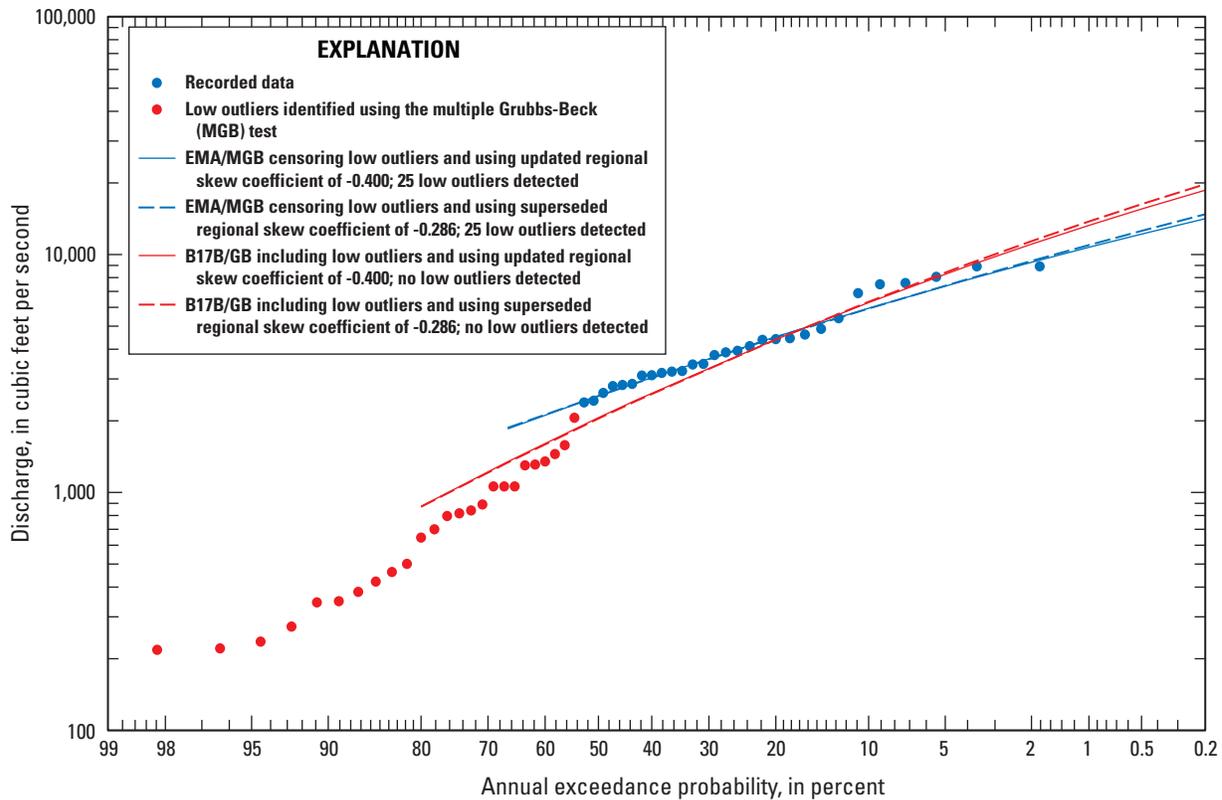


**Figure 4.** Annual exceedance-probability curves for Bloody Run Tributary near Sherrill, Iowa (05414605), showing the difference between expected moments algorithm (EMA/MGB) and Bulletin 17B/GB annual exceedance-probability analyses for a crest-stage gage (CSG) with four annual peak discharges below the minimum recording threshold.

historical flood discharge,  $Q_{hist}$ , of 30,000 ft<sup>3</sup>/s in 1885 was the largest during the historical period (1885–1917) before streamgage operation began in 1918 and during the historical period (1927–40) when streamgage operation was discontinued before reactivation in 1941. Each missing annual peak discharge from 1885–1917 and from 1927–40 can be characterized as a censored discharge for which the value is assumed not to have exceeded the  $Q_{hist}$ ; estimates for those censored discharges bounded between 0 and  $Q_{hist}$  were used in the LP3 exceedance-probability analysis. As listed in [table 1](#) (link to Excel file), the lower bound of the perception threshold for the EMA/MGB analysis for this streamgage was set at 29,500 ft<sup>3</sup>/s, or just below the historic 1885 peak discharge. The basis for the assumption of setting the lower bound of the perception threshold just below  $Q_{hist}$  for the two historical periods, is that if a flood larger than  $Q_{hist}$  had happened during the historical period, it would have been documented. Streamgages that used lower perception thresholds for missing years for historic periods outside of the systematic period of streamgage operation, or for missing years during periods of streamgage operation, are noted in [table 1](#) (link to Excel file).

The EMA/MGB method also allows use of interval discharges to characterize peak discharges that are known to be greater or less than some specific value or that can only be reliably estimated within a specific range in discharge

(Gotvald and others, 2012). Interval discharges commonly are used by the EMA/MGB method to characterize missing data during periods of systematic streamgage operation. For example, for the Bloody Run Tributary near Sherrill, Iowa, CSG (streamgage 05414605, map number 85, fig. 1), an exact peak discharge was not determined for 4 years (1994, 2001, 2003, and 2006) of the 20 years of annual peak-discharge record because the water level did not reach the gage base (bottom) of the CSG, which is the minimum recording threshold (thus producing a censored data record). The missing peaks for these 4 years can be characterized as interval discharges with a range that is bounded by zero and the discharge associated with the elevation of the minimum recording threshold. The discharge for the minimum recording threshold was 45.0 ft<sup>3</sup>/s during 1994 and 2001, 33.0 ft<sup>3</sup>/s during 2003, and 44.6 ft<sup>3</sup>/s during 2006. The EMA/MGB analysis allows the use of multiple discharge intervals to accommodate the changing minimum recording threshold of a CSG. The standard Bulletin 17B/GB analysis sets the gage base discharge at the largest minimum recording threshold (45.0 ft<sup>3</sup>/s) and all minimum recording threshold values and observed point discharge values less than 45.0 ft<sup>3</sup>/s are truncated in the computation of AEPDs. [Tables 4](#) and [6](#) (links to Excel files) lists AEPDs and figure 4 shows annual exceedance-probability curves computed for this CSG using EMA/MGB and



**Figure 5.** Annual exceedance-probability curves for West Branch Floyd River near Struble, Iowa (06600300), showing the effects of including or censoring potentially influential low flows identified from the multiple Grubbs-Beck test and of using the updated regional-skew-coefficient constant or the superseded regional skew-coefficient value.

Bulletin 17B/GB exceedance-probability analyses. CSGs that used lower perception thresholds for minimum recordable discharges are noted in [table 1](#) (link to Excel file).

## Multiple Grubbs-Beck (MGB) Test for Detecting Low Outliers

The standard Grubbs-Beck (GB) test is recommended in Bulletin 17B (Interagency Advisory Committee on Water Data, 1982) for detecting low outliers that depart substantially from the trend of the rest of the annual peak-discharge data. As with high outliers, the retention, modification, or deletion of low outliers can significantly affect the three statistics (mean, standard deviation, and skew coefficient) computed from the LP3 distribution, which can have a large influence on the extreme magnitude flood events that we are most interested in estimating. Therefore, identifying the unusually small potentially influential low flows (PILFs) is critical (T.A. Cohn, U.S. Geological Survey, written commun., 2012). The standard GB test uses the annual peak-discharge data for a streamgage to calculate a one-sided, 10-percent significance-level critical value based on a log-normal distribution of the data. Annual peak discharges identified as low outliers by the standard GB test are truncated from the record and a conditional probability adjustment is applied in the annual exceedance-probability

analysis (Interagency Advisory Committee on Water Data, 1982). Although more than one recorded annual peak discharge may be below the standard GB critical value, typically only one nonzero recorded peak discharge is identified from the test as being a low outlier (Gotvald and others, 2012).

EMA/MGB exceedance-probability analyses computed for this study used the MGB test for the development of the regional skew analysis and the regional regression equations. The MGB test also calculates a one-sided, 10-percent significance-level critical value based on a log-normal distribution of the data. The MGB test is performed so that groups of ordered data are examined (for example, the six smallest values in the record) and excluded from the dataset when the critical value is calculated. If the critical value is greater than the sixth smallest value in the example, then all six values are determined to be low outliers (Gotvald and others, 2012). The MGB test can identify low outliers for as much as 50 percent of the annual peak-discharge record (T.A. Cohn, U.S. Geological Survey, written commun., 2011). The number of low outliers identified by the MGB test for each of the 518 streamgages evaluated for use in the development of regression equations is listed in [table 1](#) (link to Excel file).

Figure 5 shows the resulting annual exceedance-probability curves computed for the West Branch Floyd River near Struble (streamgage 06600300, map number 386, fig. 1) using

EMA/MGB and standard Bulletin 17B/GB analyses. The Bulletin 17B/GB analysis excluded six annual peak discharges because they were below the gage base (349 ft<sup>3</sup>/s). Unfortunately, four of the six peaks are actual observed discharges less than 349 ft<sup>3</sup>/s. The critical low-outlier value calculated by the standard GB test was 204.8 ft<sup>3</sup>/s and no low outliers were detected below the critical value. In contrast, the critical low-outlier value calculated by the MGB test was 2,390.0 ft<sup>3</sup>/s; 25 PILFs (almost one-half of the 54 annual peak discharges) were detected below the critical value. As shown in figure 5, retention of the PILFs has a large influence on the fitting of the right-hand tail of the annual-exceedance-probability distribution. Censoring flows below the low-outlier threshold value with the EMA/MGB method leads to a better fit of the upper portion of the distribution. For this streamgage, use of EMA/MGB lowered the  $Q_{1\%}$  estimate by almost 20 percent and the  $Q_{0.2\%}$  estimate by about 24 percent when compared to the Bulletin 17B/GB estimates (*tables 4 and 6*, links to Excel files).

## Regional Skew Analysis

During the development of flood-frequency techniques in the late 1970s and early 1980s, a nationwide generalized-skew study was performed and documented in Bulletin 17B (Interagency Advisory Committee on Water Data, 1982). Station skew coefficients for long-term streamgages throughout the Nation were computed and used to produce a map of isolines of generalized skew; however, the map was prepared at a national scale using data and methods that are now more than 30 years old. To generate more accurate generalized-skew coefficients, three methods are described in Bulletin 17B for developing generalized skews using skew coefficients computed from streamgages with long-term annual peak-discharge records (25 or more years of record): (1) plot station skew coefficients on a map and construct skew isolines, (2) use regression techniques to develop a skew-prediction equation relating station skew coefficients to some set of basin characteristics, or (3) use the arithmetic mean of station skew coefficients from long-term streamgages in the area (Interagency Advisory Committee on Water Data, 1982).

A generalized-skew coefficient analysis using method 2 was performed as part of this study to update regional skew coefficients for streamgages in the study area. Reis and others (2005), Gruber and others (2007), and Gruber and Stedinger (2008) developed a Bayesian generalized least-squares (B-GLS) regression model for determining regional skew. Recently, a Bayesian weighted least-squares/generalized least-squares (B-WLS/B-GLS) methodology that relates observed skewness coefficient estimators to basin characteristics in conjunction with diagnostic statistics was developed as an extension of the B-GLS methodology (Veilleux and others, 2012; Veilleux, 2011). B-WLS/B-GLS has been shown to be effective in two California studies (Parrett and others, 2011; Lamontagne and others, 2012). B-WLS/B-GLS uses B-WLS to generate stable estimators of model parameters and B-GLS

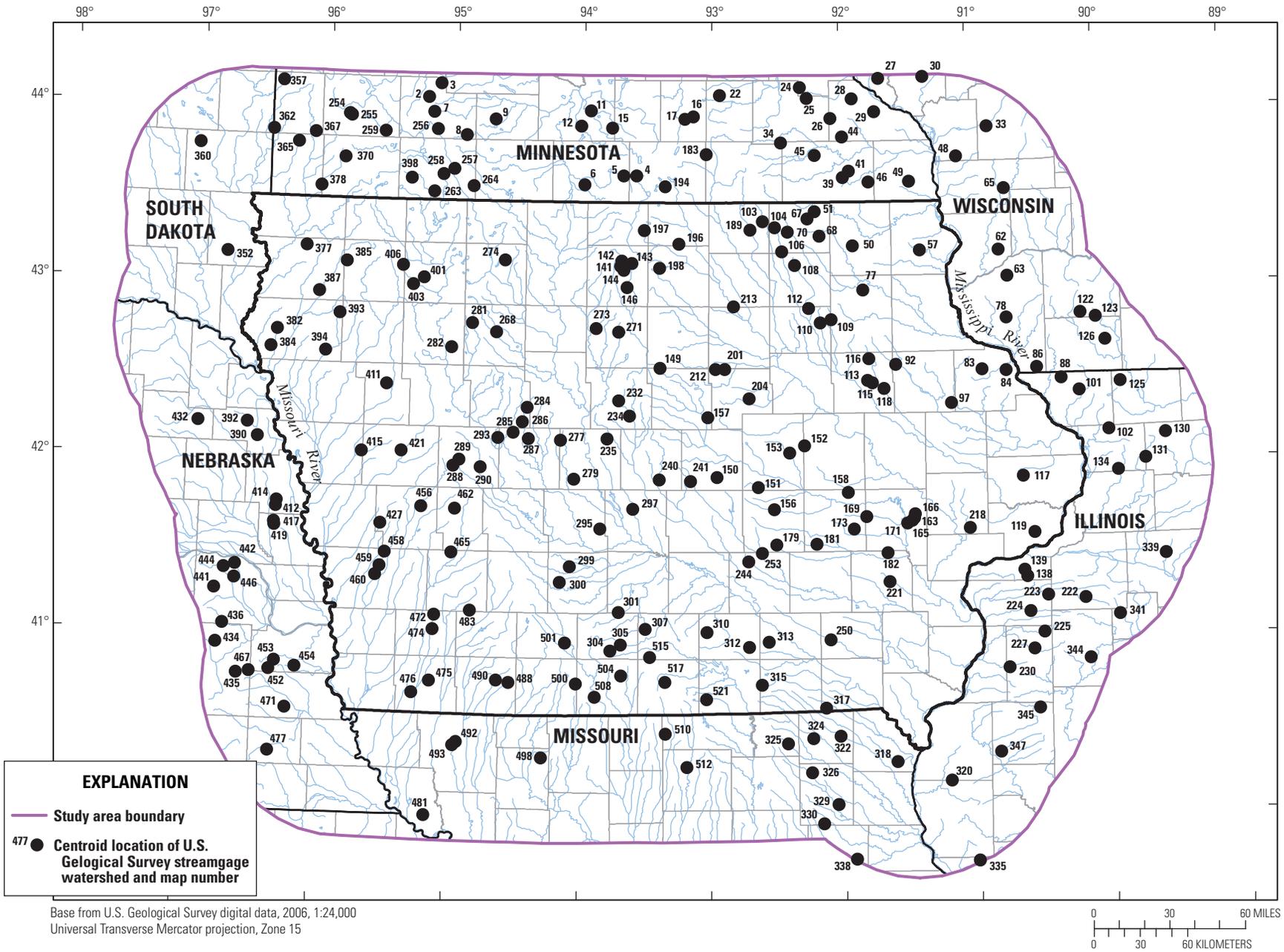
to estimate the precision of those B-WLS regression parameters, as well as the precision of the model. The Iowa regional skew study described here uses the B-WLS/B-GLS methodology, which is described in detail in the appendix.

Annual peak discharges of basins are cross-correlated because a single large storm event can cause the annual peak in several basins. One advantage of using a GLS method is that it takes this cross-correlation among the basins into account. The GLS statistical analysis depends on the estimated cross-correlations of the peak discharges at different pairs of streamgages. The cross-correlation generally is estimated as a function of distances between the centroids of gaged basins. Details on the cross-correlation model can be found in the appendix of this report.

If watersheds of two streamgages are nested such that one is contained within the other, the cross-correlation between the concurrent peak discharges would be larger than if the basins were not nested. This leads to errors in the estimation of cross-correlations for non-nested basins, resulting in incorrect model errors. A screening metric was developed to determine the redundant streamgage pairs that represent the same watershed. Details on the screening metric used can be found in the appendix of this report.

After running the screening metric on the 330 streamgages initially evaluated for inclusion in the regional skew study for Iowa, 55 streamgages subsequently were removed from the generalized-skew coefficient regression analysis because of redundancy. An additional two streamgages were removed because of backwater effects. Also, an additional 33 streamgages with EMA/MGB estimates of the MSE of skew ( $MSE_G$ ) greater than 0.4 were removed because of the inability of EMA/MGB to correctly estimate  $MSE_G$  when a streamgage has a short record length and a large percentage of censored observations. Thus, a total of 240 streamgages with at least 20 years of recorded, annual peak discharges were used for the final B-WLS/B-GLS regional skew regression analysis. Figure 6 shows the location of the basin centroids for these 240 streamgages. Details on the data analysis and streamgage removal can be found in the appendix of this report.

Based on the B-WLS/B-GLS regression analysis, a constant generalized-skew value of -0.40 was determined to be the best model to predict the generalized skew in the study area. More complicated B-WLS/B-GLS models with additional explanatory variables were evaluated, but resulted in modest improvements in accuracy. A detailed description of the accuracy of these models is available in the appendix. The modest improvements in the more complicated models were not justified because of the increased complexity associated with the additional explanatory variables. The MSE associated with the constant generalized-skew model is 0.16, and is equivalent to a record length of 50 years. This is a significant improvement over the skew map MSE value of 0.302 in Bulletin 17B, which is equivalent to 17 years of record. The generalized-skew value of -0.40 with an associated MSE of



**Figure 6.** Location of basin centroids for 240 streamgages used for regional skew analysis for Iowa.

0.16 was used to compute the weighted skew value, shown in equation 2, for the 518 streamgages with 10 or more years of record. Annual exceedance-probability curves computed for streamgage 06600300 (map number 386, fig. 1) using superseded (-0.286) and updated (-0.400) regional skew coefficients are shown in figure 5. For this example (fig. 5), use of the updated regional-skew-coefficient constant decreased the  $Q_{1\%}$  flood discharge by almost 3 percent using EMA/MGB analyses (tables 4 and 6, links to Excel files). Percentage differences in estimates of AEPDs vary from site-to-site between annual exceedance-probability analyses computed using superseded and updated regional skew coefficients, and figure 5 only illustrates the difference for one site. Tables 4 and 6 list AEPDs computed for Iowa streamgages using superseded and updated regional skew coefficients, and the section Comparison of Annual Exceedance-Probability Discharges summarizes relative percentage changes between AEPDs computed for Iowa streamgages using superseded and updated regional skew coefficients. The appendix of this report provides additional details of the generalized-skew coefficient regression analysis used for this study.

## Regional Regression Analyses to Estimate Annual Exceedance-Probability Discharges for Ungaged Stream Sites

In a regional regression study, subdividing a large study area into subregions that are more homogeneous in terms of flood hydrology typically helps to reduce error in the regression equations. Because different flood regions have been determined for Iowa in previous studies and different GIS data layers, basin-characteristic measurement methods, and annual exceedance-probability analyses were used in this study, preliminary statewide regression equations were initially developed for the  $Q_{1\%}$  flood discharge using 510 streamgages. Table 7 lists 13 streamgages that were removed from the original 523 streamgages listed in table 1 (link to Excel file) because of significant trends, urbanization, channelization, replacement of a discontinued streamgage with a nearby active streamgage, backwater, or questions concerning accuracy or determination of some annual peak discharges in the record. Table 7 also lists the five streamgages (map numbers 92, 331, 364, 484, and 496, fig. 1) that initially were included for evaluation in the regional skew analysis, but subsequently were removed before the regional regression analysis. As noted in the previous section, Regional Skew Analysis, redundant streamgages and streamgages with EMA/MGB estimates of the  $MSE_G$  greater than 0.4 were removed from the regional

**Table 7.** Streamgages removed from regional-regression analyses.

[no., number; USGS, U.S. Geological Survey. Streamgage locations are shown in figure 1]

| Map no. | USGS streamgage number | Streamgage name   | Streamgage removed prior to regional-regression analyses | Reason for removal of streamgage from regression analyses                 |
|---------|------------------------|---|--|---|
| 81      | 05414400               | Middle Fork Little Maquoketa River near Rickardsville, Iowa | No   | Significant trend.  |
| 92      | 05417000               | Maquoketa River near Manchester, Iowa                       | Yes  | Used active streamgage 05416900.  |
| 118     | 05422470               | Crow Creek at Bettendorf, Iowa                              | No   | Urbanization.   |
| 206     | 05464130               | Fourmile Creek near Lincoln, Iowa                           | No   | Channelized/significant trend.  |
| 207     | 05464133               | Half Mile Creek near Gladbrook, Iowa                        | No   | Channelized.  |
| 208     | 05464137               | Fourmile Creek near Traer, Iowa                             | No   | Channelized/significant trend.  |
| 284     | 05482800               | Happy Run at Churdan, Iowa                                  | No   | Significant trend.  |
| 294     | 05484800               | Walnut Creek at Des Moines, Iowa                            | No   | Urbanization.   |
| 331     | 05501000               | North River at Palmyra, Missouri                            | Yes  | Backwater, some peaks in peak-flow file need to be reviewed for accuracy. |
| 364     | 06482933               | Chanarambie Creek near Edgerton, Minnesota                  | Yes  | Discharge not determined for highest peak stage of record.                |
| 457     | 06807470               | Indian Creek near Emerson, Iowa                             | No   | Significant trend.  |
| 484     | 06817500               | Nodaway River near Burlington Junction, Missouri            | Yes  | Some peaks in peak-flow file need to be reviewed for accuracy.            |
| 496     | 06897000               | East Fork Big Creek near Bethany, Missouri                  | Yes  | Backwater, some peaks in peak-flow file need to be reviewed for accuracy. |

skew analysis because their inclusion was believed to significantly affect the results of the analysis (see appendix). For the regional regression analysis to estimate annual exceedance-probability discharges for ungaged sites, streamgages were not removed because of redundancy or  $MSE_G$  greater than 0.4. Although redundancy could have been addressed in the regional regression analysis, the inclusion of potentially redundant sites is not believed to significantly affect the development of regional regression equations for estimating AEPDs for ungaged stream sites in Iowa. In this study, the  $Q_{1\%}$  flood discharge was selected for optimizing predictive accuracies for the development of regression equations because this AEPD is used most often by water managers, engineers, and planners. The same explanatory variables determined to be the most significant for the development of  $Q_{1\%}$  flood-discharge regression equations also were used for the other seven exceedance probabilities for the development of statewide and regional regression equations to minimize the possibility of predictive inconsistencies between estimates of different probabilities. Predictive inconsistencies result when the discharge estimate for a larger probability is greater than the discharge estimate for a smaller probability; for example, when a  $Q_{2\%}$  flood-discharge estimate is greater than a  $Q_{1\%}$  flood-discharge estimate. Table 8 lists the significant variables identified and the predictive accuracies obtained for preliminary statewide regression equations initially developed for each of the eight annual exceedance probabilities using ordinary least-squares (OLS) and subsequently finalized using generalized least-squares (GLS) multiple-linear regression analyses (see following sections for further discussion of OLS and GLS regression). The preliminary statewide AEPD equations provided base-level predictive accuracies that RREs can be compared against to evaluate improvement in accuracy. Because RREs provided improved accuracies, the statewide equations were

not developed further and are not listed in this report; they are summarized in table 8 to provide a reference indicating the improvement obtained through regionalization.

## Definition of Flood Regions

Residual values (differences between AEPDs computed from observed annual peak discharges and those predicted from the regression equations) from the preliminary statewide regression analyses were mapped at streamgage locations to identify spatial trends in the predictive accuracy of the regression equations. Differences in plotted residual values for the streamgages were grouped to define general flood regions within the study area. Streamgages were grouped into regression subsets on the basis of the flood regions, and OLS multiple-linear regression analyses were performed for each region. Because of the amount of variability in the residual mapping for the  $Q_{1\%}$  flood discharge, a cluster analysis method also was used to help define flood regions. A cluster analysis method, called partitioning around medoids using Spotfire S+ statistical software (TIBCO Software Inc., 2008), was used to help define flood regions in Iowa. Cluster analysis is a statistical technique that was used to partition streamgages into groups (clusters) with similar streamflow or basin characteristics. The cluster analyses were based on the basin characteristics previously identified as significant variables in the preliminary statewide regression equations developed for the  $Q_{1\%}$  flood discharge (table 8). Drainage area was not included in the analyses because it is not a unique characteristic for any one cluster. The partitioning around medoids method of cluster analysis uses medoids instead of centroids to form groups for which average dissimilarity of basin-characteristic values in each group are minimal ([http://www.unesco.org/webworld/idams/advguide/Chapt7\\_1\\_1.htm](http://www.unesco.org/webworld/idams/advguide/Chapt7_1_1.htm)). Cluster analyses and the

**Table 8.** Significant explanatory variables and predictive accuracies of preliminary statewide regression equations.

[SEP, average standard error of prediction; Pseudo- $R^2$ , pseudo coefficient of determination; %, percent; DRNAREA, geographic information system drainage area; (+), explanatory variable has positive relation to the response variable; DESMOIN, percent of basin within Des Moines Lobe landform region; (-), explanatory variable has negative relation to the response variable; CCM, constant of channel maintenance]

| Preliminary generalized least-squares regression analyses results |  |               |                         |
|---|--|---------------|-------------------------|
| Annual exceedance probability                                     | Most significant explanatory variables identified for the preliminary equation and explanatory-variable relation signs | SEP (percent) | Pseudo- $R^2$ (percent) |
| 510 streamgages used in regression analyses                       |  |               |                         |
| $Q_{50\%}$  | DRNAREA <sup>0.011</sup> (+), DESMOIN (-), CCM <sup>0.55</sup> (-)   | 53.9          | 90.6                    |
| $Q_{20\%}$  | DRNAREA <sup>-0.011</sup> (+), DESMOIN (-), CCM <sup>0.55</sup> (-)  | 42.5          | 92.9                    |
| $Q_{10\%}$  | DRNAREA <sup>-0.023</sup> (+), DESMOIN (-), CCM <sup>0.55</sup> (-)  | 41.3          | 92.7                    |
| $Q_{4\%}$   | DRNAREA <sup>-0.035</sup> (+), DESMOIN (-), CCM <sup>0.55</sup> (-)  | 42.0          | 91.9                    |
| $Q_{2\%}$   | DRNAREA <sup>-0.043</sup> (+), DESMOIN (-), CCM <sup>0.55</sup> (-)  | 43.1          | 91.2                    |
| $Q_{1\%}$   | DRNAREA <sup>-0.050</sup> (+), DESMOIN (-), CCM <sup>0.55</sup> (-)  | 45.8          | 89.9                    |
| $Q_{0.5\%}$   | DRNAREA <sup>-0.056</sup> (+), DESMOIN (-), CCM <sup>0.55</sup> (-)  | 47.7          | 88.8                    |
| $Q_{0.2\%}$   | DRNAREA <sup>-0.064</sup> (+), DESMOIN (-), CCM <sup>0.55</sup> (-)  | 51.0          | 87.0                    |

residual mapping defined eight initial flood regions for the study area.

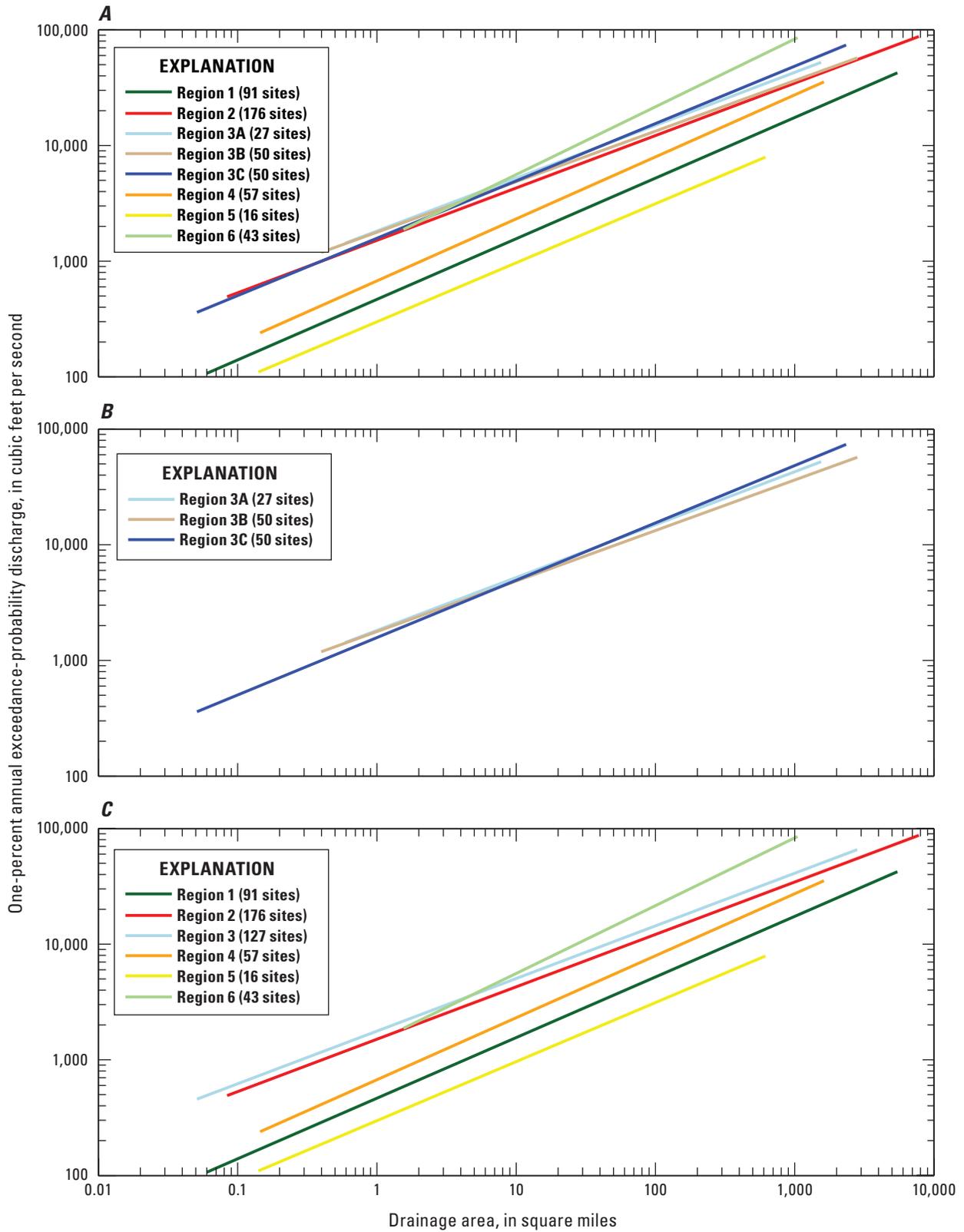
Streamgages were then grouped into eight regional datasets. Three of the regions (Des Moines Lobe, Paleozoic Plateau, and Loess Hills) were defined primarily on the basis of Iowa's landform regions (fig. 3) and a fourth region was defined on the basis of a soil region, which was used to define a hydrologic region in a previous study (region 3 in Eash, 2001). These four flood regions (three landform regions and one soil region) initially were defined with the cluster analysis and then subsequently defined as flood regions by the boundaries of their respective landform and soil regions. Analysis-of-covariance regression (Helsel and Hirsch, 2002) was used to test each flood region for statistically significant differences by comparing the intercept for each region's regression model to that for the rest of the study area by assigning a location variable for each region. Each location-indicator variable was set at one if the streamgage was in a particular region, or zero if the streamgage was not in a particular region. A two-variable OLS regression analysis that included drainage area and the location-indicator variable was performed statewide for each of the eight exceedance probabilities for each of the eight flood regions being tested. Statistical significance for each region was determined using a 95-percent confidence level. Statistical significance for the location-indicator variable indicates a difference in the regression intercept between streamgages in that region and streamgages in the rest of the study area. All eight flood regions were determined to be significantly different from each other, and preliminary RREs were developed for the  $Q_{1\%}$  flood discharge for each of the eight regions.

