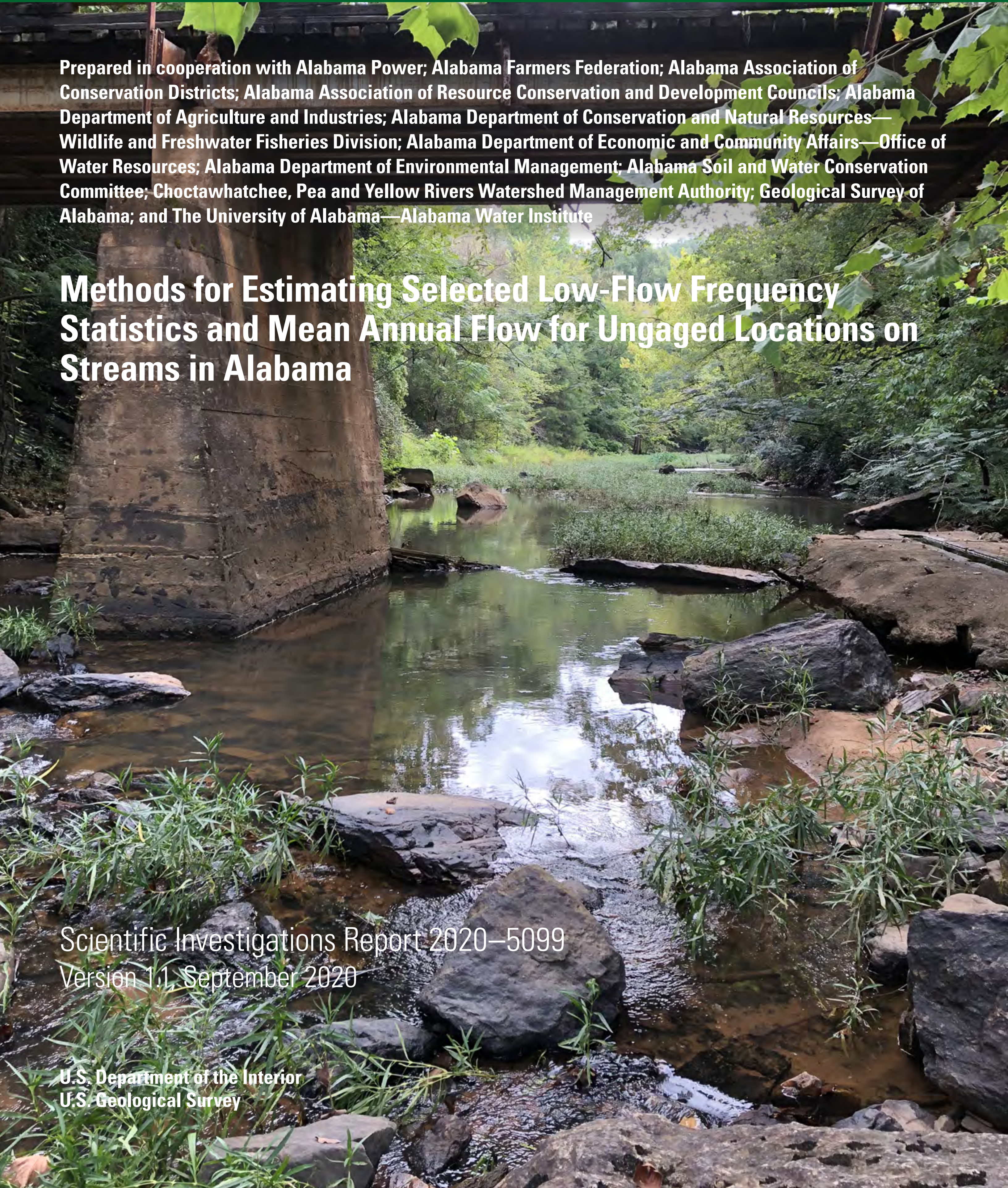
![USGS science for a changing world]



Prepared in cooperation with Alabama Power; Alabama Farmers Federation; Alabama Association of Conservation Districts; Alabama Association of Resource Conservation and Development Councils; Alabama Department of Agriculture and Industries; Alabama Department of Conservation and Natural Resources—Wildlife and Freshwater Fisheries Division; Alabama Department of Economic and Community Affairs—Office of Water Resources; Alabama Department of Environmental Management; Alabama Soil and Water Conservation Committee; Choctawhatchee, Pea and Yellow Rivers Watershed Management Authority; Geological Survey of Alabama; and The University of Alabama—Alabama Water Institute

Methods for Estimating Selected Low-Flow Frequency Statistics and Mean Annual Flow for Ungaged Locations on Streams in Alabama

Scientific Investigations Report 2020–5099
Version 1.1, September 2020

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

Cover. Photograph showing Chewacla Creek near U.S. Geological Survey streamflow-gaging station 02418760 Chewacla Creek at Chewacla State Park near Auburn, Alabama, August 22, 2019.

Methods for Estimating Selected Low-Flow Frequency Statistics and Mean Annual Flow for Ungaged Locations on Streams in Alabama

By Toby D. Feaster, Katharine R. Kolb, Jaime A. Painter, and Jimmy M. Clark

Prepared in cooperation with Alabama Power; Alabama Farmers Federation; Alabama Association of Conservation Districts; Alabama Association of Resource Conservation and Development Councils; Alabama Department of Agriculture and Industries; Alabama Department of Conservation and Natural Resources—Wildlife and Freshwater Fisheries Division; Alabama Department of Economic and Community Affairs—Office of Water Resources; Alabama Department of Environmental Management; Alabama Soil and Water Conservation Committee; Choctawhatchee, Pea and Yellow Rivers Watershed Management Authority; Geological Survey of Alabama; and The University of Alabama—Alabama Water Institute

Scientific Investigations Report 2020–5099
Version 1.1, September 2020

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

U.S. Department of the Interior
DAVID BERNHARDT, Secretary

U.S. Geological Survey
James F. Reilly II, Director

U.S. Geological Survey, Reston, Virginia
First release: 2020, online and in print
Revised: September 29, 2020 (ver. 1.1), online

Suggested citation:
Feaster, T.D., Kolb, K.R., Painter, J.A., and Clark, J.M., 2020, Methods for estimating selected low-flow frequency statistics and mean annual flow for ungaged locations on streams in Alabama (ver. 1.1, September 29, 2020): U.S. Geological Survey Scientific Investigations Report 2020–5099, 21 p., https://doi.org/10.3133/sir20205099.

