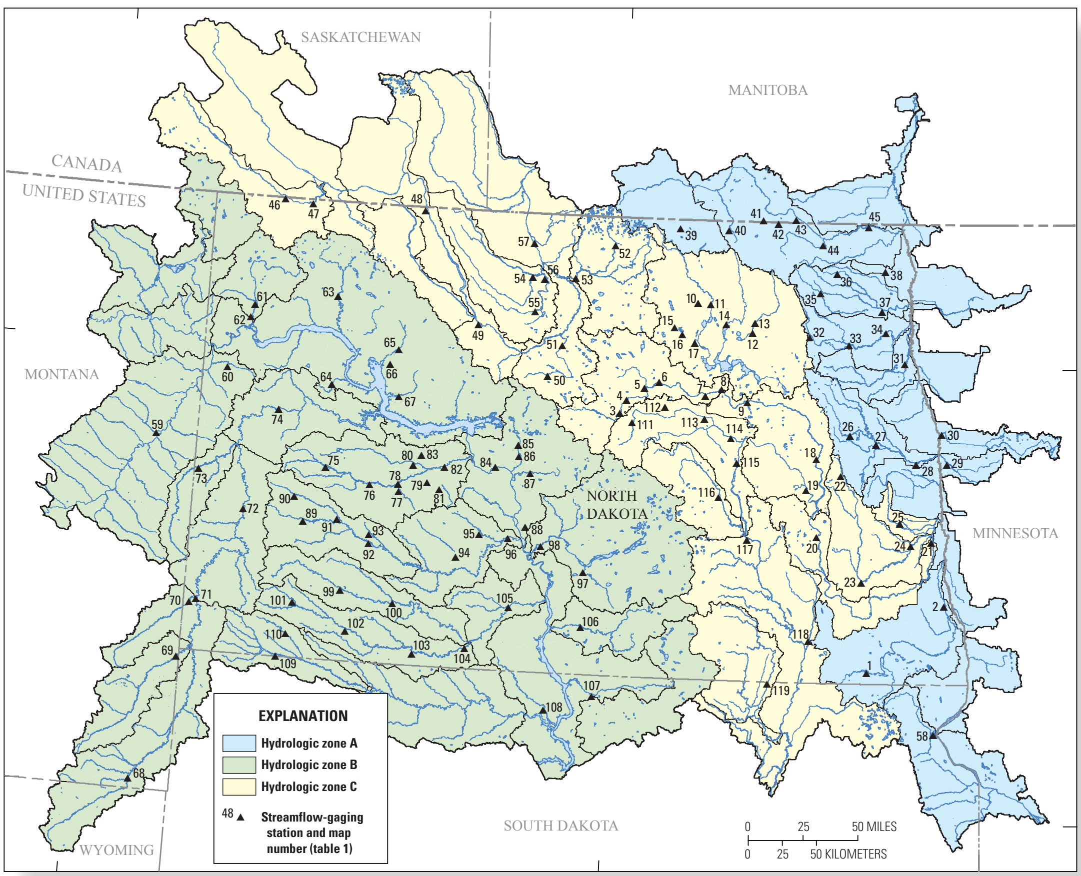
Prepared in cooperation with the North Dakota State Water Commission, North Dakota Department of Transportation, North Dakota Department of Health, Red River Joint Water Resources Board, and Devils Lake Basin Joint Water Resource Board

Regression Equations to Estimate Seasonal Flow Duration, *N*-Day High-Flow Frequency, and *N*-Day Low-Flow Frequency at Sites in North Dakota Using Data through Water Year 2009



Scientific Investigations Report 2015–5184

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

Cover illustration: Modified version of figure 1 from this report.

Regression Equations to Estimate Seasonal Flow Duration, *N*-Day High-Flow Frequency, and *N*-Day Low-Flow Frequency at Sites in North Dakota Using Data through Water Year 2009

By Tara Williams-Sether and Tara A. Gross

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

Scientific Investigations Report 2015–5184

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

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

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

U.S. Geological Survey, Reston, Virginia: 2016

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, and Gross, T.A., 2016, Regression equations to estimate seasonal flow duration, *n*-day high-flow frequency, and *n*-day low-flow frequency at sites in North Dakota using data through water year 2009: U.S. Geological Survey Scientific Investigations Report 2015–5184, 12 p., http://dx.doi.org/10.3133/sir20155184.

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)
	Area	
square mile (mi^2)	259.0	hectare (ha)
square mile (mi^2)	2.590	square kilometer (km^2)
	Flow rate	
inch per hour (in/h)	0 .0254	meter per hour (m/h)

Datum

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

Regression Equations to Estimate Seasonal Flow Duration, *N*-Day High-Flow Frequency, and *N*-Day Low-Flow Frequency at Sites in North Dakota Using Data through Water Year 2009

By Tara Williams-Sether and Tara A. Gross

Abstract

Seasonal mean daily flow data from 119 U.S. Geological Survey streamflow-gaging stations in North Dakota; the surrounding states of Montana, Minnesota, and South Dakota; and the Canadian provinces of Manitoba and Saskatchewan with 10 or more years of unregulated flow record were used to develop regression equations for flow duration, *n*-day high flow and *n*-day low flow using ordinary least-squares and Tobit regression techniques. Regression equations were developed for seasonal flow durations at the 10th, 25th, 50th, 75th, and 90th percent exceedances; the 1-, 7-, and 30-day seasonal mean high flows for the 10-, 25-, and 50-year recurrence intervals; and the 1-, 7-, and 30-day seasonal mean low flows for the 2-, 5-, and 10-year recurrence intervals. Basin and climatic characteristics determined to be significant explanatory variables in one or more regression equations included drainage area, percentage of basin drainage area that drains to isolated lakes and ponds, ruggedness number, stream length, basin compactness ratio, minimum basin elevation, precipitation, slope ratio, stream slope, and soil permeability. The adjusted coefficient of determination for the *n*-day high-flow regression equations ranged from 55.87 to 94.53 percent. The Chi^2 values for the duration regression equations ranged from 13.49 to 117.94, whereas the Chi^2 values for the *n*-day low-flow regression equations ranged from 4.20 to 49.68.