Figure 7A shows the relation between 1-percent AEPDs and drainage area for the eight initial flood regions defined for the study area. Figure 7B shows that the relation for flood regions 3A (27 streamgages located in the Paleozoic Plateau landform region), 3B (50 streamgages located in the Loess Hills landform region and southeastward), and 3C [50 streamgages located in the region 3 area from Eash (2001)] are similar. A comparison of preliminary OLS regression analyses for the  $Q_{1\%}$  flood discharge between a combined flood region 3 with 127 streamgages and three separate flood regions (3A, 3B, and 3C) indicated little difference in predictive accuracies regarding average standard errors of estimate or adjusted coefficients of determination; therefore, these three initial regions were combined into a single flood region. Because the analysis-of-covariance tests compared the regression intercept for each of these three regions individually to the regression intercept for the rest of the study area, each of these three regions were determined individually to be statistically significant compared to the rest of the study area. Figure 7C shows the relation between 1-percent AEPDs and drainage area for the six final flood regions defined for the study area. Analysis-of-covariance regression tests determined that all six final flood regions were significantly different from each other and figure 7C shows that the relation for these six final regions is unique.

The goal of the regionalization analyses was to define the best overall regions for the  $Q_{1\%}$  flood discharge and to have an adequate number of streamgages (preferably, at least 30) in each regional dataset for the regression analyses. Streamgages flagged as outliers (high leverage or high influence points) in the GLS regression analyses, using a weighted-multiple-linear regression program (WREG) (Eng and others, 2009), were reviewed for inaccurate data and for possible effects that may bias the development of regression equations. In addition to the five streamgages previously removed from the original dataset of 523 streamgages (*table 1*, link to Excel file), another eight streamgages that were flagged as outliers in the GLS regression analyses also were removed from the 518 streamgages included in the regression datasets (*tables 3 and 4*, links to Excel files) because of significant trends, urbanization, or channelization (*table 7*). All other streamgages that were flagged as outliers were kept in the regression datasets because there was no justification for removing them. Thus, a total of 510 streamgages were considered to have unaltered annual peak discharges for the regional regression analyses. All 510 of these streamgages have at least 10 years of record that can be used for the development of the eight selected annual exceedance-probability equations.

Three flood regions (regions 1, 2, and 3, fig. 1) were defined for Iowa after testing a number of different regional regression-model combinations to define the initial eight, and then the final six, flood regions for the study area (fig. 7). Because three of the final six flood regions defined for the study area are completely outside of Iowa, regression equations were not developed further for flood regions 4, 5, and 6. Flood regions 4–6 are not shown in figures 1–3, but *table 1*, (link to Excel file) lists streamgages located in flood regions 4–6 and figure 1 shows the location of all streamgages listed in *table 1* (link to Excel file). Region 4 includes 57 streamgages located in Illinois, Wisconsin, and southeastern Minnesota; region 5 includes 16 streamgages located in South Dakota and southwestern Minnesota; and region 6 includes 43 streamgages located in Nebraska (*tables 1, 3, and 4*). Although predictive accuracies were optimized by subdividing the 510 streamgages in the study area into the final six flood regions that are more homogeneous in terms of flood hydrology, it is important to note that flood-region boundaries are not actually distinct lines, but the boundaries are transition zones where the hydrologic characteristics of one region transition to the hydrologic characteristics of another region. Figure 1 shows the three flood regions defined for Iowa for the development of final RREs. Flood-region boundaries were defined along 12-digit WBD HUC boundaries to avoid drawing a flood-region boundary through a HUC polygon. For a 12-digit HUC that overlies a landform region boundary, the flood-region boundary was drawn to include the landform region that comprises the most of the 12-digit HUC area.

Flood region 1 is defined primarily by the Des Moines Lobe landform region and contains approximately 23 percent of the total land area of the State (figs. 1–3). Flood region 1 generally comprises low-relief terrain, accentuated by natural



**Figure 7.** Relation between one-percent annual exceedance-probability discharges and drainage area for *A*, eight initial; *B*, three combined; and *C*, six final flood regions defined for study area.

lakes, potholes, and marshes, where surface-water drainage typically is poorly defined and sluggish. For a given size of drainage area, estimates of AEPDs from region 1 generally are lower than those from flood regions 2 and 3 (fig. 7C).

Flood region 2 is the largest flood region in the State and contains approximately 45 percent of the total land area (figs. 1–3). Flood region 2 partially is defined by the Des Moines Lobe and Paleozoic Plateau landform-region boundaries, although flood region 2 only includes small areas of either landform region. Region 2 includes most of the Iowan Surface, the Northwest Iowa Plains, and the East-Central Iowa Drift Plain landform regions; the entire Iowa-Cedar Lowland landform region; and nearly one-half of the Southern Iowa Drift Plain landform region. Region 2 was defined as the remaining area following the definition of flood regions 1 and 3, and it generally comprises gently rolling topography and well-established drainage systems. For a given size of drainage area, estimates of AEPDs from region 2 are generally in between those from flood regions 1 and 3 (fig. 7C).

Flood region 3 is a combined regression-model region (fig. 7B) and contains approximately 32 percent of the total land area of the State (figs. 1–3). Flood region 3 includes two non-contiguous areas of the State, one that includes most of the Paleozoic Plateau landform region and the other that includes most of the Loess Hills landform region and more than one-half of the Southern Iowa Drift Plain landform region. As noted earlier in this section, flood region 3 was formed by combining three initial regression-model flood regions: (1) the Paleozoic Plateau landform region, (2) the Loess Hills landform region and areas of deep loess southeast of the Loess Hills, and (3) hydrologic region 3 in Eash (2001). Region 3 generally comprises steeply to gently rolling topography and well-established drainage in the area of the Southern Iowa Drift Plain landform region; a bedrock-dominated, erosional topography of integrated drainage networks with steep gradients, and deeply entrenched valleys in the area of the Paleozoic Plateau landform region; and topography that is sharp-featured, with a dense drainage network forming tight hollows, narrow ravines, and steep gullies in the area of the Loess Hills landform region; and an area of low soil-permeability rates in the southeast area of the State (fig. 2, soil regions 21 and 22; Eash, 2001). For a given size of drainage area, estimates of AEPDs from region 3 generally are greater than those from flood regions 1 and 2 (fig. 7C).

## Development of Regional Regression Equations

GLS multiple-linear regression analyses, weighted on the basis of streamgage record length and the variance and cross-correlation of the annual peak discharges, were used to develop equations to estimate AEPDs for ungaged stream sites in Iowa.

## Multiple-Linear Regression

Multiple-linear-regression analysis is the most common method used to develop equations for the estimation of streamflow statistics at ungaged sites. Multiple-linear regression models the relation between two or more basin characteristics (called explanatory or independent variables) and a streamflow statistic (called a response or dependent variable) by fitting a linear equation to the data. Every value of each basin characteristic is associated with the value of the streamflow statistic. Upon the development of regression equations, measurements of the basin characteristics at ungaged stream locations can be used to estimate the streamflow statistic.

The general form of equations developed from multiple-linear-regression analysis is shown in the following equation:

$$Y_i = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n + e_i \quad (3)$$

where

|                |  |
|----------------|--|
| $Y_i$          | is the response variable (estimate of the streamflow statistic computed from observed streamflow) for site $i$ ,     |
| $X_1$ to $X_n$ | are the $n$ explanatory variables (basin characteristics) for site $i$ ,   |
| $b_0$ to $b_n$ | are the $n+1$ regression model coefficients, and   |
| $e_i$          | is the residual error (difference between the observed and predicted values of the response variable) for site $i$ . |

Assumptions for the use of regression analyses are (1) the model adequately describes the linear relation between the response and explanatory variables, (2) the mean of  $e_i$  is zero, (3) the variance of  $e_i$  is constant and independent of the values of  $X_n$ , (4) the values of  $e_i$  are distributed normally, and (5) the values of  $e_i$  are independent of each other (Iman and Conover, 1983). Because streamflow data are naturally correlated spatially and temporally, the last assumption is not completely satisfied with the use of OLS. As a result, GLS regression was used to develop the final equations for estimating AEPDs for Iowa. A general overview of the OLS and GLS multiple-linear regression techniques used to develop the initial and final equations is presented in the following two sections.

## Ordinary-Least-Squares Regression

OLS regression analyses were used to develop initial multiple-linear regression equations, or models, for estimating statewide and regional annual  $Q_{1\%}$  flood discharges for Iowa. Final equations were developed using GLS regression procedures. OLS regression analyses were used to identify the best combinations of basin characteristics to use as explanatory variables in the development of regression models and to define the flood regions.

Logarithmic transformations (base 10) were performed for all response variables and for selected explanatory variables used in the OLS and GLS regression analyses. Data transformations, other than logarithmic transformations, also

were used for selected explanatory variables to obtain a more constant variance of the residuals about the regression line and to linearize the relation between the response variable and the explanatory variables. The response variable is assumed to be a linear function of one or more explanatory variables. The form of a base-10 logarithmic transformation is shown in the following equation:

$$\log Y_i = b_0 + b_1 \log X_1 + b_2 \log X_2 + \dots + b_n \log X_n + e_i \quad (4)$$

When equation 4 is retransformed back to its original units, it is algebraically equivalent to the following equation:

$$Y_i = 10^{b_0} X_1^{b_1} X_2^{b_2} \dots X_n^{b_n} 10^{e_i} \quad (5)$$

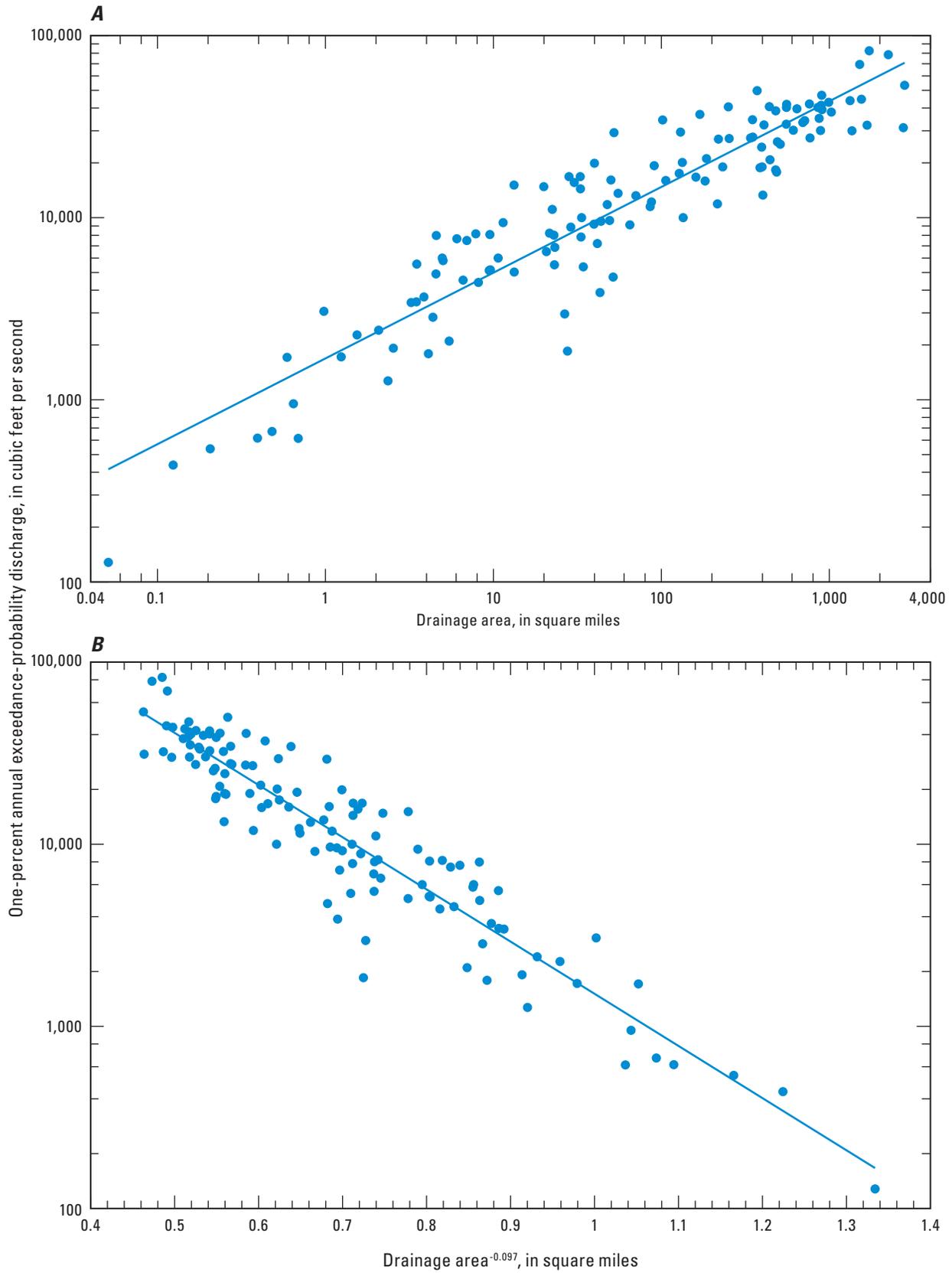
Several basin characteristics were deleted from the original regression dataset of 59 basin characteristics because of multicollinearity. Multicollinearity is the condition wherein at least one explanatory variable is related closely to (that is, not independent of) one or more other explanatory variables. Regression models that include variables with multicollinearity may be unreliable because coefficients in the models may be unstable. Correlation coefficients greater than 0.5, or less than -0.5, and plots of the data were used as guides in identifying variables with multicollinearity. The hydrologic validity of variables with multicollinearity in the context of annual peak discharge was the principal criterion used in determining which basin characteristics were deleted from the dataset.

OLS regression analyses were performed using Spotfire S+ statistical software (TIBCO Software Inc., 2008). Initial selections of significant explanatory variables for the OLS regression models were performed using the Efroymson stepwise-selection method (Efroymson, 1960). The Efroymson method is an automatic procedure for regression model selection when there are a large number of potential explanatory variables. The procedure is similar to forward selection, which tests basin characteristics one by one and identifies those that are statistically significant, except as each new basin characteristic is identified as being significant, partial correlations are checked to see if any previously identified variables can be deleted (Ahearn, 2010). When basin characteristics were determined to be highly correlated to each other, only one basin characteristic at a time was tested in the Efroymson selection process.

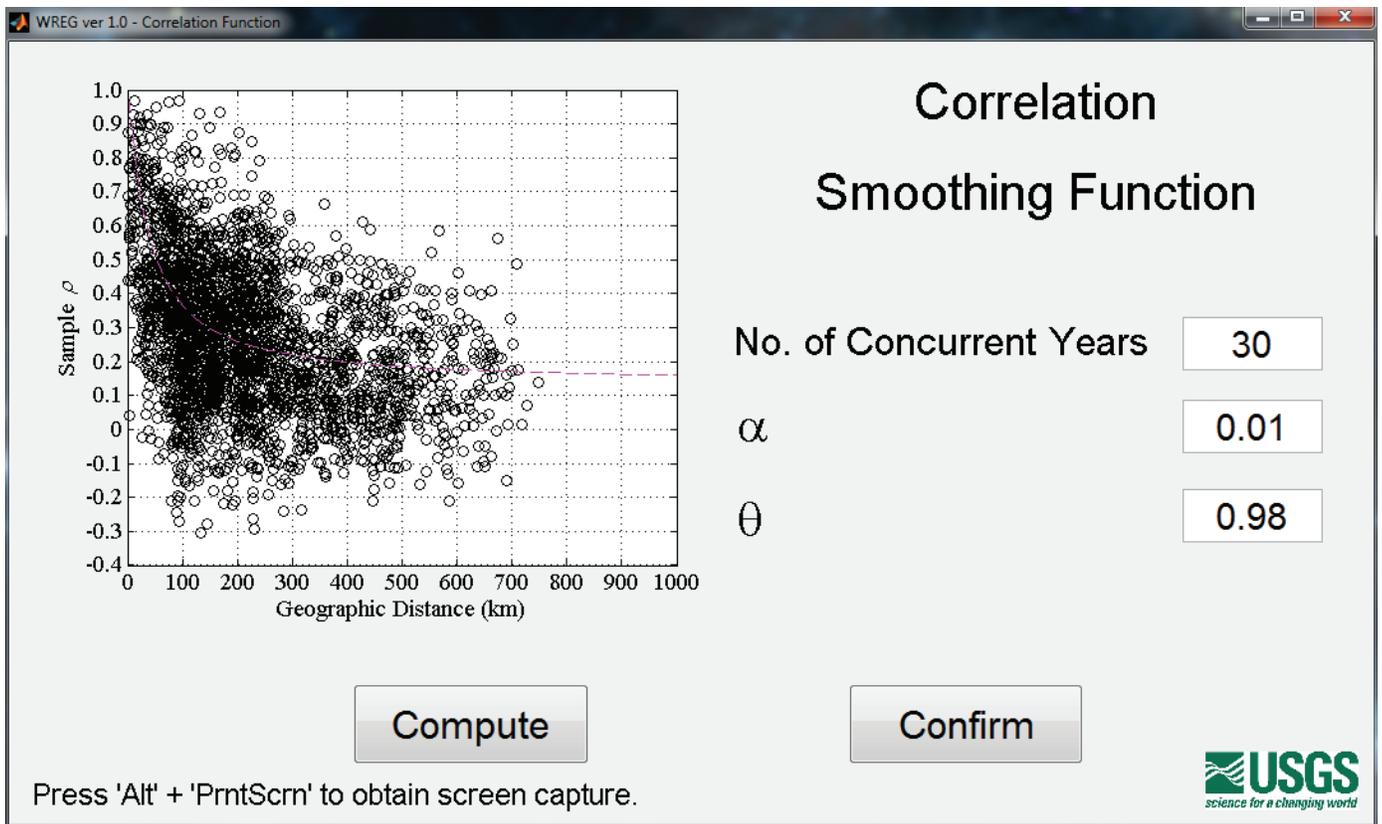
The Efroymson analyses produced a subset of potential significant basin characteristics for the  $Q_{1\%}$  flood discharge for each flood region. Each subset of basin characteristics was then iteratively tested using standard OLS regression analyses to identify several sets of the best equations (regression models) that contained no more than three significant explanatory variables (basin characteristics). A limit of three explanatory variables per equation was used to minimize overfitting of the regression models. Results of the OLS models were evaluated to determine their adequacy, including graphical relations and residual plots, variance inflation factor (VIF), Cook's D statistic (Cook, 1977; Helsel and Hirsch, 2002), high-leverage

points, the average standard error of estimate (SEE), and the adjusted coefficient of determination ( $\text{adj-R}^2$ ) (Helsel and Hirsch, 2002). The selection of explanatory variables, and the signs and magnitudes of their respective regression coefficients, were each evaluated to ensure hydrologic validity in the context of AEPDs. This criterion takes precedence for all other criteria. All explanatory variables selected by OLS regression in this study were statistically significant at the 95-percent confidence level. Explanatory variables were selected to minimize SEE and to maximize the  $\text{adj-R}^2$ . SEE is a measure of the fit of the observed data to the regression model (difference between the value of the observed AEPD and the value of the predicted AEPD) and of the error inherent in the regression model; SEE also is referred to as the root mean square error (RMSE).  $\text{Adj-R}^2$  is a measure of the proportion of the variation in the response variable that is explained by the explanatory variables and adjusted for the number of streamgages and explanatory variables used in the analysis. Correlation between explanatory variables and VIF (Marquardt, 1970; Helsel and Hirsch, 2002) was used to assess multicollinearity in the regression models. Multicollinearity problems were identified with a regression-diagnostics tool implemented in the USGS library version 4.0 (Lorenz and others, 2011) for Spotfire S+ statistical software (TIBCO Software Inc., 2008) by checking each explanatory variable for VIF greater than two.

A logarithmic graphical relation between observed  $Q_{1\%}$  flood discharges and drainage area fitted with a one-variable OLS regression equation indicates a discernible curvilinear relation for flood region 3 and a poor fit of the data with the regression line for drainage areas less than about 1 mi<sup>2</sup> (fig. 8A). Traditionally,  $\log_{10}$ -transformations have been used for all explanatory variables (basin characteristics) for the development of annual exceedance-probability regression equations for which the equations are of the form shown by equation 5 (Eash, 2001); however, for some regional datasets, exclusive use of  $\log_{10}$ -transformations does not fully linearize the relations between the variables. As a result, some systematic bias remains because of a discernible curvilinear relation between AEPDs and drainage area in  $\log_{10}$ -space (fig. 8A). A procedure developed for Texas (Asquith and Thompson, 2008) was followed in this study to determine whether a  $\log_{10}$ -transformation or a power transformation of drainage area provided the best fit of the data for each of the three flood regions defined for Iowa. The Texas procedure is a statistical framework based on the minimization of the predicted residual sums of squares (PRESS) statistic through a test of power transformations of drainage area. The PRESS statistic is regarded as a measure of regression performance when the model is used to predict new data (Montgomery and others, 2001). The PRESS statistic is a validation-type statistic (Helsel and Hirsch, 2002) and optimal power transformations were selected on the basis of minimizing the PRESS statistic. Minimization of PRESS is appropriate for annual exceedance-probability analysis because the regression equations are used in hydrologic engineering practice to predict new data (Asquith and Thompson,



**Figure 8.** Relation between one-percent annual exceedance-probability discharges and drainage area for 127 streamgages in flood region 3 for *A*, log 10 transformed drainage area and *B*, power-transformed drainage area.



**Figure 9.** Screen capture of the weighted-multiple-linear regression program (WREG) smoothing function for generalized-least-squares (GLS) correlation of the time series of annual peak discharges as a function of distance between 176 streamgages in flood region 2 with 30 years of concurrent discharge.

2008). Compared to equations using  $\log_{10}$ -transformations of drainage area, lower values for PRESS statistics and average standard error of predictions (SEPs) were obtained from equations using power transformations of drainage area derived from the PRESS minimization procedure for two of the three flood regions tested. Use of a power transformation of drainage area, instead of a  $\log_{10}$ -transformation of drainage area, improved the fit of the regression model for flood regions 2 and 3, and removed the discernible curvilinear relation for flood region 3 (fig. 8B). The PRESS-minimization procedure computed a different optimal power transformation of drainage area for each of the eight annual exceedance probabilities for flood regions 2 and 3. Because there was no difference between the PRESS statistic and SEP values for either the  $\log_{10}$  or power transformation of drainage area for the  $Q_{50\%}$  flood-discharge equation for flood region 3, the  $\log_{10}$ -transformation was selected for the final regression model.

### Generalized-Least-Squares Regression

GLS multiple-linear regression was used to develop 24 regression equations for estimating AEPDs for the three flood regions defined for Iowa. GLS regression analyses were performed using the WREG program (Eng and others, 2009). GLS regression, as described by Stedinger and Tasker (1985),

Tasker and Stedinger (1989), and Griffis and Stedinger (2007), is a method that weights streamgages in the regression according to differences in streamflow reliability (record lengths) and variability (record variance), and according to spatial cross correlations of concurrent streamflow among streamgages. Compared to OLS regression, GLS regression provides improved estimates of streamflow statistics and improved estimates of the predictive accuracy of the regression equations (Stedinger and Tasker, 1985). The correlation smoothing function used by the WREG program to compute a weighting matrix for the 176 streamgages included in the development of the GLS regression equation for estimating AEPDs for flood region 2 with 30 years of concurrent flow is shown in figure 9. The smoothing function relates the correlation between annual peak-discharge time series at two streamgages to the geographic distance between the streamgages for every paired combination of the 176 streamgages with 30 years of concurrent flow. Strong evidence of cross correlation is shown in figure 9 because of the abundance of paired points for 30 years of concurrent flow that form the tail extending towards the right side of the graph. Final GLS regression models were selected primarily on the basis of minimizing values of the standard error of model (SEM) and the SEP, and maximizing values of the pseudo coefficient of determination (pseudo- $R^2$ ).

**Table 9.** Regression equations for estimating annual exceedance-probability discharges for unregulated streams in flood region 1 of Iowa.

[SEP, average standard error of prediction; Pseudo-R<sup>2</sup>, pseudo coefficient of determination; SEM, average standard error of model; AVP, average variance of prediction; %, percent; DRNAREA, geographic information system drainage area; I24H10Y, maximum 24-hour precipitation that happens on average once in 10 years; CCM, constant of channel maintenance]

| Annual exceedance-probability equation  | SEP (percent) | Pseudo-R <sup>2</sup> (percent) | SEM (percent) | AVP (log ft <sup>3</sup> /s) <sup>2</sup> |
|---|---------------|---------------------------------|---------------|---|
| 91 streamgages used to develop equations  |               |                                 |               |   |
| $Q_{50\%} = DRNAREA^{0.675} 10^{(-0.355+0.601 \times I24H10Y-0.595 \times CCM^{0.55})}$ | 41.6          | 94.8                            | 39.4          | 0.030                                     |
| $Q_{20\%} = DRNAREA^{0.626} 10^{(0.094+0.594 \times I24H10Y-0.609 \times CCM^{0.55})}$  | 32.6          | 96.2                            | 30.4          | 0.019                                     |
| $Q_{10\%} = DRNAREA^{0.604} 10^{(0.340+0.588 \times I24H10Y-0.636 \times CCM^{0.55})}$  | 31.8          | 96.2                            | 29.4          | 0.018                                     |
| $Q_{4\%} = DRNAREA^{0.585} 10^{(0.604+0.579 \times I24H10Y-0.679 \times CCM^{0.55})}$   | 33.2          | 95.6                            | 30.4          | 0.020                                     |
| $Q_{2\%} = DRNAREA^{0.574} 10^{(0.769+0.573 \times I24H10Y-0.712 \times CCM^{0.55})}$   | 35.6          | 94.7                            | 32.6          | 0.023                                     |
| $Q_{1\%} = DRNAREA^{0.566} 10^{(0.917+0.567 \times I24H10Y-0.742 \times CCM^{0.55})}$   | 38.0          | 93.8                            | 34.7          | 0.025                                     |
| $Q_{0.5\%} = DRNAREA^{0.559} 10^{(1.06+0.560 \times I24H10Y-0.771 \times CCM^{0.55})}$  | 41.0          | 92.6                            | 37.5          | 0.029                                     |
| $Q_{0.2\%} = DRNAREA^{0.553} 10^{(1.22+0.550 \times I24H10Y-0.808 \times CCM^{0.55})}$  | 45.2          | 90.8                            | 41.5          | 0.035                                     |

**Table 10.** Regression equations for estimating annual exceedance-probability discharges for unregulated streams in flood region 2 of Iowa.

[SEP, average standard error of prediction; Pseudo-R<sup>2</sup>, pseudo coefficient of determination; SEM, average standard error of model; AVP, average variance of prediction; %, percent; DRNAREA, geographic information system drainage area; DESMOIN, percent of area within Des Moines Lobe landform region; BSHAPE, measure of basin shape]

| Annual exceedance-probability equation  | SEP (percent) | Pseudo-R <sup>2</sup> (percent) | SEM (percent) | AVP (log ft <sup>3</sup> /s) <sup>2</sup> |
|---|---------------|---------------------------------|---------------|---|
| 176 streamgages used to develop equations   |               |                                 |               |   |
| $Q_{50\%} = 10^{(-49.0+51.2 \times DRNAREA^{0.005}) \times (DESMOIN+1)^{-0.069}}$                 | 46.8          | 91.5                            | 45.7          | 0.037                                     |
| $Q_{20\%} = 10^{(30.4-27.8 \times DRNAREA^{-0.009} -0.001 \times DESMOIN-0.016 \times BSHAPE)}$   | 25.7          | 96.7                            | 24.4          | 0.012                                     |
| $Q_{10\%} = 10^{(17.4-14.6 \times DRNAREA^{-0.017} -0.002 \times DESMOIN -0.019 \times BSHAPE)}$  | 20.8          | 97.7                            | 19.3          | 0.008                                     |
| $Q_{4\%} = 10^{(13.7-10.7 \times DRNAREA^{-0.023} -0.002 \times DESMOIN -0.022 \times BSHAPE)}$   | 19.4          | 97.9                            | 17.6          | 0.007                                     |
| $Q_{2\%} = 10^{(12.2-9.10 \times DRNAREA^{-0.027} -0.002 \times DESMOIN -0.024 \times BSHAPE)}$   | 20.4          | 97.6                            | 18.5          | 0.008                                     |
| $Q_{1\%} = 10^{(11.1-7.92 \times DRNAREA^{-0.031} -0.002 \times DESMOIN -0.025 \times BSHAPE)}$   | 22.3          | 96.9                            | 20.3          | 0.009                                     |
| $Q_{0.5\%} = 10^{(10.5-7.20 \times DRNAREA^{-0.034} -0.002 \times DESMOIN -0.026 \times BSHAPE)}$ | 24.9          | 96.0                            | 22.9          | 0.011                                     |
| $Q_{0.2\%} = 10^{(9.95-6.60 \times DRNAREA^{-0.037} -0.002 \times DESMOIN -0.028 \times BSHAPE)}$ | 28.2          | 94.7                            | 26.1          | 0.014                                     |

**Table 11.** Regression equations for estimating annual exceedance-probability discharges for unregulated streams in flood region 3 of Iowa.

[SEP, average standard error of prediction; Pseudo-R<sup>2</sup>, pseudo coefficient of determination; SEM, average standard error of model; AVP, average variance of prediction; %, percent; DRNAREA, geographic information system drainage area; KSATSSUR, average saturated hydraulic conductivity of soil; BSHAPE, measure of basin shape]

| Annual exceedance-probability equation  | SEP (percent) | Pseudo-R <sup>2</sup> (percent) | SEM (percent) | AVP (log ft <sup>3</sup> /s) <sup>2</sup> |
|---|---------------|---------------------------------|---------------|---|
| 127 streamgages used to develop equations   |               |                                 |               |   |
| $Q_{50\%} = DRNAREA^{0.600} \times 10^{(2.43-0.009 \times KSATSSUR-0.017 \times BSHAPE)}$         | 43.1          | 92.4                            | 41.6          | 0.032                                     |
| $Q_{20\%} = 10^{(10.6-7.78 \times DRNAREA^{-0.035} -0.009 \times KSATSSUR-0.024 \times BSHAPE)}$  | 30.4          | 95.3                            | 29.0          | 0.017                                     |
| $Q_{10\%} = 10^{(8.14-5.17 \times DRNAREA^{-0.054} -0.008 \times KSATSSUR-0.028 \times BSHAPE)}$  | 27.0          | 96.0                            | 25.6          | 0.013                                     |
| $Q_{4\%} = 10^{(7.07-3.92 \times DRNAREA^{-0.073} -0.008 \times KSATSSUR-0.031 \times BSHAPE)}$   | 26.5          | 95.9                            | 24.8          | 0.013                                     |
| $Q_{2\%} = 10^{(6.65-3.39 \times DRNAREA^{-0.086} -0.009 \times KSATSSUR-0.033 \times BSHAPE)}$   | 27.8          | 95.3                            | 26.0          | 0.014                                     |
| $Q_{1\%} = 10^{(6.41-3.06 \times DRNAREA^{-0.097} -0.009 \times KSATSSUR-0.035 \times BSHAPE)}$   | 29.1          | 94.7                            | 27.2          | 0.015                                     |
| $Q_{0.5\%} = 10^{(6.23-2.80 \times DRNAREA^{-0.108} -0.009 \times KSATSSUR-0.037 \times BSHAPE)}$ | 30.5          | 94.1                            | 28.4          | 0.017                                     |
| $Q_{0.2\%} = 10^{(6.10-2.59 \times DRNAREA^{-0.119} -0.009 \times KSATSSUR-0.039 \times BSHAPE)}$ | 33.7          | 92.6                            | 31.5          | 0.020                                     |

## Final Regression Equations

Final regression equations developed for the three flood regions defined for Iowa are listed in tables 9–11, along with the number of streamgages included in each regression analysis and several performance metrics. A total of 394 streamgages were included in the development of the final regression equations for flood regions 1–3. StreamStats variable names are used for the explanatory variables in the final regression equations (tables 9–11); definitions of the variables and the units of measure are listed in tables 2 and 3. Six basin characteristics are used as explanatory variables in the final regression equations (table 12); these include three morphometric characteristics [DRNAREA, constant of channel maintenance (CCM), and a measure of basin shape (BSHAPE)], two geologic/pedologic characteristics [percent of basin within the Des Moines Lobe landform region (DESMOIN) and average saturated hydraulic conductivity of soil (KSATSSUR)], and one climatic characteristic [maximum 24-hour precipitation that happens on average once in 10 years (I24H10Y)] (table 2, link to Excel file). GIS software is required to measure the basin characteristics included as explanatory variables in the final regression equations. All explanatory variables included in the final regression equations were statistically significant at the 95-percent confidence level and were not correlated with other explanatory variables used in the same

equation. The performance metrics in tables 9–11 indicate the predictive accuracy of the final regression equations.