Associated data for this publication:
Kolb, K.R., Clark, J.M., Feaster, T.D., and Painter, J.A., 2020, Supporting data for estimating selected low-flow frequency statistics and mean annual flow for ungaged locations on streams in Alabama: U.S. Geological Survey data release, https://doi.org/10.5066/P994UFS7.

ISSN 2328-031X (print)
ISSN 2328-0328 (online)

ISBN 978-1-4113-4378-8

Acknowledgments

The authors would like to thank the numerous cooperators who helped make this investigation possible through their cooperative funding. The streamflow data included in this investigation were computed from data collected at U.S. Geological Survey (USGS) streamgages operated in cooperation with a variety of Federal, State, and local agencies. The authors acknowledge the commitment of those cooperators in understanding the importance of having such a long-term data-collection program.

We also acknowledge Athena Clark who was the Director of the USGS Alabama Water Science Center when this project began and is now the Science Advisor and Tribal Liaison for the USGS Southeast Region. Although she was not obligated to do so, Athena continued to meet with cooperators and give oversight and guidance throughout the duration of the project. In addition, we acknowledge the commitment and vital work of the USGS field-office personnel who collect, process, and store the water-resources data needed for this and many other such investigations.

Contents

Figures

Tables

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
	Length	
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	
foot per second (ft/s)	0.3048	meter per second (m/s)
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)
inch per hour (in/h)	0.0254	meter per hour (m/h)

Datum

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

Abbreviations

1Q10	annual minimum 1-day average flow with a 10-year recurrence interval
7Q2	annual minimum 7-day average flow with a 2-year recurrence interval
7Q10	annual minimum 7-day average flow with a 10-year recurrence interval
APS	all-possible-subsets (regression analysis)
EPA	U.S. Environmental Protection Agency
GIS	geographic information system
MOVE.1	Maintenance of Variance Extension, Type 1
OLS	ordinary least squares
SEE	standard error of estimate
SVI	streamflow-variability index
USGS	U.S. Geological Survey
VIF	variance inflation factor

Methods for Estimating Selected Low-Flow Frequency Statistics and Mean Annual Flow for Ungaged Locations on Streams in Alabama

By Toby D. Feaster, Katharine R. Kolb, Jaime A. Painter, and Jimmy M. Clark

Abstract

Streamflow data and statistics are vitally important for proper protection and management of the water quality and water quantity of Alabama streams. Such data and statistics are generally available at U.S. Geological Survey streamflow-gaging stations, also referred to as streamgages or stations, but are often needed at ungaged stream locations. To address this need, the U.S. Geological Survey, in cooperation with numerous Alabama State agencies and organizations, developed regional regression equations for estimating selected low-flow frequency statistics and mean annual flow for ungaged locations on streams in Alabama that are not substantially affected by tides, regulation, diversions, or other anthropogenic influences. A small percentage of the streamgages included in this study experience zero flows during certain periods; thus, the final low-flow frequency regression equations were developed by using weighted left-censored regression analyses to analyze the flow data in an unbiased manner, with weights based on number of years of record.

The equations developed include the annual minimum 1- and 7-day average streamflows with a 10-year recurrence interval (referred to as the 1Q10 and 7Q10 flows), the annual minimum 7-day average streamflow with a 2-year recurrence interval (referred to as the 7Q2 flow), and the mean annual flow using data from 174 streamgages from Alabama and surrounding States. For the 1Q10, 7Q2, and 7Q10 low-flow frequency statistics, the regional regression equations are functions of drainage area, streamflow-variability index, mean annual precipitation, and percentage of the drainage basin located in the Piedmont and Southeastern Plains ecoregions. The mean annual flow regression equation is a function of drainage area, mean annual precipitation, and percentage of the drainage basin located in the Southeastern Plains ecoregion. For the mean annual flow regression equation, the average standard error of estimate was 12.3 percent. For the selected low-flow frequency equations, the average standard errors of estimate ranged from 18.7 to 38.8 percent.

The regional regression equations developed from this investigation have been incorporated into the U.S. Geological Survey StreamStats application for Alabama. StreamStats (https://streamstats.usgs.gov/ss/) is a web-based geographic information system application that delineates drainage basins at selected stream locations and then generates the needed basin characteristics for available regional regression equations. Along with the low-flow frequency equations developed in this investigation, the StreamStats application also has regional regression equations for estimating flood-frequency statistics at locations on rural and urban streams in Alabama.

Introduction

As part of their mission to protect public health and aquatic ecosystems, Alabama State agencies need accurate and representative streamflow statistics to establish realistic and applicable criteria for water quality and water quantity. Historically, low-flow statistics, such as the annual minimum 7-day average flow that likely will occur once, on average, every 10 years (7Q10), have been used by water-resource managers and planners as a threshold criterion for applying the chronic aquatic life criteria for such things as determining waste-load allocations for point sources, total maximum daily loads for streams, and the quantity of water that can be safely withdrawn from a particular stream (Alabama Department of Environmental Management, 2012). Because of the importance of these applications, it is critical to effectively measure and document base-flow data for use in updating low-flow frequency relations on a regular basis, preferably every 10 years (Riggs, 1972), and especially after periods of extreme low flow as have occurred over the last decade or so in the southeastern United States (Feaster and Lee, 2017).

In 2017, the U.S. Geological Survey (USGS) completed an investigation to update selected low-flow frequency and daily duration characteristics at 210 continuous-record streamflow-gaging stations (also referred to as streamgages or stations) in Alabama and 67 streamgages in the adjoining States of Florida, Georgia, Mississippi, and Tennessee by using available daily mean streamflow data through March 2014 (Feaster and Lee, 2017). If low-flow statistics are needed at a streamgage location or a nearby ungaged location, these updated statistics can provide critical information for water-resources managers and planners. However, at locations where streamgages are not available or the record length is insufficient to provide reliable estimates, multiple linear regression techniques may provide a viable option for estimating streamflow statistics (Hedgecock and Feaster, 2007; Farmer and others, 2019). By using multiple linear regression techniques, streamflow statistics computed at USGS streamgages can be related to basin and climatic characteristics in a region of interest, resulting in regional regression equations that can be used to estimate the streamflow statistic at an ungaged location.

