Prepared in cooperation with the North Dakota State Water Commision, the North Dakota Department of Transportation, the North Dakota Department of Health, the Red River Joint Water Resources Board, and the Devils Lake Basin Joint Water Resource Board

Regional Regression Equations to Estimate Peak-Flow Frequency at Sites in North Dakota Using Data through 2009

Scientific Investigations Report 2015–5096

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

Regional Regression Equations to Estimate Peak-Flow Frequency at Sites in North Dakota Using Data through 2009

By Tara Williams-Sether

Prepared in cooperation with the North Dakota State Water Commision, the North Dakota Department of Transportation, the North Dakota Department of Health, the Red River Joint Water Resources Board, and the Devils Lake Basin Joint Water Resource Board

Scientific Investigations Report 2015–5096

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: 2015

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/.

Suggested citation:
Williams-Sether, Tara, 2015, Regional regression equations to estimate peak-flow frequency at sites in North Dakota using data through 2009: U.S. Geological Survey Scientific Investigations Report 2015–5096, 12 p., http://dx.doi.org/10.3133/sir20155096.

ISSN 2328-0328 (online)

Contents

Figures

Tables

Conversion Factors

Inch/Pound to International System of Units

Multiply	By	To obtain
	Length	
inch (in.)	2.54	centimeter (cm)
inch (in.)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
yard (yd)	0.9144	meter (m)
	Area	
acre	4,047	square meter (m^2)
acre	0.4047	hectare (ha)
acre	0.4047	square hectometer (hm^2)
acre	0.004047	square kilometer (km^2)
square foot (ft^2)	929.0	square centimeter (cm^2)
square foot (ft^2)	0.09290	square meter (m^2)
square inch (in^2)	6.452	square centimeter (cm^2)
section (640 acres or 1 square mile)	259.0	square hectometer (hm^2)
square mile (mi^2)	259.0	hectare (ha)
square mile (mi^2)	2.590	square kilometer (km^2)
	Flow rate	
acre-foot per day (acre-ft/d)	0.01427	cubic meter per second (m^3/s)
acre-foot per year (acre-ft/yr)	1,233	cubic meter per year (m^3/yr)
acre-foot per year (acre-ft/yr)	0.001233	cubic hectometer per year (hm^3/yr)
foot per second (ft/s)	0.3048	meter per second (m/s)
foot per minute (ft/min)	0.3048	meter per minute (m/min)
foot per hour (ft/hr)	0.3048	meter per hour (m/hr)
foot per day (ft/d)	0.3048	meter per day (m/d)
foot per year (ft/yr)	0.3048	meter per year (m/yr)
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)
cubic foot per second per square mile [(ft^3/s)/mi^2]	0.01093	cubic meter per second per square kilometer [(m^3/s)/km^2]
cubic foot per day (ft^3/d)	0.02832	cubic meter per day (m^3/d)
inch per hour (in/h)	0 .0254	meter per hour (m/h)
inch per year (in/yr)	25.4	millimeter per year (mm/yr)

Datums

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

Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83).

Regional Regression Equations to Estimate Peak-Flow Frequency at Sites in North Dakota Using Data through 2009

By Tara Williams-Sether

Abstract

Annual peak-flow frequency data from 231 U.S. Geological Survey streamflow-gaging stations in North Dakota and parts of Montana, South Dakota, and Minnesota, with 10 or more years of unregulated peak-flow record, were used to develop regional regression equations for exceedance probabilities of 0.5, 0.20, 0.10, 0.04, 0.02, 0.01, and 0.002 using generalized least-squares techniques. Updated peak-flow frequency estimates for 262 streamflow-gaging stations were developed using data through 2009 and log-Pearson Type III procedures outlined by the Hydrology Subcommittee of the Interagency Advisory Committee on Water Data. An average generalized skew coefficient was determined for three hydrologic zones in North Dakota. A StreamStats web application was developed to estimate basin characteristics for the regional regression equation analysis. Methods for estimating a weighted peak-flow frequency for gaged sites and ungaged sites are presented.

Introduction

Information on the magnitude and frequency of peak streamflows is an essential component of effective flood-plain mapping, general water-resources planning, management and permitting, instream flow determinations for pollution control and habitat protection, and design of hydraulic structures on or near streams. Streamflow statistics for streamflow-gaging stations (hereafter referred to as "gaging stations") can be obtained from annual surface-water data reports or by analysis of existing data in the U.S. Geological Survey (USGS) National Water Information System (NWISWeb) database (U.S. Geological Survey, 2014). Peak-flow statistics used by managers and designers are commonly needed at unregulated sites where no streamflow data are available. Regional regression equations are used to estimate peak-flow statistics at these ungaged sites. Regression equations for computing the magnitude and frequency of peak flows were last developed for North Dakota by the USGS using streamflow data collected through water year 1988 (Williams-Sether, 1992). Since that time, North Dakota has experienced relatively wet conditions that have contributed to record peak flows across the State, necessitating the need for updated peak-flow frequency regression equations.

The USGS, in cooperation with the North Dakota State Water Commission, the North Dakota Department of Transportation, the North Dakota Department of Health, the Red River Joint Water Resources Board, and the Devils Lake Basin Joint Water Resource Board, developed updated regional peak-flow frequency regression equations for ungaged streamflow sites in North Dakota using the procedures described in "Guidelines for Determining Flood Flow Frequency" (hereafter referred to as "Bulletin 17B"; Hydrology Subcommittee of the Interagency Advisory Committee on Water Data, 1982).