Data transformations were tested for each explanatory variable included in the final regression equations. As previously noted, logarithmic and power transformations were tested and used for DRNAREA. Power transformations for DRNAREA for flood regions 2 and 3 are negative for all annual exceedance probabilities, with the exception of  $Q_{50\%}$  (tables 10 and 11). Logarithmic and nonlogarithmic data transformations were tested for each of the other explanatory variables in the regression equations (tables 9–11) to linearize the relation between the response variable and the explanatory variables. For flood region 1, an exponent of 0.55 for CCM was determined to provide the best data transformation. A range of exponent values were tested for the transformation of CCM and the relation between  $\log Q_{1\%}$  and CCM was plotted for each exponent value tested. The data transformation that visually provided the most linear relation for  $Q_{1\%}$  and  $Q_{10\%}$  was then modeled with OLS and GLS regression to verify the selected data transformation also was providing the best predictive accuracy.

For the 24 GLS regression equations developed for estimating AEPDs for flood regions 1–3, an SEP (in percent), a pseudo- $R^2$  (in percent), an SEM (in percent), and an average variance of prediction [AVP, in  $(\log \text{ft}^3/\text{s})^2$  units] are reported in tables 9–11. SEP represents the sum of the model error and the sampling error. SEP is the square root of the GLS

**Table 12.** Range of basin-characteristic values used to develop annual exceedance-probability regression equations for unregulated streams in Iowa.

[GIS, geographic information system; DRNAREA, GIS drainage area; I24H10Y, maximum 24-hour precipitation that happens on average once in 10 years; CCM, constant of channel maintenance;  $\text{mi}^2$ , square mile; mi, mile; DESMOIN, percent of area within Des Moines Lobe landform region; BSHAPE, measure of basin shape, dimensionless; KSATSSUR, average saturated hydraulic conductivity of soil;  $\mu\text{m/s}$ , micrometer per second; NA, not applicable, basin characteristic not used to develop regional regression equations]

|                 | GIS drainage area,<br>DRNAREA ( $\text{mi}^2$ ) | I24H10Y (inches) | CCM ( $\text{mi}^2/\text{mi}$ ) | DESMOIN<br>(percent) | BSHAPE | KSATSSUR<br>( $\mu\text{m/s}$ ) |
|-----------------|---|------------------|---------------------------------|----------------------|--------|---------------------------------|
| Region 1        |   |                  |                                 |                      |        |                                 |
| Minimum         | 0.06  | 3.580            | 0.110                           | NA                   | NA     | NA                              |
| Maximum         | 5,463.88  | 4.500            | 2.220                           | NA                   | NA     | NA                              |
| Mean            | 353.74  | 4.175            | 0.906                           | NA                   | NA     | NA                              |
| Median          | 40.72   | 4.115            | 0.875                           | NA                   | NA     | NA                              |
| Number of sites | 91  | 91               | 91                              | NA                   | NA     | NA                              |
| Region 2        |   |                  |                                 |                      |        |                                 |
| Minimum         | 0.08  | NA               | NA                              | 0                    | 0.806  | NA                              |
| Maximum         | 7,782.62  | NA               | NA                              | 100.000              | 13.941 | NA                              |
| Mean            | 467.84  | NA               | NA                              | 8.175                | 3.826  | NA                              |
| Median          | 45.55   | NA               | NA                              | 0                    | 3.303  | NA                              |
| Number of sites | 176   | NA               | NA                              | 176                  | 176    | NA                              |
| Region 3        |   |                  |                                 |                      |        |                                 |
| Minimum         | 0.05  | NA               | NA                              | NA                   | 0.339  | 1.883                           |
| Maximum         | 2,809.05  | NA               | NA                              | NA                   | 13.523 | 33.572                          |
| Mean            | 323.29  | NA               | NA                              | NA                   | 4.514  | 7.867                           |
| Median          | 52.26   | NA               | NA                              | NA                   | 4.019  | 6.153                           |
| Number of sites | 127   | NA               | NA                              | NA                   | 127    | 127                             |

AVP (Tasker and Stedinger, 1989; Eng and others, 2009). The pseudo- $R^2$  is a measure of the percentage of the variation explained by the basin characteristics (explanatory variables) included in the model. The pseudo- $R^2$  value is calculated on the basis of the degrees of freedom in the regression. Griffis and Stedinger (2007) describe how the pseudo- $R^2$  is more appropriate than the traditional  $R^2$  or adjusted  $R^2$  in measuring the true variation explained by the explanatory variables in the GLS model. SEM measures the error of the model itself and does not include sampling error. SEM is the square root of the GLS model error variance (MEV) (Tasker and Stedinger, 1989). AVP is a measure of the average accuracy of prediction for all sites used in the development of the regression model and assumes that the explanatory variables for the streamgages included in the regression analysis are representative of all streamgages in the region (Verdi and Dixon, 2011).

The performance metrics pseudo- $R^2$  and SEM indicate how well the equations perform on the streamgages used in the regression analyses. SEP and AVP are measures of the accuracy that GLS regression models can predict AEPDs at ungaged sites. The same explanatory variables were used to develop all eight annual exceedance-probability equations for each region to minimize the possibility of predictive inconsistencies between estimates of different probabilities, so that estimates will increase with decreasing probabilities. For example, maintaining the same regression-model form (same explanatory variables) helps to maximize the chances an estimate for the  $Q_{1\%}$  flood discharge is greater than an estimate for the  $Q_{2\%}$  flood discharge, an estimate for the  $Q_{2\%}$  flood discharge is greater than an estimate for the  $Q_{4\%}$  flood discharge, and so forth for all eight exceedance probabilities. The one exception to maintaining the same regression-model form for all eight annual exceedance probabilities is the  $Q_{50\%}$  regression equation for flood region 2. A two-variable regression model was developed instead of a three-variable model for this annual exceedance probability because the basin characteristic BSHAPE was not a statistically significant explanatory variable at the 95-percent confidence level and the three explanatory variables DRNAREA, DESMOIN, and BSHAPE provided the best GLS regression model for the seven other annual exceedance probabilities. Flood-discharge estimates computed for streamgages using the two-variable  $Q_{50\%}$  and three-variable  $Q_{20\%}$  regression equations for flood region 2 listed in table 10 indicated that  $Q_{20\%}$  estimates should exceed  $Q_{50\%}$  estimates.

Output from the WREG program (Eng and others, 2009) for GLS regression models identifies streamgages that are possible outliers in the dataset as plotted points or tabulated values that exceed a leverage threshold value or an influence threshold value. Leverage points (Eng and others, 2009) are outliers that may unduly influence the estimation of regression constants and are a measure of how large or small explanatory variables (basin characteristics) are for a specific streamgage as compared to the centroid of values of the same explanatory variables at all other streamgages. Influence points are measured by Cook's D statistic (Cook, 1977; Helsel and Hirsch,

2002), and these are outliers that have unduly influenced the estimation of regression constants. As previously noted in an earlier section Definition of Flood Regions, eight streamgages identified by the WREG program as outliers were removed from the regression datasets because of significant trends, urbanization, or channelization (table 7).

DRNAREA, drainage area, is the most significant explanatory variable for each set of RREs and is related positively to AEPDs; ungaged basins with larger drainage areas will produce greater estimates of AEPDs and ungaged basins with smaller drainage areas will produce lower estimates of AEPDs. The second RRE variable for flood region 1, I24H10Y, also is related positively to AEPDs. I24H10Y values for flood region 1 decrease in a northerly direction and basins within the southern part of the Des Moines Lobe will produce greater estimates of AEPDs than basins within the northern part of the Des Moines Lobe given the same values for DRNAREA and CCM. The third RRE variable for flood region 1, CCM, is related negatively to AEPDs. CCM is a measure of drainage density calculated as a ratio of drainage area divided by the total length of all mapped streams (1:24,000-scale) in the basin. Because most of flood region 1 is within the Des Moines Lobe landform region, an area where surface-water drainage typically is defined poorly, watersheds with fewer miles of stream length per square mile of drainage area will result in larger values for CCM and lower estimates of AEPDs. The second RRE variable for flood region 2, DESMOIN, is related negatively to AEPDs. Basins with areas within the Des Moines Lobe landform region will produce lower estimates of AEPDs than basins outside the Des Moines Lobe or than basins with a smaller percentage of area within the Des Moines Lobe (fig. 3). The third RRE variable for flood regions 2 and 3, BSHAPE, also is related negatively to AEPDs. BSHAPE is calculated as a ratio of basin length squared divided by drainage area. The more elongated the shape of a basin, the larger the value for BSHAPE and the longer the lag-time for tributary flows to peak at the basin outlet, which will produce lower estimates of AEPDs; conversely, the more rounded the shape of the basin, the smaller the value for BSHAPE and the shorter the lag-time for tributary flows to peak at the basin outlet, which will produce greater estimates of AEPDs. The second RRE variable for flood region 3, KSATSSUR, is related negatively to AEPDs. KSATSSUR, or saturated hydraulic conductivity, is a quantitative measure of a saturated soil's ability to transmit water when subjected to a hydraulic gradient, which can be thought of as the ease with which pores of a saturated soil permit water movement (<http://soils.usda.gov/technical/technotes/note6.html>). Basins with smaller average values of KSATSSUR have lower saturated hydraulic-conductivity and infiltration rates resulting in more surface-water runoff and larger estimates of AEPDs; conversely, basins with larger average values of KSATSSUR have higher saturated hydraulic-conductivity and infiltration rates leading to less surface-water runoff and lower estimates of AEPDs.

## Accuracy and Limitations of Regression Equations

The RREs developed in this study only apply to stream sites in Iowa where peak discharges are not affected significantly by regulation, diversion, channelization, backwater, or urbanization. The applicability and accuracy of the regional equations depend on if the basin characteristics measured for an ungaged stream site are within the range of the characteristic values used to develop the regression equations. The acceptable range of basin-characteristic values used to develop each RRE (tables 9–11) are tabulated as minimum and maximum values in table 12. The applicability of the regional equations is unknown when any characteristic value measured for an ungaged site is outside the acceptable range. In addition, basin-characteristic measurements at ungaged sites should be computed using the same GIS datasets and measurement methods used in this study; the USGS StreamStats Web-based GIS tool includes the same GIS data layers and measurement methods as used to develop the regression equations in this study.

The AEPD regression equations presented in this report should be used with caution for ungaged stream sites with basin-characteristic values approaching the minimum or maximum limits (table 12) because the predictive errors of the equations increase with distance from the mean or median values of the explanatory variables and inconsistencies in the estimates may result. For different annual exceedance probabilities, the AEPD estimate for a larger probability may be greater than the AEPD estimate for a smaller probability; for example, a  $Q_{2\%}$  flood discharge estimate may be greater than a  $Q_{1\%}$  flood discharge estimate. Although no inconsistencies in RRE estimates resulted for any of the eight AEPDs for the 394 streamgages listed in [table 4](#) (link to Excel file), there is the possibility that inconsistencies in RRE estimates may result for ungaged sites. If inconsistencies in RRE estimates are obtained for an ungaged stream site, a comparison of all AEPDs for the site and a check of streamgage data or other published data may help to determine which AEPD is inconsistent.

In general, predictive accuracies tend to be the best for flood region 2, second best for flood region 3, and poorest for flood region 1. For the AEPD equations, SEP for flood region 1 ranges from 31.8 to 45.2 percent (table 9), SEP for flood region 2 ranges from 19.4 to 46.8 percent (table 10), and SEP for flood region 3 ranges from 26.5 to 43.1 percent (table 11). The percentage of variation in the response variables explained by the explanatory variables (pseudo- $R^2$ ) for the AEPD equations developed for flood region 1 ranges from 90.8 to 96.2 percent, for flood region 2 ranges from 91.5 to 97.9 percent, and for flood region 3 ranges from 92.4 to 96.0 percent. Of the eight AEPD equations developed for each region, the  $Q_{10\%}$ ,  $Q_{4\%}$ , and  $Q_{2\%}$  flood-discharge regression equations generally have the best predictive accuracy and the  $Q_{50\%}$  and  $Q_{0.2\%}$  flood-discharge equations generally have the poorest accuracy. The natural variability of streamflow may be

an important factor associated with the predictive accuracy of AEPD equations. Estimation of AEPDs that have greater variability will have poorer predictive accuracies than estimation of AEPDs with less variability.

The regression equations also should be used with caution for streams within the Mississippi River and Missouri River Alluvial Plains landform regions (fig. 3) because streamgage data representing these landform regions were not included in the development of the regression equations. Although two streamgages are located within the Missouri River Alluvial Plain landform region, West Fork Ditch at Hornick, Iowa (streamgage 06602020, map number 394, fig. 1), and Little Sioux River near Kennebec, Iowa (streamgage 06606700, map number 407, fig. 1), neither streamgage is considered representative of this landform region because nearly the entire watershed for each streamgage is located outside of the Missouri River Alluvial Plain landform region.

Because the precision of response- and explanatory-variable data used to develop the equations was often limited to three significant figures, AEPDs estimated from the regression equations also should be limited to three significant figures.

Figure 10 shows the relation between observed and predicted flood discharges for  $Q_{1\%}$  for each of the three flood regions. The uncertainty of regression estimates can be seen graphically as a greater scatter of observed in relation to predicted points along the 1:1 line.

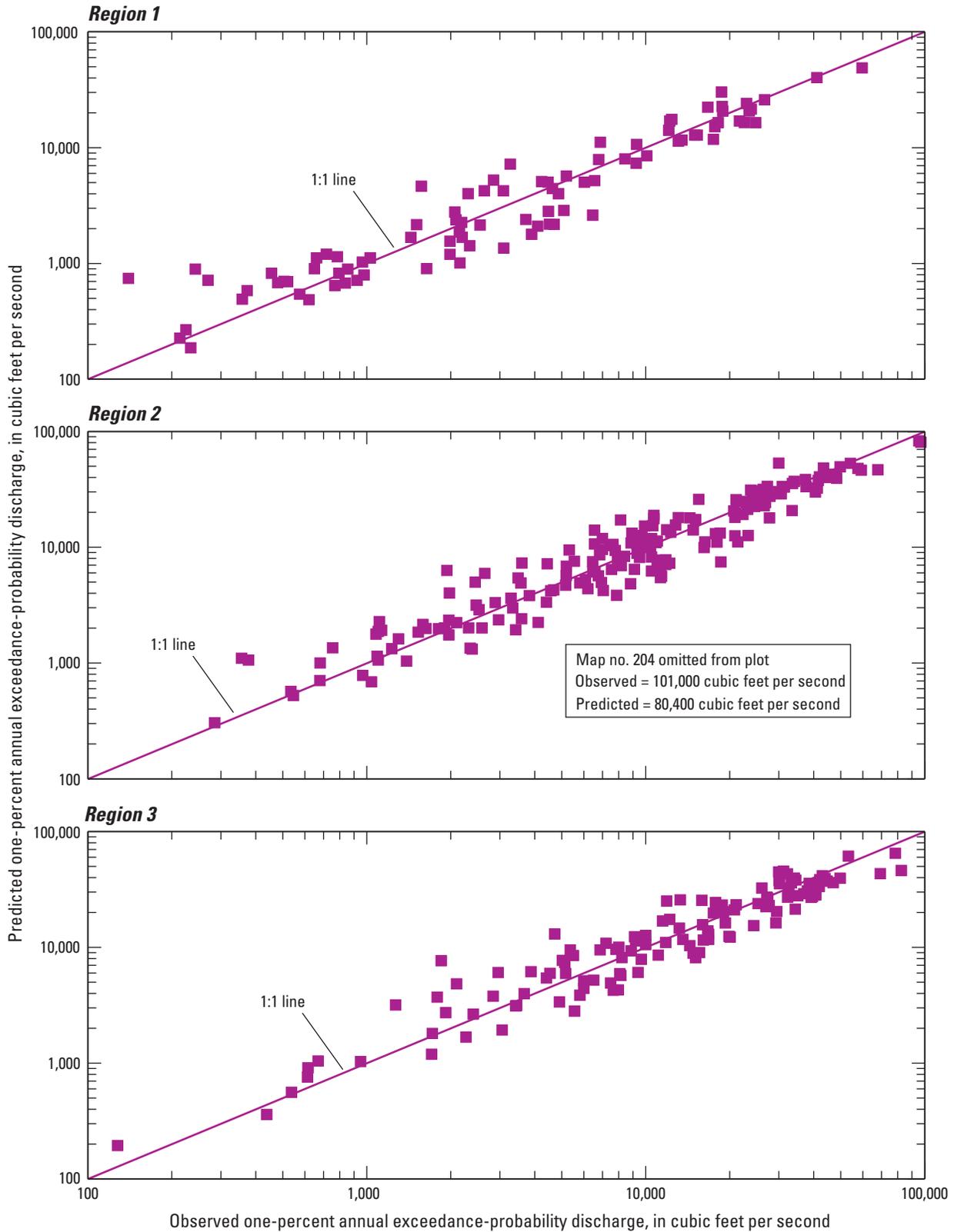
## Prediction Intervals

Although regression equations presented in tables 9–11 can be used to estimate AEPDs, the true values of the AEPDs are unknown. A measure of the uncertainty associated with the regression estimate of an AEPD is the prediction interval. The interval is the estimated value plus or minus a given margin of error. A prediction interval is the probability that the actual value of the estimated AEPD will be within this margin of error (Helsel and Hirsch, 2002). The prediction interval determines the range of discharge values predicted for selected statistics given a confidence level and the SEP. For a 90-percent prediction interval, the true AEPD has a 90-percent probability of being within the margin of error. The USGS StreamStats Web-based GIS tool (<http://water.usgs.gov/osw/streamstats/index.html>) uses the 90-percent prediction interval estimates as part of the computation of AEPD estimates for ungaged stream sites. The following equation, modified from Tasker and Driver (1988), can be used to compute the 90-percent prediction interval for the true value of an AEPD for an ungaged site:

$$\frac{Q}{T} < Q < QT \quad (6)$$

where

$Q$  is the AEPD predicted for the ungaged site from the regression equation, and the



**Figure 10.** Relation between one-percent annual exceedance-probability discharges computed from observed streamflow and those predicted from regression equations for flood regions in Iowa.

following equation can be used to compute  $T$ :

$$T = 10^{[t_{(\alpha/2, n-p)} S_i]} \tag{7}$$

where

- $t_{(\alpha/2, n-p)}$  is the critical value,  $t$ , from the student's  $t$ -distribution at alpha level  $\alpha$  [ $\alpha=0.10$  for 90-percent prediction intervals, critical values may be obtained in many statistics textbooks, Iman and Conover (1983), or from the World Wide Web];
- $n-p$  is the degrees of freedom with  $n$  streamgages included in the regression analysis and  $p$  parameters in the equation (the number of explanatory variables plus one); and
- $S_i$  is the standard error of prediction for site  $i$ , and the following equation can be used to compute  $S_i$ :

$$S_i = [MEV + X_i U X_i']^{0.5} \tag{8}$$

where

- $MEV$  is the model error variance from GLS regression equations developed in this study;
- $X_i$  is the row vector for the streamgage  $i$ , starting with the number 1, followed by the logarithmic values of the basin characteristics used in the regression;
- $U$  is the covariance matrix for the regression coefficients; and
- $X_i'$  is the matrix algebra transpose of  $X_i$  (Ludwig and Tasker, 1993; Ries and Friesz, 2000).

Similar to the SEP,  $S_i$  represents the sum of the model error and the sampling error for a single site  $i$ . The  $X_i U X_i'$  term in equation 8 also is referred to as the sampling error variance. The values of  $t_{(\alpha/2, n-p)}$  and  $U$  needed to determine prediction intervals for estimates obtained by the regression equations in tables 9–11 are presented in [table 13](#) (link to Excel file).

### Application of Regression Equations

Methods for applying the RREs listed in tables 9–11 are described in the following examples:

#### Example 1

This example is a calculation of the  $Q_{0.2\%}$  flood discharge (500-year recurrence-interval flood discharge) for a stream site in flood region 1. Figure 1 shows the location of the streamgage 05482500 North Raccoon River near Jefferson, Iowa, as map number 283. This watershed is located entirely within flood region 1. Using the USGS StreamStats Web-based GIS tool, DRNAREA (drainage area) is measured as

1,609.35 mi<sup>2</sup>, I24H10Y (maximum 24-hour precipitation that happens on average once in 10 years) is measured as 4.321 inches, and CCM (constant of channel maintenance) is measured as 1.067 square miles per mile (mi<sup>2</sup>/mi) ([table 3](#), link to Excel file). Because all three basin-characteristic values are within the range of values listed in table 12, the GLS regression equation is applicable for estimating the  $Q_{0.2\%}$  flood discharge. The GLS regression equation for estimating the  $Q_{0.2\%}$  flood discharge from table 9 is

$$Q_{0.2\%} = DRNAREA^{0.553} 10^{(1.22+0.550 \times I24H10Y - 0.808 \times CCM^{0.55})}$$

$$Q_{0.2\%} = 1,609.35^{0.553} 10^{(1.22+0.550 \times 4.321 - 0.808 \times 1.067^{0.55})}$$

$$Q_{0.2\%} = 59.33 \times 10^{(1.22+2.377-0.8373)}$$

$$Q_{0.2\%} = 34,100 \text{ ft}^3/\text{s}$$

To calculate a 90-percent prediction interval for this  $Q_{0.2\%}$  flood-discharge estimate using equation 6, the  $X_i$  vector is  $X_i = \{1, \log_{10}(1,609.35), 4.321, 1.067^{0.55}\}$ , the MEV from [table 13](#) (link to Excel file) is 0.029998, and the following table lists the covariance matrix ( $U$ ):

|           | Intercept    | DRNAREA      | I24H10Y      | CCM          |
|-----------|--------------|--------------|--------------|--------------|
| Intercept | 0.579267620  | -0.000733524 | -0.141069630 | 0.014150952  |
| DRNAREA   | -0.000733524 | 0.000744762  | 0.000042403  | -0.001270299 |
| I24H10Y   | -0.141069630 | 0.000042403  | 0.035474003  | -0.007395614 |
| CCM       | 0.014150952  | -0.001270299 | -0.007395614 | 0.020494642  |

Using matrix algebra, the product of  $X_i U X_i'$  is determined in two steps: (1) by multiplying  $X_i'$  (the transpose of  $X_i$ ) by the covariance matrix,  $U$ , to obtain  $U X_i'$ ; and (2) by multiplying  $U X_i'$  by  $X_i$ . In this example, the value of  $X_i U X_i'$  is 0.00327303.

The standard error of prediction for this site as computed from equation 8 is

$$S_i = [0.029998 + 0.00327303]^{0.5} = 0.182403,$$

and  $T$  from equation 7 is

$$T = 10^{(1.6626)(0.182403)} = 2.0103,$$

where the critical value ( $t_{(\alpha/2, n-p)}$ ) from the student's  $t$ -distribution for the 90-percent prediction interval is 1.6626 ([table 13](#), link to Excel file).

The 90-percent prediction interval is estimated from equation 6 as

$$34,100/2.0103 < Q_{0.2\%} < (34,100)(2.0103), \text{ or}$$

$$17,000 < Q_{0.2\%} < 68,600 \text{ ft}^3/\text{s}.$$

#### Example 2

This example is a calculation of the  $Q_{1\%}$  flood discharge (100-year recurrence-interval flood discharge) for a stream site in flood region 2. Figure 1 shows the location of the streamgage 05455500 English River at Kalona, Iowa, as map number 181. This watershed is located entirely within flood region 2. Using the USGS StreamStats Web-based GIS tool, DRNAREA (drainage area) is measured as 574.10 mi<sup>2</sup>, DESMOIN (percent area of basin within the Des Moines Lobe

landform region) is measured as 0.000 percent, and BSHAPE (measure of basin shape) is measured as 6.155 (table 3, link to Excel file). Because all three basin-characteristic values are within the range of values listed in table 12, the GLS regression equation is applicable for estimating the  $Q_{1\%}$  flood discharge. The GLS regression equation for estimating the  $Q_{1\%}$  flood discharge from table 10 is

$$Q_{1\%} = 10^{(11.1-7.92 \times \text{DRNAREA}^{-0.031} - 0.002 \times \text{DESMOIN} - 0.025 \times \text{BSHAPE})}$$

$$Q_{1\%} = 10^{(11.1-7.92 \times 574.10^{-0.031} - 0.002 \times 0.000 - 0.025 \times 6.155)}$$

$$Q_{1\%} = 10^{(11.1-6.504 - 0.000 - 0.1539)}$$

$$Q_{1\%} = 27,700 \text{ ft}^3/\text{s}$$

To calculate a 90-percent prediction interval for this  $Q_{1\%}$  flood-discharge estimate using equation 6, the  $X_i$  vector is

$X_i = \{1, 574.10^{-0.031}, 0.000, 6.155\}$ , the MEV from table 13 (link to Excel file) is 0.007617, and the following table lists the covariance matrix ( $U$ ):

|           | Intercept    | DRNAREA      | DESMOIN      | BSHAPE       |
|-----------|--------------|--------------|--------------|--------------|
| Intercept | 0.041672405  | -0.045784820 | -0.000039051 | -0.000518851 |
| DRNAREA   | -0.045784820 | 0.052558399  | 0.000038686  | 0.000436101  |
| DESMOIN   | -0.000039051 | 0.000038686  | 0.000000287  | 0.000000161  |
| BSHAPE    | -0.000518851 | 0.000436101  | 0.000000161  | 0.000025378  |

Using matrix algebra, the product of  $X_i U X_i'$  is determined in two steps: (1) by multiplying  $X_i'$  (the transpose of  $X_i$ ) by the covariance matrix,  $U$ , to obtain  $U X_i'$ ; and (2) by multiplying  $U X_i'$  by  $X_i$ . In this example, the value of  $X_i U X_i'$  is 0.000902081.

The standard error of prediction for this site as computed from equation 8 is

$$S_i = [0.007617 + 0.000902081]^{0.5} = 0.0922989, \text{ and } T$$

from equation 7 is

$$T = 10^{(1.6538)(0.0922989)} = 1.4212,$$

where the critical value ( $t_{(\alpha/2, n-p)}$ ) from the student's  $t$ -distribution for the 90-percent prediction interval is 1.6538 (table 13, link to Excel file).

The 90-percent prediction interval is estimated from equation 6 as

$$27,700/1.4212 < Q_{1\%} < (27,700)(1.4212), \text{ or} \\ 19,500 < Q_{1\%} < 39,400 \text{ ft}^3/\text{s}.$$

## Weighted Method to Estimate Annual Exceedance-Probability Discharges for Streamgages

The Interagency Advisory Committee on Water Data (1982) recommends that improved estimates of AEPDs at streamgages can be obtained by weighting the annual exceedance-probability LP3 estimate (EMA/MGB) with the RRE estimate using the variance of prediction for each of these two

estimates. The variance of prediction can be thought of as a measure of the uncertainty in either the EMA/MGB estimate or the RRE estimate. If the two estimates are assumed to be independent and are weighted inversely proportional to their associated variances, the variance of the weighted estimate will be less than the variance of either of the independent estimates. Optimal weighted estimates of AEPDs were computed for this study using the Weighted Independent Estimates (WIE) computer program available at <http://water.usgs.gov/usgs/osw/swstats/freq.html>. Information on this computer program is presented by Cohn and others (2012).

The variance of prediction corresponding to the EMA/MGB estimate from the LP3 analysis is computed using the asymptotic formula given in Cohn and others (2001) with the addition of the MSE of the generalized skew (Griffis and others, 2004). This variance varies as a function of the length of record, the fitted LP3 distribution parameters (mean, standard deviation, and weighted skew), and the accuracy of the procedure used to determine the regional skew component of the weighted skew (Verdi and Dixon, 2011). The variance of prediction for the EMA/MGB LP3 estimate generally decreases with increasing record length and the quality of the LP3 distribution fit. The variance of prediction values for the EMA/MGB LP3 estimates for 394 streamgages included in this study are listed in table 14 (link to Excel file). The variance of prediction from the RREs is a function of the regression equations and the values of the explanatory variables (basin characteristics) used to compute the AEPDs from the regression equations. This variance generally increases as the values of the explanatory variables move further from the mean or median values of the explanatory variables. The variance of prediction values for the RREs used in this study also are listed in table 14 (link to Excel file).

Once the variances have been computed, the two independent discharge estimates can be weighted using the following equation (Verdi and Dixon, 2011; Cohn and others, 2012; Gotvald and others, 2012):

$$\log Q_{P(g)w} = \frac{VP_{P(g)r} \log Q_{P(g)s} + VP_{P(g)s} \log Q_{P(g)r}}{VP_{P(g)s} + VP_{P(g)r}}, \quad (9)$$

where

- $Q_{P(g)w}$  is the WIE estimate of flood discharge for the selected P-percent annual exceedance probability for a streamgage, g, in cubic feet per second;
- $VP_{P(g)r}$  is the variance of prediction at the streamgage derived from the applicable RREs for the selected P-percent annual exceedance probability (from table 14, link to Excel file), in log units;
- $Q_{P(g)s}$  is the estimate of flood discharge at the streamgage from the EMA/MGB LP3 analysis for the selected P-percent annual exceedance probability (from table 4, link to Excel file), in cubic feet per second;
- $VP_{P(g)s}$  is the variance of prediction at the streamgage

from the EMA/MGB LP3 analysis for the selected P-percent annual exceedance probability (from [table 14](#), link to Excel file), in log units; and

$Q_{P(g)r}$  is the flood-discharge estimate for the selected P-percent annual exceedance probability at the streamgage derived from the applicable RREs (from [table 4](#), link to Excel file), in cubic feet per second.

When the variance of prediction corresponding to one of the estimates is high, the uncertainty also is high, for which the weight of the estimate is relatively small. Conversely, when the variance of the prediction is low, the uncertainty also is low, for which the weight is correspondingly large. The variance of prediction associated with the weighted estimate,  $VP_{P(g)w}$ , is computed using the following equation (Verdi and Dixon, 2011; Gotvald and others, 2012):

$$VP_{P(g)w} = \frac{VP_{P(g)s} VP_{P(g)r}}{VP_{P(g)s} + VP_{P(g)r}}, \quad (10)$$

[Table 4](#) (link to Excel file) lists the improved AEPDs that were weighted using equation 9 along with the variance of prediction values from [table 14](#) (link to Excel file) for 394 streamgages included in this study.