Introduction

Flow statistics, such as flow duration, *n*-day high-flow frequency, and *n*-day low-flow frequency, can be used to characterize flow of a certain magnitude at a location of interest on a stream and are important to Federal, State, and local agencies for water-supply planning and management and water-quality regulatory activities. Flow statistics are also used in design and management decisions for hydroelectric facilities, reservoir storage, fish passage, temporary control of water during construction, culverts, bridges, and irrigation. An accurate calculation of flow statistics is dependent on the availability and quantity of measured flow records on a stream. Although flow statistics can be calculated at streamflow-gaging stations, statistical techniques can be used to make estimates of flow statistics at locations where streamflow-gaging stations do not exist. If the stream location where a flow statistic is needed is close to a streamflow-gaging station, then streamflow information can be extrapolated from the streamflow-gaging station record using a drainage-area ratio method (Perry and others, 2004). For locations farther away from streamflow-gaging stations, regression equations that relate flow statistics with physical and climatic characteristics of drainage basins can be used.

The U.S. Geological Survey (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 regression equations for estimating flow duration, *n*-day high-flow frequency, and *n*-day low-flow frequency statistics for ungaged sites in North Dakota. In addition, the regression equations developed from this study also are included in the North Dakota StreamStats Web-based tool that can be accessed at http://water.usgs.gov/osw/streamstats/. StreamStats allows users to obtain flow statistics, drainage-basin characteristics, and other information for user-selected sites on a stream. Using a geographic information system based interactive map of North Dakota, the user can "point and click" at a location in a stream, and StreamStats will rapidly delineate the basin upstream from the selected location, calculate climatic and basin characteristics, and provide estimated streamflow statistics. The user also can "point and click" on USGS streamflow-gaging stations and receive flow statistics and information about those stations.

Purpose and Scope

This report presents the results of statistical analyses used to compute seasonal flow duration, seasonal n-day high-flow frequency, and seasonal n-day low-flow frequency statistics at unregulated sites in North Dakota using unregulated daily mean streamflow-gaging station data for period of record through water year 2009. A water year is the 12-month period October 1 through September 30 designated by the calendar year in which it ends. North Dakota has several streamflow-gaging stations that only operate seasonally; therefore, seasonal statistics were computed for March through September. The statistics computed include seasonal flow durations (at the 10th, 25th, 50th, 75th, and 90th percent exceedances); the 1-, 7-, and 30-day seasonal mean low flows for the 2-year (0.5 exceedance probability), 5-year (0.2 exceedance probability), and 10-year (0.1 exceedance probability) recurrence intervals; and the 1-, 7-, and 30-day seasonal mean high flows for the 10-year (0.1 exceedance probability), 25-year (0.04 exceedance probability), and 50-year (0.02 exceedance probability) recurrence intervals. In addition, this report also describes the development of regression equations that relate selected climatic and basin characteristics to these flow statistics. These equations can be used to provide estimates of flow conditions at ungaged sites. A minimum of 10 years of unregulated streamflow record were required for streamflow-gaging stations to be included in the statistical and regression analyses.

Previous Studies

Statistical flow summaries have been documented by Haffield (1981), Wiche and Williams-Sether (1997), Williams-Sether and Wiche (1998), Williams-Sether (2012) and Williams-Sether (2015) for selected sites in North Dakota. Development of regression equations for estimating flow statistics such as flow duration, *n*-day high-flow frequency, and *n*-day low-flow frequency at ungaged sites have not been attempted within North Dakota.

Calculation of Seasonal Flow-Duration, *N*-Day High-Flow, and *N*-Day Low-Flow Frequency Statistics

All data used to compute flow statistics in this study were from daily mean flow records at 119 selected streamflow-gaging stations operated in North Dakota; the surrounding states of Montana, Minnesota, and South Dakota; and the Canadian provinces of Manitoba and Saskatchewan (fig. 1; table 1, at the back of this report). Seasonal (March through September) daily mean flow records were retrieved by station number from the USGS National Water Information System (U.S. Geological Survey, 2014). Streamflow-gaging stations selected were required to have a minimum of 10 years of unregulated flow record. Period of record data through water year 2009 were used.

Flow Duration

Flow durations represent the percentage of time that a given flow, measured during a specified time interval, has been equaled or exceeded during that time interval; for example, a 90-percent flow exceedance represents a streamflow that is equaled or exceeded 90 percent of the time. Flow durations characterize the range of flows for the period during which data were collected. Flow duration statistics are computed by sorting the daily mean streamflow values for the period of record from the largest to the smallest and assigning each streamflow value a rank, starting with one for the largest value. The exceedance probabilities are then computed using the Weibull formula for computing plotting position (Helsel and Hirsch, 1992)

$$P = 100 \text{ x} \left[\frac{M}{n+1}\right] \quad (1)$$

where

P is the probability that a given flow will be equaled or exceeded (percentage of time);

M is the ranked position, from highest to lowest, of all daily mean flows for the specified period of record (dimensionless); and

n is the number of daily mean flows for the period of record (dimensionless).