Purpose and Scope

The purpose of this report is to present methods for estimating low-flow frequency statistics for the annual minimum 1- and 7-day average flows with a 10-year recurrence interval (1Q10 and 7Q10, respectively), the annual minimum 7-day average flow with a 2-year recurrence interval (7Q2), and mean annual flow for ungaged locations on Alabama streams that are not substantially affected by tides, regulation, diversions, urbanization, or other anthropogenic influences. The regional regression analysis is based on a subset of the low-flow frequency statistics published by Feaster and Lee (2017) that are related to basin and climatic characteristics that were computed for those streamgages and are presented in this report. Mean annual streamflow, which is defined as the average of the daily mean flows for a given year, was computed from daily mean flows through September 2018 based on the water year, which is defined as the 12-month period from October 1 to September 30 with the year being designated by the calendar year in which it ends. The data and geospatial datasets used to create the new low-flow and mean annual flow regression equations are available from Kolb and others (2020).

Study Area

The study area encompasses most of the State of Alabama except for coastal areas where low flows are tidally influenced. Shared basins and ecoregions from adjoining States are also included in the study area. Alabama encompasses 52,420 square miles (mi^2) in the southeastern United States (USGS, 2016) and lies within six U.S. Environmental Protection Agency (EPA) Level III ecoregions—Piedmont (9.3 percent), Southeastern Plains (59.6 percent), Ridge and Valley (8.6 percent), Southwestern Appalachians (14.6 percent), Interior Plateau (6.4 percent), and Southern Coastal Plain (1.5 percent) (fig. 1; EPA, 2016). The ecoregions represent areas of general similarity in ecosystems and type, quality, and quantity of environmental resources, and they provide a spatial framework for research, assessment, management, and monitoring of ecosystems and ecosystem components. The ecoregions were determined from an analysis of the spatial patterns and the composition of biotic and abiotic phenomena that include geology, physiography, vegetation, climate, soils, land use, wildlife, and hydrology (Omernik, 1987).

The Southwestern Appalachians ecoregion extends from Alabama to Kentucky and is composed of open, low mountains containing a mosaic of forest and woodland with some cropland and pasture. The eastern boundary of this ecoregion, along the more abrupt escarpment where it meets the Ridge and Valley ecoregion, is relatively smooth and only slightly notched by small, eastward-flowing streams. The Ridge and Valley is composed of roughly parallel ridges and valleys of various widths, heights, and geologic materials. The western boundary of the Southwestern Appalachians ecoregion is shared with the Interior Plateau ecoregion, which extends from southern Indiana and Ohio to northern Alabama. Elevations in the Interior Plateau are lower than those in the Southwestern Appalachians. Limestone, chert, sandstone, siltstone, and shale of Mississippian to Ordovician age compose the landforms of open hills, irregular plains, and tablelands of the Interior Plateau (Omernik, 1987).

The Piedmont ecoregion, which extends from Alabama to New Jersey, is a transitional area between the mountainous ecoregions of the Appalachians to the northwest and the relatively flat coastal plain to the southeast. The Piedmont is a complex mosaic of metamorphic and igneous rocks of Precambrian and Paleozoic age and contains moderately dissected irregular plains and some hills. The soils tend to be finer textured than those in the coastal plain. The Piedmont was once a largely cultivated region, but much of it has reverted to pine and hardwood woodlands, with increasing conversion to urban and suburban land cover (Omernik, 1987).

The Fall Line, which extends from Alabama to New Jersey, is a geologic feature that separates the higher elevation Interior Plateau, Southwestern Appalachians, Ridge and Valley, and Piedmont ecoregions from the lower elevation Southeastern Plains ecoregion (fig. 1). The Southeastern Plains ecoregion, which extends from Virginia to Louisiana and Tennessee, is composed of irregular plains made up of a mixture of cropland, pasture, woodland, and forest. The sands, silts, and clays of this ecoregion contrast geologically with the older rocks of the Piedmont ecoregion. Elevations and relief in the Southeastern Plains are greater than in the Southern Coastal Plain but generally are less than in much of the Piedmont. Streams in the Southeastern Plains have relatively low gradients and sandy bottoms. The Southern Coastal Plain ecoregion consists of mostly flat plains but is heterogeneous, containing barrier islands, coastal lagoons, marshes, and swampy lowlands along the Gulf and Atlantic coasts. Relative to the Southeastern Plains ecoregion, the Southern Coastal Plain ecoregion is lower in elevation, with less relief and wetter soils (Omernik, 1987).