### Example 3

This example is a calculation of a weighted estimate of the  $Q_{1\%}$  flood discharge (100-year recurrence-interval flood discharge) for a discontinued streamgage in flood region 3 with only 15 years of annual peak-discharge record available for computing AEPDs using an EMA/MGB LP3 analysis. Figure 1 shows the location of the streamgage 06610520 Mosquito Creek near Earling, Iowa, as map number 425. This watershed is located entirely within flood region 3. The estimate for the  $Q_{1\%}$  flood discharge from the EMA/MGB LP3 analysis is 14,400 ft<sup>3</sup>/s and from the RRE is 10,400 ft<sup>3</sup>/s ([table 4](#), link to Excel file). The variance of prediction from the EMA/MGB LP3 analysis is 0.0160 and from the RRE is 0.0146 ([table 14](#), link to Excel file). A WIE estimate is calculated for this streamgage using equation 9 as

$$\begin{aligned} \log Q_{P(g)w} &= \frac{VP_{P(g)r} \log Q_{P(g)s} + VP_{P(g)s} \log Q_{P(g)r}}{VP_{P(g)s} + VP_{P(g)r}}, \\ \log Q_{P(g)w} &= \frac{0.0146 \log 14,400 + 0.0160 \log 10,400}{0.0160 + 0.0146}, \\ \log Q_{P(g)w} &= 4.084 \text{ or } Q_{P(g)w} = 12,100 \text{ ft}^3/\text{s}. \end{aligned}$$

The weighted variance is calculated for this streamgage using equation 10 as

$$\begin{aligned} VP_{P(g)w} &= \frac{VP_{P(g)s} VP_{P(g)r}}{VP_{P(g)s} + VP_{P(g)r}}, \\ VP_{P(g)w} &= \frac{0.0160 \times 0.0146}{0.0160 + 0.0146}, \end{aligned}$$

$$VP_{P(g)w} = 0.0076.$$

Because of the short record for the streamgage in this example, the variance of the RRE estimate is slightly lower than the variance of the EMA/MGB LP3 estimate and slightly more weight is given to the RRE estimate in the WIE calculation in equation 9 than to the EMA/MGB LP3 estimate. The variance of prediction is lower for the WIE estimate than for either the EMA/MGB LP3 or RRE estimates indicating that the uncertainty in the estimate of the  $Q_{1\%}$  flood discharge is reduced.

## Weighted Methods to Estimate Annual Exceedance-Probability Discharges for Ungaged Sites on Gaged Streams

AEPDs at ungaged sites located on gaged streams can be improved by weighting AEPDs from a nearby streamgage. Two methods for weighting AEPDs from a nearby streamgage are applicable. Both methods require the measurement of drainage area (DRNAREA) for the ungaged site and a weighted AEPD  $Q_{P(g)w}$  from a nearby streamgage (see equation 9 in previous section). The first weighting method, presented in the following section Regression-Weighted Estimates for Ungaged Sites on Gaged Streams, requires a regional regression estimate for the ungaged site. The second weighting method, presented in the following section Area-Weighted Estimates for Ungaged Sites on Gaged Streams, does not require a regional regression estimate for the ungaged site. AEPDs calculated from the regression-weighted method are considered to provide better predictive accuracies for ungaged sites than estimates calculated from the area-weighted method.

### Regression-Weighted Estimates for Ungaged Sites on Gaged Streams

Sauer (1974) presented the following regression-weighted method to improve AEPDs for an ungaged site near a streamgage, on the same stream, with 10 or more years of annual peak-discharge record. To obtain a regression-weighted AEPD ( $Q_{P(u)rw}$ ) for P-percent annual exceedance probability at the ungaged site, the WIE estimate for an upstream or downstream streamgage ( $Q_{P(g)w}$ ) must first be determined using equation 9 presented in the previous section. The regression-weighted AEPD for the ungaged site ( $Q_{P(u)w}$ ) is then computed using the following equation (Verdi and Dixon, 2011):

$$Q_{P(u)rw} = \left[ \left( \frac{2\Delta A}{A_{(g)}} \right) + \left( 1 - \frac{2\Delta A}{A_{(g)}} \right) \left( \frac{Q_{P(g)w}}{Q_{P(g)r}} \right) \right] Q_{P(u)r}, \quad (11)$$

where

$Q_{P(u)rw}$  is the regression-weighted estimate of flood discharge for the selected P-percent annual

- exceedance probability for the ungaged site,  $u$ , in cubic feet per second;
- $\Delta A$  is the absolute value of the difference between the drainage areas of the streamgage and the ungaged site, in square miles;
- $Q_{P(g)w}$  is described for equation 9 above;
- $A_{(g)}$  is the drainage area for the streamgage, in square miles;
- $Q_{P(g)r}$  is described for equation 9 above; and
- $Q_{P(u)r}$  is the flood-discharge estimate derived from the applicable RREs for the selected P-percent annual exceedance probability at the ungaged site, in cubic feet per second.

Use of equation 11 gives full weight to the regression equation estimates when the drainage area for the ungaged site is equal to 0.5 or 1.5 times the drainage area for the streamgage, and increasing weight to the streamgage estimates as the drainage area ratio approaches 1. The regression-weighted method is not applicable when the drainage area ratio for the ungaged site and streamgage is less than 0.5 or greater than 1.5 (Verdi and Dixon, 2011; Gotvald and others, 2012). The regression-weighted method is not implemented in StreamStats (Ries and others, 2008), but this method can easily be calculated outside of StreamStats after using StreamStats to compute  $Q_{P(u)r}$ . The StreamStats computation of  $Q_{P(u)r}$  provides the drainage area for the ungaged site from which  $\Delta A$  can be calculated for use in equation 11.

### Example 4

This example is a calculation of a regression-weighted estimate for the  $Q_{2\%}$  flood discharge (50-year recurrence-interval flood discharge) statistic for a stream site in flood region 1. Streamgage 06605600 Little Sioux River at Gillett Grove is shown on figure 1 as map number 402; this streamgage will be assumed to be an ungaged site for this example. Another streamgage 06605850 Little Sioux River at Linn Grove, also is shown on figure 1 as map number 404, which is located downstream on the same stream; this site will be used as the streamgage in this example. This watershed is located entirely within flood region 1. Seven steps are required to calculate a regression-weighted estimate using equation 11:

$$Q_{2\%(u)rww} = \left[ \left( \frac{2\Delta A}{A_{(g)}} \right) + \left( 1 - \frac{2\Delta A}{A_{(g)}} \right) \left( \frac{Q_{2\%(g)w}}{Q_{2\%(g)r}} \right) \right] Q_{2\%(u)r}$$

The first step is to use the USGS StreamStats Web-based GIS tool to measure basin characteristics for the ungaged site (map number 402, fig. 1). DRNAREA (drainage area) is measured as 1,352.59 mi<sup>2</sup>, I24H10Y (maximum 24-hour precipitation that happens on average once in 10 years) is measured as 4.088 inches, and CCM (constant of channel maintenance) is measured as 0.840 mi<sup>2</sup>/mi (table 3, link to Excel file). Because all three basin characteristic values are within the range of values listed in table 12, the flood region 1 regression equation is applicable for estimating the  $Q_{2\%(u)r}$  flood discharge for the

ungaged site. The second step is to calculate the drainage area ratio between the ungaged site and the streamgage (map number 404, fig. 1) to determine whether the regression-weighted method is applicable for the ungaged site. The drainage area (DRNAREA) of the streamgage is listed as 1,567.26 mi<sup>2</sup> in table 3 (link to Excel file); therefore, the drainage area ratio is 0.863 (1,352.59 mi<sup>2</sup>/1,567.26 mi<sup>2</sup>), and the regression-weighted method is applicable for the ungaged site because the drainage area ratio is not less than 0.5 or greater than 1.5. The third step is to calculate the  $Q_{2\%(u)r}$  flood discharge for the ungaged site (map number 402, fig. 1) using the GLS regression equation from table 9:

$$Q_{2\%(u)r} = \text{DRNAREA}^{0.574} 10^{(0.769+0.573 \times \text{I24H10Y}-0.712 \times \text{CCM}^{0.55})}$$

$$Q_{2\%(u)r} = 1,352.59^{0.574} 10^{(0.769+0.573 \times 4.088-0.712 \times 0.840^{0.55})}$$

$$Q_{2\%(u)r} = 62.70 10^{(0.769+2.342-0.6469)}$$

$$Q_{2\%(u)r} = 18,300 \text{ ft}^3/\text{s}$$

The fourth step is to obtain the value of  $Q_{2\%(g)w}$  for the streamgage (map number 404, fig. 1), which is the WIE estimate listed in table 4 (link to Excel file) as 19,700 ft<sup>3</sup>/s. The fifth step is to obtain the value of  $Q_{2\%(g)r}$  for the streamgage, which is the RRE estimate listed in table 4 (link to Excel file) as 20,500 ft<sup>3</sup>/s. The sixth step is to calculate  $\Delta A$ , where  $\Delta A = |1,352.59 \text{ mi}^2 - 1,567.26 \text{ mi}^2| = 214.67 \text{ mi}^2$ . The seventh step is to calculate the regression-weighted estimate for the ungaged site,  $Q_{2\%(u)rww}$ , using equation 11:

$$Q_{2\%(u)rww} = \left[ \left( \frac{2 \times 214.67}{1,567.26} \right) + \left( 1 - \frac{2 \times 214.67}{1,567.26} \right) \left( \frac{19,700}{20,500} \right) \right] 18,300$$

$$Q_{2\%(u)rww} = 17,800 \text{ ft}^3/\text{s}.$$

### Area-Weighted Estimates for Ungaged Sites on Gaged Streams

A similar, but simpler, calculation is used in StreamStats (Ries and others, 2008; <http://streamstats.usgs.gov/ungaged2.html>) to area-weight AEPDs on the basis of the drainage area ratio between a streamgage and an ungaged site on the same stream. The weighting procedure is not applicable when the drainage area ratio is less than 0.5 or greater than 1.5, or when the flood characteristics significantly change between sites. The area-weighting method was presented in Elements of Applied Hydrology (Johnstone and Cross, 1949; Zariello and others, 2012) and the original equation from this publication is listed on page 11 in Federal Emergency Management Agency (2009). To obtain an area-weighted AEPD  $Q_{P(u)aw}$  for P-percent annual exceedance probability at the ungaged site, the WIE estimate for an upstream or downstream streamgage  $Q_{P(g)w}$  must first be determined using equation 9. The area-weighted AEPD for the ungaged site  $Q_{P(u)aw}$  is then computed using the following equation:

$$Q_{P(u)aw} = \left(\frac{A_{(u)}}{A_{(g)}}\right)^b Q_{P(g)w}, \tag{12}$$

where

- $Q_{P(u)aw}$  is the area-weighted estimate of flood discharge for the selected P-percent annual exceedance probability for the ungaged site,  $u$ , in cubic feet per second;
- $A_{(u)}$  is the drainage area of the ungaged site, in square miles;
- $A_{(g)}$  is described for equation 11 above;
- $Q_{P(g)w}$  is described for equation 9 above; and
- $b$  is the exponent of drainage area from the appropriate regional exponent in table 15.

Regional exponents derived from WREG using a GLS analysis of log-10 drainage area (DRNAREA) range from 0.506 to 0.641 for flood region 1, from 0.436 to 0.579 for flood region 2, and from 0.438 to 0.579 for flood region 3 (table 15). The exponent for a selected P-percent annual exceedance probability (table 15) is recommended for use for exponent  $b$  for this study to obtain an area-weighted estimate of an AEPD at an ungaged site on a gaged stream, although an average exponent  $b$  for the range of exceedance probabilities (0.554 for flood region 1, 0.484 for flood region 2, and 0.487 for flood region 3 in this study) is used by some in equation 12.

### Example 5

This example is a calculation of an area-weighted estimate for the  $Q_{2\%}$  flood discharge (50-year recurrence-interval flood discharge) for the same ungaged stream site (map number 402, fig. 1) and streamgage (map number 404, fig. 1) as illustrated in example 4. Values for  $A_{(u)}$ ,  $A_{(g)}$ , and  $Q_{P(g)w}$  from example 4 are used to solve equation 12:

$$Q_{P(u)aw} = \left(\frac{1,352.594}{1,567.265}\right)^{0.535} 19,700$$

$$Q_{P(u)aw} = 18,200 \text{ ft}^3/\text{s}.$$

### Estimates for Ungaged Sites on Gaged Streams Between Two Streamgages

For an ungaged site that is located between two streamgages on the same stream, two AEPDs can be estimated for the ungaged site using equations 11 or 12 by substituting either streamgage into the equation. StreamStats uses the area-weighted estimates (eq. 12) from both streamgages to then weight the individual estimates based on the proximity of the streamgages to the ungaged site to obtain final weighted estimates for the ungaged site (Ries and others, 2008; <http://streamstats.usgs.gov/ungaged2.html>). Additional hydrologic judgment may be necessary to determine which of the two estimates (or an average or some interpolation thereof) is most appropriate, including an evaluation of possible differences in flood-region characteristics of the two streamgages in comparison to the ungaged site. Other factors that might be considered when evaluating the two estimates include differences in the length or quality of annual peak-discharge records between the two streamgages, and the hydrologic conditions that occurred during the data-collection period for each streamgage (for example, were the annual peak discharges collected during a predominately wet or dry period?) (Verdi and Dixon, 2011; Gotvald and others, 2012).

### Weighted Method to Estimate Annual Exceedance-Probability Discharges for Ungaged Sites Draining More Than One Flood Region

For an ungaged stream site with a watershed that drains more than one flood region, the RREs can be applied separately for each flood region using basin characteristics for the entire watershed upstream from the ungaged site. AEPDs computed for each flood region can then be weighted by the

**Table 15.** Regional exponents determined from regional regression of log-10 drainage area for area-weighting method to estimate annual exceedance-probability discharges for ungaged sites on gaged streams.

[Note, the constant is not used in the area-weighting method (eq. 11), but could be used to estimate annual exceedance-probability discharges at ungaged sites from drainage area only]

| Annual exceedance probability (percent) | Flood region 1 |          | Flood region 2 |          | Flood region 3 |          |
|---|----------------|----------|----------------|----------|----------------|----------|
|   | Exponent $b$   | Constant | Exponent $b$   | Constant | Exponent $b$   | Constant |
| 50                                      | 0.641          | 1.66     | 0.579          | 2.15     | 0.579          | 2.31     |
| 20                                      | 0.592          | 2.06     | 0.525          | 2.57     | 0.528          | 2.69     |
| 10                                      | 0.569          | 2.26     | 0.499          | 2.78     | 0.503          | 2.87     |
| 4                                       | 0.548          | 2.45     | 0.476          | 2.97     | 0.479          | 3.05     |
| 2                                       | 0.535          | 2.57     | 0.463          | 3.08     | 0.466          | 3.16     |
| 1                                       | 0.524          | 2.67     | 0.453          | 3.18     | 0.455          | 3.25     |
| 0.5                                     | 0.516          | 2.75     | 0.445          | 3.26     | 0.447          | 3.33     |
| 0.2                                     | 0.506          | 2.85     | 0.436          | 3.35     | 0.438          | 3.41     |

proportion of drainage area within each flood region to calculate final AEPDs for the ungaged site. For example, if 30 percent of the drainage area for an ungaged site is in the upstream region and 70 percent is in the downstream region, the AEPD computed from an equation for the upstream region is multiplied by 0.30 and is added to 0.70 times the AEPD computed from an equation for the downstream region. The variance of prediction for this weighted method also can be approximated by using the same weighting method based on proportional drainage areas (Gotvald and others, 2012). StreamStats uses this weighted method to estimate AEPDs and prediction errors for ungaged sites draining more than one flood region (Ries and others, 2008; <http://streamstats.usgs.gov/ungaged2.html>).

## Region-of-Influence Method to Estimate Annual Exceedance-Probability Discharges for Ungaged Stream Sites

The region-of-influence (RoI) method has been used to estimate AEPDs at ungaged sites by relating basin characteristics to AEPDs for a unique subset of streamgages (Burn, 1990; Eng and others, 2005, 2007). The RoI method was tested as part of this study using WREG (Eng and others, 2009) to determine if predictive accuracies for AEPDs may be improved using RoI compared to the traditional regional-regression method. The RoI method defines a unique subset, or region of influence, for each ungaged site determined by selecting streamgages with basin characteristics similar to those measured for the ungaged site. The RoI is defined as the number of streamgages (N) “nearest” to the ungaged site, where “nearest” is measured by similarity of basin

characteristics in Euclidean space. An advantage of this method is extrapolation errors tend to be small because the predictions naturally result near the center of the space of the basin characteristics.

To investigate the RoI method for this study, basin characteristics identified as the most significant in the statewide OLS regression analyses were selected and compiled into an RoI dataset that included the same 510 streamgages as used for the development of statewide regression equations (table 8). The RoI method in WREG allows three approaches for defining hydrologic similarity among streamgage basins: independent or predictor-variable space RoI, geographic space RoI, and a combination of predictor-variable and geographic spaces called hybrid RoI. Preliminary RoI analyses were performed to determine the best combination of three input parameters required by the RoI program in WREG: (1) the best set of basin characteristics must be selected for use as explanatory variables, (2) the number of streamgages (N) must be selected to compose the specific region of influence for the statewide study area, and (3) the predictor-variable space RoI, geographic space RoI, or hybrid RoI approach must be selected.

RMSEs were evaluated for the preliminary RoI analyses to determine the best combination of the three required input parameters for WREG. Table 16 lists the best combinations of explanatory variables with the lowest RMSEs that were identified statewide, and by flood regions, for each of the eight selected annual exceedance probabilities through iterative RoI analyses using WREG. Although statewide and regional RMSE (RoI method) and SEP (RRE method) performance metrics are not directly comparable, overall, RMSEs for RoI were poorer than SEPs for statewide GLS regression equations (table 8) or SEPs for RREs (tables 9–11). RMSE (SEE) is not appropriate for evaluating GLS regressions because of the unequal weighting given to the streamgages in GLS regression (Risley and others, 2008). The resulting unequally

**Table 16.** Significant explanatory variables and predictive accuracies of preliminary region-of-influence equations in Iowa.

[RoI, region of influence; RMSE, root mean-square error; %, percent; DRNAREA, geographic information system drainage area; (+), explanatory variable has a positive relation with the response variable; DESMOIN, percent of basin within Des Moines Lobe landform region; (-), explanatory variable has a negative relation with the response variable; CCM, constant of channel maintenance; GRoI, geographic space RoI]

| Statistic   | Most significant explanatory variables identified for the preliminary equation and explanatory-variable relation signs | N, number of streamgages used to form RoI | RoI method | Statewide RMSE (percent) | Region 1 RMSE (percent) | Region 2 RMSE (percent) | Region 3 RMSE (percent) |
|---|--|---|------------|--------------------------|-------------------------|-------------------------|-------------------------|
| Preliminary RoI analysis results                      |  |   |            |                          |                         |                         |                         |
| Q <sub>50%</sub>                                      | DRNAREA <sup>0.011</sup> (+), DESMOIN (-), CCM <sup>0.55</sup> (-)   | 55  | GRoI       | 50.3                     | 46.7                    | 46.9                    | 48.9                    |
| Q <sub>20%</sub>                                      | DRNAREA <sup>-0.011</sup> (+), DESMOIN (-), CCM <sup>0.55</sup> (-)  | 55  | GRoI       | 44.0                     | 39.7                    | 34.2                    | 42.8                    |
| Q <sub>10%</sub>                                      | DRNAREA <sup>-0.023</sup> (+), DESMOIN (-), CCM <sup>0.55</sup> (-)  | 55  | GRoI       | 45.3                     | 40.4                    | 33.8                    | 43.7                    |
| Q <sub>4%</sub>                                       | DRNAREA <sup>-0.035</sup> (+), DESMOIN (-), CCM <sup>0.55</sup> (-)  | 55  | GRoI       | 49.1                     | 43.8                    | 36.8                    | 46.9                    |
| Q <sub>2%</sub>                                       | DRNAREA <sup>-0.043</sup> (+), DESMOIN (-), CCM <sup>0.55</sup> (-)  | 55  | GRoI       | 52.6                     | 47.0                    | 40.0                    | 49.7                    |
| Q <sub>1%</sub>                                       | DRNAREA <sup>-0.050</sup> (+), DESMOIN (-), CCM <sup>0.55</sup> (-)  | 55  | GRoI       | 56.3                     | 50.5                    | 43.5                    | 52.7                    |
| Q <sub>0.5%</sub>                                     | DRNAREA <sup>-0.056</sup> (+), DESMOIN (-), CCM <sup>0.55</sup> (-)  | 55  | GRoI       | 60.0                     | 54.1                    | 47.1                    | 55.8                    |
| Q <sub>0.2%</sub>                                     | DRNAREA <sup>-0.064</sup> (+), DESMOIN (-), CCM <sup>0.55</sup> (-)  | 55  | GRoI       | 64.9                     | 59.1                    | 51.7                    | 59.9                    |
| <b>Number of streamgages included in RoI analysis</b> |  |   |            | <b>510</b>               | <b>91</b>               | <b>176</b>              | <b>127</b>              |

weighted GLS residuals produce inflated RMSE values that are not comparable to RMSE from the RoI regression analyses. Because RREs provided improved predictive accuracies, the RoI method was not developed further and RoI equations are not listed in this report but are summarized in table 16 to provide a reference for indicating the improvement obtained using RREs.

## Comparison of Annual Exceedance-Probability Discharges

To better understand how AEPDs computed using the new EMA/MGB analyses used in this study compare to those computed using the standard Bulletin 17B/GB analyses, how AEPDs computed using the updated regional-skew-coefficient constant compared to those computed using the superseded regional skew coefficients, and how AEPDs computed using the updated RREs developed in this study compare to those computed in the previous study (Eash, 2001), relative percentage changes were calculated for these different AEPD estimates. Relative percentage change is calculated using the following equation:

$$RP_{change} = 100 \left( \frac{Q_{new} - Q_{old}}{Q_{old}} \right) \quad (13)$$

where

- $RP_{change}$  is the relative percentage change, which represents the relative change between the old AEPD estimate value and the new AEPD estimate value;
- $Q_{new}$  is the new AEPD estimate value, in cubic feet per second; and
- $Q_{old}$  is the old AEPD estimate value, in cubic feet per second.

$RP_{change}$  values that are positive imply that  $Q_{new}$  is greater than  $Q_{old}$ , and  $RP_{change}$  values that are negative imply that  $Q_{old}$  is greater than  $Q_{new}$ .

## Estimates from Annual Exceedance-Probability Analyses

Two new elements were included in the computation of AEPDs for this study. First, as described in the section Regional Skew Analysis, a regional skew study was performed for Iowa from which an updated regional-skew-coefficient constant was developed for the study area and was used in annual exceedance-probability analyses for all streamgages included in this study. Second, as described in the section Expected Moments Algorithm (EMA/MGB) Analyses, a new annual exceedance-probability analysis method was used to compute AEPDs for all streamgages included in this study. To better understand the effects of these two new elements in

the computation of AEPDs for this study, three variations of annual exceedance-probability analyses were also computed for 283 streamgages (table 6, link to Excel file). These 283 streamgages are a subset of the 523 streamgages listed in table 1, (link to Excel file). All 283 streamgages are located in Iowa and were used to develop regional regression equations for this study, as noted in table 4 (link to Excel file) with an asterisk next to the acronym RRE. Table 17 lists relative percentage change between AEPDs based on data through the 2010 water year for four variations of annual exceedance-probability analyses: (1) EMA/MGB analyses computed using the updated regional-skew-coefficient constant (table 4, link to Excel file), (2) EMA/MGB analyses computed using superseded regional skew coefficients (table 6, link to Excel file), (3) standard Bulletin 17B/GB analyses computed using the updated regional-skew-coefficient constant (table 6, link to Excel file), and (4) Bulletin 17B/GB analyses computed using superseded regional skew coefficients (table 6, link to Excel file). Table 17 lists relative percentage change calculated for five sets of comparisons between different combinations of these four sets of AEPD estimates for 283 streamgages in Iowa.

The first and second sets of comparisons listed in table 17 summarize relative percentage changes from the use of superseded regional skew coefficients to the use of the updated regional-skew-coefficient constant for Bulletin 17B/GB and EMA/MGB analyses. Mean and median relative percentage changes indicate that AEPDs generally decreased statewide with the use of the updated regional-skew-coefficient constant for probabilities of 10 percent and lower (for 10-year recurrence-interval floods and greater) for Bulletin 17B/GB and EMA/MGB analyses, and that AEPDs generally increased statewide for probabilities greater than 10 percent (2- and 5-year recurrence-interval floods). For example, the first set of comparisons indicate that mean and median differences for  $Q_{1\%}$  flood discharge estimates are 6.0 and 5.1 percent lower, respectively, for Bulletin 17B/GB analyses using the updated-regional-skew coefficient constant, and the second set of comparisons indicate similar results with mean and median differences that are 7.1 and 5.7 percent lower, respectively, for EMA/MGB analyses using the updated regional-skew-coefficient constant. These relative percentage changes appear to be reasonable considering that the updated statewide regional-skew-coefficient constant of -0.400 is lower than superseded regional skew coefficients for most of the State (fig. 2 and table 2 (link to Excel file); Eash, 2001). Also, mean and median relative percentage changes indicate that AEPDs decrease more for the smaller exceedance probabilities using the updated regional-skew-coefficient constant, compared to using the superseded regional skew coefficients, and the decrease is slightly greater using EMA/MGB analyses compared to using Bulletin 17B/GB analyses. As the first and second sets of comparisons listed in table 17 indicate for the 283 streamgages in Iowa that were compared, relative percentage change from AEPD estimates computed using superseded regional skew coefficients to those computed using

**Table 17.** Relative percentage change between annual exceedance-probability discharge estimates based on data through the 2010 water year for 283 streamgages in Iowa using expected moments algorithm (EMA/MGB) and Bulletin 17B/GB analyses and using updated and superseded regional skew coefficient values.

|  | Annual exceedance probability* |            |            |           |           |           |             |             |
|--|--------------------------------|------------|------------|-----------|-----------|-----------|-------------|-------------|
|  | 50-percent                     | 20-percent | 10-percent | 4-percent | 2-percent | 1-percent | 0.5-percent | 0.2-percent |
| Relative percentage change between Bulletin 17B/GB estimates computed using the updated regional skew coefficient value and Bulletin 17B/GB estimates computed using superseded regional skew coefficient values |                                |            |            |           |           |           |             |             |
| Maximum  | 10.9                           | 20.3       | 22.3       | 22.7      | 21.7      | 19.4      | 17.6        | 19.7        |
| Minimum  | -10.1                          | -9.9       | -14.3      | -19.0     | -23.6     | -28.9     | -33.7       | -41.5       |
| Mean   | 1.5                            | 0.4        | -1.0       | -3.0      | -4.5      | -6.0      | -7.4        | -9.3        |
| Median   | 1.2                            | 0.2        | -0.8       | -2.5      | -3.8      | -5.1      | -6.5        | -8.2        |
| Relative percentage change between EMA/MGB estimates computed using the updated regional skew coefficient value and EMA/MGB estimates computed using superseded regional skew coefficient values                 |                                |            |            |           |           |           |             |             |
| Maximum  | 10.5                           | 8.1        | 6.1        | 4.5       | 8.3       | 12.1      | 16.2        | 21.6        |
| Minimum  | -4.4                           | -2.4       | -4.4       | -13.1     | -22.4     | -30.8     | -38.9       | -48.6       |
| Mean   | 1.9                            | 1.4        | -0.3       | -2.9      | -5.0      | -7.1      | -9.0        | -11.6       |
| Median   | 1.3                            | 0.8        | 0.0        | -2.1      | -4.1      | -5.7      | -7.4        | -9.9        |
| Relative percentage change between EMA/MGB and Bulletin 17B/GB estimates computed using superseded regional skew coefficient values  |                                |            |            |           |           |           |             |             |
| Maximum  | 43.3                           | 33.1       | 76.6       | 141.7     | 186.9     | 225.2     | 259.3       | 295.4       |
| Minimum  | -71.7                          | -42.8      | -23.2      | -33.9     | -39.4     | -43.8     | -47.0       | -50.5       |
| Mean   | -5.2                           | 2.4        | 6.2        | 9.8       | 11.8      | 13.5      | 14.8        | 16.1        |
| Median   | -1.3                           | 1.1        | 2.8        | 3.1       | 3.5       | 3.1       | 2.8         | 2.4         |
| Relative percentage change between EMA/MGB and Bulletin 17B/GB estimates computed using the updated regional skew coefficient value  |                                |            |            |           |           |           |             |             |
| Maximum  | 42.1                           | 33.1       | 76.0       | 141.1     | 186.4     | 225.7     | 261.1       | 298.7       |
| Minimum  | -73.5                          | -40.0      | -20.7      | -29.7     | -34.1     | -37.4     | -39.9       | -42.4       |
| Mean   | -4.8                           | 3.4        | 6.9        | 9.8       | 11.1      | 11.8      | 12.3        | 12.3        |
| Median   | -0.7                           | 1.8        | 2.9        | 2.7       | 2.7       | 2.0       | 1.6         | 1.0         |
| Relative percentage change between EMA/MGB estimates computed using the updated regional skew coefficient value and Bulletin 17B/GB estimates computed using superseded regional skew coefficient values         |                                |            |            |           |           |           |             |             |
| Maximum  | 42.8                           | 33.1       | 76.4       | 143.2     | 190.5     | 232.0     | 269.9       | 311.8       |
| Minimum  | -72.9                          | -39.8      | -23.2      | -34.8     | -41.1     | -45.8     | -49.8       | -54.1       |
| Mean   | -3.4                           | 3.8        | 5.8        | 6.5       | 6.0       | 5.0       | 3.8         | 1.8         |
| Median   | 0.2                            | 2.4        | 2.2        | 0.7       | -0.9      | -2.1      | -3.3        | -5.6        |

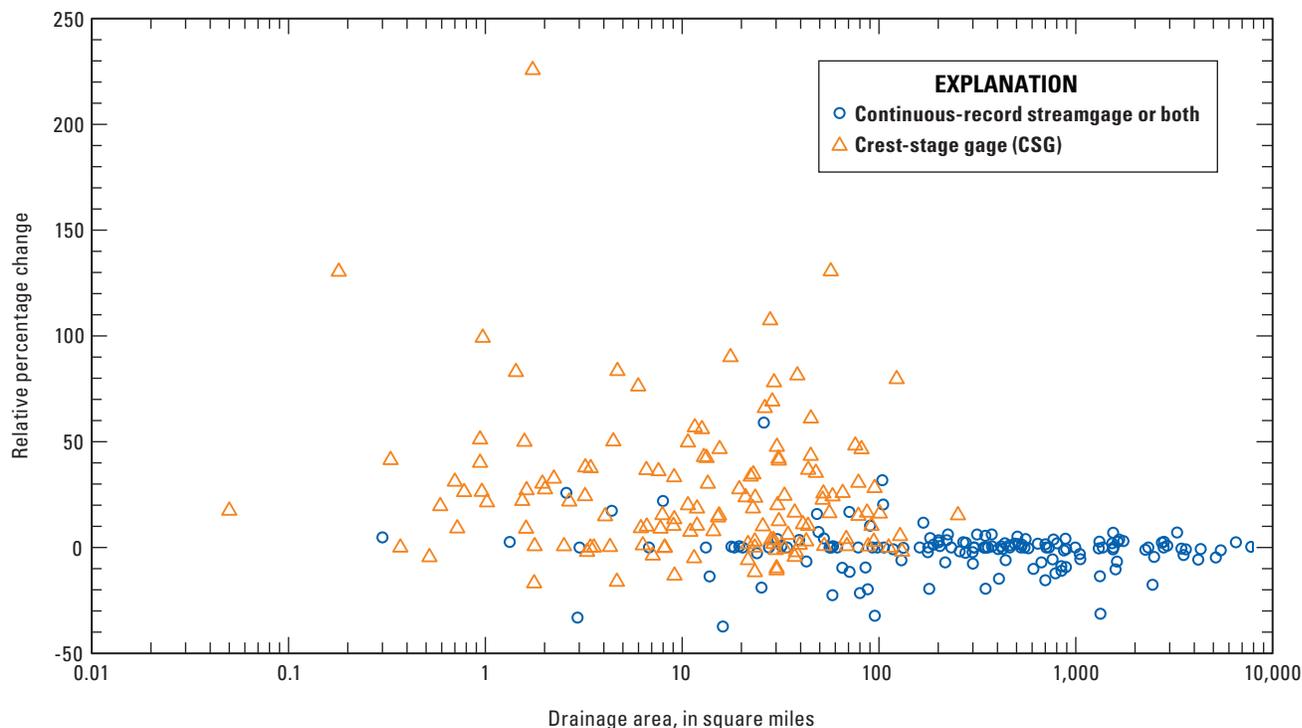
\*EMA/MGB estimates using the updated regional skew coefficient value developed in this study are listed in [table 4](#) (link to Excel file) and all other exceedance-probability estimates are listed in [table 6](#) (link to Excel file).

the updated regional-skew-coefficient constant have decreased on average about 4 percent for the eight annual exceedance probabilities.