Flow duration statistics are usually computed on an annual basis; however, for this study, flow duration statistics were computed for a seasonal period of March through September to avoid having too sparse of a database. In others words, there would have been too few streamflow-gaging stations with complete annual daily flow records. An internal (not publicly available) USGS computer program, DVSTAT, part of the USGS Automated Data Processing System (U.S. Geological Survey, 2003), was used to compute the seasonal flow durations. The DVSTAT completes these computations by tallying flows in 35 classes, with class boundaries based on the range of data analyzed. The 10th-, 25th-, 50th-, 75th-, and 90th-percent seasonal duration flows for streamflow-gaging stations used in this study are listed in table 1 (at the back of this report).

High-Flow Frequency and Low-Flow Frequency

N-day high-flow frequency data are determined from a series of the highest mean discharges for some specified time period of n consecutive days; for example, an annual series of 7-day high flows consists of the highest mean discharge during any 7-day consecutive period during each year of record. The *n*-day high-flows were computed for a seasonal

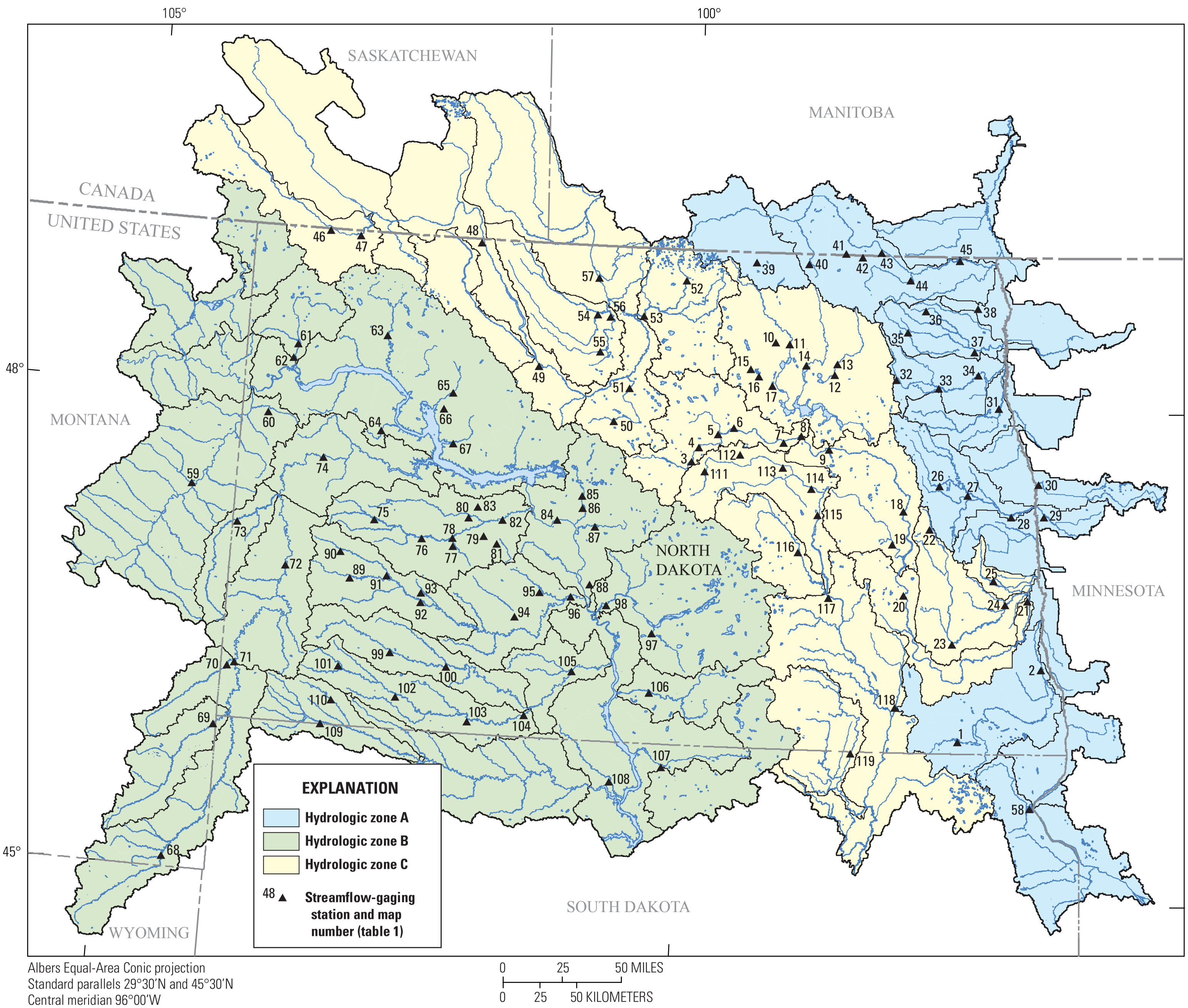


Figure 1. Streamflow-gaging stations and hydrologic zones used in the development of seasonal flow duration, *n*-day high-flow frequency, and *n*-day low-flow frequency regression equations for North Dakota streams.

period of March through September. Likewise, *n*-day low-flow frequency data are determined from a series of the lowest mean discharges for some specified time period of n consecutive days; for example, an annual series of 7-day low flows consists of the lowest mean discharge during any 7-day consecutive period during each year of record. The *n*-day low-flows were computed for a seasonal period of March through September. In both cases, the series was ranked in order of magnitude, and then the recurrence interval for each value was computed and a plot of the streamflows against their respective recurrence interval was prepared (Riggs, 1972). Recurrence intervals typically are computed by fitting a log-Pearson Type III distribution to the data.