Purpose and Scope

The purpose of this report is to describe the methods and results of the regression analysis that were used to develop regional regression equations using expected peak-flow frequencies and basin characteristics for selected exceedance probabilities for ungaged streamflow sites in North Dakota. Peak-flow frequency regression equations were developed using unregulated peak streamflow data from 231 selected gaging stations in North Dakota and parts of Montana, South Dakota, and Minnesota with 10 or more years of peak-flow data through water year 2009. Peak-flow data for the the 0.5, 0.20, 0.10, 0.04, 0.02, 0.01, and 0.002 exceedance probabilities are presented. A North Dakota StreamStats web application tool was developed to generate basin characteristics used in this study. Basin characteristics that were explored and used during the peak-flow frequency regression analysis for each gaging station are presented and the locations of the stations used in this study are shown in figure 1.

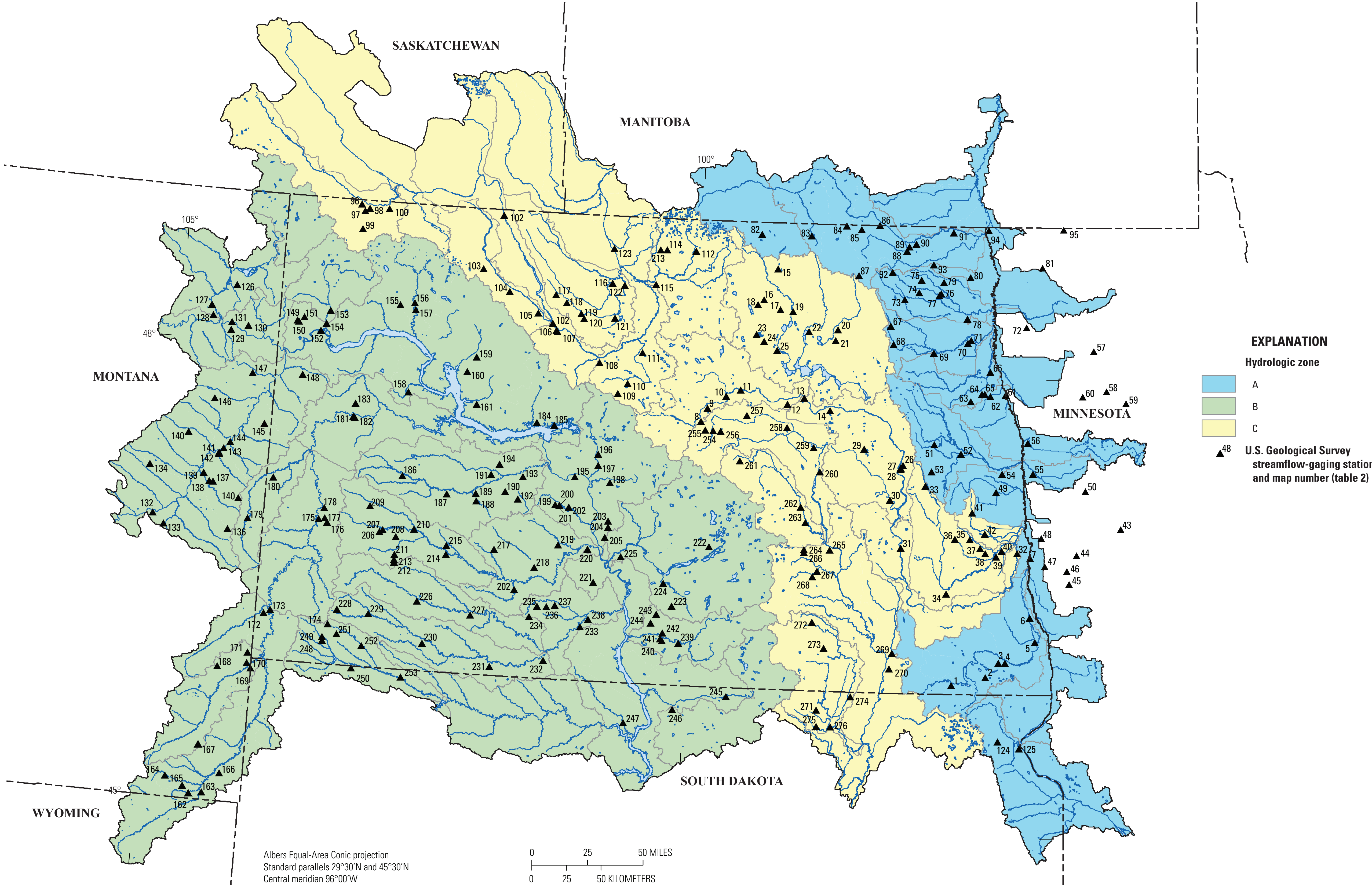

Figure 1. Streamflow-gaging stations and hydrologic zones used in the development of peak-flow regional regression equations for North Dakota streams.

Previous Studies

A study of the magnitude and frequency of floods in North Dakota and South Dakota using data through 1955 was completed by McCabe and Crosby (1959). A limited analysis of the magnitude and frequency of floods in North Dakota was made by Crosby (1970, 1975), in which the magnitude and frequency relations were updated for small drainage basins of 100 square miles (mi^2) or less. Regression equations for estimating the magnitude of peak flows in North Dakota for selected recurrence intervals were last published in 1992 using data through 1988 (Williams-Sether, 1992).