The third and fourth sets of comparisons listed in table 17 summarize relative percentage changes from the use of Bulletin 17B/GB analyses to the use of EMA/MGB analyses for the use of the updated regional-skew-coefficient constant and of superseded regional skew coefficients. Mean and median relative percentage changes indicate that AEPDs generally increased statewide with the use of EMA/MGB analyses for probabilities of 20 percent and lower (for 5-year recurrence-interval floods and larger) for the use of the updated regional-skew-coefficient constant and the use of superseded regional

skew coefficients, and that AEPDs generally decreased statewide for the 50-percent probability (2-year recurrence-interval flood). For example, the third set of comparisons indicate that mean and median differences for  $Q_{1\%}$  flood discharge estimates are 13.5 and 3.1 percent greater, respectively, for EMA/MGB analyses using the superseded regional skew coefficients, and the fourth set of comparisons indicate similar results with mean and median differences that are 11.8 and 2.0 percent greater, respectively, for EMA/MGB analyses using the updated regional-skew-coefficient constant.

Figure 11 is a plot of the  $Q_{1\%}$  flood-discharge data used in the fourth set of comparisons. This plot shows that relative percentage changes from Bulletin 17B/GB to

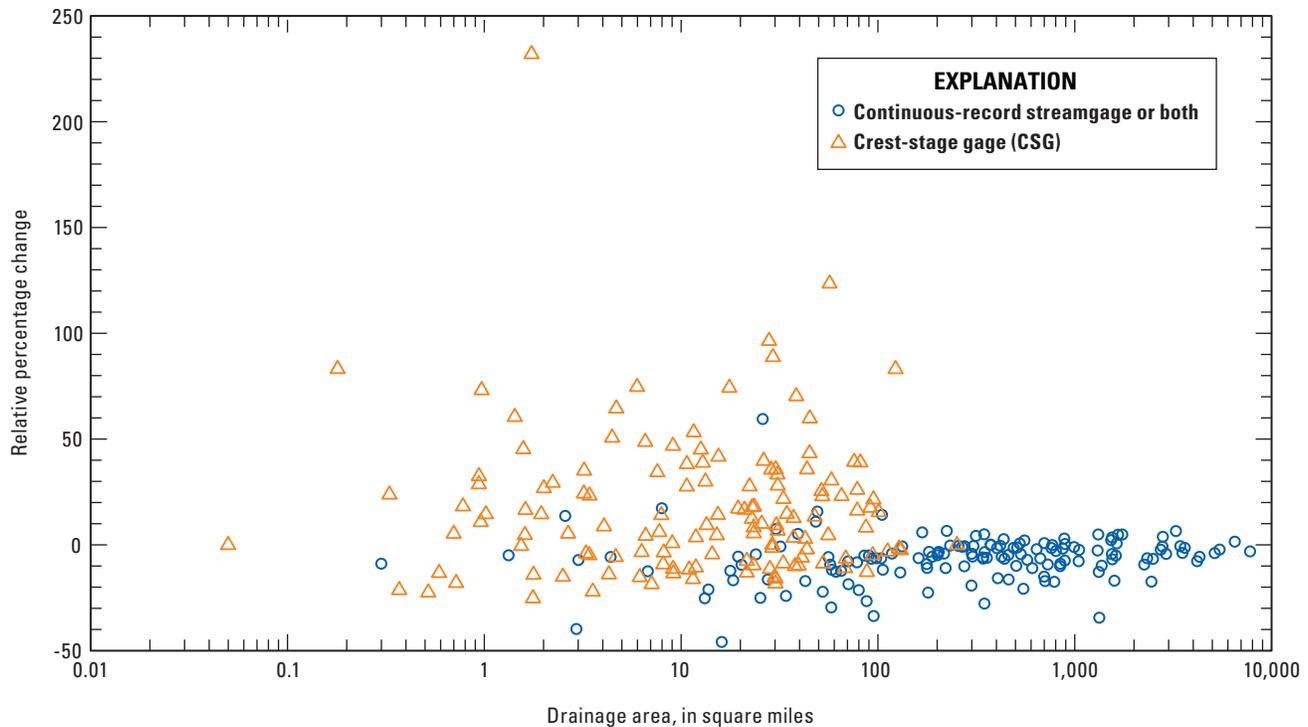


**Figure 11.** Relative percentage change, by drainage area and type of streamgauge, between expected moments algorithm (EMA/MGB) and standard Bulletin 17B/GB analyses computed using the updated regional-skew-coefficient constant for one-percent annual exceedance-probability discharges for 283 streamgages in Iowa.

EMA/MGB analyses for the larger drainage area, continuous-record streamgages generally are small values, and that differences for a few of these streamgages have negative values greater than 20 percent. Figure 11 also shows that relative percentage changes for the smaller drainage-area CSGs generally are positive values greater than 20 percent and that differences for several of these CSGs are positive values greater than 50 percent. This plot indicates that the type of streamgauge, or the type of streamgauge record, results in the greatest differences between EMA/MGB and Bulletin 17B/GB estimates of the  $Q_{1\%}$  flood discharge for this comparison of 283 streamgages. Of these 283 streamgages, 147 of them are continuous-record streamgages, or streamgages that have been operated as a continuous-record streamgauge and as a CSG, and 136 of them are CSGs. For the fourth set of comparisons, mean and median relative percentage changes for  $Q_{1\%}$  flood discharge estimates for the 147 streamgages that have been operated as a continuous-record streamgauge, or as a continuous-record and CSG streamgauge, are -1.1 and 0 percent, respectively, and mean and median relative percentage changes for the 136 CSGs are 25.8 and 17.9 percent, respectively. Because most of the CSGs included in this comparison have censored data records, the larger AEPDs computed for CSGs using EMA/MGB analyses, compared to Bulletin 17B/GB analyses, are believed to result mainly from the ability of the EMA/MGB analysis to use a specific discharge interval for data that is censored by the standard Bulletin 17B/GB analysis when it is below the largest minimum recordable

threshold discharge. As the third and fourth sets of comparisons listed in table 17 indicate for the 283 streamgages in Iowa that were compared, relative percentage change from AEPD estimates computed using standard Bulletin 17B/GB analyses to those computed using EMA/MGB analyses have increased on average about 8 percent for the eight annual exceedance probabilities.

The fifth set of comparisons listed in table 17 summarizes the relative percentage changes from Bulletin 17B/GB analyses computed using superseded regional skew coefficients to EMA/MGB analyses computed using the updated regional-skew-coefficient constant. This comparison includes both of the new elements that were applied in the computation of AEPDs for this study. Mean relative percentage changes indicate that AEPDs generally increased statewide with the use of EMA/MGB analyses using the updated regional-skew-coefficient constant for probabilities of 20 percent and lower (for 5-year recurrence-interval floods and larger), and that estimates generally decreased statewide for the 50-percent probability (2-year recurrence-interval flood). Median relative percentage changes indicate that AEPDs generally decreased statewide with the use of EMA/MGB analyses using the updated regional-skew-coefficient constant for the 2-percent probability and lower (50-year recurrence-interval floods and larger), and that estimates generally increased statewide for the 4-percent probability and higher (25-year recurrence-interval floods and smaller). For example, the fifth set of comparisons indicate that the mean difference for  $Q_{1\%}$  flood discharge



**Figure 12.** Relative percentage change, by drainage area and type of streamgage, between expected moments algorithm (EMA/MGB) analyses computed using the updated regional-skew-coefficient constant and Bulletin 17B/GB analyses computed using superseded regional skew coefficients for one-percent annual exceedance-probability discharges for 283 streamgages in Iowa.

estimates is 5.0 percent greater, and that the median difference is 2.1 percent lower for EMA/MGB analyses using the updated regional-skew-coefficient constant. Figure 12 is a plot of the  $Q_{1\%}$  flood-discharge data used in fifth set of comparisons. This plot shows similar results to figure 11, except there are six fewer streamgages with relative percentage changes greater than 50 percent in figure 12. The difference between figures 11 and 12 is that relative percentage changes from superseded to updated regional skew coefficients also are shown in figure 12. Similar to figure 11, figure 12 indicates that regardless of the regional skew coefficients used in the analyses, the greatest differences between EMA/MGB and Bulletin 17B/GB estimates of the  $Q_{1\%}$  flood discharge are because of the type of streamgage, or streamgage record, for this comparison of 283 streamgages in Iowa. As the fifth set of comparisons listed in table 17 indicates for the 283 streamgages in Iowa that were compared, relative percentage change from AEPD estimates computed using standard Bulletin 17B/GB analyses and superseded regional skew coefficients to those computed using EMA/MGB analyses and the updated regional-skew-coefficient constant have increased on average about 4 percent for the eight annual exceedance probabilities. A comparison of EMA/MGB and standard Bulletin 17B/GB analysis estimates for the  $Q_{1\%}$  annual exceedance probability for 283 streamgages in Iowa indicated a median flood-discharge relative percentage change of zero percent for 147 continuous-record streamgages and a median

flood-discharge relative percentage change increase of 18 percent for 136 crest-stage gages. Because most of the crest-stage gages included in this comparison have censored data records, the larger annual exceedance-probability discharges computed for crest-stage gages using EMA/MGB analyses, compared to Bulletin 17B/GB analyses, are believed to result mainly from the ability of the EMA/MGB analyses to use a specific discharge interval for data that is censored by standard Bulletin 17B/GB analyses when it is below the largest minimum-recording-threshold discharge.

## Estimates from Regional Regression Equations

Table 18 lists relative percentage change from RRE estimates computed in a previous study (Eash, 2001) based on data through the 1997 water year to RRE estimates computed in this study based on data through the 2010 water year for 185 streamgages in Iowa. Mean and median relative percentage changes indicate that RRE estimates generally increased statewide ranging from 4.4 to 13.3 percent and from 3.0 to 14.1 percent, respectively, with the development of the RREs for this study. For example, table 18 shows that mean and median relative percentage changes for the  $Q_{1\%}$  flood discharge estimate are 7.9 and 6.4 percent greater using the updated RREs developed in this study. In part, the larger RRE estimates computed for this study, compared to the previous study (Eash, 2001), can be attributed to the additional annual

**Table 18.** Relative percentage change from regional-regression-equation estimates computed in the previous study to those computed in this study for 185 streamgages in Iowa.

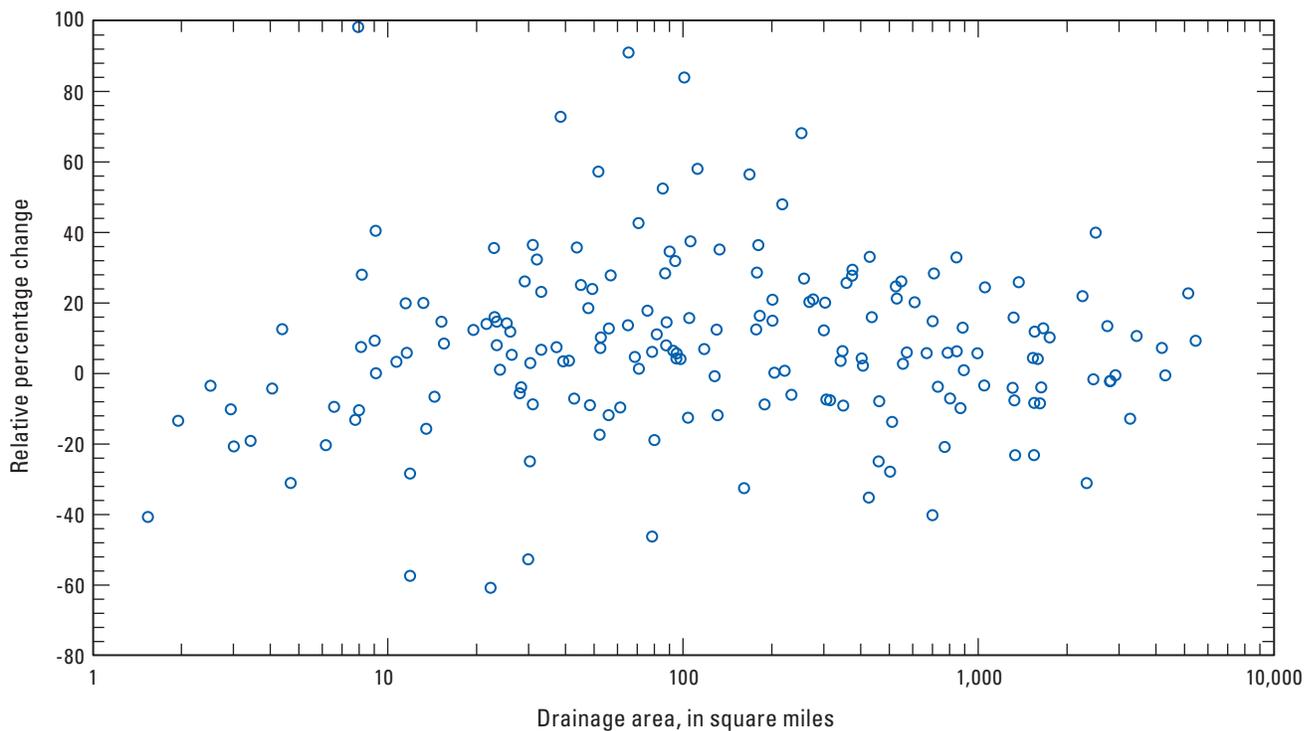
|         | Relative percentage change for annual exceedance probability <sup>1</sup> |            |            |           |           |           |             |             |
|---------|---|------------|------------|-----------|-----------|-----------|-------------|-------------|
|         | 50-percent  | 20-percent | 10-percent | 4-percent | 2-percent | 1-percent | 0.5-percent | 0.2-percent |
| Maximum | 68.2  | 70.0       | 74.7       | 83.6      | 87.7      | 98.3      | 109.3       | 115.4       |
| Minimum | -62.7   | -63.4      | -62.8      | -61.7     | -61.3     | -60.8     | -60.1       | -60.8       |
| Mean    | 8.9   | 6.0        | 4.4        | 8.5       | 8.2       | 7.9       | 13.3        | 6.8         |
| Median  | 10.6  | 4.4        | 3.0        | 8.1       | 7.6       | 6.4       | 14.1        | 6.2         |

<sup>1</sup>Regional-regression-equation estimates are listed in [table 4](#) (link to Excel file) for this study and are listed in Eash (2001) for the previous study.

peak-discharge record collected from 1998 to 2010 that was included in annual exceedance-probability analyses computed in this study. Large flood events in Iowa during 1998 (Fischer, 1999), 1999 (Ballew and Fischer, 2000; Ballew and Eash, 2001), 2002 (Eash, 2004), during 2004 (Eash, 2006), 2008 (Fischer and Eash, 2010; Linhart and Eash, 2010; Buchmiller and Eash, 2010; Holmes and others, 2010), and during 2010 (Eash, 2012; Barnes and Eash, 2012) have contributed to increased AEPDs and RRE estimates computed in this study. Relative percentage changes shown in table 18 highlight the need to periodically update AEPDs for streamgages and RREs for Iowa to obtain reliable estimates of AEPDs for ungaged stream sites.

Figure 13 shows relative percentage change by drainage area for  $Q_{1\%}$  flood-discharge estimates computed using RREs from a previous study (Eash, 2001) to those computed using RREs developed in this study for the same 185 streamgages in

Iowa as summarized in table 18. Figure 13 shows a fairly uniform relative percentage change by drainage area between the two sets of RRE estimates, with a few sites indicating positive and negative relative percentage changes greater than 40 percent for drainage areas generally less than about 300 mi<sup>2</sup>. This plot indicates that there does not appear to be a bias because of drainage area between AEPD estimates computed from RREs developed in this study compared to AEPD estimates computed from RREs developed in Eash (2001). As table 18 and figure 13 indicate for the 185 streamgages in Iowa that were compared, relative percentage change from AEPD estimates computed using RREs developed in the previous study (Eash, 2001) to those computed using RREs developed in this study have increased on average about 8 percent for the eight annual exceedance probabilities. As indicated by the fifth set of comparisons in table 17, about 4 percent of this increase likely



**Figure 13.** Relative percentage change by drainage area between one-percent annual exceedance-probability discharges computed using regional regression equations developed in this study and those developed in the previous study for 185 streamgages in Iowa.

is because of the use of EMA/MGB analyses with the updated regional-skew-coefficient constant.

During the early 1990's, 35 new CSGs with drainage areas of 11 mi<sup>2</sup> or less began operation statewide (May and others, 1996) replacing older CSGs with either adequate, or inadequate, annual peak-discharge records for the computation of AEPDs. These new CSGs were not included in the development of RREs for a previous study (Eash, 2001) because they did not have 10 years of annual peak-discharge record through the 1997 water year, whereas most of them were included in the development of RREs in this study. Thus, although these newer CSGs are included in figures 11 and 12 and in table 17, they are not included in figure 13 or in table 18. Therefore, because figures 11 and 12 indicate that a few of these newer CSGs with drainage areas of 11 mi<sup>2</sup> or less have large positive relative percentage changes, if these newer CSGs were included in figure 13 there could be a few more data points plotting as positive relative percentage changes above 40 percent.

## StreamStats

StreamStats is a USGS Web-based GIS tool (<http://water.usgs.gov/osw/streamstats/index.html>) that allows users to obtain streamflow statistics, drainage-basin characteristics, and other information for user-selected sites on streams. Users can select stream site locations of interest from an interactive map and can obtain information for these locations. If a user selects the location of a USGS streamgage, the user will receive previously published information for the streamgage from a database. If a stream site location is selected where no data are available (an ungaged site), a GIS program will estimate information for the site. The GIS program determines the boundary of the drainage basin upstream from the stream site, measures the basin characteristics of the drainage basin, and solves the appropriate regression equations to estimate streamflow statistics for that site. The results are presented in a table and a map showing the basin-boundary outline. The estimates are applicable for stream sites not significantly affected by regulation, diversions, channelization, backwater, or urbanization. In the past, it could take an experienced person more than a day to estimate this information at an ungaged site. StreamStats reduces the effort to only a few minutes.

StreamStats makes the process of computing streamflow statistics for ungaged sites much faster, more accurate, and more consistent than previously used manual methods. It also makes streamflow statistics for streamgages available without the need to locate, obtain, and read the publications in which they were originally provided. Examples of streamflow statistics that can be provided by StreamStats include the  $Q_{1\%}$  flood discharge, the median annual flow, and the mean 7-day, 10-year low flow. Examples of basin characteristics include the drainage area, basin shape, mean annual precipitation, percent of area underlain by hydrologic soil types, and so forth. Basin characteristics provided by StreamStats are the physical, geologic, and climatic properties that have been statistically

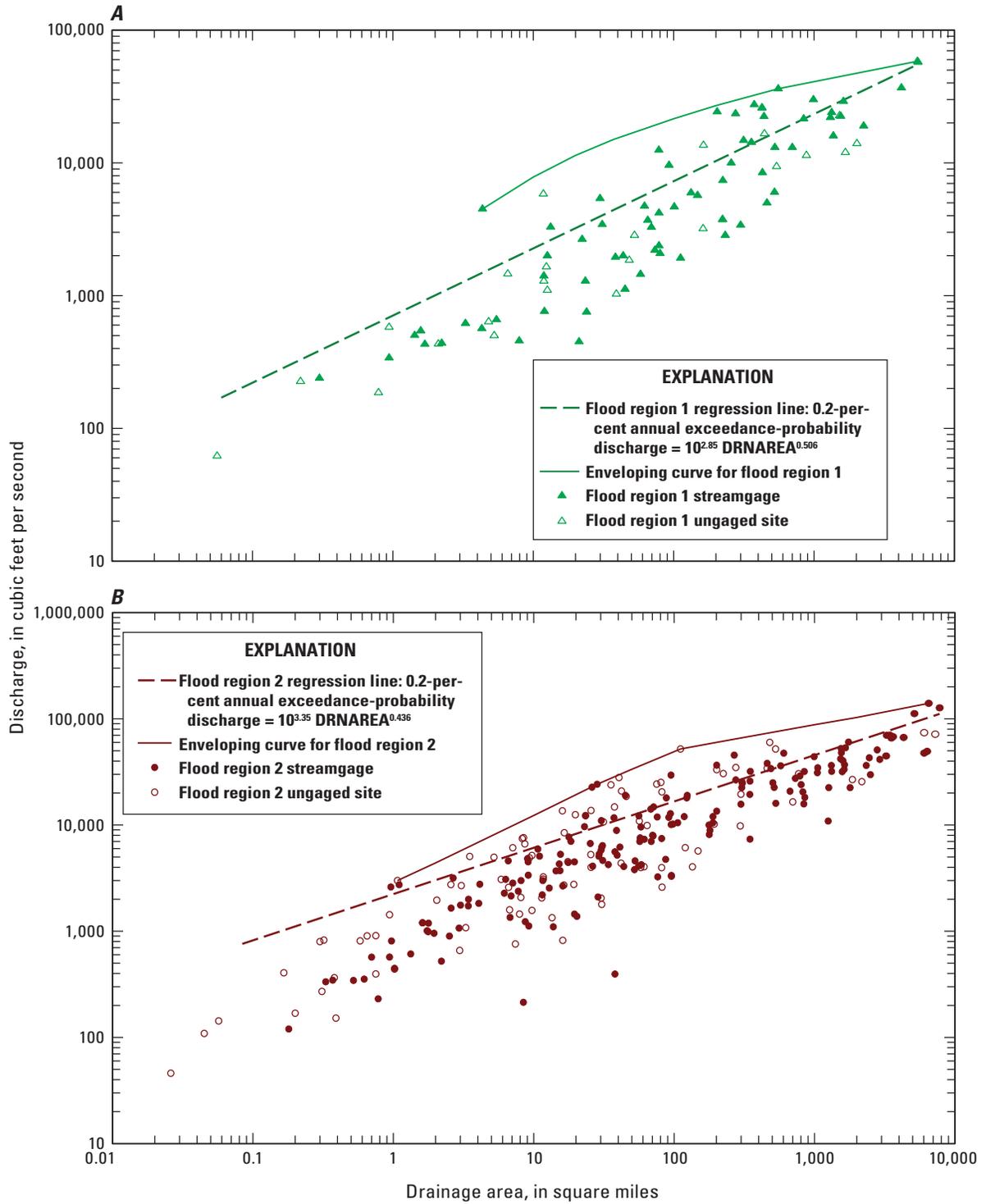
related to movement of water through a drainage basin to a stream site.

Streamflow statistics can be needed at any location along a stream and can assist with water-resources planning, management, and permitting; design and permitting of facilities such as wastewater-treatment plants and water-supply reservoirs; and design of structures such as roads, bridges, culverts, dams, locks, and levees. In addition, planners, regulators, engineers, and hydrologists often need to know the physical and climatic characteristics (basin characteristics) of the drainage basins upstream from locations of interest to help them understand the processes that control water availability and water quality at these locations. StreamStats will be a valuable tool to assist with these needs.

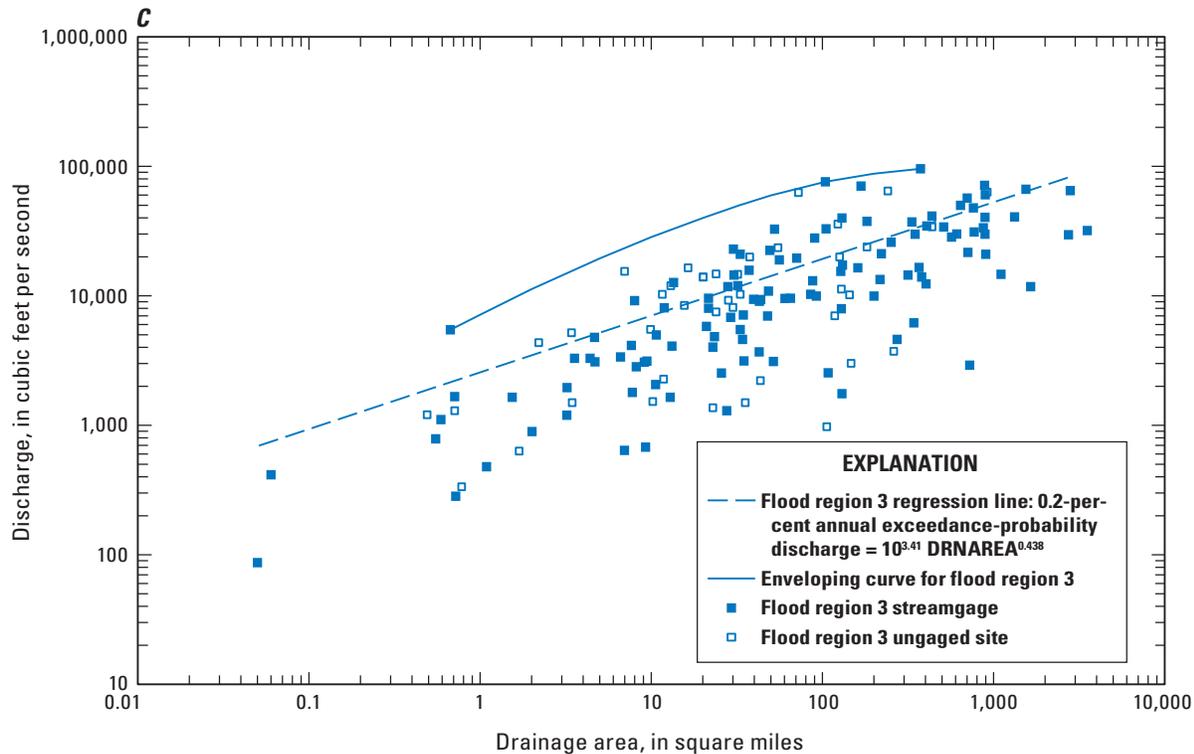
The regression equations presented in this report will be incorporated in the USGS StreamStats Web-based GIS tool (<http://water.usgs.gov/osw/streamstats/index.html>). StreamStats will provide users the ability to estimate eight AEPDs, and 90-percent prediction intervals for ungaged stream sites in Iowa.

## Maximum Floods in Iowa

For certain high-risk flood-plain developments or for evaluation of the reasonableness of unusually large flood-discharge estimates, data on maximum known floods may be considered in addition to AEPDs. Maximum floods in Iowa and their estimated annual exceedance-probability ranges are listed in [table 1](#), (link to Excel file) for streamgages included in this study. Figure 14 shows the relation between maximum flood discharge and drainage area for each of the three flood regions for 516 streamgages in Iowa. A total of 360 of these sites are active or discontinued, unregulated streamgages with annual peak-discharge records and 156 of these sites are ungaged sites. Flood-peak discharges were determined at the ungaged sites using indirect measurement methods (Benson and Dalrymple, 1967). Regression lines for the  $Q_{0.2\%}$  flood discharge (one-variable equations from table 15) and enveloping curves for the maximum known floods are shown for each flood region in figure 14. The enveloping curves indicate maximum flood-discharge potential for a range of drainage areas for each flood region. Figure 14 shows that about 115 of the 516, or about 22 percent, of the data points for streamgages and ungaged sites are present between the enveloping curves and the regional regression lines for the  $Q_{0.2\%}$  flood discharge. Most of these maximum floods happened as the result of rare storm phenomena. Maximum differences between the regression lines and the enveloping curves occur in the drainage area range from approximately 5 to 50 mi<sup>2</sup> for flood region 1, from approximately 20 to 500 mi<sup>2</sup> for flood region 2, and from approximately 5 to 100 mi<sup>2</sup> for flood region 3. These maximum differences may indicate that maximum flood-discharge potential, as unit runoff per square mile [(ft<sup>3</sup>/s)/mi<sup>2</sup>], may be greatest within these drainage area ranges for watersheds in these three flood regions.



**Figure 14.** Relation between maximum flood discharge and drainage area for streams in *A*, flood region 1, *B*, flood region 2, and *C*, flood region 3.



**Figure 14.** Relation between maximum flood discharge and drainage area for streams in *A*, flood region 1, *B*, flood region 2, and *C*, flood region 3.—Continued

## Summary

Reliable estimates of annual exceedance-probability discharges (AEPDs) are essential for the economic planning and safe design of bridges, dams, levees, and other structures located along rivers and streams, and for the effective management of flood plains. In response to the need to update and improve the predictive accuracy of estimates of AEPDs for ungaged stream sites in Iowa, the U.S. Geological Survey (USGS), in cooperation with the Iowa Department of Transportation and the Iowa Highway Research Board, initiated a statewide study in 2006.

Methods described in this report for estimating selected AEPDs are applicable to streams in Iowa that are not significantly affected by regulation, diversion, channelization, backwater, or urbanization. Estimation equations were developed for eight selected annual exceedance-probability statistics for flood discharges with 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent annual exceedance probabilities, which are equivalent to annual flood-frequency recurrence intervals of 2, 5, 10, 25, 50, 100, 200, and 500 years, respectively. Major accomplishments of the study included (1) performing a Bayesian weighted least-squares/generalized least-squares regression analysis to update regional skew coefficients for Iowa; (2) computing eight selected AEPDs at 518 streamgages within Iowa and adjacent States with at least 10 years of annual peak-discharge record, based on data

through September 30, 2010; (3) measuring 59 basin characteristics for each streamgage; (4) defining three flood regions for the State and developing 24 regional regression equations (RREs) to estimate eight selected AEPDs at ungaged stream sites based on basin characteristics; (5) calculating weighted AEPDs at 394 streamgages using the weighted independent estimates method; and (6) calculating AEPD relative percentage change for streamgages in Iowa between estimates from different annual exceedance-probability analyses based on data through the 2010 water year and between RREs developed in this study and a previous study (Eash, 2001).

Kendall's tau tests were performed for 518 streamgages included in the regression study because trends in annual peak-discharge data could introduce a bias into the annual exceedance-probability analyses. Results of the Kendall's tau tests indicated statistically significant trends for 25 streamgages when tested using 94 percent of the entire record length. The entire record length was not used because of the sensitivity of the Kendall's tau tests to multiyear sequences of larger or smaller discharges if the sequences happen near the beginning or end of the period of record. Twenty-two of these streamgages were included in the regional regression analyses because of uncertainty in the trends because of short or broken records. The remaining three streamgages represent actual trends or anomalies and were omitted from the regression analyses. The number of uncensored annual peak-discharge record lengths used in the study for the 518 streamgages ranged from 9 to 108 years with a mean

of 35.4 years and a median of 28 years. Drainage areas of the 518 streamgages ranged from 0.05 to 7,783 square miles.