The USGS computer program Surface-Water Statistics (SWSTAT) accessed through the U.S. Environmental Protection Agency Better Assessment Science Integrating point & Non-point Sources program (BASINS) (http://www2.epa.gov/exposure-assessment-models/basins) was used to load data and compute the *n*-day high-flow and *n*-day low-flow frequency statistics. The SWSTAT program produces the series of *n*-day values, determines the maximum or minimum values for each year, ranks the time period series, fits the time period series to a log-Pearson type III distribution, plots the resulting curves, and extracts the specified high- or low-flow frequency statistics. The *n*-day high-flow frequencies are expressed in terms of exceedance probability and the *n*-day low-flow frequencies are expressed in terms of non-exceedance probability. The SWSTAT program provides adjusted frequency values if the *n*-day low-flow statistics contain zero values. The values for the 1-, 7-, and 30-day seasonal mean high flows at 10-, 25-, and 50-year recurrence intervals; and the values for the 1-, 7-, and 30-day seasonal mean low flows at 2-, 5-, and 10-year recurrence intervals used in this study are listed in table 1 (at the back of this report).

Basin and Climatic Characteristics

The basin and climatic characteristics used as possible explanatory variables in the development of regression equations were generated by the North Dakota StreamStats Web tool (Williams-Sether, 2015). The 17 basin and climatic characteristics that were generated by the North Dakota StreamStats Web tool are described in table 2. The hydrologic zones A, B, and C (fig. 1) previously defined by Williams-Sether (1992, 2015) are used in this study.

To determine a subset of basin characteristics that might best explain the variations in the flow statistics, the REG procedure within the Statistical Analysis System (SAS) program was used (SAS Institute, Inc., 1990). The REG procedure fits linear regression models using a least-squares method. Usefulness of multiple independent variables was tested for significance in a model using the STEPWISE selection method, which adds variables one by one to the model if the assigned F-statistic significance level is attained. A significance level of 0.08 was assigned for variable entry into a model, but a significance level of 0.05 was chosen to determine which variables were selected to stay in a model.

Development of Regression Equations

A regression model can be developed using flow statistics and basin and climatic characteristics of streamflow-gaging stations to estimate flow statistics at ungaged sites where basin and climatic characteristics can be measured. Multiple linear regression analyses were used to determine which basin and climatic characteristics (the independent variables) best explain, statistically, the variations in the flow statistic (the dependent variable). Regression analyses were also used to develop the final equations that relate the dependent and independent variables.

Ordinary-least-squares (OLS) regression within the commercial statistics and data-management software S–Plus (MathSoft, 1999) was used to develop the final regression equations for the 1-, 7-, and 30-day seasonal mean high flow for the 10-year (0.1 exceedance probability), 25-year (0.04 exceedance probability), and 50-year (0.02 exceedance probability) recurrence intervals for hydrologic zones A, B, and C (table 3). The basin characteristics used in the final *n*-day high-flow frequency regression equations were drainage area (DRNAREA) for zones A, B, and C (fig. 1), percentage of basin drainage area that drains to isolated lakes and ponds (ISOLAKEDA) for zones A and B, ruggedness number (RUGGED) for zone B, and sum of length of all mapped streams (STREAMLENGTH) for zone C. The coefficient of determination (R^2) is used to quantify how well the data fits a regression equation. The R^2 value will range between 0 (indicating no fit) and 1 (for a perfect fit). An adjusted R^2 is a modified version of the coefficient of determination (R^2) that has been adjusted for the number of predictors in the model. The adjusted R^2 increases only if the new term improves the model more than would be expected by chance. It decreases when a predictor improves the model by less than expected by chance. The adjusted R^2 can be negative, but it's usually not. It is always lower than the R^2. For a regression equation with more than one variable, the adjusted R^2 value can also be expressed as a percentage; for example, an adjusted R^2 of 0.90 indicates that 90 percent of the variation in the data can be explained by the regression equation. Adjusted R^2 values for the *n*-day high-flow equations ranged from 80.37 to 94.53 percent for zone A, 84.64 to 93.56 percent for zone B, and 55.87 to 81.26 percent for zone C.

The presence of zero values in the flow durations and *n*-day low-flow frequencies made using OLS regression impractical because of the need to use logarithmic transformations in the regression equation development. To include zero values in a logarithmic transformation analysis, the Tobit analysis was used. Tobit analysis is a widely accepted maximum-likelihood method for estimating a regression-like model when adjusted data are present (Tobin, 1958; Judge and others,

Table 2. Basin and climatic characteristics generated by the North Dakota Streamstats Web tool used in the development of seasonal flow duration, *n*-day high-flow frequency, and *n*-day low-flow frequency 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.

1985; Cohn, 1988). Adjusted data are either censored or have had a discrete value added to them. Censored data are less than a threshold value that is increased to the censoring value (for example, all values less than 0.7 are increased to 0.7). Discrete values are added to all data before logarithmic transformation and then subtracted from all final regression model values.

The Survival Regression Procedure in the S–Plus 2000 software package (MathSoft, 1999) was used in this study to fit the Tobit models. A Tobit analysis was completed for each of the flow durations and for the *n*-day low-flow data sets using a censored value (tables 4 and 5, respectively). The censored values used were determined from graphical analyses of the non-zero data. Regression equations were unable to be developed in hydrologic zone A for the low-flow 7-day 5-year, 7-day 10-year, and 1-day 10-year frequencies because of lack of data. The basin characteristics used in the final flow duration regression equations were drainage area (DRNAREA) for zones B and C, percentage of basin drainage area that drains to isolated lakes and ponds (ISOLAKEDA) for zone B, compactness ratio (COMPRAT) for zone A, sum of length of all mapped streams (STREAMLENGTH) for zones A and C, minimum basin elevation (MINBELEV) for zone C, mean annual precipitation (PRECIP) for zone C, and slope ratio (SLOPERAT) for zone C. The measure of fit of the Tobit analysis model is quantified by the value of Chi^2 statistic (the higher the value, the better the fit). Values of Chi^2 ranged from 13.49 to 47.78 for zone A, 36.96 to 117.94 for zone B, and 22.76 to 61.22 for zone C. The basin characteristics used in the final *n*-day low-flow frequency regression equations were drainage area (DRNAREA) for zones A, B, and C, sum of

Table 3. Regression equations for seasonal *n*-day high-flow frequency estimates in North Dakota.