Peak-Flow Frequency Analysis at Gaging Stations

Peak-flow frequency analysis is a statistical technique used to estimate exceedance probabilities associated with floods. Through analysis of past floods, a relation between peak-flow magnitude and frequency can be estimated. The peak-flow frequency analysis for this study is based on guidelines published in Bulletin 17B (Hydrology Subcommittee of the Interagency Advisory Committee on Water Data, 1982). These guidelines promote a uniform technique to estimate peak-flow frequencies for gaging stations. The frequency of a flood (annual peak flow) is described in terms of exceedance probabilities. Exceedance probability is the chance or likelihood that a given flow of specific magnitude will happen in any 1-year period. Exceedance probabilities formerly were reported as flood recurrence intervals expressed in years. For example, a flood magnitude that has a 1-percent chance (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 exceedance probability is the inverse of the recurrence interval multiplied by 100. Although the exceedance probability is an estimate of the likelihood in any 1-year period, more than one flood discharge with a specific magnitude and exceedance probability could happen in the same 1-year period. The USGS computer program PEAKFQ (Flynn and others, 2006), which is based on Bulletin 17B guidelines, was used in this study to provide estimates of the magnitude of annual peak flows having exceedance probabilities of 0.50, 0.20, 0.10, 0.04, 0.02, 0.01, and 0.002. Gaging station selection for this report was based on stations having a minimum of 10 years of unregulated peak-flow record and using data through 2009, which resulted in 262 selected stations (table 1; available at http://dx.doi.org/10.3133/sir20155096). The initial peak-flow frequency curves generated using PEAKFQ were reviewed to determine how well the annual peak flows fit the theoretical (log-Pearson Type III) distribution and were adjusted for low outliers when appropriate. The low outlier adjustments were determined by visual examination of the plotted theoretical log-Pearson Type III distribution on a log-probability scale and the systematic data. No adjustments for high outliers were made due to the lack of accurate historical information.

Generalized Skew Coefficient

Guidelines in Bulletin 17B (Hydrology Subcommittee of the Interagency Advisory Committee on Water Data, 1982) recommend using a weighted skew coefficient to reduce the uncertainty in peak-flow frequency estimates. The weighted skew coefficient is calculated by mathematically weighting the gaging station skew coefficient and generalized skew coefficient. The generalized skew coefficient can provide some regional geographic continuity by combining the coefficient of skew from many nearby gaging stations. The coefficient of skew calculated from the gaging station data affects the shape of the peak-flow frequency curve and is sensitive to extreme events. Three methods for developing a generalized skew coefficient are described in Bulletin 17B: (1) plot computed gaging station skews on a map and construct skew isolines on the map, (2) develop a skew prediction equation relating basin characteristics to station skews, and (3) compute the arithmetic mean of station skew coefficients from regional stations with 25 or more years of record.

Attempts were made to develop a new generalized skew map for North Dakota using methods 1 and 3 noted above using data from 262 selected gaging stations in North Dakota and parts of Montana, South Dakota, and Minnesota. The gaging stations that were used had unregulated peak-flow data and at least 10 years of record, and station skews were computed using guidelines published in Bulletin 17B. Method 2 was not attempted because basin characteristics were not available at the time the skew analysis was made. For method 1, the computed gaging station skews were plotted using station latitude and longitude and evaluated for any geographic or topographic patterns in a Geographic Information Sysytem (GIS). No definable patterns were evident. Station skews were initially restricted to those that had at least 25 years of data. Also, datasets of 10 years or more and 15 years or more were considered. Station skew values from all runs (25 years or more, 10 years or more, and 15 years or more) exhibited large variance in values and did not generate any discernible patterns across the State or within the three hydrologic zones A, B, and C defined in Williams-Sether (1992) (fig. 1). Because it was a possibility that the lack of similarity between the data periods of record was causing generalization problems, a common period (from 1960 to 2009) was used to screen the initial stations used. The 262 stations were screened to allow 25 percent of each station record to be missing; therefore, record lengths could be as much as 50 years but no less than 37 years. This common period approach reduced the number of gaging stations available for analysis from 262 to 60 (fig. 2) and methods 1 and 3 were repeated.

Again, method 1 resulted in non-patterning of skew values across the State; therefore, the three hydrologic zones A, B, and C previously defined by Williams-Sether (1992) were used for method 3. The arithmetic mean, variance, and standard error of the skews were computed using data from 25 stations in zone A, 19 stations in zone B, and 16 stations in zone C. The arithmetic mean, variance, and standard error of the skew coefficients for each hydrologic zone was computed after the gaging station skew values were adjusted for record length bias as suggested in Tasker and Stedinger (1989). The averaged skew (mean generalized skew), variance, and standard error values for each hydrologic zone are listed table 2.

The Bulletin 17B skew map (Hydrology Subcommittee of the Interagency Advisory Committee on Water Data, 1982), generated in 1976 using data through 1973, has a generalized skew value for North Dakota of -0.40 with a variance of 0.302 and a standard error value of 0.550. The skew standard error values determined for the three hydrologic zones are noticeably smaller than the Bulletin 17B value, and were determined using current data and consideration for hydrologic and climatic differences across North Dakota. Thus, the mean generalized skew values listed in table 2 were used instead of the generalized skew estimate from Bulletin 17B skew map for the peak-flow frequency estimates. The final gaging station Log Pearson Type III peak-flow frequency estimates are based on station skews weighted with the new generalized skew values and are presented in table 1.

Basin Characteristics

The USGS StreamStats Program was created by the USGS in cooperation with Environmental Systems Research Institute, Inc. (Esri) to make GIS based estimation of streamflow statistics easier, faster, and more consistent than previously used manual techniques. The USGS StreamStats Program is a map-based internet interface designed for national application, with each state, territory, or group of states responsible for creating unique geospatial datasets and regression equations to compute streamflow statistics. Further information about StreamStats usage and limitations can be accessed at http://water.usgs.gov/osw/streamstats/. The primary purpose of the North Dakota StreamStats web application is to provide estimates of basin characteristics and streamflow frequency statistics for user-selected ungaged sites on North Dakota streams. The North Dakota StreamStats application covers 55 processed hydrologic units (HUs) (U.S. Geological Survey and others, 2012) (fig. 3). Because the StreamStats application determines characteristics for a basin from a selected point, verification of delineated drainage area was compared to previously published values listed in the NWISWeb database for each gaging station used (U.S. Geological Survey, 2014). Any delineated drainage area values that differed by more than 10 percent from the previously published values were screened for possible errors and corrected if necessary. The main problem causing drainage area differences between delineated and previously reported values was gaging station location errors (a station not being located on the stream channel). Other differences that were determined seemed to be delineation interpretation differences between the computer data and past "hand-delineations" from paper maps. For this study, the delineated drainage areas were used to maintain consistency in computation methods of the basin characterisitcs. Updates of the drainage area values reported in NWISWeb with those delineated by the StreamStats program have not been made and are pending approval.