Three regionalization approaches were investigated in this study for estimating AEPDs at ungaged sites in Iowa: statewide, regional, and region-of-influence regression. Regression analyses were used to relate physical and climatic characteristics of drainage basins to AEPDs. Data collected for 518 streamgages were compiled into statewide, regional, and region-of-influence datasets for regression analyses. Root mean square errors and average standard errors of prediction (SEP) calculated for each equation for each AEPD were compared for each regression to evaluate the predictive accuracy. Because regional regression provided the best predictive accuracy of the three approaches investigated, preliminary equations developed for the statewide and region-of-influence methods are not listed in this report. Regional regression analyses included 510 streamgages after 8 additional streamgages were removed from the regression dataset. These additional eight streamgages were flagged as outliers in the regression analyses and were removed because of significant trends, urbanization, or channelization.

The study area, which included all of Iowa and adjacent areas (within 50 miles of the State border) of neighboring States, was divided into three flood regions on the basis of landform regions and soil regions. The three flood regions (regions 1, 2, and 3) were defined for Iowa after testing a number of different regional combinations to define six final flood regions for the study area. Because three of the six final flood regions defined for the study area are completely outside of Iowa, regression equations were not developed further for flood regions 4, 5, and 6. Generalized least-squares (GLS) multiple-linear regression analyses, weighted on the basis of streamgage record length and the variance and cross-correlation of the annual peak discharges, were performed to develop the final equations that included 394 streamgages for the three flood regions. Preliminary multiple-linear-regression analyses, using ordinary-least-squares regression, were performed to test for significant differences among the flood regions and to identify the most significant basin characteristics for inclusion in the GLS regressions.

Fifty-nine basin characteristics measured for each streamgage were determined from digital databases using geographic information system (GIS) software. Six basin characteristics are used as explanatory variables in the final regression equations; these include three morphometric characteristics: drainage area (DRNAREA), constant of channel maintenance (CCM), and basin shape (BSHAPE); two geologic/pedologic characteristics: percent area within the Des Moines Lobe landform region (DESMOIN) and average saturated hydraulic conductivity of soil (KSATSSUR); and one climatic characteristic: maximum 24-hour precipitation that happens on average once in 10 years (I24H10Y). Predictive accuracies for the annual exceedance-probability equations developed for each region are indicated by several performance metrics. SEPs range from 31.8 to 45.2 percent for flood region 1, from 19.4 to 46.8 percent for flood region 2,

and from 26.5 to 43.1 percent for flood region 3. The pseudo coefficients of determination (pseudo- $R^2$ ) for the GLS equations range from 90.8 to 96.2 percent for flood region 1, range from 91.5 to 97.9 percent for flood region 2, and range from 92.4 to 96.0 percent for flood region 3. In general, predictive accuracies tend to be the best for flood region 2, second best for flood region 3, and poorest for flood region 1. Of the eight annual exceedance-probability equations developed for each region, the  $Q_{10\text{-percent } (\%)}$ ,  $Q_{4\%}$ , and  $Q_{2\%}$  flood-discharge regression equations generally have the best predictive accuracy and the  $Q_{50\%}$  and  $Q_{0.2\%}$  flood-discharge equations generally have the poorest accuracy.

The regional-regression equations developed in this study are not intended for use at ungaged stream sites in which the basin characteristics are outside the range of those used to develop the equations. Inconsistencies in estimates may result for the annual exceedance-probability equations if basin-characteristic values approach the minimum or maximum limits of the range. GIS software is required to measure the basin characteristics included as explanatory variables in the regression equations.

To better understand the effects of the new expected moments algorithm (EMA), with the multiple Grubbs-Beck (MGB) test for detecting low outliers, method of annual exceedance-probability analysis and the new updated regional-skew-coefficient constant used in this study have on the estimation of AEPDs for Iowa, AEPDs computed using these two new elements were compared to AEPDs computed using the standard Bulletin 17B/GB annual exceedance-probability analysis and superseded regional skew coefficient. Results of this comparison for 283 streamgages in Iowa indicate that on average for the eight annual exceedance probabilities, AEPDs are lower by about 4 percent using the updated regional-skew-coefficient constant and are greater by about 8 percent using the EMA/MGB analysis method, and overall, AEPDs are about 4 percent greater using the EMA/MGB analysis method and the updated regional-skew-coefficient constant. The larger estimates computed on average for these 283 streamgages by the EMA/MGB analysis, compared to the Bulletin 17B/GB analysis, primarily are because of the type of streamgage. Comparison results for the  $Q_{1\%}$  flood discharge indicate on average that AEPDs are about 1 percent lower for 147 continuous-record streamgages and are about 26 percent greater for 136 crest-stage gages (CSGs). Because most of the CSGs included in this comparison have censored data records, the larger AEPDs computed for CSGs are believed to result mainly from the ability of the EMA/MGB analysis to use a specific discharge interval for data that is censored by the Bulletin 17B/Grubbs-Beck analysis when it is below the largest minimum recordable threshold discharge. A comparison between regional-regression-equation estimates computed in this study and those computed in a previous 2001 study also is presented. Results of this comparison for 185 streamgages in Iowa indicate that on average AEPDs are greater by about 8 percent for the eight annual exceedance probabilities using the regional-regression equations developed in this study.

About 4 percent of this increase likely is because of the use of EMA/MGB analyses with the updated regional-skew-coefficient constant.

All 24 regression equations developed for this study are to be included in the USGS StreamStats Web-based GIS tool. StreamStats will provide users with a set of AEPDs for ungaged stream sites within Iowa in addition to the basin characteristics for the sites. Ninety-percent prediction intervals also are automatically calculated. A 90-percent prediction interval denotes there is 90-percent certainty that the true value of an AEPD at an ungaged stream site will be within a plus or minus interval around the predicted AEPD.

## Acknowledgments

The authors gratefully acknowledge the following U.S. Geological Survey (USGS) personnel: William Asquith, for his assistance with information on how to apply a program to optimize power transformations for drainage area on the basis of minimization of the predicted residuals sum of squares statistic; Kris Lund, for her work to prepare geographic information system (GIS) base data layers required for StreamStats; Rochelle Galer, for her work to measure basin characteristics for streamgages, to prepare base GIS data layers required for StreamStats and for areas in neighboring States, and to prepare Natural Resources Conservation Service Soil Survey Geographic database data layers required for use with the Soil Data Viewer; Gabe Ritter, for his work to prepare GIS base data layers required for StreamStats, and to create programs for calculating the streamflow-variability index basin characteristic and for improving the efficiency of calculating the base-flow index basin characteristic; Ken Eng, for his assistance with the measurement of the annual base-flow-recession time constant basin characteristic; and Ed Fischer, for his work to create scripts for automating the computations from the weighted independent estimates and hydrograph separation and analysis programs, for plotting eight of the figures in this report, and for his work to develop statewide region of influence regression equations.

The authors also would like to express their appreciation to the many other USGS personnel who assisted with collection and analysis of flood data used in this study. The flood data often were collected during adverse conditions, and the efforts of these individuals made this study possible. The information contained herein also is based on data collected by the U.S. Army Corps of Engineers, the National Weather Service, and several State and local agencies who were involved with collection of flood data.

## References Cited

- Ahearn, E.A., 2003, Peak-flow frequency estimates for U.S. Geological Survey streamflow-gaging stations in Connecticut: U.S. Geological Survey Water-Resources Investigations Report 03-4196, 29 p. (Also available at <http://pubs.usgs.gov/wri/wri034196/wrir03-4196.pdf>.)
- Ahearn, E.A., 2010, Regional regression equations to estimate flow-duration statistics at ungaged stream sites in Connecticut: U.S. Geological Survey Scientific Investigations Report 2010-5052, 45 p. (Also available at <http://pubs.usgs.gov/sir/2010/5052/>.)
- Arneson, L.A., Zevenbergen, L.W., Lagasse, P.F., and Clopper, P.E., 2012, Evaluating scour at bridges (5th ed.): Federal Highway Administration, Publication No. FHWA-HIF-12-003, Hydraulic Engineering Circular No. 18, 340 p., accessed March 15, 2013, at <http://www.fhwa.dot.gov/engineering/hydraulics/pubs/hif12003.pdf>.
- Asquith, W.H., and Thompson, D.B., 2008, Alternative regression equations for estimation of annual peak-streamflow frequency for undeveloped watersheds in Texas using PRESS minimization: U.S. Geological Survey Scientific Investigations Report 2008-5084, 40 p. (Also available at <http://pubs.usgs.gov/sir/2008/5084/>.)
- Ballew, J.L., and Fischer, E.E., 2000, Floods of May 17-20, 1999, in the Volga and Wapsipinicon River Basins, north-east Iowa: U.S. Geological Survey Open-File Report 00-237, 36 p. (Also available at <http://pubs.usgs.gov/of/2000/0237/report.pdf>.)
- Ballew, J.L., and Eash, D.A., 2001, Floods of July 19-25, 1999, in the Wapsipinicon and Cedar River Basins, north-east Iowa: U.S. Geological Survey Open-File Report 01-13, 45 p. (Also available at <http://pubs.usgs.gov/of/2001/0013/report.pdf>.)
- Barnes, K.K., and Eash, D.A., 2012, Floods of August 11-16, 2010, in the South Skunk River Basin, Central and South-east Iowa: U.S. Geological Survey Open-File Report 2012-1202, 27 p. with appendix. (Also available at <http://pubs.usgs.gov/of/2012/1202/>.)
- Benson, M.A., and Dalrymple, Tate, 1967, General field and office procedures for indirect discharge measurements: U.S. Geological Survey Techniques of Water-Resources Investigations, book 3, chap. A1, 30 p. (Also available at: <http://pubs.usgs.gov/twri/twri3-a1/>.)
- Buchmiller, R.C., and Eash, D.A., 2010, Floods of May and June 2008 in Iowa: U.S. Geological Survey Open-File Report 2010-1096, 10 p. (Also available at: <http://pubs.usgs.gov/of/2010/1096/>.)

- Burn, D.H., 1990, Evaluation of regional flood frequency analysis with region of influence approach: *Water Resources Research*, v. 26, no. 10, p. 2257–2265, accessed March 15, 2013, at <http://www.agu.org/journals/wr/v026/i010/WR026i010p02257/WR026i010p02257.pdf>.
- Cohn, T.A., 2011, PeakfqSA Version 0.960 (software): U.S. Geological Survey, [Information on PeakfqSA available at [http://www.timcohn.com/TAC\\_Software/PeakfqSA/](http://www.timcohn.com/TAC_Software/PeakfqSA/).]
- 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, accessed March 15, 2013, at <http://onlinelibrary.wiley.com/doi/10.1029/97WR01640/pdf>.
- 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, accessed March 15, 2013, at <http://timcohn.com/Publications/CohnLaneSted2001WR900016.pdf>.
- Cohn, T.A., Berenbrock, Charles, Kiang, J.E., and Mason, R.R., Jr., 2012, Calculating weighted estimates of peak streamflow statistics: U.S. Geological Survey Fact Sheet 2012–2038, 4 p. (Also available at: <http://pubs.usgs.gov/fs/2012/3038/>.) [also see <http://water.usgs.gov/usgs/osw/swstats/freq.html>.]
- Cook, R.D., 1977, Detection of influential observation in linear regression: *Technometrics*, v. 19, p. 15–18, accessed March 15, 2013, at <http://www.ime.usp.br/~abe/lista/pdfWiH1zqnMHo.pdf>.
- Eash, D.A., 1993, Estimating design-flood discharges for streams in Iowa using drainage-basin and channel-geometry characteristics: U.S. Geological Survey Water-Resources Investigations Report 93–4062, 96 p. (Also available at <http://pubs.er.usgs.gov/publication/wri934062>.)
- Eash, D.A., 2001, Techniques for estimating flood-frequency discharges for streams in Iowa: U.S. Geological Survey Water-Resources Investigations Report 00–4233, 88 p. (Also available at [http://ia.water.usgs.gov/pubs/reports/WRIR\\_00-4233.pdf](http://ia.water.usgs.gov/pubs/reports/WRIR_00-4233.pdf).)
- Eash, D.A., 2004, Flood of June 4–5, 2002, in the Maquoketa River Basin, East-Central Iowa: U.S. Geological Survey Open-File Report 2004–1250, 29 p. (Also available at: <http://pubs.usgs.gov/of/2004/1250/>.)
- Eash, D.A., 2006, Flood of May 23, 2004, in the Turkey and Maquoketa River Basins, Northeast Iowa: U.S. Geological Survey Open-File Report 2006–1067, 35 p. (Also available at: <http://pubs.usgs.gov/of/2006/1067/>.)
- Eash, D.A., 2012, Floods of July 23–26, 2010, in the Little Maquoketa River and Maquoketa River Basins, Northeast Iowa: U.S. Geological Survey Open-File Report 2011–1301, 45 p. with appendix. (Also available at: <http://pubs.usgs.gov/of/2011/1301/>.)
- Eash, D.A., and Barnes, K.K., 2012, Methods for estimating low-flow frequency statistics and harmonic mean flows for streams in Iowa: U.S. Geological Survey Scientific Investigations Report 2012–5171, 94 p. (Also available at <http://pubs.usgs.gov/sir/2012/5171/>.)
- Efroymson, M.A., 1960, Multiple regression analysis, in Ralston, A., and Wilf, H.S., eds., *Mathematical methods for digital computers*: New York, John Wiley and Sons, Inc., p. 191–203.
- Eng, Ken, Tasker, G.D., and Milly, P.C.D., 2005, An analysis of region-of-influence methods for flood regionalization in the Gulf-Atlantic Rolling Plains: *Journal of American Water Resources Association*, v. 41, no. 1, p. 135–143, accessed March 15, 2013, at [http://water.usgs.gov/nrp/proj.bib/Publications/2005/eng\\_tasker\\_etal\\_2005.pdf](http://water.usgs.gov/nrp/proj.bib/Publications/2005/eng_tasker_etal_2005.pdf).
- Eng, K., and Milly, P.C.D., 2007, Relating low-flow characteristics to the base flow recession time constant at partial record stream gauges: *Water Resources Research*, v. 43, W01201, doi:10.1029/2006WR005293, accessed March 15, 2013, at <http://www.agu.org/journals/wr/wr0701/2006WR005293/2006WR005293.pdf>.
- Eng, Ken, Milly, P.C.D., and Tasker, G.D., 2007, Flood regionalization—A hybrid geographic and predictor-variable region-of-influence regression method: *Journal of Hydrologic Engineering*, ASCE, v. 12, no. 6, p. 585–591, accessed March 15, 2013, at <http://ascelibrary.org/doi/pdf/10.1061/%28ASCE%291084-0699%282007%2912%3A6%28585%29>.
- 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 Methods, book 4, chap. A8, 21 p. (Also available at <http://pubs.usgs.gov/tm/tm4a8/>.)
- Environmental Systems Research Institute, Inc., 2009, ArcGIS desktop help, accessed March 15, 2013, at <http://webhelp.esri.com/arcgisdesktop/9.3>.
- Feaster, T.D., Gotvald, A.J., and Weaver, J.C., 2009, Magnitude and frequency of rural floods in the Southeastern United States, 2006—Volume 3, South Carolina: U.S. Geological Survey Scientific Investigations Report 2009–5156, 226 p. (Also available at: <http://pubs.usgs.gov/sir/2009/5156/>.)

- Federal Emergency Management Agency, 2002, National flood insurance program, program description: FEMA Federal Insurance and Mitigation Administration, 41 p., accessed March 15, 2013, at <http://www.fema.gov/library/viewRecord.do?id=1480>.
- Federal Emergency Management Agency, 2009, Flood insurance study for Kent County, Rhode Island: Flood Insurance Study Number 44003CV001A, 126 p., accessed March 15, 2013, at <http://www.warwickri.gov/pdfs/planning/mapmod/Kent%20County,%20RI%20Preliminary%20FIS%20Report%20-%20DRAFT.pdf>.
- Fischer, E.E., 1995, Potential-scour assessments and estimates of maximum scour at selected bridges in Iowa: U.S. Geological Survey Water-Resources Investigations Report 95-4051, 75 p.
- Fischer, E.E., 1999, Flood of June 15-17, 1998, Nishnabotna and East Nishnabotna Rivers, southwest Iowa: U.S. Geological Survey Open-File Report 99-70, 15 p. (Also available at <http://pubs.usgs.gov/of/1999/0070/report.pdf>.)
- Fischer, E.E., and Eash, D.A., 2010, Flood of June 8-9, 2008, Upper Iowa River, Northeast Iowa: U.S. Geological Survey Open-File Report 2010-1087, 17 p. with appendix. (Also available at: <http://pubs.usgs.gov/of/2010/1087/>.)
- Flynn, K.M., Kirby, W.H., and Hummel, P.R., 2006, User's manual for program PeakFQ annual flood-frequency analysis using Bulletin 17B guidelines: U.S. Geological Survey, Techniques and Methods, book 4, chap. B4, 42 p. (Also available at <http://pubs.usgs.gov/tm/2006/tm4b4/tm4b4.pdf>.)
- Gesch, D.B., 2007, The national elevation dataset, in Maune, D., ed., Digital elevation model technologies and applications—The DEM users manual (2d ed.): Bethesda, Maryland, American Society for Photogrammetry and Remote Sensing, p. 99-118, accessed March 15, 2013, at [http://topotools.cr.usgs.gov/pdfs/Gesch\\_Chp\\_4\\_Nat\\_Elev\\_Data\\_2007.pdf](http://topotools.cr.usgs.gov/pdfs/Gesch_Chp_4_Nat_Elev_Data_2007.pdf) [see also <http://ned.usgs.gov/>].
- Gotvald, A.J., Feaster, T.D., and Weaver, J.C., 2009, Magnitude and frequency of rural floods in the southeastern United States, 2006—Volume 1, Georgia: U.S. Geological Survey Scientific Investigations Report 2009-5043, 120 p. (Also available at <http://pubs.usgs.gov/sir/2009/5043/>.)
- Gotvald, A.J., Barth, N.A., Veilleux, A.G., and Parrett, Charles, 2012, Methods for determining magnitude and frequency of floods in California, based on data through water year 2006: U.S. Geological Survey Scientific Investigations Report 2012-5113, 38 p., 1 pl. (Also available at <http://pubs.usgs.gov/sir/2012/5113/>.)
- Griffis, V.W., Stedinger, J.R., and Cohn, T.A., 2004, Log-Pearson type 3 quantile estimators with regional skew information and low outlier adjustments: Water Resources Research, v. 40, W07503, 17 p., accessed March 15, 2013, at <http://onlinelibrary.wiley.com/doi/10.1029/2003WR002697/pdf>.
- Griffis, V.W., and Stedinger, J.R., 2007, The use of GLS regression in regional hydrologic analyses: Journal of Hydrology, v. 344, p. 82-95, accessed March 15, 2013, at <http://www.sciencedirect.com/science/article/pii/S0022169407003848>.
- Gruber, A.M., Reis, D.S., Jr., and Stedinger, J.R., 2007, Models of regional skew based on Bayesian GLS regression, in Kabbes, K.C., ed., Proceedings of the World Environmental and Water Resources Congress, Restoring our Natural Habitat, May 15-18, 2007: Tampa, Florida, American Society of Civil Engineers, paper 40927-3285.
- Gruber, A. M., and Stedinger, J.R., 2008, Models of LP3 Regional Skew, Data Selection, and Bayesian GLS Regression, in, Babcock, R.W. and Watson, R. eds, World Environmental and Water Resources Congress 2008 Ahupua'a, May 12-16, 2008: Honolulu, Hawaii, paper 596.
- Harvey, C.A., and Eash, D.A., 1996, Description, instructions, and verification for Basinsoft, a computer program to quantify drainage-basin characteristics: U.S. Geological Survey Water-Resources Investigations Report 95-4287, 25 p. (Also available at <http://pubs.er.usgs.gov/publication/wri954287>.)
- Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. A3, 510 p. (Also available at [http://pubs.usgs.gov/twri/twri4a3/html/pdf\\_new.html](http://pubs.usgs.gov/twri/twri4a3/html/pdf_new.html).)
- Holmes, R.R., Jr., Koenig, T.A., and Karstensen, K.A., 2010, Flooding in the United States Midwest, 2008: U.S. Geological Survey Professional Paper 1775, 64 p. (Also available at: <http://pubs.usgs.gov/pp/1775/>.)
- Homer, Collin, Huang, Chengquan, Yang, Limin, Wylie, Bruce, and Coan, Michael, 2004, Development of a 2001 National Land-Cover Database for the United States: Photogrammetric Engineering and Remote Sensing, v. 70, no. 7, p. 829-840, accessed March 15, 2013, at [http://www.mrlc.gov/pdf/July\\_PERS.pdf](http://www.mrlc.gov/pdf/July_PERS.pdf), also see <http://www.mrlc.gov/index.php>.
- Hydrologic Frequency Analysis Work Group, 2012, minutes of March 19, 2012, meeting, accessed March 15, 2013, at [http://acwi.gov/hydrology/Frequency/minutes/Minutes\\_HFAWG\\_meeting\\_mar19\\_2012\\_040212.pdf](http://acwi.gov/hydrology/Frequency/minutes/Minutes_HFAWG_meeting_mar19_2012_040212.pdf).

- Huff, F.A., and Angel, J.R., 1992, Rainfall frequency atlas of the Midwest: Champaign, Illinois, State Water Survey, Bulletin 71, 141 p., accessed March 15, 2013, at <http://www.isws.illinois.edu/pubdoc/B/ISWSB-71.pdf>.
- Iman, R.L., and Conover, W.J., 1983, A modern approach to statistics: New York, John Wiley and Sons, Inc., 497 p.
- Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood flow frequency: Reston, Virginia, Hydrology Subcommittee Bulletin 17B, 28 p. and appendixes, accessed March 15, 2013, at [http://water.usgs.gov/osw/bulletin17b/dl\\_flow.pdf](http://water.usgs.gov/osw/bulletin17b/dl_flow.pdf).
- Johnstone, Don, and Cross, W.P., 1949, Elements of applied hydrology: New York, Ronald Press Co., 276 p.
- Koltun, G.F., and Whitehead, M.T., 2002, Techniques for estimating selected streamflow characteristics of rural, unregulated streams in Ohio: U.S. Geological Survey Water-Resources Investigations Report 02-4068, 50 p. (Also available at <http://oh.water.usgs.gov/reports/wrir/wrir02-4068.pdf>.)
- Lamontagne, J.R., Stedinger, J.R., Berenbrock, Charles, Veilleux, A.G., Ferris, J.C., and Knifong, D.L., 2012, Development of regional skews for selected flood durations for the Central Valley Region, California, based on data through water year 2008: U.S. Geological Survey Scientific Investigations Report 2012-5130, 60 p. (Also available at <http://pubs.usgs.gov/sir/2012/5130/>)
- Lara, O.G., 1973, Floods in Iowa—Technical manual for estimating their magnitude and frequency: Iowa Natural Resources Council Bulletin 11, 56 p.
- Lara, O.G., 1987, Methods for estimating the magnitude and frequency of floods at ungaged sites on unregulated rural streams in Iowa: U.S. Geological Survey Water-Resources Investigations Report 87-4132, 34 p. (Also available at <http://pubs.er.usgs.gov/publication/wri874132>.)
- Linhart, S.M., and Eash, D.A., 2010, Floods of May 30 to June 15, 2008, in the Iowa and Cedar River Basins, Eastern Iowa: U.S. Geological Survey Open-File Report 2010-1190, 99 p. with appendixes. (Also available at <http://pubs.usgs.gov/of/2010/1190/>.)
- Lorenz, D.L., and others, 2011, USGS library for S-PLUS for windows—Release 4.0: U.S. Geological Survey Open-File Report 2011-1130, accessed March 15, 2013, at <http://water.usgs.gov/software/S-PLUS/>.
- Ludwig, A.H., and Tasker, G.D., 1993, Regionalization of low flow characteristics of Arkansas streams: U.S. Geological Survey Water-Resources Investigations Report 93-4013, 19 p. (Also available at <http://pubs.usgs.gov/wri/1993/4013/report.pdf>.)
- Marquardt, D.W., 1970, Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation: *Technometrics*, v. 12, no. 3, p. 591–612, accessed March 15, 2013, at <http://www.jstor.org/stable/1267205>.
- May, J.E., Gorman, J.G., Goodrich, R.D., Bobier, M.W., and Miller, V.E., 1996, Water resources data, Iowa, water year 1995: U.S. Geological Survey Water-Data Report IA-95-1, 387 p.
- Montgomery, D.C., Peck, E.A., and Vining, G.G., 2001, Introduction to linear regression analysis (3d ed.): New York, Wiley, 641 p.
- Multi-Resolution Land Characteristics Consortium (MRLC), 2012, National Land Cover Database (NLCD): U.S. Geological Survey, accessed March 15, 2013, at <http://www.mrlc.gov/index.php>.
- National Climatic Data Center, Climate of Iowa, accessed March 15, 2013, at [http://www.crh.noaa.gov/images/dvn/downloads/Clim\\_IA\\_01.pdf](http://www.crh.noaa.gov/images/dvn/downloads/Clim_IA_01.pdf).
- National Cooperative Soil Survey and Natural Resources Conservation Service, Iowa Soil Regions Based on Parent Materials: accessed March 15, 2013, at <ftp://ftp-fc.sc.egov.usda.gov/IA/technical/IowaSoilRegionsMap.html>.
- Natural Resources Conservation Service, Soil Survey Geographic (SSURGO) Database: accessed March 15, 2013, at <http://soildatamart.nrcs.usda.gov/>.
- Natural Resources Conservation Service, Saturated hydraulic conductivity, water movement concept and class history, accessed March 15, 2013, at <http://soils.usda.gov/technical/technotes/note6.html>.
- Oschwald, W.R., Riecken, F.F., Dideriksen, R.I., Scholtes, W.H., and Schaller, F.W., 1965, Principal soils of Iowa: Ames, Iowa, Iowa State University, Department of Agronomy, Special Report no. 42, 77 p.
- Parrett, C., Veilleux, A.G., Stedinger, J.R., Barth, N.A., Knifong, D.L., and Ferris, J.C., 2011, Regional skew for California, and flood frequency for selected sites in the Sacramento–San Joaquin River Basin, based on data through water year 2006: U.S. Geological Survey Scientific Investigations Report 2010-5260, 94 p. (Also available at <http://pubs.usgs.gov/sir/2010/5260/>.)
- Prior, J.C., 1991, Landforms of Iowa: Iowa City, University of Iowa Press, 154 p., accessed March 15, 2013, at <http://www.igsb.uiowa.edu/Browse/landform.htm>.
- Prior, J.C., Kohrt, C.J., and Quade, D.J., 2009, The landform regions of Iowa, vector digital data: Iowa City, Iowa, Iowa Geological Survey, Iowa Department of Natural Resources, accessed March 15, 2013, at [ftp://ftp.igsb.uiowa.edu/gis\\_library/ia\\_state/geologic/landform/landform\\_regions.zip](ftp://ftp.igsb.uiowa.edu/gis_library/ia_state/geologic/landform/landform_regions.zip).

- Parameter-Elevation Regressions on Independent Slopes Model Climate Group, 2008, Normal annual precipitation grid for the conterminous United States: Oregon State University, accessed March 15, 2013, at [http://www.prism.oregonstate.edu/state\\_products/maps.phtml?id=US](http://www.prism.oregonstate.edu/state_products/maps.phtml?id=US), <http://www.prism.oregonstate.edu/pub/prism/docs/prisguid.pdf>.
- Rantz, S.E., and others, 1982, Measurement and computation of streamflow—Volume 1, Measurement of stage and discharge, and volume 2, Computation of discharge: U.S. Geological Survey Water-Supply Paper 2175, 631 p. (Also available at <http://pubs.usgs.gov/wsp/wsp2175/>.)
- Reis, D.S., Jr., Stedinger, J.R., and Martins, E.S., 2005, Bayesian generalized least squares regression with application to the log Pearson type 3 regional skew estimation: *Water Resources Research*, v. 41, W10419, doi:10.1029/2004WR003445, accessed March 15, 2013, at <http://www.agu.org/journals/wr/wr0510/2004WR003445/2004WR003445.pdf>.
- Ries, K.G., and Friesz, P.J., 2000, Methods for estimating low-flow statistics for Massachusetts streams: U.S. Geological Survey Investigations Report 00–4135, 81 p. (Also available at <http://pubs.usgs.gov/wri/wri004135/>.)
- Ries, K.G., III, Guthrie, J.D., Rea, A.H., Steeves, P.A., and Stewart, D.W., 2008, StreamStats—A water resources web application: U.S. Geological Survey Fact Sheet 2008–3067, 6 p. (Also available at <http://pubs.usgs.gov/fs/2008/3067/>.)
- Risley, John, Stonewall, Adam, and Haluska, Tana, 2008, Estimating flow-duration and low-flow frequency statistics for unregulated streams in Oregon: U.S. Geological Survey Scientific Investigations Report 2008–5126, 22 p. (Also available at <http://pubs.usgs.gov/sir/2008/5126/>.)
- Ryberg, K.R., 2008, PFReports—A program for systematic checking of annual peaks in NWISWeb: U.S. Geological Survey Open-File Report 2008–1284, 17 p. (Also available at: <http://pubs.usgs.gov/of/2008/1284/>.)
- Sauer, V.B., 1974, Flood characteristics of Oklahoma streams—Techniques for calculating magnitude and frequency of floods in Oklahoma, with compilations of flood data through 1971: U.S. Geological Survey Water-Resources Investigations Report 52–73, 301 p.
- Schwob, H.H., 1953, Iowa floods—Magnitude and frequency: Iowa Highway Research Board Bulletin 1, 171 p.
- Schwob, H.H., 1966, Magnitude and frequency of Iowa floods: Iowa Highway Research Board Bulletin 28, part I, 47 p.
- Simley, J.D., and Carswell, W.J., Jr., 2009, The National Map—Hydrography: U.S. Geological Survey Fact Sheet 2009–3054, 4 p. (Also available at <http://pubs.usgs.gov/fs/2009/3054/>, also see <http://nhd.usgs.gov/>.)
- Sloto, R.A., and Crouse, M.Y., 1996, HYSEP—A computer program for streamflow hydrograph separation and analysis: U.S. Geological Survey Water-Resources Investigations Report 96–4010, 46 p. (Also available at <http://pubs.usgs.gov/wri/1996/4040/report.pdf>.)
- Soenksen, P.J., and Eash, D.A., 1991, Iowa floods and droughts, in Paulson, R.W., ed., National Water Summary 1988–89—Hydrologic events and floods and droughts: U.S. Geological Survey Water-Supply Paper 2375, 591 p. (Also available at <http://pubs.er.usgs.gov/publication/wsp2375>.)
- Soil Survey Staff, 2012, Natural Resources Conservation Service, U.S. Department of Agriculture, Soil Survey Geographic (SSURGO) Database [for all counties included in the study area shown in figure 1]: accessed March 15, 2013, at <http://soildatamart.nrcs.usda.gov/>.
- Stedinger, J.R., and Tasker, G.D., 1985, Regional hydrologic analysis 1—Ordinary, weighted, and generalized least square compared: *Water Resources Research*, v. 21, no. 9, p. 1421–1432, accessed March 15, 2013, at <http://www.agu.org/journals/wr/v021/i009/WR021i009p01421/WR021i009p01421.pdf>.
- Tasker, G.D., and Driver, N.E., 1988, Nationwide regression models for predicting urban runoff water quality at unmonitored sites: *Water Resources Bulletin*, v. 24, no. 5, p. 1091–1101, accessed March 15, 2013, at <http://onlinelibrary.wiley.com/doi/10.1111/j.1752-1688.1988.tb03026.x/pdf>.
- Tasker, G.D., and Stedinger, J.R., 1989, An operational GLS model for hydrologic regression: *Journal of Hydrology*, v. 111, p. 361–375, accessed March 15, 2013, at [http://www.sciencedirect.com/science?\\_ob=ArticleURL&\\_udi=B6V6C-487DD3J-SK&\\_user=696292&\\_coverDate=12%2F31%2F1989&\\_rdoc=1&\\_fmt=high&\\_orig=gateway&\\_origin=gateway&\\_sort=d&\\_docanchor=&view=c&\\_searchStrId=1710988392&\\_rerunOrigin=google&\\_acct=C000038819&\\_version=1&\\_urlVersion=0&\\_userid=696292&md5=f5f3116e25045e16b991f7d08c4f4603&searchtype=a](http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6V6C-487DD3J-SK&_user=696292&_coverDate=12%2F31%2F1989&_rdoc=1&_fmt=high&_orig=gateway&_origin=gateway&_sort=d&_docanchor=&view=c&_searchStrId=1710988392&_rerunOrigin=google&_acct=C000038819&_version=1&_urlVersion=0&_userid=696292&md5=f5f3116e25045e16b991f7d08c4f4603&searchtype=a).
- TIBCO Software Inc., 2008, TIBCO Spotfire S+ 8.1 for Windows® User's Guide: Palo Alto, California, 582 p., accessed March 15, 2013, at <http://www.msi.co.jp/splus/support/download/V81/getstart.pdf>.
- United Nations Educational, Scientific and Cultural Organization, IDAMS Statistical Software, Partitioning around medoids, accessed March 15, 2013, at [http://www.unesco.org/webworld/idams/advguide/Chapt7\\_1\\_1.htm](http://www.unesco.org/webworld/idams/advguide/Chapt7_1_1.htm).
- U.S. Army Corps of Engineers, 2009, Iowa River regulated flow study, final report: Rock Island District, 78 p.