[R^2, coefficient of determination, in percent; Q_{ndxyr}, n consecutive day mean high flow at the x-year return interval; ≤, less than or equal to; DRNAREA, drainage area, in square miles; ISOLAKEDA, percentage of basin drainage area that drains to isolated lakes and ponds, in percent; RUGGED, ruggedness number, in feet per mile; STREAMLENGTH, sum of length of all mapped streams, in miles]

Regression equation	Range of explanatory variables	Residual standard error, in log units	R^2, in percent	Adjusted R^2, in percent
Hydrologic zone A				
log Q_{1d10yr}=1.761+0.622xlog(DRNAREA)	12.83≤DRNAREA≤3,393; 0.07≤ISOLAKEDA≤100	0.2052	81.75	80.88
log Q_{1d25yr}=1.895+0.645xlog(DRNAREA)		0.2093	82.23	81.39
log Q_{1d50yr}=1.964+0.663xlog(DRNAREA)		0.2222	81.27	80.37
Log Q_{7d10yr}=1.290+0.731xlog(DRNAREA)		0.1832	88.59	88.04
log Q_{7d25yr}=1.361+0.830xlog(DRNAREA)-0.006xlog(ISOLAKEDA)		0.1479	93.34	92.63
log Q_{7d50yr}=1.417+0.858xlog(DRNAREA)-0.007xlog(ISOLAKEDA)		0.1543	93.23	92.51
log $Q_{30d10yr}$=0.590+0.910xlog(DRNAREA)-0.006xlog(ISOLAKEDA)		0.1469	94.51	93.94
log $Q_{30d25yr}$=0.651+0.971xlog(DRNAREA)-0.007xlog(ISOLAKEDA)		0.1478	95.05	94.53
log $Q_{30d50yr}$=0.670+1.015xlog(DRNAREA)-0.008xlog(ISOLAKEDA)		0.1556	94.98	94.45
Hydrologic zone B				
log Q_{1d10yr}=2.892+0.862xlog(DRNAREA)-0.018xlog(ISOLAKEDA)-0.473xlog(RUGGED)	9.08≤DRNAREA ≤8,342; 0.05≤ISOLAKEDA ≤79.67; 153.22≤ RUGGED≤7,820	0.1679	91.24	90.68
log Q_{1d25yr}=3.214+0.850xlog(DRNAREA)-0.018xlog(ISOLAKEDA)-0.507xlog(RUGGED)		0.1855	88.88	88.17
log Q_{1d50yr}=3.417+0.844xlog(DRNAREA)-0.018xlog(ISOLAKEDA)-0.533xlog(RUGGED)		0.2110	85.56	84.64
log Q_{7d10yr}=2.478+0.984xlog(DRNAREA)-0.018xlog(ISOLAKEDA)-0.531xlog(RUGGED)		0.1678	92.78	92.31
log Q_{7d25yr}=2.802+0.977xlog(DRNAREA)-0.018xlog(ISOLAKEDA)-0.567xlog(RUGGED)		0.1831	91.16	90.59
log Q_{7d50yr}=3.019+0.976xlog(DRNAREA)-0.018xlog(ISOLAKEDA)-0.597xlog(RUGGED)		0.2062	88.75	88.04
log $Q_{30d10yr}$=1.864+1.002xlog(DRNAREA)-0.017xlog(ISOLAKEDA)-0.496xlog(RUGGED)		0.1561	93.95	93.56
log $Q_{30d25yr}$=2.164+0.987xlog(DRNAREA)-0.017xlog(ISOLAKEDA)-0.519xlog(RUGGED)		0.1724	92.40	91.91
log $Q_{30d50yr}$=2.376+0.981xlog(DRNAREA)-0.017xlog(ISOLAKEDA)-0.544xlog(RUGGED)		0.1961	90.13	89.50
Hydrologic zone C				
log Q_{1d10yr}=1.572+0.561xlog(DRNAREA)	16.26≤DRNAREA≤7,440; 4.67≤STREAMLENGTH ≤4,776	0.2941	56.89	55.87
log Q_{1d25yr}=1.688+0.591xlog(DRNAREA)		0.2838	61.17	60.24
log Q_{1d50yr}=1.754+0.645xlog(STREAMLENGTH)		0.2691	67.34	66.56
log Q_{7d10yr}=1.219+0.682xlog(STREAMLENGTH)		0.2604	71.09	70.40
log Q_{7d25yr}=1.343+0.711xlog(STREAMLENGTH)		0.2539	73.79	73.16
log Q_{7d50yr}=1.400+0.734xlog(STREAMLENGTH)		0.2644	73.42	72.79
log $Q_{30d10yr}$=0.588+0.802xlog(STREAMLENGTH)		0.2359	80.22	79.73
log $Q_{30d25yr}$=0.724+0.827xlog(STREAMLENGTH)		0.2318	81.71	81.26
log $Q_{30d50yr}$=0.790+0.846xlog(STREAMLENGTH)		0.2418	81.12	80.66

Table 4. Regression equations for seasonal flow duration estimates in North Dakota.