Basin characteristics investigated as potential explanatory variables in the regression analysis were selected on the basis of the previous study (Williams-Sether, 1992), theoretical relation to peak flows, and the ability to generate characteristics using GIS technology and digital datasets. Basin characteristics, used as possible explanatory variables in the peak-flow frequency regional regression equations development for this study, were generated by the North Dakota StreamStats web tool and are listed in tables 1 and 3.

Development of Regional Regression Equations

Multiple-linear-regression techniques were used to develop regional regression equations relating gaging station peak-flow frequencies to various basin characteristics for selected exceedance probabilities. Regression equations can be developed using ordinary-least-squares (OLS) and generalized-least-squares (GLS) techniques. The OLS technique gives equal weight to peak flows at all gaging stations, regardless of record length and the possible correlation among concurrent flows at different sites, and only provides a rough estimate of model error. The GLS technique accounts for unequal record length as well as cross correlation of concurrent flows at different stations, and provides better estimates of the predictive accuracy of peak-flow estimates that are computed by the regression equations and nearly unbiased estimates of the variance of the underlying regression model error (Stedinger and Tasker, 1985). The USGS weighted-multiple-linear-regression computer program, WREG, was used to develop initial and final peak-flow frequency regional regression equations. For further detailed explanations about the OLS and GLS regression techniques, refer to the WREG user's guide (Eng and others, 2009; Stedinger and Tasker, 1985; Tasker and Stedinger, 1989).

Various basin characteristics (tables 1 and 3) were tested as explanatory variables in the regression analysis. Scatterplot matrices of the log-transformed (base 10) peak-flow discharges, log-transformed (base 10) explanatory variables, and untransformed explanatory variables were generated to evaluate whether log-transformation of the explanatory variables were needed and to check for correlation of the explanatory variables with peak flow. Explanatory variables that indicated poor correlation with peak flow were eliminated prior to regression anlaysis.

To simplify the variable selection process, models were initially selected using OLS regression. In addition, to reduce

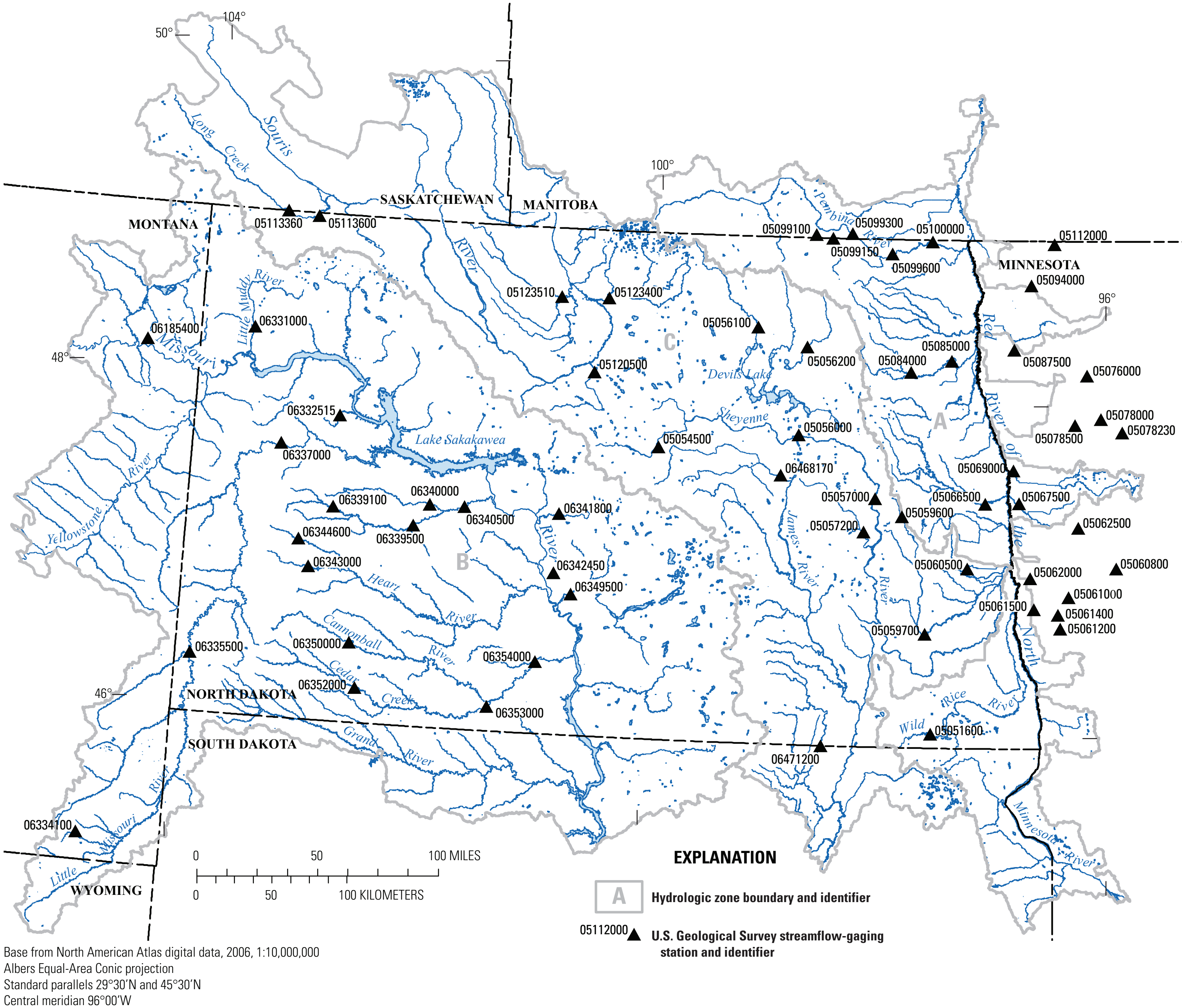

Figure 2. Streamflow-gaging stations used to determine a generalized skew coefficient for hydrologic zones A, B, and C (shown in fig. 1), North Dakota.