- U.S. Army Corps of Engineers, 2010, Des Moines River regulated flow frequency study: Rock Island District, 82 p., accessed March 15, 2013, at <http://www.mvr.usace.army.mil/Missions/FloodRiskManagement/DesMoinesFlowFrequencyStudy.aspx>.
- U.S. Department of Agriculture, Natural Resources Conservation Service, 2012, Geospatial data gateway: U.S. Department of Agriculture, accessed March 15, 2013, at <http://datagateway.nrcs.usda.gov/>.
- U.S. Geological Survey, 2012, National Water Information System data available on the World Wide Web (USGS Water Data for the Nation), accessed March 15, 2013, at <http://nwis.waterdata.usgs.gov/usa/nwis/peak>.
- U.S. Geological Survey, Iowa Water Science Center, Flood information at selected bridge and culvert sites, accessed March 15, 2013, at <http://ia.water.usgs.gov/projects/ia006.html>.
- U.S. Geological Survey, National hydrography dataset: U.S. Geological Survey, accessed March 15, 2013, at <http://nhd.usgs.gov/>.
- U.S. Geological Survey, National elevation dataset: U.S. Geological Survey, accessed March 15, 2013, at <http://ned.usgs.gov/>.
- U.S. Geological Survey, Welcome to StreamStats: U.S. Geological Survey, accessed March 15, 2013, at <http://water.usgs.gov/osw/streamstats/index.html>. [also see <http://water.usgs.gov/osw/streamstats/bcdefinitions1.html> and <http://streamstats.usgs.gov/ungaged2.html>.]
- U.S. Geological Survey and U.S. Department of Agriculture, Natural Resources Conservation Service, 2009, Federal guidelines, requirements, and procedures for the National Watershed Boundary Dataset: U.S. Geological Survey Techniques and Methods, book 11, chap. A3, 55 p. (Also available at <http://pubs.usgs.gov/tm/tm11a3/>.) [also see <http://datagateway.nrcs.usda.gov/>.]
- Veilleux, A.G., 2011, Bayesian GLS regression, leverage and influence for regionalization of hydrologic statistics: Cornell, Cornell University, Ph.D. dissertation.
- Veilleux, A.G., Stedinger, J.R., and Eash, D.A., 2012, Bayesian WLS/GLS regression for regional skewness analysis for crest stage gage networks, in Loucks, E.D., ed., Proceedings World Environmental and Water Resources Congress, Crossing boundaries, May 20–24, 2012: Albuquerque, New Mexico, American Society of Civil Engineering, paper 227, p. 2253–2263. (Also available at <http://ia.water.usgs.gov/media/pdf/report/Veilleux-Stedinger-Eash-EWRI-2012-227R.pdf>.)
- Verdi, R.J., and Dixon, J.F., 2011, Magnitude and frequency of floods for rural streams in Florida, 2006: U.S. Geological Survey Scientific Investigations Report 2011–5034, 69 p., 1 pl. (Also available at <http://pubs.usgs.gov/sir/2011/5034/>.)
- Wahl, K.L., 1998, Sensitivity of non-parametric trend analyses to multi-year extremes, in Proceeding of the Western Snow Conference, April 20–23, 1998: Snowbird, Utah, Western Snow Conference, p. 157–160.
- Weaver, J.C., Feaster, T.D., and Gotvald, A.J., 2009, Magnitude and frequency of rural floods in the southeastern United States, 2006—Volume 2, North Carolina: U.S. Geological Survey Scientific Investigations Report, 2009–5158. (Also available at <http://pubs.usgs.gov/sir/2009/5158/>.)
- Zarriello, P.J., Ahearn, E.A., and Levin, S.B., 2012, Magnitude of flood flows for selected annual exceedance probabilities in Rhode Island through 2010: U.S. Geological Survey Scientific Investigations Report 2012–5109, 81 p. (Also available at <http://pubs.usgs.gov/sir/2012/5109/>.)



# Appendix

---

# Regional Skewness Regression

By Andrea G. Veilleux, U.S. Geological Survey and Jerry R. Stedinger, Cornell University, Ithaca, New York

## Introduction to Statistical Analysis of Regional Skew

For the log-transformation of annual peak discharges, Bulletin 17B (Interagency Advisory Committee on Water Data, 1982) recommends using a weighted average of the station skew coefficient and a regional skew coefficient to help improve estimates of annual exceedance-probability discharges (AEPDs) (eq. 2 in report). Bulletin 17B supplies a national map, but also encourages the hydrologist to develop more specific local relations. Since the first map was published in 1976, some 35 years of additional information has accumulated, and better spatial estimation procedures have been developed (Stedinger and Griffis, 2008).

Tasker and Stedinger (1986) developed a weighted least-squares (WLS) procedure for estimating regional skew coefficients based on sample skew coefficients for the logarithms of annual peak-discharge data. Their method of regional analysis of skewness estimators accounts for the precision of the skew-coefficient estimate for each streamgage or station, which depends on the length of record for each streamgage and the accuracy of an ordinary least-squares (OLS) regional mean skewness. More recently, 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 Bayesian methodology allows for the computation of a posterior distribution of the regression parameters and the model error variance. As documented in Reis and others (2005), for cases in which the model error variance is small compared to the sampling error of the station estimates, the Bayesian posterior distribution provides a more reasonable description of the model error variance than the GLS method-of-moments and maximum likelihood point estimates (Veilleux, 2011). Although WLS regression accounts for the precision of the regional model and the effect of the record length on the variance of skew-coefficient estimators, GLS regression also considers the cross-correlations among the skew-coefficient estimators. In some studies, the cross-correlations have had a large effect on the precision attributed to different parameter estimates (Parrett and others, 2011; Feaster and others, 2009; Gotvald and others, 2009; Weaver and others, 2009).

Because of complications introduced by the use of the expected moments algorithm (EMA/MGB) with multiple Grubbs-Beck censoring of low outliers (Cohn and others, 1997) 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 skewness (Veilleux and others, 2012; Veilleux, 2011; Lamontagne and others, 2012). This alternate procedure is referred to as the Bayesian WLS/Bayesian GLS (B-WLS/B-GLS) regression framework (Veilleux and others, 2012; Veilleux, 2011; Veilleux and others, 2011). It uses an OLS analysis to fit an initial regional skewness model; that OLS model is then used to generate a stable regional skew-coefficient estimate for each site. That stable regional estimate is the basis for computing the variance of each station skew-coefficient estimator employed in the 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.

To provide cost effective peak-discharge data for smaller drainage basins in the study area, the U.S. Geological Survey (USGS) operates a large network of crest-stage gages (CSGs) that only measure discharges above a minimum recording threshold (thus producing a censored data record). CSGs are different from continuous-record streamgages, which measure almost all discharges and have been used in previous B-GLS and B-WLS/B-GLS regional skew studies. Thus, although the Iowa regional skew study described here did not exhibit large cross-correlations between annual peak discharges, it did make extensive use of EMA/MGB to estimate the station skew and its mean square error. Because EMA/MGB allows for the censoring of low outliers, as well as the use of estimated interval discharges for missing, censored, and historic data, it complicates the calculations of effective record length (and effective concurrent record length) used to describe the precision of sample estimators because the peak discharges are no longer solely represented by single values. To properly account for these complications, the new B-WLS/B-GLS procedure was employed. The steps of this alternative procedure are described in the following sections.

## Methodology for Regional Skewness Model

This section provides a brief description of the B-WLS/B-GLS methodology (as it appears in Veilleux and others, 2012). Veilleux and others (2011) and Veilleux (2011) provide a more detailed description.

## OLS Analysis

The first step in the B-WLS/B-GLS regional skewness analysis is the estimation of a regional skewness model using OLS. The OLS regional regression yields parameters  $\hat{\beta}_{OLS}$  and a model that can be used to generate unbiased and relatively stable regional estimates of the skewness for all streamgages:

$$\tilde{y}_{OLS} = X\hat{\beta}_{OLS} \quad (A1)$$

where

- $\tilde{y}_{OLS}$  are the estimated regional skewness values,
- $X$  is an  $(n \times k)$  matrix of basin characteristics,
- $n$  is the number of streamgages, and
- $k$  is the number of basin parameters including a column of ones to estimate the constant.

These estimated regional skewness values  $\tilde{y}_{OLS}$  are then used to calculate unbiased station-regional skewness variances using the equations reported in Griffis and Stedinger (2009). These station-regional skewness variances are based on the regional OLS estimator of the skewness coefficient instead of the station skewness estimator, thus making the weights in the subsequent steps relatively independent of the station skewness estimates.

## WLS Analysis

A B-WLS analysis is used to develop estimators of the regression coefficients for each regional skewness model (Veilleux, 2011; Veilleux and others, 2011). The WLS analysis explicitly reflects variations in record length, but intentionally neglects cross correlations thereby avoiding the problems experienced with GLS parameter estimators (Veilleux, 2011; Veilleux and others, 2011).

## GLS Analysis

After the regression model coefficients,  $\hat{\beta}_{WLS}$ , are determined with a WLS analysis, the precision of the fitted model and the precision of the regression coefficients are estimated using a B-GLS analysis (Veilleux, 2011; Veilleux and others, 2011). Precision metrics include the standard error of the regression parameters,  $SE(\hat{\beta}_{WLS})$ , and the model error variance,  $\sigma_{\delta, B-GLS}^2$ , pseudo  $R_{\delta}^2$  as well as the average variance of prediction at a streamgage not used in the regional model,  $AVP_{new}$ .

## Data Analysis

The statistical analysis of the data requires several steps. This section describes a redundant site analysis, the calculations for pseudo record length for each site given the number of censored observations and concurrent record lengths, as well as the development of the model of cross-correlations of concurrent annual peak discharges.

## Data for Iowa Regional Skew Study

This study is based on annual peak-discharge data from 330 streamgages in Iowa and the surrounding states. The annual peak-discharge data were downloaded from the USGS National Water Information System (NWIS) database (U.S. Geological Survey, 2012). In addition to the peak-discharge data, over 65 basin characteristics for each of the 330 sites were available as explanatory variables in the regional study. The basin characteristics available include percent of basin contained within different hydrologic regions, as well as the more standard morphometric parameters such as location of the basin centroid, drainage area, main channel slope, and basin shape among others.

## Redundant Sites

Redundancy results when the drainage basins of two streamgages are nested, meaning that one is contained inside the other and the two basins are of similar size. Then, instead of providing two independent spatial observations, depicting how drainage-basin characteristics are related to skew (or AEPDs), these two basins will have the same hydrologic response to a given storm, and thus represent only one spatial observation. When sites are redundant, a statistical analysis using both streamgages incorrectly represents the information in the regional dataset (Gruber and Stedinger, 2008). To determine if two sites are redundant and thus represent the same hydrologic experience, two pieces of information are considered: (1) if their watersheds are nested, and (2) the ratio of the basin drainage areas.

The standardized distance (SD) is used to determine the likelihood the basins are nested. The standardized distance between two basin centroids, SD, is defined as

$$SD_{ij} = \frac{D_{ij}}{\sqrt{0.5(DA_i + DA_j)}}, \quad (A2)$$

where

- $D_{ij}$  is the distance between centroids of basin  $i$  and basin  $j$ , and
- $DA_i$  and  $DA_j$  are the drainage areas at sites  $i$  and  $j$ .

The drainage area ratio, DAR, is used to determine if two nested basins are sufficiently similar in size to conclude that they are essentially, or are at least in large part, the same watershed for the purposes of developing a regional hydrologic model. The drainage area ratio of two basins, DAR, is defined as (Veilleux, 2009)

$$DAR = \text{Max} \left[ \frac{DA_i}{DA_j}, \frac{DA_j}{DA_i} \right], \quad (A3)$$

where

- $DA_i$  and  $DA_j$  have already been defined in equation A2.

Two basins might be expected to have possible redundancy if the basin sizes are similar and the basins are nested. Previous studies suggest that site-pairs having SD less than or equal to 0.50 and DAR less than or equal to 5 likely were to

have possible redundancy problems for purposes of determining regional skew. If DAR is large enough, even if the sites are nested, they will reflect different hydrologic experiences because they respond to storms of different sizes and durations.

*Table A1* (link to Excel file) shows the results of the redundant site screening on the Iowa regional skew data. All possible combinations of site-pairs from the 330 streamgages were considered in the redundancy analysis (including all types of streamgages: continuous-record streamgages, CSGs, and both continuous-record streamgages/CSGs). To be conservative, all site-pairs with SD less than 0.75 and DAR less than 8 were identified as possible redundant site-pairs. All sites identified as redundant were then investigated to determine if in fact one site of the pair was nested inside the other. For site-pairs that were nested, one site from the pair was removed from the regional skew analysis. Sites removed from the Iowa regional skew study because of redundancy are identified in *table A1* (link to Excel file) as “no - R.”

From the 95 identified possible redundant site-pairs, 87 were found to be redundant and 55 sites were removed. Two sites also were removed because of backwater effects and they are identified in *table A1* (link to Excel file) as “no - B.” Thus, of the 330 sites, 57 have been removed because of redundancy and backwater, leaving 273 sites for the Iowa regional skew study.

## Station Skewness Estimators

The EMA/MGB analysis method was used to estimate the station log10 skew coefficient,  $G$ , and its mean square error,  $MSE_G$  (Cohn and others, 1997; Griffis and others, 2004). EMA/MGB provides a straightforward and efficient method for the incorporation of historical information and censored data, such as those from a CSG, contained in the record of annual peak discharges for a streamgage. PeakfqSA, an EMA/MGB software program developed by Cohn (2011), is used to generate the station log10 estimates of  $G$  and its  $MSE_G$ , assuming a log-Pearson Type III distribution and employing a multiple Grubbs-Beck test for low-outlier screening. EMA/MGB estimates, based on annual peak-discharge data through September 30, 2008, of  $G$  and its  $MSE_G$  are listed in *table A1* (link to Excel file) for the 330 streamgages evaluated for the Iowa regional skew study. [See sections Expected Moments Algorithm (EMA/MGB) Analyses and Multiple Grubbs-Beck Test for Detecting Low Outliers for more detail regarding EMA/MGB.]

## Pseudo Record Length

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 should be 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 following calculations provide a pseudo record length,  $P_{RL}$ , which appropriately accounts for all peak-discharge data types available for a site.  $P_{RL}$  equals the systematic record length if such a complete record is all that is available for a site.

The first step is to run EMA/MGB with all available information, including historic information and censored peaks (denoted EMA/MGB<sub>c</sub>, for EMA/MGB complete). From the EMA/MGB run, the station skewness without regional information  $\hat{G}_c$  and the MSE of that skewness estimator  $MSE(\hat{G}_c)$  are extracted, as well as the year the historical period begins,  $YB_c$ , the year the historical period ends,  $YE_c$ , and the length of the historical period  $H_c$ . ( $YB_c$ ,  $YE_c$ , and  $H_c$  are used in equation A12.)

The second step is to run EMA/MGB with only the systematic peaks (denoted EMA/MGB<sub>s</sub>, for EMA/MGB systematic). From the EMA/MGB<sub>s</sub> analysis, the station skewness without regional information  $\hat{G}_s$  and the MSE of that skewness estimator,  $MSE(\hat{G}_s)$ , are extracted, as well as the number of peaks  $P_s$ . ( $P_s$  is used in equation A6.)

The third step is to represent, from EMA/MGB<sub>c</sub> and EMA/MGB<sub>s</sub>, the precision of the skewness estimators as two record lengths,  $RL_c$  and  $RL_s$ , based on the estimated skew and MSE. The corresponding record lengths are calculated using equation A4 below from Griffis and others (2004) and Griffis and Stedinger (2009):

$$MSE(\hat{G}) = \left[ \frac{6}{RL} + a(RL) \right] * \left[ 1 + \left( \frac{9}{6} + b(RL) \right) \hat{G}^2 + \left( \frac{15}{48} + c(RL) \right) \hat{G}^4 \right] \quad (A4)$$

$$a(RL) = -\frac{17.75}{RL^2} + \frac{50.06}{RL^3}$$

$$b(RL) = \frac{3.93}{RL^{0.3}} - \frac{30.97}{RL^{0.6}} + \frac{37.1}{RL^{0.9}}$$

$$c(RL) = -\frac{6.16}{RL^{0.56}} + \frac{36.83}{RL^{1.12}} - \frac{66.9}{RL^{1.68}}$$

where

$$\begin{array}{ll} RL_c & \text{uses } \hat{G}_c \text{ and } MSE(\hat{G}_c), \text{ and} \\ RL_s & \text{uses } \hat{G}_s \text{ and } MSE(\hat{G}_s). \end{array}$$

Next, the difference between  $RL_c$  and  $RL_s$  is employed as a measure of the extra information provided by the historic or censored information, or both, that was included in the EMA/MGB<sub>c</sub> analysis, but not in the EMA/MGB<sub>s</sub> analysis.

$$RL_{diff} = RL_c - RL_s \quad (A5)$$

The pseudo record length for the entire record at the streamgage,  $P_{RL}$ , is calculated using  $RL_{diff}$  from equation A5 and the number of systematic peaks  $P_s$ :

$$P_{RL} = RL_{diff} + P_s \quad (A6)$$

$P_{RL}$  must be non-negative. If  $P_{RL}$  is greater than  $H_C$ , then  $P_{RL}$  should be set to equal  $H_C$ . Also if  $P_{RL}$  is less than  $P_S$ , then  $P_{RL}$  is set to  $P_S$ . This ensures that the pseudo record length will not be larger than the complete historical period or less than the number of systematic peaks.

## Unbiasing the Station Estimators

The station skewness estimates are unbiased by using the correction factor developed by Tasker and Stedinger (1986) and employed in Reis and others (2005). The unbiased station skewness estimator using the pseudo record length is

$$\hat{Y}_i = \left[ 1 + \frac{6}{P_{RL,i}} \right] G_i \quad (A7)$$

where

- $\hat{Y}_i$  is the unbiased station sample skewness estimate for site  $i$ ,
- $P_{RL,i}$  is the pseudo record length for site  $i$  as calculated in equation A6, and
- $G_i$  is the traditional biased station skewness estimator for site  $i$  from EMA/MGB.

The variance of the unbiased station skewness includes the correction factor developed by Tasker and Stedinger (1986):

$$Var[\hat{Y}_i] = \left[ 1 + \frac{6}{P_{RL,i}} \right]^2 Var[G_i] \quad (A8)$$

where

$Var[G_i]$  is calculated using (Griffis and Stedinger, 2009).

$$Var(\hat{G}) = \left[ \frac{6}{P_{RL}} + a(P_{RL}) \right] * \left[ 1 + \left( \frac{9}{6} + b(P_{RL}) \right) \hat{G}^2 + \left( \frac{15}{48} + c(P_{RL}) \right) \hat{G}^4 \right] \quad (A9)$$

$$a(P_{RL}) = -\frac{17.75}{P_{RL}^2} + \frac{50.06}{P_{RL}^3}$$

$$b(P_{RL}) = \frac{3.92}{P_{RL}^{0.3}} - \frac{31.10}{P_{RL}^{0.6}} + \frac{34.86}{P_{RL}^{0.9}}$$

$$c(P_{RL}) = -\frac{7.31}{P_{RL}^{0.59}} + \frac{45.90}{P_{RL}^{1.18}} - \frac{86.50}{P_{RL}^{1.77}}$$

## Estimating the Mean Square Error of the Skewness Estimator

There are several possible ways to estimate  $MSE_G$ . The approach used by EMA/MGB [taken from eq. 55 in Cohn and others (2001)] generates a first order estimate of the  $MSE_G$ , which should perform well when interval data are present. Another option is to use the Griffis and Stedinger (2009) formula in equation 8 (the variance is equated to the MSE), employing either the systematic record length or the length of the whole historical period; however, this method does not account for censored data, and thus can lead to inaccurate and

underestimated  $MSE_G$ . This issue has been addressed by using the pseudo record length instead of the length of the historical period; the pseudo record length reflects the effect of the censored data and the number of recorded systematic peaks. Figure A1 compares the unbiased  $MSE_G$  estimates from the Griffis and Stedinger (2009) approach based on pseudo record lengths and regional skewness estimates, and the unbiased EMA/MGB  $MSE_G$  estimates based on the estimated station skewness.

As shown in figure A1, for those streamgages with  $MSE_G$  less than about 0.4, the two methods generate similar  $MSE_G$ ; however, for 33 streamgages, EMA/MGB generates unreasonably large  $MSE_G$  with values greater than about 0.4. For these sites, the Griffis and Stedinger (2009) formula does not generate a  $MSE_G$  greater than 0.5. For these 33 streamgages EMA/MGB is having trouble estimating the  $MSE_G$  due at least in part to the number of censored observations. Of these 33 sites with EMA/MGB unbiased  $MSE_G$  greater than 0.4, 45-percent of the sites had 50 percent or more of their record comprised of censored observations, whereas 81 percent of the sites had 20 percent or more of their record comprised of censored observations. Also, the average  $P_{RL}$  for all 273 sites in the Iowa study is 49 years; however, the longest record of the 33 sites with EMA/MGB unbiased  $MSE_G$  greater than 0.4 is 43 years, with 85 percent of the 33 sites having  $P_{RL}$  less than or equal to 35 years and 42 percent of the 33 sites have  $P_{RL}$  less than or equal to 25 years. Thus, it appears that for those sites with shorter record lengths and a large percentage of their record comprised of censored observations, EMA/MGB has trouble estimating the  $MSE_G$ . For this reason, these 33 sites with EMA/MGB unbiased  $MSE_G$  greater than 0.4 were removed from the analysis. Thus, there are 240 streamgages remaining from which to build a regional skewness model for the Iowa study area. Figure 6 shows the location of the basin centroids for these 240 streamgages. The unbiased Griffis and Stedinger (2009)  $MSE_G$  is used in the regional skewness model because it is more stable and relatively independent of the station skewness estimator. The 33 sites removed from the Iowa regional skew study because of a  $MSE_G$  greater than 0.4 as estimated by EMA/MGB are identified in [table A1](#) (link to Excel file) as “no – E.”

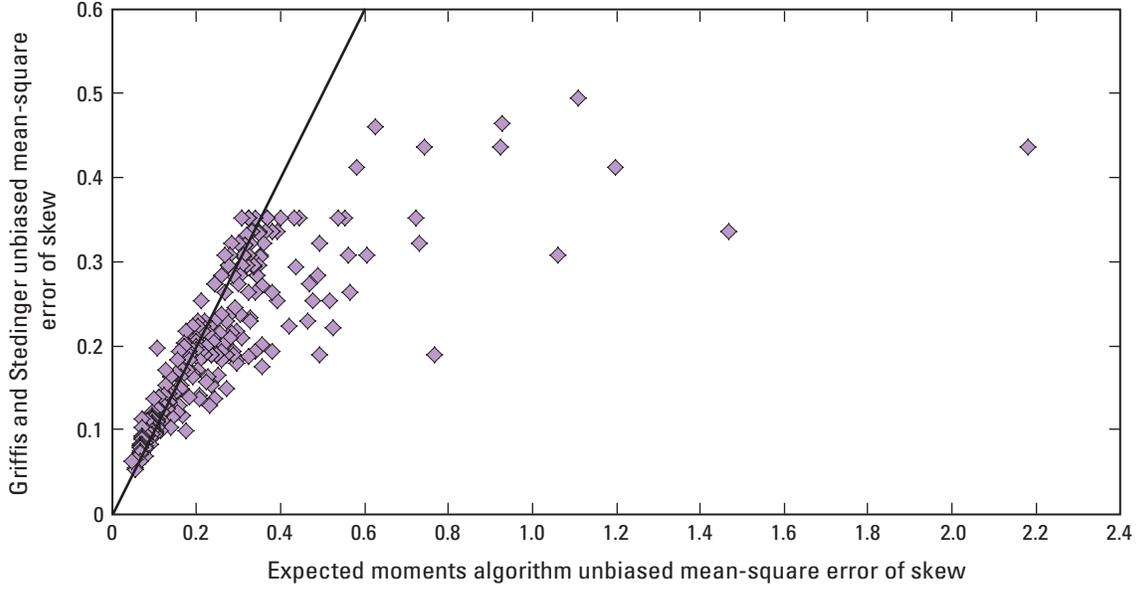
## Cross-Correlation Models

A critical step for a GLS analysis is estimation of the cross-correlation of the skewness coefficient estimators. Martins and Stedinger (2002) used Monte Carlo experiments to derive a relation between the cross-correlation of the skewness estimators at two stations  $i$  and  $j$  as a function of the cross-correlation of concurrent annual maximum flows,  $\rho_{ij}$ :

$$\hat{\rho}(\hat{y}_i, \hat{y}_j) = Sign(\hat{\rho}_{ij}) c_{ij} |\hat{\rho}_{ij}|^k \quad (A10)$$

where

$\hat{\rho}_{ij}$  is the cross-correlation of concurrent annual



**Figure A1.** Comparison of EMA/MGB and Griffis and Stedinger (2009)  $MSE_g$  estimates of station skewness estimators for each of the 273 streamgages in the Iowa regional skewness study.

peak discharge for two streamgages,  $k$  is a constant between 2.8 and 3.3, and  $cf_{ij}$  is a factor that accounts for the sample size difference between stations and their concurrent record length, is defined in the following equation:

$$cf_{ij} = CY_{ij} / \sqrt{(P_{RL,i})(P_{RL,j})} \quad (A11)$$

where

$CY_{ij}$  is the pseudo record length of the period of concurrent record, and  $P_{RL,i}$ ,  $P_{RL,j}$  is the pseudo record length corresponding to sites  $i$  and  $j$ , respectively (see equation A6)

### Pseudo Concurrent Record Length

After calculating the  $P_{RL}$  for each streamgage in the study, the pseudo concurrent record length between pairs of sites can be calculated. Because of the use of censored data and historic data, the effective concurrent record-length calculation is more complex than determining in which years the two streamgages have recorded systematic peaks.

The years of historical record in common between the two streamgages is first determined. For the years in common, with beginning year  $YB_{ij}$  and ending year  $YE_{ij}$ , the following equation is used to calculate the concurrent years of record between site  $i$  and site  $j$ .

$$CY_{ij} = (YE_{ij} - YB_{ij} + 1) \left( \frac{P_{RL,i}}{H_{C,i}} \right) \left( \frac{P_{RL,j}}{H_{C,j}} \right) \quad (A12)$$

The computed pseudo concurrent record length depends on the years of historical record in common between the two streamgages, as well as the ratios of the pseudo record length to the historical record length for each of the two streamgages.