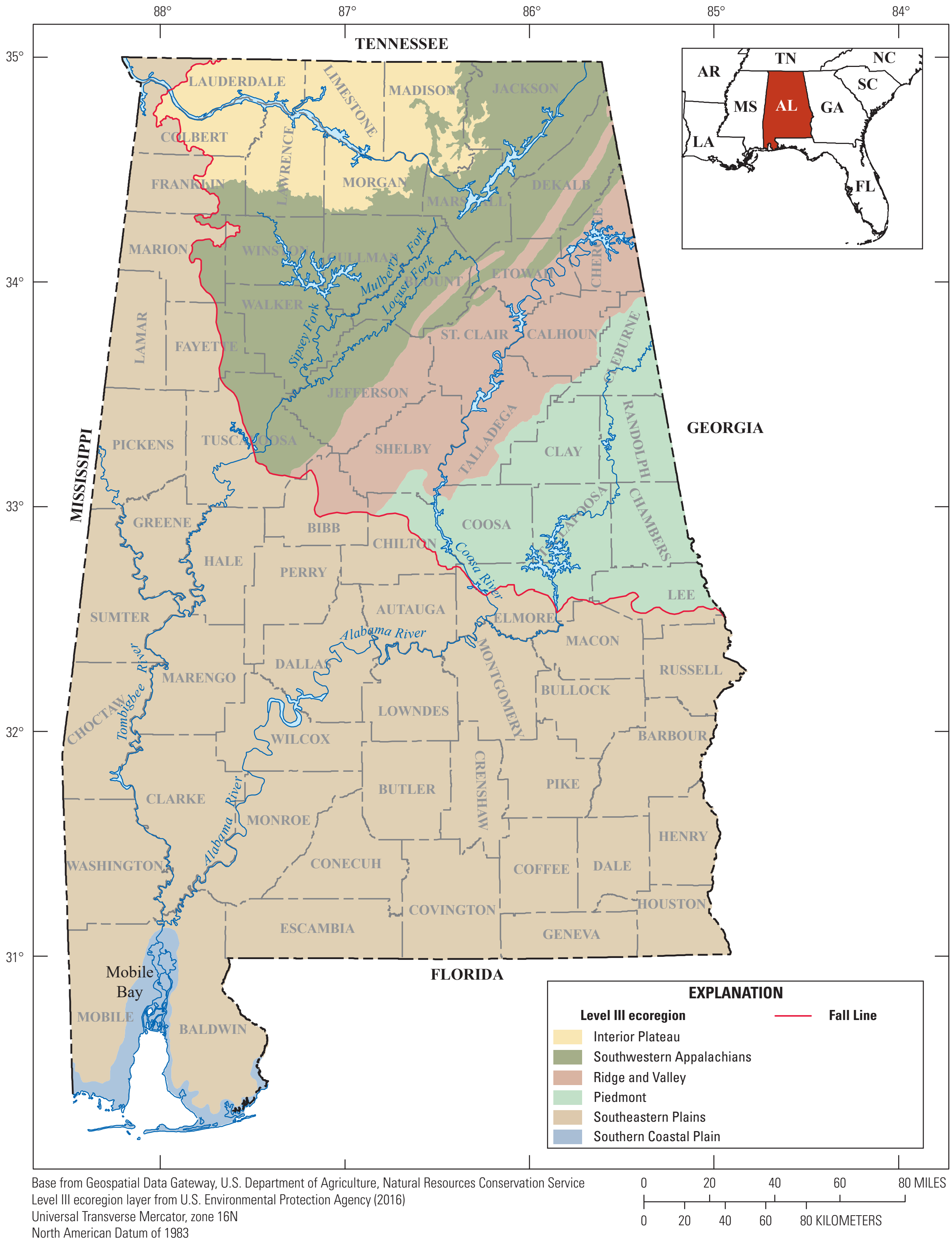


Figure 1. U.S. Environmental Protection Agency Level III ecoregions (U.S. Environmental Protection Agency, 2016) in Alabama.

Previous Studies

Feaster and Lee (2017) detailed previous studies that provided low-flow frequency and flow duration statistics at selected USGS streamgages in Alabama going back to the late 1950s. With respect to previous regionalization studies of low-flow frequency statistics, Bingham (1982) developed low-flow frequency regression equations for the 7Q2 and 7Q10 low-flow frequency statistics. The equations were applicable statewide for natural streams that were not substantially altered by anthropogenic activities. The equations were functions of drainage area, mean annual precipitation, and streamflow recession index. The streamflow-recession index is defined in days per log cycle, which is the number of days required for streamflow to decline one complete log cycle. To compute the streamflow recession index at a streamgaging location, streamflow recess curves were plotted on semi-log graph paper for several different recessions. Streamflow records for recession periods during November through February were generally used, and the average slope of the recession curves of base flow was used to determine the streamflow-recession index for that site. The streamflow-recession index values computed for the USGS streamgage locations were used to delineate areas determined to have similar values and were provided on a map that could be used to estimate the streamflow-recession index at ungaged locations.

Selection of Streamgages

As previously noted, the streamgages used in the Alabama low-flow regional regression analysis were a subset of the 277 streamgages (210 within Alabama and 67 in adjoining States) included in Feaster and Lee (2017), which included streamflow data through March 2014 and only included nontidal streamgages (fig. 2). For ease of reference, the map index numbers used in this report match the map index numbers used by Feaster and Lee (2017). Streamgages with low-flow frequency statistics computed for regulated or urbanized basins were excluded (appendix 1). In addition, the following USGS stations in Alabama were excluded because they are springs with indeterminate drainage areas: 02403500 W 12 Coldwater Spring near Anniston, and 03590500 Tuscumbia Spring at Tuscumbia. The following streamgages in Alabama were excluded because they have large drainage basins (17,095, 25,610, and 30,810 mi^2, respectively): 02423000 Alabama River at Selma, 03575500 Tennessee River at Whitesburg, and 03589500 Tennessee River at Florence.

Determining a true natural condition in a stream or accurately accounting for all diversions in a basin is generally problematic. Feaster and Lee (2017) noted that diversions from natural flow in a stream can occur for a variety of reasons. In some cases, diversions in a stream may only affect the flows from a short distance along the stream. For example, water may be removed from the river channel, passed through a manufacturing plant for use in processing, cooling, or dilution of wastes, and then returned to the river. In such cases, consumptive losses from diversions may be negligible (Ries, 1994). Feaster and Lee (2017) completed a series of quality assurance and quality control reviews of the streamflow data included in their investigation to assess potential changes that might be the result of anthropogenic influences. Based on those reviews, if the low flows at a streamgage were considered to be substantially affected by anthropogenic influences, the streamgage was excluded from this regionalization study.

Another resource that was used to help assess the streamgages for inclusion in the regionalization study was the Geospatial Attributes of Gages for Evaluating Streamflow, version II (GAGES-II) dataset (Falcone, 2011). The GAGES-II dataset provides geospatial data classifications for 9,322 streamgages maintained by the USGS and consists of streamgages that have 20 or more years of record since 1950 or are currently active. The GAGES-II dataset classifies streamgages considered “near natural” as “reference” gages. GAGES-II also has a disturbance index to help assess if the gage is a reference or non-reference streamgage. For many of the streamgages in GAGES-II, screening comments are included to provide some insight into the classification. The GAGES-II documentation notes that the classifications presented are not intended to be definitive and, as such, are considered another resource to assist in determining whether to include a streamgage in a regional regression analysis.

Additional information on potential influences on streamflow at the USGS streamgages was obtained from annual water data reports (USGS, 2019a). Station description information available from the internal USGS site information management system also was reviewed to help determine potential influences on streamflow.