Table 2. Mean generalized skew, variance, and standard error values for hydrologic zones A, B, and C, North Dakota.

Range of station record length (years)	Number of stations used	Mean generalized skew	Variance of skew	Standard error of skew
		Zone A		
42 to 50	25	-0.509	0.136	0.368
		Zone B		
42 to 50	19	-0.730	0.067	0.259
		Zone C		
38 to 50	16	-0.585	0.084	0.289

the potential complexity of the models and maintain similarity among the models for all of the exceedance probabilities, only exceedance probabilities 0.5, 0.01, and 0.002 were used for model selection. The variables selected for exceedance probabilities 0.5, 0.01, and 0.002 were the same, and those variables were also used for exceedance probabilities of 0.20, 0.10, 0.04, and 0.02. Step-wise regression was used to evaluate statistically significant basin characteristics to use as explanatory variables in the development of the final peak-flow frequency regression models for hydrologic zones A, B, and C. The resulting OLS regression models were restricted to no more than three explanatory variables using a 5-percent (0.05) significance level, selecting models with the largest coefficient of determination (R^2), and examining scatterplots of residuals versus fitted values and each of the selected explanatory variables.

The explanatory variables and transformations determined in the OLS regression models were then used to form GLS regression models. Gaging stations that were “flagged” by the WREG program as having large influence or leverage values were further examined for elimination. Residual scatterplots versus fitted values and explanatory variables were examined as well to determine if “flagged” gaging stations with large influence and leverage were isolated hydrologic outliers and could be removed from the analysis. The gaging stations removed from the final regression analyses for hydrologic zones A, B, and C are noted in table 1. Final GLS regression equations were selected on the basis of minimizing values of the standard model error (SME) and the standard error of prediction (S_p), and maximizing values of the pseudo R^2 (Eng and others, 2009).

The final peak-flow frequency regression equations for exceedance probabilities 0.50, 0.20, 0.10, 0.04, 0.02, 0.01, and 0.002 developed for hydrologic zones A, B, and C are shown in table 4 (available at http://dx.doi.org/10.3133/sir20155096). The characteristics used in the final regression equations are drainage area (hydrologic zones A, B, and C), stream slope computed using the longest flow path (CSL1085LFP, table 3) (zone A), ruggedness number (zones B and C), and compactness number (zone B). The S_p for the various exceedance probabilities ranges from 51.03 to 58.84 percent for zone A, 57.95 to 75.32 percent for zone B, and 52.85 to 82.90 percent for zone C (table 4) and are lower than those reported by Williams-Sether (1992). The pseudo R^2 ranges from 75.32 to 83.47 percent for zone A, 87.40 to 91.21 percent for zone B, and 66.85 to 84.89 percent for zone C and are higher than those reported by Williams-Sether (1992).

Limitations of the Regional Regression Equations

The following limitations should be considered when using the regression equations to compute peak-flow frequencies for North Dakota streams: (1) the streams sites should be located in rural watersheds and not significantly affected by urbanization or regulation, (2) the explanatory variables should be computed using the same GIS techniques that were used to develop the regression equations, and (3) the explanatory variables should stay within the range of the data used to develop the regression equations (table 4).

Web Application for Solving Regional Regression Equations

The North Dakota StreamStats web application incorporates the new peak-flow frequency regression equations and provide peak-flow frequency estimates for most unregulated sites in the State. Peak-flow regression estimates will not be available for unregulated sites that have some part of a drainage basin located outside of a processed HU code. The web application includes (1) a mapping tool to specify a location on a stream where peak-flow statistics are desired; (2) a database that includes peak-flow frequency statistics, hydrologic characteristics, location, and descriptive information for all USGS gaging stations used in this study; and (3) an automated GIS procedure that measures the required basin characteristics and solves the regression equations to estimate peak-flow statistics for user-selected sites.