[ft^3/s, cubic foot per second; chi^2, used to measure goodness of fit and to test hypotheses and obtain confidence intervals for the variance of a normally distributed variable; Q_{px}, duration flow for *x* exceedance probability; ≤, less than or equal to; STREAMLENGTH, sum of length of all mapped streams, in miles; COMPRAT, compactness ratio, dimensionless; DRNAREA, drainage area, in square miles; ISOLAKEDA, percentage of basin drainage area that drains to isolated lakes and ponds, in percent; MINBELEV, minimum basin elevation, in feet (North American Vertical Datum of 1988 [NAVD 88]); PRECIP, mean annual precipitation, in inches (normal period 1961–90); SLOPERAT, slope ratio, dimensionless]

Regression equation	Range of explanatory variables	Censor value, in ft^3/s	Number of stations	Number of stations censored	Estimated residual standard error, in log units	Chi^2
Hydrologic zone A						
log Q_{p90}=-5.422+1.646xlog(STREAMLENGTH)	13.91≤STREAMLENGTH ≤3,373;1.89≤COMPRAT ≤4.77	0.100	23	14	0.9885	13.49
log Q_{p75}=-6.118+2.110xlog(STREAMLENGTH)		0.100	23	12	0.9370	21.62
log Q_{p50}=-4.099+1.437xlog(STREAMLENGTH)+ 2.276xlog(COMPRAT)		0.500	23	6	0.3545	47.78
log Q_{p25}=-2.507+1.257xlog(STREAMLENGTH)+ 1.350xlog(COMPRAT)		0.500	23	1	0.3393	47.18
log Q_{p10}=-1.320+1.029xlog(STREAMLENGTH)+ 1.455xlog(COMPRAT)		0.500	23	0	0.2813	47.60
Hydrologic zone B						
log Q_{p90}=-4.281+1.387xlog(DRNAREA)- 0.013xISOLAKEDA	9.08≤DRNAREA≤8,342; 0.05≤ISOLAKEDA ≤79.67	0.032	52	18	0.8573	36.96
log Q_{p75}=-3.074+1.226xlog(DRNAREA)- 0.012xISOLAKEDA		0.100	52	11	0.5205	63.62
log Q_{p50}=-2.251+1.139xlog(DRNAREA)- 0.013xISOLAKEDA		0.100	52	5	0.4948	66.52
log Q_{p25}=-1.410+1.039xlog(DRNAREA)- 0.009xISOLAKEDA		0.500	52	2	0.3282	91.06
log Q_{p10}=-0.799+1.020xlog(DRNAREA)- 0.009xISOLAKEDA		0.500	52	0	0.2436	117.94
Hydrologic zone C						
log Q_{p90}=-1.308+0.979xlog(DRNAREA)- 0.002xMINBELEV	16.26≤DRNAREA≤7,440; 884≤MINBELEV≤1,883 ;15.84≤PRECIP≤21.34; 0.01≤SLOPERAT≤0.16; 4.67 ≤STREAMLENGTH ≤4,776	0.316	44	34	0.6221	22.76
log Q_{p75}=-1.027+1.068xlog(DRNAREA)- 0.002xMINBELEV		1.000	44	32	0.4629	35.16
log Q_{p50}=-7.754+0.865xlog(DRNAREA)+ 0.214xPRECIP-1.185xlog(SLOPERAT)		1.000	44	18	0.4054	59.26
log Q_{p25}=-3.819+0.588xlog(STREAMLENGTH)+ 0.120xPRECIP-0.876xlog(SLOPERAT)		1.000	44	2	0.3354	59.00
log Q_{p10}=-2.169+0.621xlog(STREAMLENGTH)+ 0.086xPRECIP-0.575xlog(SLOPERAT)		1.000	44	0	0.2904	61.22

Table 5. Regression equations for seasonal *n*-day low-flow frequency estimates in North Dakota.

[ft^3/s, cubic foot per second; chi^2, used to measure goodness of fit and to test hypotheses and obtain confidence intervals for the variance of a normally distributed variable; Q_{ndxyr}, *n* consecutive day mean low flow at the *x*-year return interval; ≤, less than or equal to; STREAMLENGTH, sum of length of all mapped streams, in miles; COMPRAT, compactness ratio, dimensionless; CSL1085LFP, stream slope, change in elevation between points 10 and 85 percent of length along the longest flow path, in feet per mile; DRNAREA, drainage area, in square miles; ISOLAKEDA, percentage of basin drainage area that drains to isolated lakes and ponds, in percent; SOILPERM, average soil permeability, in inches per hour; SLOPERAT, slope ratio, dimensionless]