For streams with multiple streamgages, an assessment was made to determine potential redundancy. Redundancy occurs when the drainage basins of two streamgages are nested, which is when one basin is contained inside another basin and most or all of the streamflow records at the two streamgages represent concurrent periods of time (Feaster and others, 2014). For this investigation, a streamgage was considered redundant if it was on the same stream as another streamgage and the drainage area of one streamgage represented more than 50 percent of the drainage area of the second streamgage, unless the periods of record for the two streamgages do not overlap. In cases where one record represented a relatively long period of time and the other record represented a relatively short period of time and was completely or mostly concurrent with the same period in the longer record, the streamgage with the short record was excluded. In addition, if a streamgage record was extended by Feaster and Lee (2017) based on the MOVE.1 record-extension method (Hirsch, 1982), the streamgage with the extended record was excluded (appendix 1). Most of the streamgages that were excluded because the low-flow statistics were based on the MOVE.1 record-extension method also were nested with the

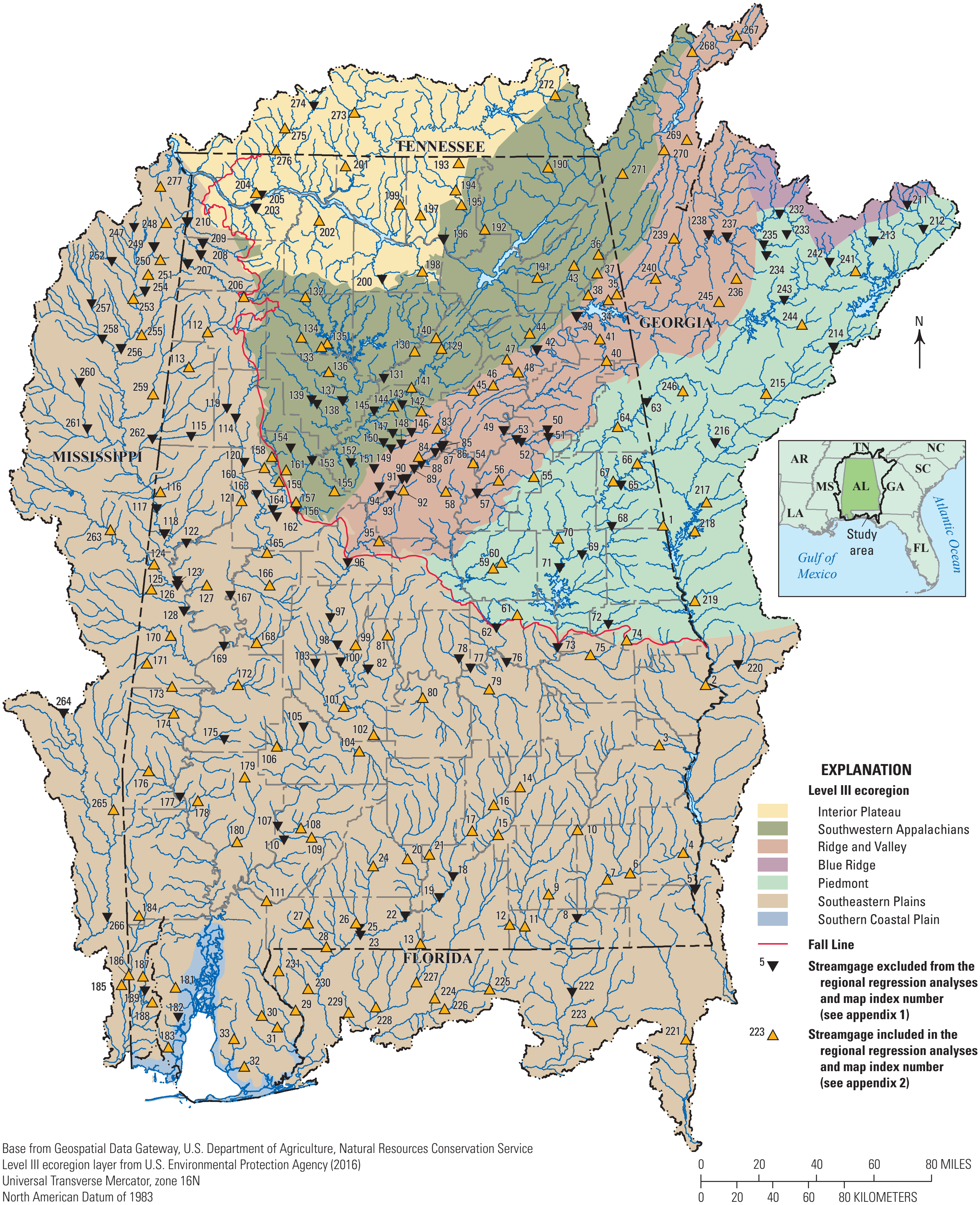


Figure 2. Level III ecoregions and streamgages considered for inclusion in the regional regression analyses for ungaged locations on streams in Alabama.

long-term streamgage used in the correlation procedure and, therefore, would have been considered redundant even if the correlation techniques had not been applied.

Along with using the aforementioned resources, final decisions on inclusion or exclusion of streamgages were made by using hydrologic judgment. Based on these criteria and reviews, 90 streamgages were excluded from the initial dataset (appendix 1). An additional 11 streamgages were excluded because all or substantial parts of the drainage basin are in the Blue Ridge ecoregion, which is not one of the ecoregions in Alabama. Also, USGS station 02341800 Upatoi Creek near Columbus, Ga., was excluded because it drains 100 percent from the Sand Hills ecoregion in Georgia, which is not an ecoregion in Alabama. During the final regression analysis, one additional station (USGS station 02454055 Lost Creek above Parrish, Ala.) was excluded because it was flagged in the regression diagnostics as having high influence and (or) leverage. Upon review, it was concluded that the low flows at USGS station 02454055 were likely being influenced enough by anthropogenic sources that it should be removed from the regression analysis. After excluding these 103 streamgages, 174 streamgages were available for inclusion in the regional regression analyses.

Physical and Climatic Basin Characteristics

Streamflow statistics can be estimated at ungaged sites by using multiple linear regression techniques that relate streamflow characteristics computed at gaged sites (such as the 7Q10 flow) to selected physical and (or) climatic basin characteristics computed for the gaged drainage basins (table 1). Determining the drainage-basin boundaries for each streamgage is the first step for generating other basin or climatic characteristics. For this investigation, basin boundaries were generated by using data from The National Map 3D Elevation Program (USGS, 2019b). Basin characteristics chosen for testing as potential explanatory variables were selected based on their potential theoretical relation to low flows, results of previous studies of low-flow regionalization, and the ability to quantify the basin characteristics by using geographic information systems (GIS). The use of GIS enables automation of the determination of basin characteristics and solution of the regional regression equations when using the USGS StreamStats application (USGS, 2019c).