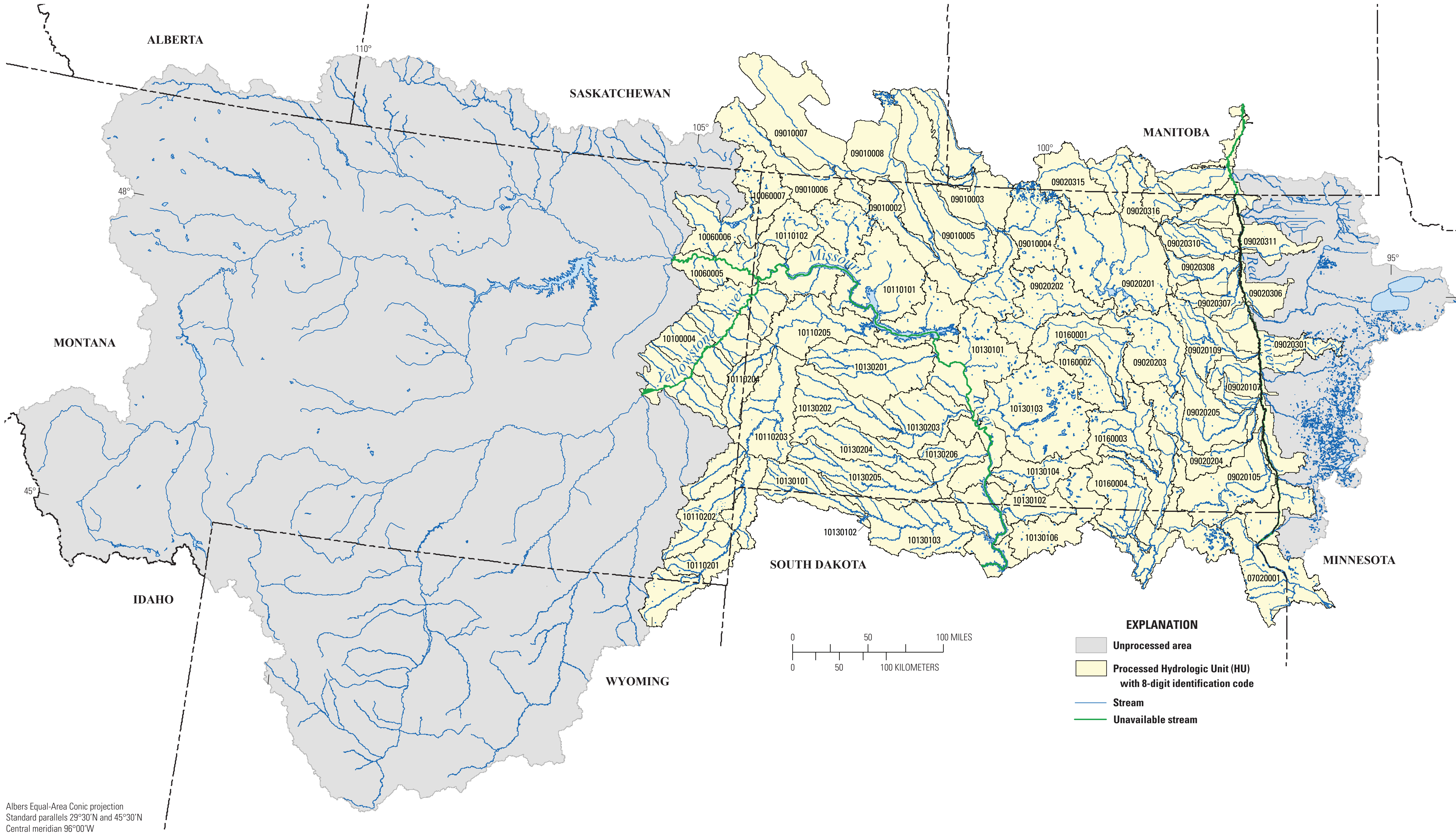

Figure 3. Processed hydrologic units used in the StreamStats web application for North Dakota.

Table 3. Basin characteristics generated by the North Dakota Streamstats web tool used in the development of peak-flow frequency regional regression equations for North Dakota streams.

Characteristic name	Characteristic label	Characteristic definition	Characteristic unit of measure
Ag_Land_Percentage	AG_OF_DA	Agricultural land in percentage of drainage area (Hortness, 2006)	Percent.
Basin_Perimeter	BASINPERIM	Perimeter of the drainage basin as defined in Gingerich (2005)	Miles.
Mean_Basin_Slope_from_10m_DEM	BSLDEM10M	Mean basin slope computed from 10-meter digital elevation model (DEM)	Percent.
Compactness_Ratio	COMPRAT	A measure of basin shape related to basin perimeter and drainage area. Computed as basin perimeter divided by two times the square root of pi times drainage area.	Dimensionless.
Stream_Slope_10_and_85_Longest_Flow_Path	CSL1085LFP	Change in elevation between points 10 and 85 percent of length along the longest flow path determined by a geographic information system (GIS) divided by length between points	Feet per mile.
Drainage_Area	DRNAREA	Area that drains to a point on a stream	Square miles.
Mean_Basin_Elevation	ELEV	Mean basin elevation	Feet.
Maximum_Basin_Elevation	ELEVMAX	Maximum basin elevation	Feet.
Percent_Isolated_Lake_and_Ponds_Drainage	ISOLAKEDA	Percentage of basin drainage area that drains to isolated lakes and ponds	Percent.
Percent_Lakes_and_Ponds	LAKEAREA	Percentage of basin drainage area that are lakes and ponds	Percent.
LFP_length	LFPLENGTH	Length of longest flow path	Miles.
Minimum_Basin_Elevation	MINBELEV	Minimum basin elevation	Feet.
Mean_Annual_Precipitation	PRECIP	Mean annual precipitation	Inches.
Ruggedness_Number	RUGGED	Ruggedness number computed as stream density times basin relief; where stream density is the stream length divided by the drainage area, and basin relief is the maximum basin elevation minus the minimum basin elevation.	Feet per mile.
Slope_Ratio	SLOPERAT	Slope ratio computed as longest flow path slope divided by basin slope	Dimensionless.
Average_Soil_Permeability	SOILPERM	Average soil permeability	Inches per hour.
Stream_length	STREAMLENGTH	Sum of length of all mapped streams	Miles.

Application of Regional Regression Equations

The developed regression equations can be used to provide peak-flow frequency estimates for sites on ungaged or gaged streams. Peak-flow frequency estimates for gaging stations can be improved by computing a weighted-average value of two independent estimates: the at-site log-Pearson Type III frequency curve estimate and the appropriate regression equation estimate. By weighting each estimate with an appropriate weighting factor, the resulting weighted-average value will represent an improved estimate (Hydrology Subcommittee of the Interagency Advisory Committee on Water Data, 1982). Peak-flow frequency estimates for an ungaged site located on a stream with a gaging station can be improved by combining the appropriate regression equation estimate for the ungaged site with the log-Pearson Type III frequency-curve estimate for the gaging station. Peak-flow frequency estimates can be computed using the appropriate regression equation for an ungaged site on a stream without a gaging station.

Estimating the Weighted Peak-Flow Frequency for a Gaging Station

Two estimates of peak-flow frequency for a gaging station are available: one from the at-site log-Pearson Type III frequency curve and the other from the appropriate peak-flow frequency regression equation developed in this study. A theoretically improved estimate can be calculated if the individual estimates are independent and the variances of the individual estimates can be determined. If the independent estimates are weighted inversely proportional to their respective variances, then the variance of the weighted-average estimate will be less than the variances associated with each individual estimate (Tasker, 1975; Hydrology Subcommittee of the Interagency Advisory Committee on Water Data, 1982).