### Iowa Study Area Cross-Correlation Model of Concurrent Annual Peak Discharge

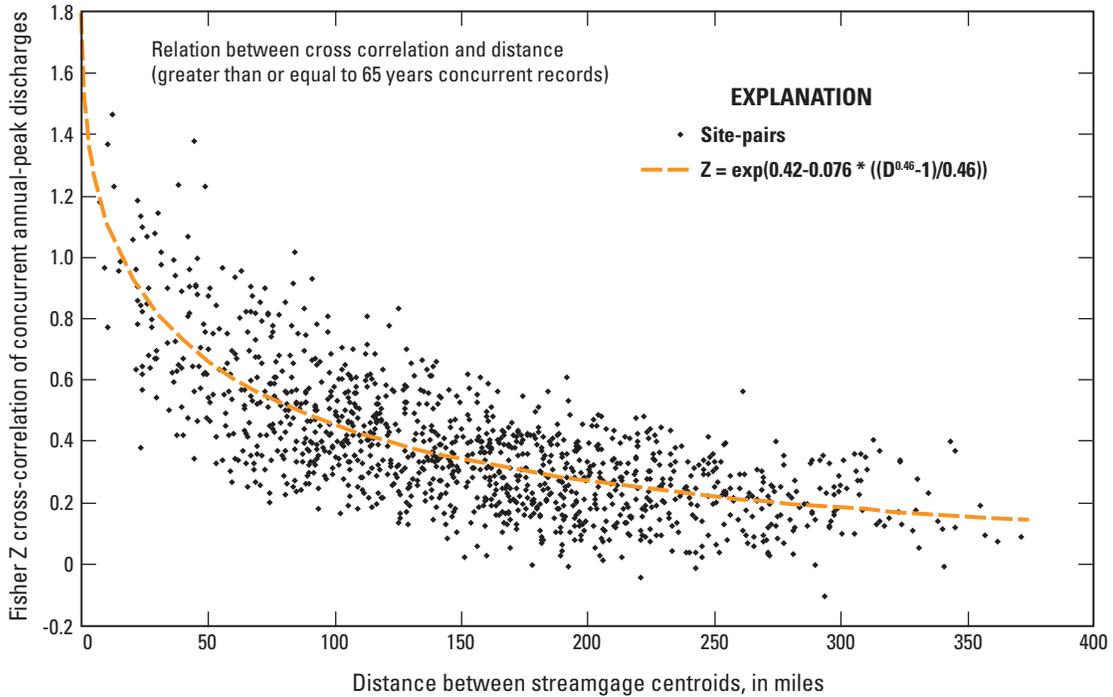
A cross-correlation model for the log annual peak discharges in the Iowa study area were developed using 53 sites with at least 65 years of concurrent systematic peaks (zero flows not included). Various models relating the cross-correlation of the concurrent annual peak discharge at two sites,  $p_{ij}$ , to various basin characteristics were considered. A logit model, termed the Fisher Z Transformation ( $Z = \log[(1+r)/(1-r)]$ ), provided a convenient transformation of the sample correlations  $r_{ij}$  from the (-1, +1) range to the  $(-\infty + \infty)$  range. The adopted model for estimating the cross-correlations of concurrent annual peak discharge at two stations, which used the distance between basin centroids,  $D_{ij}$ , as the only explanatory variable, is

$$\rho_{ij} = \left( \frac{\exp(2Z_{ij}) - 1}{\exp(2Z_{ij}) + 1} \right) \quad (A13)$$

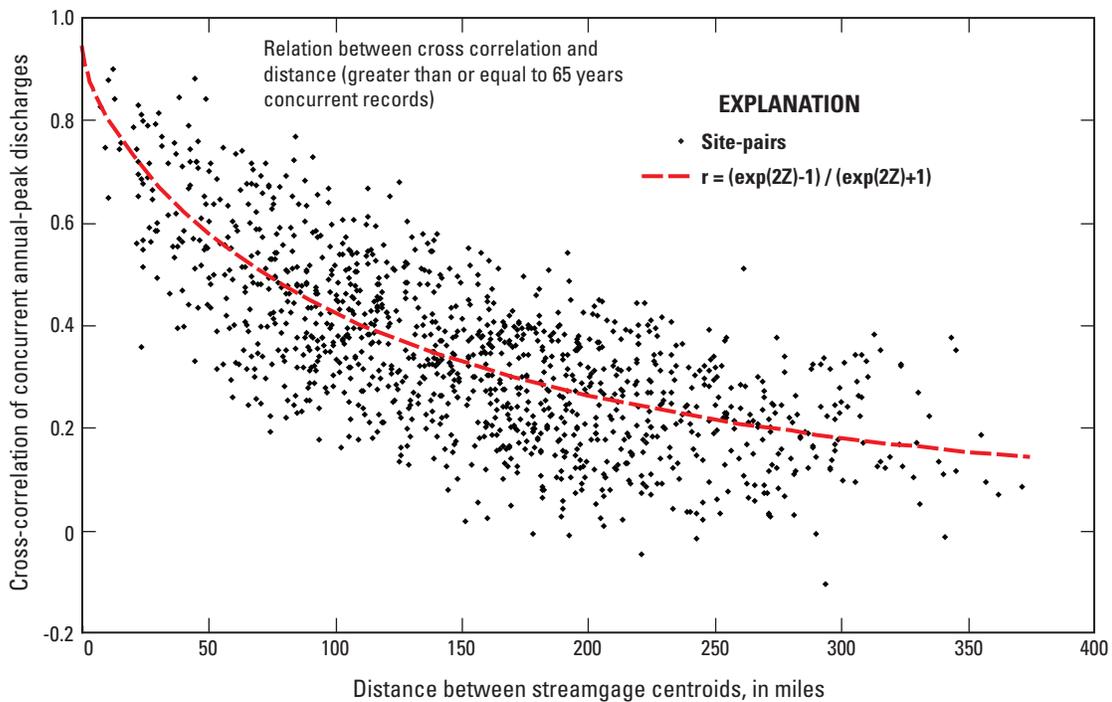
where

$$Z_{ij} = \exp \left( 0.42 - 0.076 \left( \frac{D_{ij}^{0.46} - 1}{0.46} \right) \right) \quad (A14)$$

An OLS regression analysis based on 1,164 station-pairs indicated that this model is as accurate as having 152 years of concurrent annual peaks from which to calculate cross-correlation. Figure A2 shows the fitted relation between  $Z$  and



**Figure A2.** Relation between Fisher Z transformed cross-correlation of logs of annual peak discharge and distance between basin centroids for 1,164 station-pairs with concurrent record lengths greater than or equal to 65 years from 53 streamgages in Iowa and neighboring states.



**Figure A3.** Relation between un-transformed cross-correlation of logs of annual peak discharge and distance between basin centroids based for 1,164 station-pairs with concurrent record lengths greater than or equal to 65 years from 53 streamgages in Iowa and neighboring states.

distance between basin centroids together with the plotted sample data from the 1,164 station pairs of data. Figure A3 shows the functional relation between the untransformed cross correlation and distance between basin centroids together with the plotted sample data from the 1,164 station pairs of data. The cross correlation model was used to estimate site-to-site cross correlations for concurrent annual peak discharges at all pairs of sites in the regional skew study.

## Iowa Regional Skew Study Results

The results of the Iowa regional skew study using the B-WLS/B-GLS regression methodology are provided below. All of the available basin characteristics initially were considered as explanatory variables in the regression analysis for regional skew. Available basin characteristics include precipitation (mean annual, mean monthly, maximum 24 hours for a number of years), soil (hydrologic soil types, soil type, soil permeability), stream characteristics (stream density, ruggedness, number of first order streams, total stream length), basin measures (drainage area, slope, relief, length, perimeter, shape factor), hydrologic parameters (base flow index, base flow recession), and hydrologic regions. A few basin characteristics statistically were significant in explaining the site-to-site variability in skewness, including slope, drainage area, basin length, and the total length of mapped streams in the basin. The best model, as classified by having the smallest model error variance,  $\sigma_\delta^2$ , and largest pseudo  $R_\delta^2$ , which included a constant and a parameter (or combination of parameters), was the model that included drainage area. Table A2 provides the final results for the constant skewness model denoted “Constant,” and the model that uses a linear relation between skewness and log10 drainage area.

Table A2 includes the pseudo  $R_\delta^2$  value for both models; pseudo  $R_\delta^2$  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  $R_\delta^2$  equals 0. The “DA” model has a pseudo  $R_\delta^2$  of 19 percent. The posterior mean of the model error variance,  $\sigma_\delta^2$ , for the

DA model is 0.12, which is smaller than that for the constant model for which  $\sigma_\delta^2 = 0.15$ . This indicates that the inclusion of drainage area as an explanatory variable in the regression helps explain some of the variability in the true skewness; however, this small gain in precision does not warrant the increased model complexity. Thus, the constant model is selected as the best regional model for Iowa study area skewness. The average sampling error variance (ASEV) in table A2 is the average error in the regional skewness estimator at the sites in the dataset. The average variance of prediction at a new site ( $AVP_{new}$ ) corresponds to the mean square error (MSE) used in Bulletin 17B to describe the precision of the generalized skewness. The constant model has an  $AVP_{new}$  equal to 0.16, which corresponds to an effective record length of 50 years. An  $AVP_{new}$  of 0.16 is a marked improvement over the Bulletin 17B national skew map, whose reported MSE is 0.302 (Interagency Committee on Water Data, 1982) for a corresponding effective record length of only 17 years. Thus, the new regional model has three times the information content (as measured by effective record length) of that calculated for the Bulletin 17B map.

Figure A4 shows the relation between the unbiased station skewness and drainage area; the marker selected for each streamgage represents the station pseudo record length. The sites with the largest drainage area generally have the longest pseudo record lengths. It is not apparent from the data that the upward trend, suggested by the DA model, occurs between the unbiased station skewness and drainage area. Thus, for this study, the simpler model is selected, in other words the constant model.

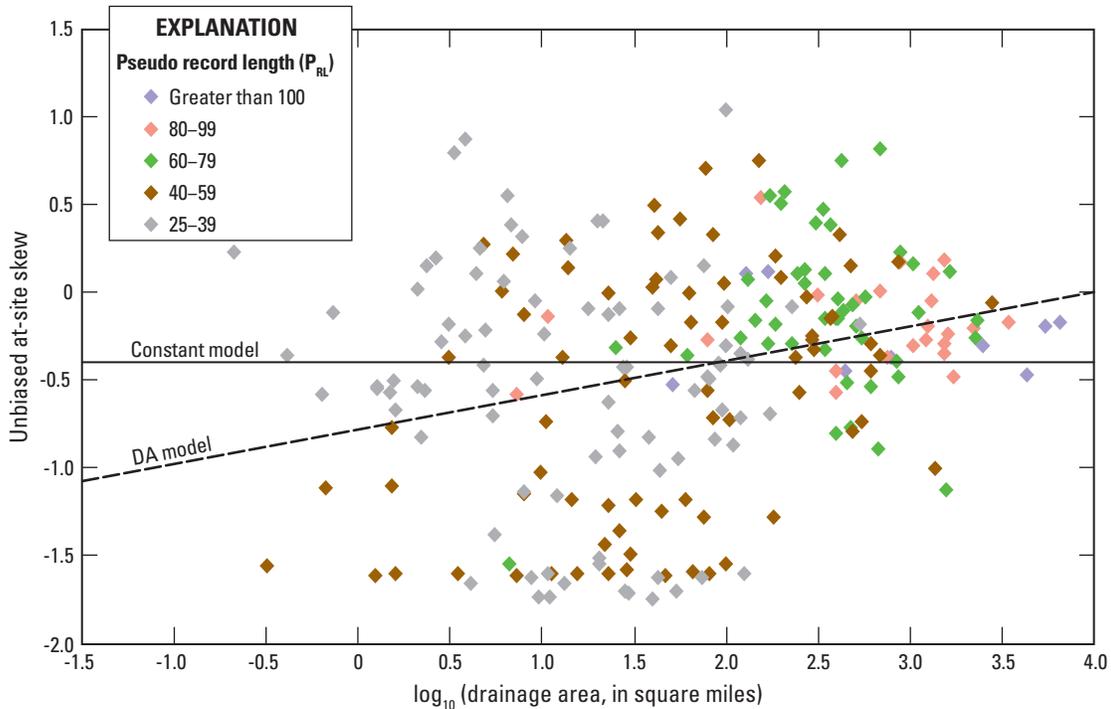
## B-WLS/B-GLS Regression Diagnostics

To determine if a model is a good representation of the data and which regression parameters, if any, should be included in a regression model, diagnostic statistics have been developed to evaluate how well a model fits a regional hydrologic dataset (Griffis, 2006; Gruber and others, 2008). In this study, the goal was to determine the set of possible explanatory variables that best fit annual peak discharges for the Iowa study area affording the most accurate skew prediction while

**Table A2.** Regional skewness models for Iowa study area.

[ $\sigma_\delta^2$  is the model error variance; ASEV, is the average sampling error variance;  $AVP_{new}$ , is the average variance of prediction for a new site; Pseudo  $R_\delta^2$  (%), describes the fraction of the variability in the true skews explained by each model (Gruber and others, 2007); %, percent;  $\hat{y}$ , unbiased regional skewness estimate; DA, drainage area. Standard deviations are in parentheses]

| Model                                    | Regression parameters |                      |                   |      |             |                     |
|--|-----------------------|----------------------|-------------------|------|-------------|---------------------|
|  | $b_1$                 | $b_2$                | $\sigma_\delta^2$ | ASEV | $AVP_{new}$ | Pseudo $R_\delta^2$ |
| Constant: $\hat{y} = b_1$                | -0.40<br>(0.09)<br>0% | -                    | 0.15<br>(0.03)    | 0.01 | 0.16        | 0%                  |
| DA: $\hat{y} = b_1 + b_2[\log_{10}(DA)]$ | -0.78<br>(0.16)<br>0% | 0.20<br>(0.05)<br>0% | 0.12<br>(0.02)    | 0.01 | 0.13        | 19%                 |



**Figure A4.** Relations between the unbiased station skew and drainage area for the 240 sites in the Iowa regional skew study. Each of the diamonds represents one of the 240 sites with each of the five different colors of the diamonds representing five different groupings of station pseudo record length ( $P_{RL}$ ); purple/violet = record length greater than 100 years, pink = record length between 80–99 years, green = record length between 60–79 years, brown = record length between 40–59 years, gray = record length between 25–39 years. The solid black line represents the constant model from table A2, while the dashed black line represents the DA model from table A2.

also keeping the model as simple as possible. This section presents the diagnostic statistics for a B-WLS/B-GLS analysis, and discusses the specific values obtained for the Iowa regional skew study.

Table A3 presents a Pseudo Analysis of Variance (Pseudo ANOVA) table for the Iowa regional skew analysis containing regression diagnostics/goodness of fit statistics.

In particular, the table describes how much of the variation in the observations can be attributed to the regional model, and how much of the residual variation can be attributed to model error and sampling error, respectively. Difficulties arise in determining these quantities. The model errors cannot be resolved because the values of the sampling errors  $n_i$  for each site  $i$ , are not known; however, the total sampling error sum of squares can be described by its mean value,  $\sum_{i=1}^n Var(\hat{y}_i)$ . Because there are  $n$  equations, the total variation because of the model error  $\delta$  for a model with  $k$  parameters has a mean equal to  $n\sigma_\delta^2(k)$ . Thus, the residual variation attributed to the sampling error is  $\sum_{i=1}^n Var(\hat{y}_i)$ , and the residual variation attributed to the model error is  $n\sigma_\delta^2(k)$ .

For a model with no parameters other than the mean (that is, the constant skew model), the estimated model error variance  $\sigma_\delta^2(0)$  describes all of the anticipated variation in  $y_i = \mu + \delta_i$ , where  $\mu$  is the mean of the estimated station sample skews. Thus, the total expected sum of squares variation because

of model error  $\delta_i$  and because of sampling error  $n_i = \hat{y}_i - y_i$  in expectation should equal  $n\sigma_\delta^2(0) + \sum_{i=1}^n Var(y\hat{y}_i)$ . Therefore, the expected sum of squares attributed to a regional skew model with  $k$  parameters equals  $n[\sigma_\delta^2(0) - \sigma_\delta^2(k)]$ , because the sum of the model error variance  $n\sigma_\delta^2(k)$  and the variance explained by the model must sum to  $n\sigma_\delta^2(0)$ . Table A3 considers models with  $k = 0$  and 1.

This division of the variation in the observations is referred to as a Pseudo ANOVA because the contributions of the three sources of error are estimated or constructed, rather than being determined from the computed residual errors and the observed model predictions, while also ignoring the effect of correlation among the sampling errors.

Table A3 compares the Pseudo ANOVA results for the constant model and log10(DA) model described in the report text. The log10(DA) model contains a constant and one explanatory variable, a linear function of drainage area. As described previously, the first step of the B-WLS/B-GLS regression procedure is to perform an OLS analysis to generate smoothed estimates of the mean square error of the station skew. Thus, the constant model used a constant OLS regression to generate the  $MSE_G$ , while the log10(DA) model contains a constant and log10(DA) as an explanatory variable, which indicates the estimates of the  $MSE_G$  vary between the two models.

**Table A3.** Pseudo ANOVA table for the Iowa regional skew study for the constant model and the log10(DA) model.

[Pseudo ANOVA, pseudo analysis of variance; DA, drainage area; *EVR*, error variance ratio; *MBV\**, misrepresentation of the beta variance; Pseudo  $R_\delta^2$ , fraction of variability in the true skews explained by the model]

| Source         | Degrees-of-freedom   |     | Equations | Sum of squares  |          |    |
|----------------|--|-----|-----------|---|----------|----|
|                | A  | B   |           | A:<br>Constant  | B:<br>DA |    |
| Model          | $k$  | 0   | 1         | $n[\sigma_\delta^2(0) - \sigma_\delta^2(k)]$          | 0        | 6  |
| Model error    | $n-k-1$  | 272 | 271       | $n[\sigma_\delta^2(0)]$                               | 37       | 30 |
| Sampling error | $n$  | 273 | 273       | $\sum_{i=1}^n Var(\hat{y}_i)$                         | 46       | 47 |
| Total          | $2n-1$   | 545 | 545       | $n[\sigma_\delta^2(0)] + \sum_{i=1}^n Var(\hat{y}_i)$ | 83       | 83 |
| <i>EVR</i>     | $EVR = \frac{\sum_{i=1}^n Var(\hat{y}_i)}{n[\sigma_\delta^2(k)]}$        |     |           | 1.3   | 1.6      |    |
| <i>MBV*</i>    | $MBV^* = \frac{w^T Aw}{\sum_{i=1}^n w_i}$ where $w_i = \frac{1}{A_{ii}}$ |     |           | 6.5   | 6.9      |    |
| $R_\delta^2$   | $R_\delta^2 = 1 - \frac{\sigma_\delta^2(k)}{\sigma_\delta^2(0)}$         |     |           | 0%  | 19%      |    |

The constant model and the log10(DA) model have sampling error variances larger than their model error variances; however, it is important to note that the model error attributed to the log10(DA) model  $\sigma_\delta^2(1)$  is about one-sixth of the model error variance for the constant model  $\sigma_\delta^2(0)$ . This difference in model error is accounted for by the variation in the sample that the log10(DA) model appears to explain. Because the constant model does not have any explanatory variables, the variation attributed to that model is zero. On the other hand, the log10(DA) model has one explanatory variable. The analysis attributes a variance of six to the log10(DA) model. This accounts for the reduction of the model error variance from the constant model to the log10(DA) model; however, the addition of the drainage area explanatory variable in the log10(DA) model does not significantly improve the ability of the model to describe the variation in observed skew coefficients. This is reflected in the pseudo  $R_\delta^2$ , which in this case has a value of only 19 percent because the log10(DA) model explains only 19 percent of the estimated variation  $\sigma_\delta^2(0)$  in the true skew from site-to-site. Thus, the constant model is the model selected for the Iowa regional skew model.

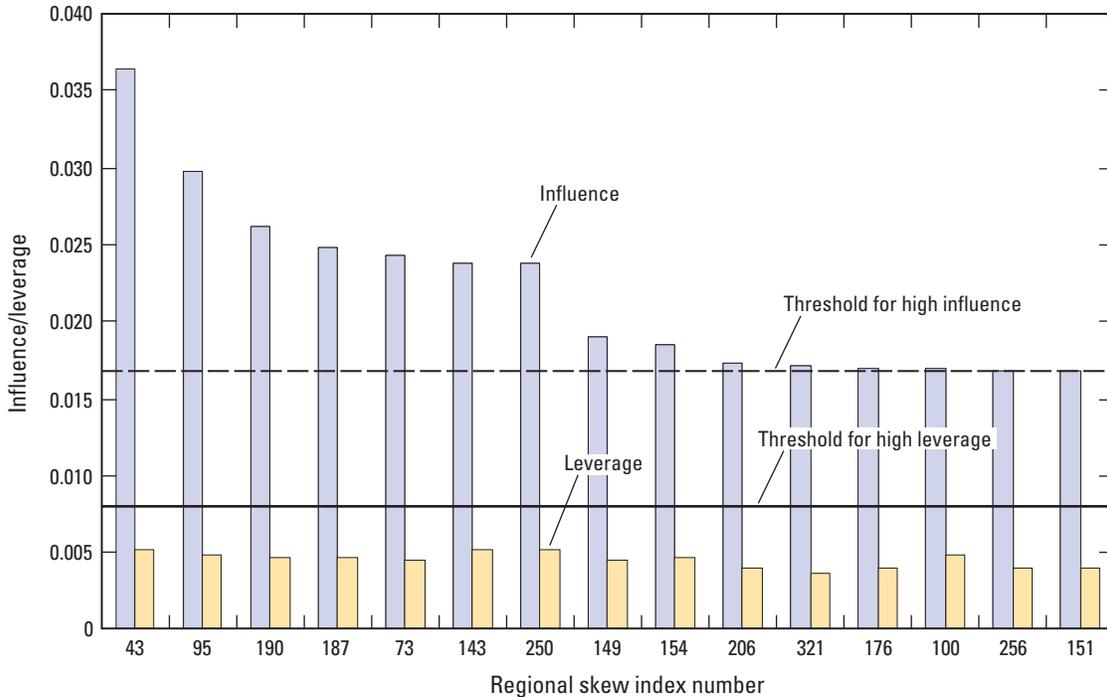
The Error Variance Ratio (EVR) is a modeling diagnostic used to evaluate if a simple OLS regression is sufficient, or if a more sophisticated WLS or GLS analysis is appropriate. EVR is the ratio of the average sampling error variance to the model

error variance. Generally, an EVR greater than 0.20 indicates that the sampling variance is not negligible when compared to the model error variance, suggesting the need for a WLS or GLS regression analysis. The EVR is calculated as

$$EVR = \frac{SS(\text{sampling error})}{SS(\text{model error})} = \frac{\sum_{i=1}^n Var(\hat{y}_i)}{n\sigma_\delta^2(k)} \quad (A15)$$

For the Iowa study-area data, EVR had a value of 1.3 for the constant model and 1.6 for the log10(DA) model. The sampling variability in the sample skewness estimators was larger than the error in the regional model. Thus, an OLS model that neglects sampling error in the station skewness estimators may not provide a statistically reliable analysis of the data. Given the variation of record lengths from site-to-site, it is important to use a WLS or GLS analysis to evaluate the final precision of the model, rather than a simpler OLS analysis.

The Misrepresentation of the Beta Variance (MBV\*) statistic is used to determine if a WLS regression is sufficient, or if a GLS regression is appropriate to determine the precision of the estimated regression parameters (Veilleux, 2011; Griffis, 2006). The MBV\* describes the error produced by a WLS regression analysis in its evaluation of the precision of  $b_0^{WLS}$ , which is the estimator of the constant  $\beta_0^{WLS}$ , because the



**Figure A5.** Regression Diagnostics: Leverage and Influence for the Iowa study area B-WLS/B-GLS constant model. The solid line represents the threshold for high leverage, while the dotted line represents the threshold for high influence.

covariance among the estimated station skews  $\hat{y}_i$  generally has its greatest effect on the precision of the constant term (Stedinger and Tasker, 1985). If the MBV\* substantially is greater than 1, then a GLS error analysis should be employed. The MBV\* is calculated as

$$MBV^* = \frac{Var[b_o^{WLS}|GLS\ analysis]}{Var[b_o^{WLS}|WLS\ analysis]} = \frac{w^T Aw}{\sum_{i=1}^n w_i} \quad \text{where } w_i = \frac{1}{\sqrt{A_{ii}}} \quad (A16)$$

For the Iowa regional skew study, the MBV\* is equal to 6.5 for the constant model and 6.9 for the log10(DA) model. This is a large value indicating the cross-correlation among the skewness estimators has had an effect on the precision with which the regional average skew coefficient can be estimated; if a WLS precision analysis were used for the estimated constant parameter in the constant model, the variance would be underestimated by a factor of 6.5. Thus, a WLS analysis would seriously misrepresent the variance of the constant in the constant model and in the log10(DA) model of regional skew. Moreover, a WLS model would have resulted in underestimation of the variance of prediction, given that the sampling error in the constant term in both models sufficiently was large enough to make an appreciable contribution to the average variance of prediction.

### Leverage and Influence

Leverage and influence diagnostic statistics can be used to identify rogue observations and to effectively address lack-of-fit when estimating skew coefficients. Leverage identifies those streamgages in the analysis where the observed values have a large effect on the fitted (or predicted) values (Hoaglin and Welsch, 1978). Generally, leverage considers if an observation, or explanatory variable, is unusual, and thus likely to have a large effect on the estimated regression coefficients and predictions. Unlike leverage, which highlights points with the ability or potential to affect the fit of the regression, influence attempts to describe those points that do have an unusual effect on the regression analysis (Belsley and others, 1980; Cook and Weisberg, 1982; Tasker and Stedinger, 1989). An influential observation is one with an unusually large residual that has a disproportionate effect on the fitted regression relations. Influential observations often have high leverage. For a detailed description of the equations used to determine leverage and influence for a B-WLS/B-GLS analysis, see Veilleux and others (2011) and Veilleux (2011).

Figure A5 displays the leverage and influence values for the B-WLS/B-GLS constant regional skew model for the Iowa study area. The 15 sites included in the figure have high influence, and thus have an unusual effect on the fitted regression relation. The sites are ordered, starting from the left, by decreasing influence, as it identifies those sites that had a large effect on the analysis. No sites in the regression had high

leverage, and the differences in leverage values for the constant model reflect the variation in record lengths among sites. Streamgage 05410490 (regional skew index number 43, map number 65, fig. 1) has the highest influence value because of its large residual, the third largest positive residual in the study (in other words the largest positive unbiased station skew = 1.04), and its large drainage area (700 mi<sup>2</sup>), which is larger than all of the other sites with large influences.

## References

- Belsley, D.A., Kuh, E., and Welsch, R.E., 1980, Regression diagnostics—Identifying influential data and sources of collinearity: John Wiley & Sons, Inc., p. 6–84.
- Cohn, T.A., 2011, PeakfqSA Version 0.960 (software): U.S. Geological Survey, [Information on PeakfqSA available at [http://www.timcohn.com/TAC\\_Software/PeakfqSA/](http://www.timcohn.com/TAC_Software/PeakfqSA/).]
- 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, accessed March 15, 2013, at <http://onlinelibrary.wiley.com/doi/10.1029/97WR01640/pdf>.
- 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, accessed March 15, 2013, at <http://timcohn.com/Publications/CohnLaneSted2001WR900016.pdf>.
- Cook, R.D., and Weisberg, S., 1982, Residuals and influence in regression: New York, Chapman and Hall, 230 p.
- Feaster, T.D., Gotvald, A.J., and Weaver, J.C., 2009, Magnitude and frequency of rural floods in the southeastern United States, 2006—Volume 3, South Carolina: U.S. Geological Survey Scientific Investigations Report 2009–5156, 226 p. (Also available at: <http://pubs.usgs.gov/sir/2009/5156/>.)
- Gotvald, A.J., Feaster, T.D., and Weaver, J.C., 2009, Magnitude and frequency of rural floods in the southeastern United States, 2006—Volume 1, Georgia: U.S. Geological Survey Scientific Investigations Report 2009–5043, 120 p. (Also available at <http://pubs.usgs.gov/sir/2009/5043/>.)
- Griffis, V.W., 2006, Flood frequency analysis— Bulletin 17, regional information, and climate change: Cornell, Cornell University, Ph.D. Dissertation.
- Griffis, V.W., and Stedinger, J.R., 2009, Log-Pearson type 3 distribution and its application in flood frequency analysis, III—Sample skew and weighted skew estimators: *Journal of Hydrology*, v. 14, no. 2, p. 121–130.
- Griffis, V.W., Stedinger, J.R., and Cohn, T.A., 2004, Log-Pearson type 3 quantile estimators with regional skew information and low outlier adjustments: *Water Resources Research*, v. 40, W07503, 17 p., accessed March 15, 2013, at <http://onlinelibrary.wiley.com/doi/10.1029/2003WR002697/pdf>.
- Gruber, A.M., Reis, D.S., Jr., and Stedinger, J.R., 2007, Models of regional skew based on Bayesian GLS regression, in Kabbes, K.C., ed., Proceedings of the World Environmental and Water Resources Congress, Restoring our Natural Habitat, May 15–18, 2007: Tampa, Florida, American Society of Civil Engineers, paper 40927-3285.
- Gruber, A. M., and Stedinger, J.R., 2008, Models of LP3 regional skew, data selection, and Bayesian GLS regression, in Babcock, R.W. and Watson, R. eds, World Environmental and Water Resources Congress 2008 Ahupua'a, May 12–16, 2008: Honolulu, Hawaii, paper 596.
- Hoaglin, D.C. and Welsch, R.E., 1978, The Hat Matrix in Regression and ANOVA, *The American Statistician*, 32(1), p. 17–22.
- Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood flow frequency: Reston, Virginia, Hydrology Subcommittee Bulletin 17B, 28 p. and appendixes, accessed March 15, 2013, at [http://water.usgs.gov/osw/bulletin17b/dl\\_flow.pdf](http://water.usgs.gov/osw/bulletin17b/dl_flow.pdf).
- Lamontagne, J.R., Stedinger, J.R., Berenbrock, Charles, Veilleux, A.G., Ferris, J.C., and Knifong, D.L., 2012, Development of regional skews for selected flood durations for the Central Valley Region, California, based on data through water year 2008: U.S. Geological Survey Scientific Investigations Report 2012–5130, 60 p. (Also available at <http://pubs.usgs.gov/sir/2012/5130/>.)
- Martins, E.S., and Stedinger, J.R., 2002, Cross-correlation among estimators of shape: *Water Resources Research*, v. 38, no. 11, (Also available at <http://dx.doi.org/10.1029/2002WR001589>.)
- Parrett, C., Veilleux, A.G., Stedinger, J.R., Barth, N.A., Knifong, D.L., and Ferris, J.C., 2011, Regional skew for California, and flood frequency for selected sites in the Sacramento–San Joaquin River Basin, based on data through water year 2006: U.S. Geological Survey Scientific Investigations Report 2010–5260, 94 p. (Also available at <http://pubs.usgs.gov/sir/2010/5260/>.)
- Reis, D.S., Jr., Stedinger, J.R., and Martins, E.S., 2005, Bayesian generalized least squares regression with application to the log Pearson type 3 regional skew estimation: *Water Resources Research*, v. 41, W10419, doi:10.1029/2004WR003445, accessed March 15, 2013, at <http://www.agu.org/journals/wr/wr0510/2004WR003445/2004WR003445.pdf>.

- Stedinger, J.R., and Tasker, G.D., 1985, Regional hydrologic analysis 1—Ordinary, weighted, and generalized least square compared: *Water Resources Research*, v. 21, no. 9, p. 1421–1432, accessed March 15, 2013, at <http://www.agu.org/journals/wr/v021/i009/WR021i009p01421/WR021i009p01421.pdf>.
- Stedinger, J.R., and Cohn, T.A., 1986, Flood frequency analysis with historical and paleoflood information: *Water Resources Research*, v. 22, no. 5, p. 785–793.
- Stedinger, J.R. and Griffis, V.W., 2008, Flood frequency analysis in the United States—Time to Update: (editorial) *Journal of Hydrologic Engineering*, April, p. 199–204.
- Tasker, G.D., and Stedinger, J.R., 1986, Regional skew with weighted LS regression: *Journal of Water Resources Planning and Management*, v. 112, no. 2, p. 225–237.
- Tasker, G.D., and Stedinger, J.R., 1989, An operational GLS model for hydrologic regression: *Journal of Hydrology*, v. 111, p. 361–375.
- U.S. Geological Survey, 2012, National Water Information System data available on the World Wide Web (USGS Water Data for the Nation), accessed March 15, 2013, at <http://nwis.waterdata.usgs.gov/usa/nwis/peak>.
- Veilleux, A.G. 2009, Bayesian GLS regression for regionalization of hydrologic statistics, floods and Bulletin 17 skew: Cornell, Cornell University, M.S. Thesis.
- Veilleux, A.G., 2011, Bayesian GLS regression, leverage and influence for regionalization of hydrologic statistics: Cornell, Cornell University, Ph.D. dissertation.
- Veilleux, A.G., Stedinger, J.R., and Lamontagne, J.R., 2011, Bayesian WLS/GLS regression for regional skewness analysis for regions with large cross-correlations among flood flows, *in* World Environmental and Water Resources Congress 2011, Bearing Knowledge for Sustainability, May 22–26, 2011: Palm Springs California, ASCE, paper 3103.
- Veilleux, A.G., Stedinger, J.R., and Eash, D.A., 2012, Bayesian WLS/GLS regression for regional skewness analysis for crest stage gage networks, *in* Loucks, E.D., ed., Proceedings World Environmental and Water Resources Congress, Crossing boundaries, May 20–24, 2012: Albuquerque, New Mexico, American Society of Civil Engineering, paper 227, p. 2253–2263. (Also available at <http://ia.water.usgs.gov/media/pdf/report/Veilleux-Stedinger-Eash-EWRI-2012-227R.pdf>.)
- Weaver, J.C., Feaster, T.D., and Gotvald, A.J., 2009, Magnitude and frequency of rural floods in the southeastern United States, 2006—Volume 2, North Carolina: U.S. Geological Survey Scientific Investigations Report, 2009–5158, 111p. (Also available at: <http://pubs.usgs.gov/sir/2009/5158/>.)

**Publishing support provided by:**

Rolla Publishing Service Center

**For more information concerning this publication, contact:**

Director, USGS Iowa Water Science Center

P.O. Box 1230

Iowa City, IA 52244

(319) 337-4191

**Or visit the Iowa Water Science Center Web site at:**

*<http://ia.water.usgs.gov>*