Streamflow-Variability Index

The streamflow-variability index (SVI) was originally introduced by Lane and Lei (1950) as a useful index in analyzing hydraulic engineering projects and producing synthetic flow-duration curves. Areas with similar surface geology could be expected to correspond to similar SVI values, suggesting SVI would be a beneficial characteristic for regionalizing low flows based on geology (Ruhl and Martin, 1991). Given that low flows are a groundwater phenomenon, aquifer characteristics would be expected to influence low flows. Aquifer characteristics are diverse, and the interaction of aquifers and streamflow is complex; therefore, the flow in many streams is likely to be affected by several aquifers (Friel and others, 1989). Consequently, along with bringing a regional component related to geology, the SVI also incorporates the integrated effects of multiple aquifers on low flows within a given basin. As such, a generalized SVI has been successfully used in many low-flow regionalization studies (Friel and others, 1989; Martin and Ruhl, 1993; Koltun and Whitehead, 2002; Martin and Arihood, 2010; Koltun and Kula, 2013; Southard, 2013; and Eash and Barnes, 2017).

The SVI is a dimensionless hydrologic characteristic that provides a measure of the steepness of the slope of the flow-duration curve (Koltun and Whitehead, 2002; Southard, 2013). The flow-duration curve is a cumulative frequency curve that shows the percentage of time that a specific streamflow is equaled or exceeded (fig. 3). The SVI is the standard deviation of the logarithms of the 19 streamflow values at 5-percent class intervals from 5 to 95 percent on the flow-duration curve of daily mean flows for the analysis period (Searcy, 1959). The magnitude of the SVI is inversely related to the capacity of a basin to sustain base flow in a stream; for example, smaller SVI values are indicative of a higher sustained base flow. The SVI is computed as follows:

$$SVI\sqrt{\frac{\sum_{i=5,5}^{95}\left(\log_{10}\left(D_i\right)-\overline{\log_{10}(D)}\right)^2}{18}} \qquad (1)$$

where

SVI is the streamflow-variability index,

D_i is the ith percent duration streamflow (i=5, 10, 15, …95), and

$\overline{\log10\left(D\right)}$ is the mean of the base 10 logarithms of the 19 streamflow values at 5-percent class intervals from 5 to 95 percent on the flow-duration curve of the daily mean flow.

For two cases, the 95-percent flow value at a streamgage was zero. Because the logarithm of zero is undefined, those values were set equal to 0.01.

The SVI values computed at the streamgage locations and the ArcMap Natural Neighbor interpolation tool were used to create a grid for the State of Alabama, which can be used to estimate the SVI at an ungaged location (Esri, 2019; fig. 4). The natural neighbors of any point are those associated with neighboring Thiessen polygons. The natural neighbor method uses the closest subset of input points that surrounds a query point and weights them based on proportionate areas of the polygons.

Table 1. Basin characteristics tested as explanatory variables in the development of low-flow regression equations for ungaged locations in Alabama.

[USGS, U.S. Geological Survey; NLCD, National Land Cover Database; SSURGO, Soil Survey Geographic Database; NOAA, National Oceanic and Atmospheric Administration]

Basin characteristic	Unit of measure	Source
Drainage area	Square miles	National Map 3D Elevation Program (USGS, 2019b)
Mean elevation	Feet	National Map 3D Elevation Program (USGS, 2019b)
Minimum elevation	Feet	National Map 3D Elevation Program (USGS, 2019b)
Maximum elevation	Feet	National Map 3D Elevation Program (USGS, 2019b)
Relief (maximum elevation − minimum elevation)	Feet	National Map 3D Elevation Program (USGS, 2019b)
Relief ratio (mean elevation-minimum elevation)/(maximum elevation − minimum elevation)	Unitless	National Map 3D Elevation Program (USGS, 2019b)
Basin perimeter	Feet	National Map 3D Elevation Program (USGS, 2019b)
Percentage of basin in each level III ecoregion	Percent	U.S. Environmental Protection Agency (2016)
Percentage of developed land	Percent	Multi-Resolution Land Characteristics Consortium (version 2011) https://www.mrlc.gov/data
Percentage of impervious area	Percent	Multi-Resolution Land Characteristics Consortium (version 2011) https://www.mrlc.gov/data
Percentage of forest	Percent	Multi-Resolution Land Characteristics Consortium (version 2011) https://www.mrlc.gov/data
Percentage of planted/cultivated (agriculture)	Percent	Multi-Resolution Land Characteristics Consortium (version 2011) https://www.mrlc.gov/data
Percentage of woody wetlands (NLCD 2011)	Percent	Multi-Resolution Land Characteristics Consortium (version 2011) https://www.mrlc.gov/data
Percentage of herbaceous wetlands (NLCD 2011)	Percent	Multi-Resolution Land Characteristics Consortium (version 2011) https://www.mrlc.gov/data
Percentage of storage	Percent	Multi-Resolution Land Characteristics Consortium (version 2011) https://www.mrlc.gov/data
Base-flow index (Wolock, 2003)	Unitless	https://water.usgs.gov/GIS/metadata/usgswrd/XML/bfi48grd.xml
Groundwater head	Feet	National Map 3D Elevation Program (USGS, 2019b)
Saturated hydraulic conductivity (K-sat)	Inches per hour	SSURGO gridded data (U.S. Department of Agriculture, Natural Resources Conservation Service, 2017)
Percent clay, silt, sand	Percent	SSURGO gridded data (https://sdmdataaccess.sc.egov.usda.gov)
Soil drainage index	Unitless	SSURGO gridded data (https://sdmdataaccess.sc.egov.usda.gov)
Hydrologic soil index	Unitless	SSURGO gridded data (https://sdmdataaccess.sc.egov.usda.gov)
Mean annual precipitation	Inches	PRISM 2010 (https://prism.oregonstate.edu/normals/)

Methods for Estimating Low-Flow Frequency Statistics and Mean Annual Flow at Ungaged Locations in Alabama

Alabama currently (2020) has 129 streamgages that provide real-time daily streamflow data along with a number of inactive streamgages for which historical data are available through the USGS National Water Information System (NWIS) database (USGS, 2020). However, water-resource managers and engineers regularly need to determine selected streamflow statistics at ungaged locations. Two methods that can be used to generate such statistics are (1) a drainage-area ratio method on streams with a nearby streamgage, and (2) regional regression equations that relate streamflow statistics to selected basin characteristics.