For a particular exceedance probability, the variance associated with the at-site frequency-curve estimate can be computed using an expression for the asymptotic variance developed by Cohn and others (2001) and implemented in the Weighted Independent Estimates (WIE) program (Berenbrock and Cohn, 2008). The magnitude of the variance associated with the at-site frequency-curve estimate is dependent on the length of record; the mean, standard deviation, and skew of the fitted log-Pearson Type III frequency curve; and the accuracy of the method used to determine the generalized skew (Gotvald and others, 2009).

For a selected exceedance probability, the variance associated with the appropriate regression equation is the average variance of prediction (AVP), which can be computed from the standard error of prediction (S_p) using the following equation:

$$S_p = 100\left[10^{AVP \times \ln(10)}\right]^{\frac{1}{2}} \qquad (1)$$

An alternative variance associated with the regression equation can be determined for each gaging station, although the average variance of prediction from all stations is used in this study.

Using the variances from the two independent peak-flow frequency estimates, the weighted-average peak-flow frequency estimate is computed using the following equation (Gotvald and others, 2009):

$$\log Q_{\%gw} = \frac{V_{\%gr} \times \log Q_{\%gs} + V_{\%gs} \times \log Q_{\%gr}}{V_{\%gs} + V_{\%gr}} \qquad (2)$$

where

$Q_{\%gw}$ is the peak-flow estimate for selected exceedance probability at a gaging station, weighted-average value, in cubic feet per second;

$Q_{\%gs}$ is the peak-flow estimate for selected exceedance probability at a gaging station, value from the station log-Pearson Type III frequency curve, in cubic feet per second;

$Q_{\%gr}$ is the peak-flow estimate for selected exceedance probability at a gaging station, value from the appropriate regression equation, in cubic feet per second;

$V_{\%gs}$ is the variance of prediction of a peak-flow estimate for selected exceedance probability at a gaging station, value from station log-Pearson Type III frequency curve, in logarithm units; and

$V_{\%gr}$ is the average variance of prediction of a peak-flow estimate for selected exceedance probability at a gaging station, value associated with the appropriate regression equation, in logarithm units.

For this study, weighted-average peak-flow frequency estimates for gaging stations were computed using the WIE program (Berenbrock and Cohn, 2008). The resulting gaging station weighted-average peak-flow frequency estimates, the regression estimates, and associated variances generated from the WIE program are listed in table 1.

Estimating the Peak-Flow Frequency for an Ungaged Site

The procedure for estimating peak-flow frequency for selected exceedance probabilities for a specific ungaged site depends on whether the site is located near a gaging station on the same stream or is an ungaged site on an ungaged stream. For an ungaged site near a gaging station on the same stream, a drainage-area ratio method should be used. For an ungaged site on an ungaged stream, the regional regression equations developed for this study should be used.

Regression-Weighted and Area-Weighted Estimates for an Ungaged Site on a Gaged Stream

For an ungaged site on a stream with a gaging station that has 10 or more years of peak-flow record, the peak-flow frequency estimate from the appropriate regression equation for the ungaged site can be combined with the weighted-average peak-flow frequency estimate and regression equation peak-flow frequency estimate from the nearby station to produce an improved estimate. Sauer (1974) and Verdi and Dixon (2011) presented the following regression-weighted equation to improve the peak-flow frequency estimate for an ungaged site on a stream with a gaging station:

$$Q_{\%uw} = \left[\left(\frac{2\left|A_g - A_u\right|}{A_g} \right) + \left(\left(1 - \frac{2\left|A_g - A_u\right|}{A_g} \right) \frac{Q_{\%gw}}{Q_{\%gr}} \right) \right] Q_{\%ur} \qquad (3)$$

where

$Q_{\%uw}$ is the peak-flow estimate for selected exceedance probability at an ungaged site, weighted-average value, in cubic feet per second;

$Q_{\%gw}$ is the peak-flow estimate for selected exceedance probability at a gaging station, weighted-average value, in cubic feet per second;

$Q_{\%gr}$ is the peak-flow estimate for selected exceedance probability at a gaging station, value from the appropriate regression equation, in cubic feet per second;

$Q_{\%ur}$ is the peak-flow estimate for selected exceedance probability at an ungaged site, value from the appropriate regression equation, in cubic feet per second;

A_g is the drainage area associated with a gaging station, in square miles; and

A_u is the drainage area associated with an ungaged site, in square miles.

The following simpler area-weighted equation can be used as an alternative:

$$Q_{\%uw} = Q_{\%gw} \left(\frac{A_u}{A_g} \right)^b \qquad (4)$$

where

$Q_{\%uw}$ is the peak-flow estimate for selected exceedance probability at an ungaged site, weighted-average value, in cubic feet per second;

$Q_{\%gw}$ is the peak-flow estimate for selected exceedance probability at a gaging station, weighted-average value, in cubic feet per second;

A_g is the drainage area associated with a gaging station, in square miles;

A_u is the drainage area associated with an ungaged site, in square miles; and

b is the regional exponent of drainage area from the appropriate hydrologic zone.

The regional exponents (*b*, table 5) were derived from WREG using a GLS analysis of log-transformed (base-10) drainage area. The exponents range from 0.499 to 0.628 for zone A, from 0.577 to 0.609 for zone B, and from 0.438 to 0.543 for zone C. The exponent *b* for a selected exceedance probability is recommended for use to obtain an area-weighted peak-flow frequency estimate at an ungaged site on a gaged stream. An average exponent *b* for the range of exceedance probabilities (0.563 for zone A, 0.589 for zone B, and 0.501 for zone C) can be used in equation 4 for each of the peak-flow exceedance frequency estimates as well for a more general peak-flow frequency estimate.