Regression equation	Range of explanatory variables	Censor value, in ft^3/s	Number of stations	Number of stations censored	Estimated residual standard error, in log units	Chi^2
Hydrologic zone A						
log Q_{1d2yr}=-4.371+1.292xlog(STREAMLENGTH)	44.60≤STREAMLENGTH ≤ 3,373; 2.06≤COMPRAT ≤3.64	0.032	14	4	1.0340	4.22
log Q_{1d5yr}=-9.897+16.700xlog(COMPRAT)		0.003	14	8	1.7920	4.26
log Q_{1d10yr}=not determined						
log Q_{7d2yr}=-6.081+1.804xlog(STREAMLENGTH)	44.60≤STREAMLENGTH ≤3,373	0.010	16	5	1.2640	5.60
log Q_{7d5yr}=not determined						
log Q_{7d10yr}=not determined						
log Q_{30d2yr}=-7.813+2.162xlog(DRNAREA)+2.049xlog (CSL1085LFP)	12.83≤DRNAREA≤ 3,393; 1.02≤CSL1085LFP ≤51.92; 13.91≤ STREAMLENGTH ≤3,373	0.100	20	10	1.0030	14.61
log Q_{30d5yr}=-8.432+2.023xlog(DRNAREA)+2.422xlog (CSL1085LFP)		0.032	20	11	1.2160	10.11
log $Q_{30d10yr}$=-6.533+1.593xlog(STREAMLENGTH)		0.010	20	12	1.5470	5.33
Hydrologic zone B						
log Q_{1d2yr}=-4.641+1.406xlog(DRNAREA)	24.43≤DRNAREA≤8,342; 0.11≤ISOLAKEDA≤77.68	0.032	44	14	0.8115	25.52
log Q_{1d5yr}=-4.014-3.721xlog(ISOLAKEDA)+6.604xsqrt (log(ISOLAKEDA))		0.032	37	19	0.9925	4.70
log Q_{1d10yr}=-10.406-8.293xlog(ISOLAKEDA)+16.492x sqrt(log(ISOLAKEDA))		0.003	37	25	1.7970	8.37
log Q_{7d2yr}=-4.731+1.461xlog(DRNAREA)	15.50≤DRNAREA≤8,342	0.010	47	11	0.9136	30.46
log Q_{7d5yr}=-6.602+1.739xlog(DRNAREA)		0.010	47	25	1.2530	20.38
log Q_{7d10yr}=-6.091+1.205xlog(DRNAREA)		0.010	47	33	1.7240	4.57
log Q_{30d2yr}=-4.315+1.519xlog(DRNAREA)		0.010	48	7	0.6941	49.68
log Q_{30d5yr}=-5.383+1.618xlog(DRNAREA)		0.010	48	16	1.0370	28.54
log $Q_{30d10yr}$=-6.672+1.901xlog(DRNAREA)		0.010	49	23	1.1900	27.50
Hydrologic zone C						
log Q_{1d2yr}=-5.424+1.214xlog(DRNAREA)+2.240xlog (SOILPERM)	16.26≤DRNAREA≤7,440; 1.30≤SOILPERM≤8.57; 0.01≤SLOPERAT≤0.08	0.032	22	8	0.8512	13.57
log Q_{1d5yr}=-5.323-2.273xlog(SLOPERAT)		0.032	22	13	1.1150	5.14
log Q_{1d10yr}=-4.984-1.950xlog(SLOPERAT)		0.032	22	14	1.0680	4.20
log Q_{7d2yr}=-7.263+1.555xlog(DRNAREA)+3.336xlog (SOILPERM)		0.032	25	12	1.0610	13.74
log Q_{7d5yr}=-7.309-3.283xlog(SLOPERAT)		0.032	25	17	1.3220	7.02
log Q_{7d10yr}=-8.458-3.600xlog(SLOPERAT)		0.010	25	17	1.6290	5.83
log Q_{30d2yr}=-7.154+1.576xlog(DRNAREA)+3.376xlog (SOILPERM)	16.26≤DRNAREA≤7,440; 0.80≤SOILPERM≤8.57; 0.01≤SLOPERAT≤0.16	0.032	39	21	1.2240	20.98
log Q_{30d5yr}=-6.496+1.215xlog(DRNAREA)+2.786xlog (SOILPERM)		0.032	39	26	1.2760	12.27
log Q_{3010yr}=-7.411-2.848xlog(SLOPERAT)		0.010	39	28	1.8630	4.86

length of all mapped streams (STREAMLENGTH) for zone A, percentage of basin drainage area that drains to isolated lakes and ponds (ISOLAKEDA) for zone B, compactness ratio (COMPRAT) for zone A, slope ratio (SLOPERAT) for zone C, stream slope computed using the longest flow path (CSL1085LFP) for zone A, and average soil permeability (SOILPERM) for zone C. Values of Chi^2 ranged from 4.22 to 14.61 for zone A, 4.57 to 49.68 for zone B, and 4.20 to 20.98 for zone C. The low Chi^2 values in the flow duration and *n*-day low-flow frequency regressions are most likely because of lack of uncensored data and excessive zero flow values in the data. More nonzero data would be needed to generate better-fitting regression equations.

Limitations of the Regional Regression Equations

The following limitations should be considered when using the regression equations to compute flow durations and *n*-day flow frequencies for North Dakota streams: (1) the stream sites should be in rural drainage basins and not significantly affected by urbanization or regulation, (2) the explanatory variables should be computed using the same geographic information system techniques that were used to develop variables for the regression equations, and (3) the explanatory variables should stay within the range of the data used to develop the regression equations as shown in tables 3, 4 and 5.

Web Application for Solving Regional Regression Equations

The North Dakota StreamStats Web application (http://water.usgs.gov/osw/streamstats/) incorporates the new flow duration, *n*-day high-flow, and *n*-day low-flow regression equations and will provide streamflow estimates for most unregulated sites in the State. Streamflow estimates will not be available for sites on a stream considered to be regulated or for unregulated sites that have some part of a drainage basin outside the area covered by North Dakota StreamStats. The Web application includes (1) a mapping tool to specify a location on a stream where flow statistics are desired; (2) a database that includes peak-flow, duration, and *n*-day frequency statistics, hydrologic characteristics, location, and descriptive information for all USGS streamflow-gaging stations used in this study; and (3) an automated geographic information system procedure that measures the required basin and climatic characteristics and solves the regression equations to estimate flow statistics for user-selected sites. Using North Dakota StreamStats to compute the flow statistics discussed in this report will also help avoid “misuse” of the regression equations by alerting users when a selected site has one or more basin and climatic characteristics outside the range of values identified in tables 3, 4, and 5.