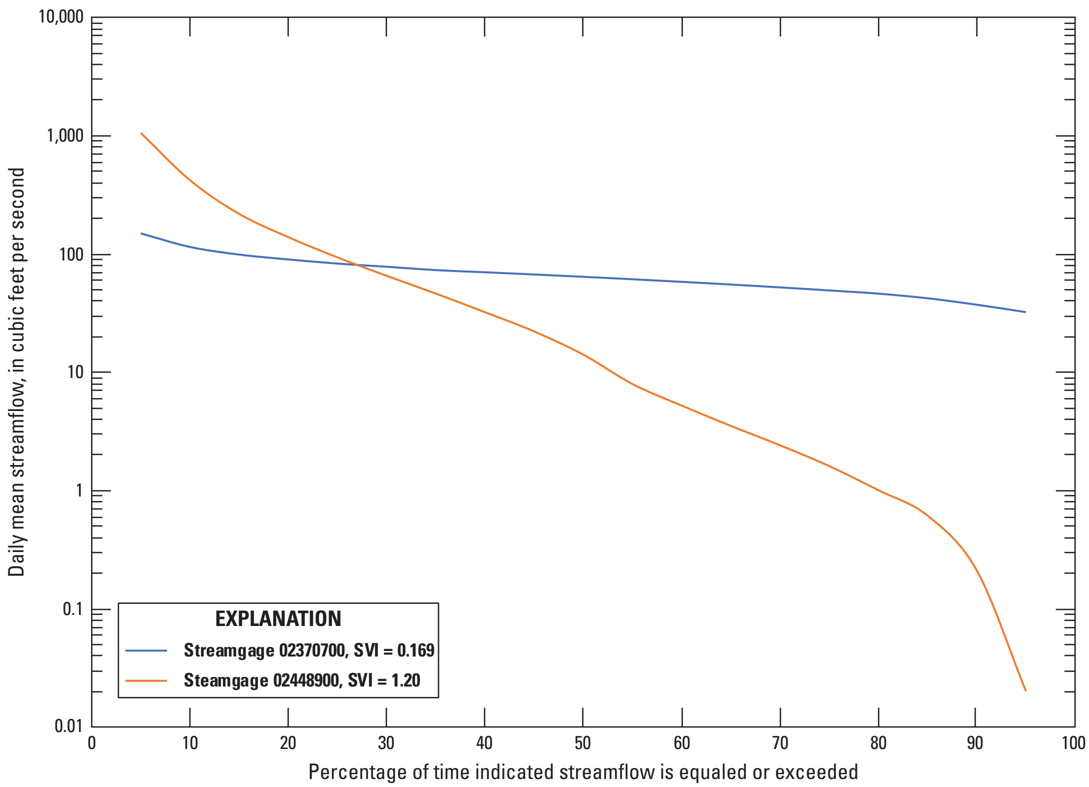


Figure 3. Flow-duration curves for U.S. Geological Survey stations 02370700 Pond Creek near Milton, Florida, and 02448900 Bodka Creek near Geiger, Alabama. [SVI, streamflow-variability index]

Drainage-Area Ratio Method

The drainage-area ratio method assumes that the streamflow characteristics at the ungaged site are the same on a per unit area basis as the nearby streamgage. Consequently, it is important to consider the proximity of the streamgage to the ungaged site as well as similarities in drainage area size, other physical and climatic characteristics, and the magnitude of anthropogenic influences (Ries and Friesz, 2000; Watson and others, 2005). As such, the drainage-area ratio method is most commonly applied, and will tend to be most accurate, when the streamgage is on the same stream as the ungaged site.

The drainage-area ratio equation is as follows:

$$Q_{DARu} = \left[\frac{DA_u}{DA_g}\right] Q_g \quad (2)$$

where

Q_{DARu} is the drainage-area ratio streamflow statistic at the ungaged site,

DA_u is the drainage area of the ungaged site,

DA_g is the drainage area of the gaged site, and

Q_g is the streamflow statistic from the gaged site.

With respect to a reasonable drainage-area ratio that is appropriate for applying this method, a few researchers have tested the procedure and recommended ranges of drainage-area ratios. Koltun and Schwartz (1986) recommended a range of 0.85 to 1.15 for the ungaged to gaged drainage areas for Ohio streams. For Pennsylvania streams, ratios between 0.33 and 3.0 were recommended (Ries and Friesz, 2000). Ries and Friesz (2000) found that for Massachusetts streams, the drainage-area ratio method estimates were generally as accurate or more accurate than regression estimates when the drainage-area ratios for the ungaged and gaged sites were between 0.3 and 1.5. In Idaho, Hortness (2006) found that the drainage-area ratios of 0.5 to 1.5 were reasonable, and in Iowa, Eash and Barnes (2017) recommended a range from 0.5 to 1.4. Southard (2013) found the drainage-area ratio range of 0.4 to 1.5 to be reasonable for Missouri streams.

Regional Regression Analysis

A regional regression analysis is an iterative process (fig. 5; Farmer and others, 2019). For this investigation, the data assembly step included updating low-flow frequency estimates at USGS streamgages (Feaster and Lee, 2017) and computing potential physical and climatic basin characteristics by using GIS tools (table 1). Model development involves exploratory data analysis to get a general understanding of relations among the variables that are to be predicted, which can be referred to as the dependent or response variable (for example, the 7Q10 streamflow statistic), and certain basin characteristics, which can be referred to as the independent or