If the drainage area associated with the ungaged site is between 50 and 150 percent of the area associated with the gaging station, equations 3 and 4 are applicable. If the drainage area associated with the ungaged site is less than 50 or greater than 150 percent of the area associated with the gaging station, then no weighting adjustment is applied to the peak-flow frequency regression estimate for the ungaged site.

Table 5. Regional exponents determined from regional regression of log-transformed (base-10) drainage area for an area-weighted ratio method to estimate peak-flow frequency for an ungaged site on a gaged stream in North Dakota.

[*b*, regional exponent of drainage area used in equation 4]

Annual exceedance probability	Hydrologic zone A	Hydrologic zone B	Hydrologic zone C
	Exponent *b*	Exponent *b*	Exponent *b*
0.50	0.499	0.609	0.438
0.20	0.525	0.597	0.475
0.10	0.544	0.591	0.493
0.04	0.566	0.587	0.511
0.02	0.582	0.584	0.521
0.01	0.597	0.581	0.529
0.002	0.628	0.577	0.543
Average	0.563	0.589	0.501

Summary

The U.S. Geological Survey, in cooperation with the North Dakota State Water Commission, the North Dakota Department of Transportation, the North Dakota Department of Health, the Red River Joint Water Resources Board, and the Devils Lake Basin Joint Water Resource Board used data from streamflow-gaging stations in North Dakota and parts of Montana, South Dakota, and Minnesota, with 10 or more years of unregulated peak-flow record, to updated regional regression equations for estimating peak-flow frequency for the 0.5, 0.20, 0.10, 0.04, 0.02, 0.01, and 0.002 exceedance probabilities. Peak-flow frequencies for 262 streamflow-gaging stations were estimated using available data through 2009. An average generalized skew coefficient was developed for three hydrologic zones in North Dakota, replacing the generalized skew values for North Dakota used in previous peak-flow studies. A StreamStats web application tool was developed to generate the basin characteristics for streamflow-gaging stations.

The updated peak-flow frequency data and basin characteristics from 231 gaged sites were used to develop the final regional regression equations using generalized least-squares regression techniques. The final regression equations were chosen based on minimizing values of the standard model error and the standard error of prediction, maximizing values of the pseudo coefficient of determination, and examination of regression residuals. Updated peak-flow frequency data, peak-flow regional regression frequency data, and weighted peak-flow frequency data for streamflow-gaging stations used in the study are provided. Methods were presented for determining weighted peak-flow frequency data for streamflow-gaging stations and at ungaged sites.

References Cited

Berenbrock, C., and Cohn, T., 2008, Weighted independent estimates, WIE: software program accessed October 1, 2009, at http://water.usgs.gov/usgs/osw/FreqReg/.

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.

Crosby, O.A., 1970, A proposed streamflow-data program for North Dakota: U.S. Geological Survey Open-File Report 70–91, 73 p.

Crosby, O.A., 1975, Magnitude and frequency of floods in small drainage basins in North Dakota: U.S. Geological Survey Water-Resources Investigations Report 75–19, 24 p.

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.]

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., accessed April 30, 2014, at http://water.usgs.gov/software/PeakFQ.

Gingerich, S.B., 2005, Median and Low-Flow Characteristics for Streams under Natural and Diverted Conditions, Northeast Maui, Hawaii: U.S. Geological Survey, Scientific Investigations Report 2004–5262, 72 p.

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.

Hortness, J.E., 2006, Estimating low-flow frequency statistics for unregulated streams in Idaho: U.S. Geological Survey Scientific Investigations Report 2006–5035, 31 p.

Hydrology Subcommittee of the Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood flow frequency: Hydrology Subcommittee Bulletin 17B, 28 p., 14 appendixes.

Lorenz, D.L., Sanocki, C.A., and Kocian, M.J., 2010, Techniques for estimating the magnitude and frequency of peak flows on small streams in Minnesota based on data through water year 2005: U.S. Geological Survey Scientific Investigations Report 2009–5250, 54 p.

McCabe, J.A., and Crosby, O.A., 1959, Floods in North Dakota and South Dakota, frequency and magnitude: U.S. Geological Survey Open-File Report 59–80, 132 p.

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 Investigation Report 73–52, 46 p.

Stedinger, J.R., and Tasker, G.D., 1985, Regional hydrologic analysis 1—Ordinary, weighted, and generalized least squares compared: Water Resources Research, v. 21, no. 9, p. 1421–1432.

Tasker, G.D., 1975, Combining estimates of low-flow characteristics of streams in Massachusetts and Rhode Island: Journal of Research of the U.S. Geological Survey, v. 3, no. 1, p. 107–112.

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, 2014, National Water Information System (NWISWeb): U.S. Geological Survey database: accessed July 28, 2014, at http://waterdata.usgs.gov/nwis/.

U.S. Geological Survey, U.S. Department of Agriculture, and Natural Resources Conservation Service, 2012, Federal standards and procedures for the National Watershed Boundary Dataset (WBD) (3rd ed.): U.S. Geological Survey Techniques and Methods book 11, chap. A3, 63 p.

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.

Williams-Sether, Tara, 1992, Techniques for estimating peak-flow frequency relations for North Dakota stream: U.S. Geological Survey Water-Resources Investigations Report 92–4020, 57 p.

Publishing support provided by:
Rolla Publishing Service Center

For more information concerning this publication, contact:
Director, USGS North Dakota Water Science Center
821 East Interstate Avenue
Bismarck, North Dakota 58503
(701) 250–7400

Or visit the North Dakota Water Science Center Web site at:
http://nd.water.usgs.gov/

USGS

Williams-Sether—**Regional Regression Equations to Estimate Peak-Flow Frequency at Sites in North Dakota Using Data through 2009**—SIR 2015–5096

ISSN 2328-0328 (online)
http://dx.doi.org/10.3133/sir20155096