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 seasonal mean daily flow data from 119 U.S. Geological Survey streamflow-gaging stations in North Dakota; the surrounding states of Montana, Minnesota, and South Dakota; and the Canadian provinces of Manitoba and Saskatchewan with 10 or more years of unregulated flow record to develop regression equations for flow duration, *n*-day high-flow, and *n*-day low-flow using ordinary least-squares and Tobit regression techniques. Regression equations were developed for seasonal flow durations at the 10th, 25th, 50th, 75th, and 90th percent exceedances; the 1-, 7-, and 30-day seasonal mean high flows for the 10-, 25-, and 50-year recurrence intervals; and the 1-, 7-, and 30-day seasonal mean low flows for the 2-, 5-, and 10-year recurrence intervals. Basin and climatic characteristics determined to be significant explanatory variables in one or more regression equations included drainage area, percentage of basin drainage area that drains to isolated lakes and ponds, ruggedness number, stream length, basin compactness ratio, minimum basin elevation, precipitation, slope ratio, stream slope, and soil permeability. The adjusted coefficient of determination for the *n*-day high-flow regression equations ranged from 55.87 to 94.53 percent. The Chi^2 values for the duration regression equations ranged from 13.49 to 117.94, whereas the Chi^2 values for the *n*-day low-flow regression equations ranged from 4.20 to 49.68.

References Cited

Cohn, T.A., 1988, Adjusted maximum likelihood estimation of the moments of log-normal population of Type I censored samples: U.S. Geological Survey Open-File Report 88–350, 34 p. [Also available at http://pubs.usgs.gov/of/1988/0350/report.pdf.]

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. [Also available at http://pubs.usgs.gov/sir/2004/5262/.]

Haffield, N.D., 1981, Statistical summaries of streamflow and water-quality data for streams of western North Dakota, 1977–80: U.S. Geological Survey Open-File Report 81–1066, 78 p.

Helsel, D.R., and Hirsch, R.M., 1992, Statistical methods in water resources: New York, Elsevier Science Publishing Company, Inc., Studies in Environmental Science, v. 49, 522 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. [Also available at http://pubs.usgs.gov/sir/2006/5035/.]

Judge, G.G., Griffiths, W.E., Hill, R.C., Lutkepohl, H., and Lee, T.C., 1985, Qualitative and limited dependent variable models, chap. 18 *of* Judge, G.G., Griffiths, W.E., Hill, R.C., Lutkepohl, H., and Lee, T.C., The theory and practice of econometrics (2d ed.): New York, Wiley, 1,064 p.

MathSoft, 1999, S–PLUS 2000 guide to statistics: Seattle, Washington, MathSoft, Inc., v. 1–3, [variously paged].

Perry, C.A., Wolock, D.M., and Artman, J.C., 2004, Estimates of flow duration, mean flow, and peak-discharge frequency values for Kansas stream locations: U.S. Geological Survey Scientific Investigations Report 2004–5033, 651 p. [Also available at http://pubs.usgs.gov/sir/2004/5033/pdf/sir2004.5033front.pdf.]

Riggs, H.C., 1972, Low-flow investigations: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. B1, 18 p. [Also available at http://pubs.usgs.gov/twri/twri4b1/.]

SAS Institute Inc., 1990, The REG procedure, chap. 73 *of* SAS/STAT user's guide (4th ed.): Cary, N.C., SAS Institute Inc., v. 2, no. 6, p. 1351–1456.

Tobin, James, 1958, Estimation of relationships for limited dependent variables: Econometrica, v. 26, no. 1, p. 24–36. [Also available at http://dx.doi.org/10.2307/1907382.]

U.S. Geological Survey, 2014, USGS water data for the Nation: U.S. Geological Survey, National Water Information System, accessed July 28, 2014, at http://waterdata.usgs.gov/nwis/.

U.S. Geological Survey, 2003, User's manual for the National Water Information System of the U.S. Geological Survey Automated Data Processing System (ADAPS) (ver. 4.3): U.S. Geological Survey Open-File Report 03–123. [Also available at http://pubs.usgs.gov/of/2003/ofr03123/.]

Wiche, G.J., and Williams-Sether, Tara, 1997, Streamflow characteristics of streams in the upper Red River of the North Basin, North Dakota, Minnesota, and South Dakota: U.S. Geological Survey Open-File Report 97–416, 374 p. [Also available at http://pubs.er.usgs.gov/publication/ofr97416.]

Williams-Sether, Tara, 1992, Techniques for estimating peak-flow frequency relations for North Dakota streams: U.S. Geological Survey Water-Resources Investigations Report 92–4020, 57 p. [Also available at http://pubs.er.usgs.gov/publication/wri924020.]

Williams-Sether, Tara, 2012, Streamflow statistics for selected streams in North Dakota, Minnesota, Manitoba, and Saskatchewan: U.S. Geological Survey Open-File Report 2012–1147, 11 p. [Also available at http://pubs.usgs.gov/of/2012/1147/.]

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. [Also available at http://dx.doi.org/10.3133/sir20155096.]

Williams-Sether, Tara, and Wiche, G.J., 1998, Streamflow statistics of selected streams in the lower Red River of the North Basin, in North Dakota, Minnesota, and Manitoba: U.S. Geological Survey Open-File Report 98–21, 407 p. [Also available at http://pubs.er.usgs.gov/publication/ofr9821.]

Table 1

Table 1. Climatic and basin characteristics and seasonal flow duration, *n*-day high-flow frequency, and *n*-day low-flow frequency discharges for streamflow-gaging stations used in the regression analyses for North Dakota.

Table 1 is an Excel file that can be accessed at http://dx.doi.org/10.3133/sir20155184.

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/

Williams-Sether and Gross—**Regression Equations to Estimate Seasonal Flow Duration, *N*-day High-flow Frequency, and *N*-day Low-flow Frequency**—Scientific Investigations Report 2015–5184

ISSN 2328-0328 (online)
http://dx.doi.org/10.3133/sir20155184