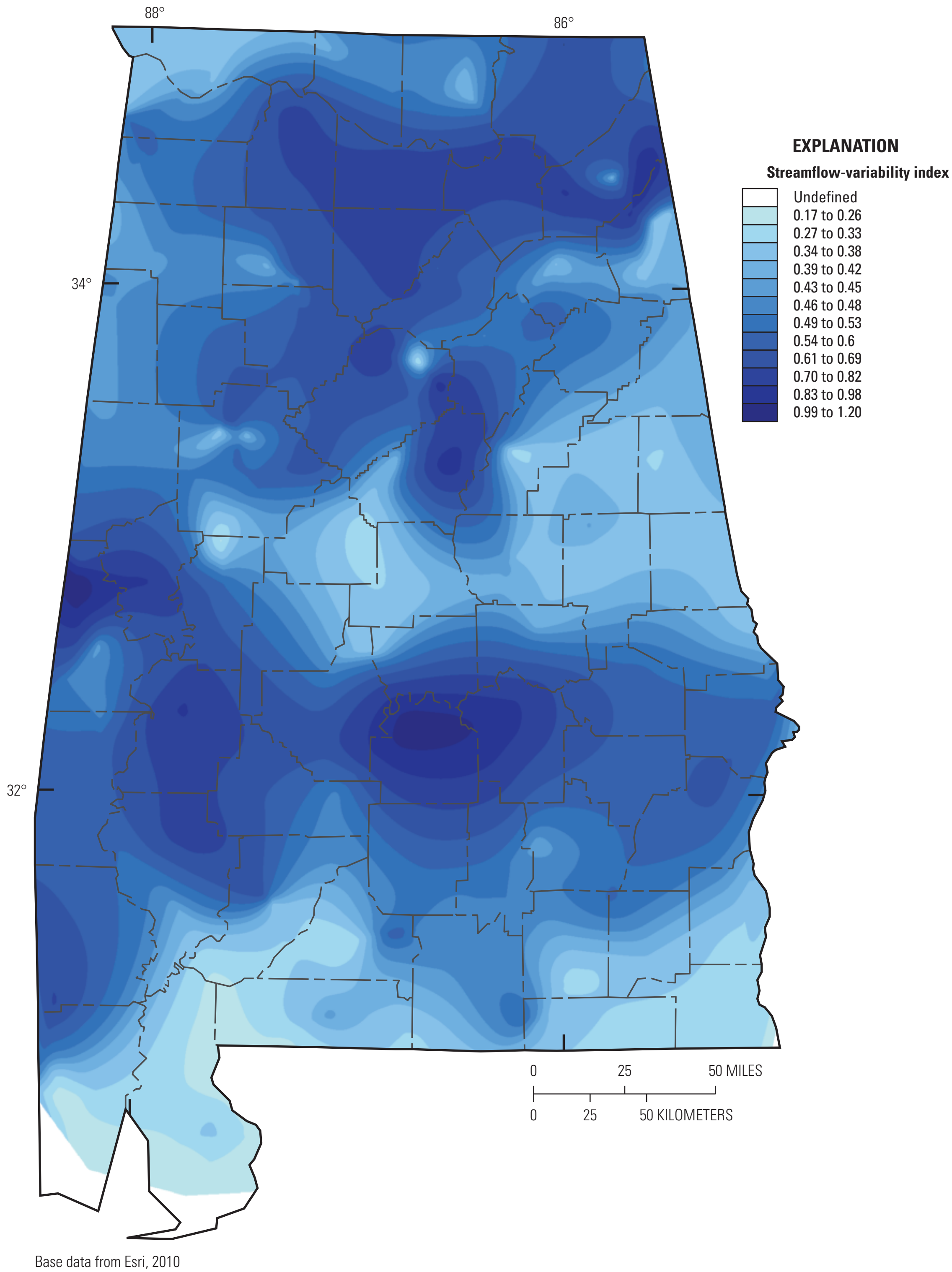


Figure 4. Streamflow-variability index grid for Alabama.

explanatory variables (for example, drainage area). Often the estimation step begins with assessing the complete study area by testing all potential independent variables with a dependent variable of choice. The results are then evaluated based on statistical significance of the independent variables, adherence to the assumptions of the regression techniques being applied, which are typically assessed based on the residuals computed from the difference in the observed and predicted dependent variable, and the geographical distribution of those residuals. Mapping the geographical distribution of the residuals provides information on potential subregions that might be warranted to improve the uncertainty in the predictions. Once a set of regression equations has been determined, the final step is to document the process with ample detail such that users of the equations have a general understanding of the proper application of the equations.

Development of Regional Regression Equations

For an exploratory regression analysis, it is best to start with the largest reasonable region, which is often the complete study area (Farmer and others, 2019). The initial regression analysis for this investigation included the complete study area, and from there, several potential regions were tested. The tested regions included dividing the study area into two regions that were differentiated by the Fall Line. The four flood regions used in the Alabama flood-frequency study by Hedgecock and Feaster (2007) also were tested along with various combinations of the EPA Level III ecoregions (fig. 1).

The preliminary regression analysis was done by using ordinary least squares (OLS) regression techniques (SAS Institute, 2020). The general model for an OLS regression analysis is of the form

$$Q_T = aA^bB^cC^d..., \quad (3)$$

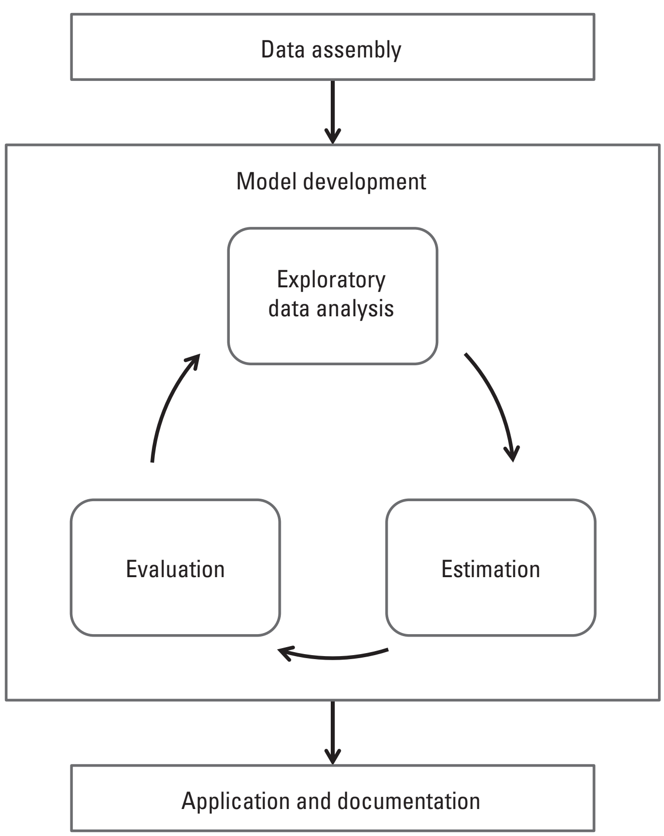


Figure 5. A generalized workflow for a regression-based regionalization study (from Farmer and others, 2019).

where

Q_T is a low-flow frequency statistic such as the 7Q10;

A, B, C are explanatory (independent) variables; and

a, b, c, and d are regression coefficients.

If the response and explanatory variables are logarithmically transformed, the regression model has the following line form:

$$\log Q_T = \log a + b(\log A) + c(\log B) + d(\log C) + ..., \quad (4)$$

where the variables are as previously defined in equation 3. The logarithmic and arithmetic relations were used in this investigation because the logarithmic transformation of some variables did not improve the linear relation with Q_T.

The OLS regression is an efficient way to explore possible linear relations between the explanatory (basin characteristics) and response (selected low-flow statistics) variables. An all-possible-subsets (APS) regression analysis was done including all streamgages and the potential basin characteristics (table 1). The initial OLS regression analysis was completed by using the SAS statistical software package 9.4 (SAS Institute, 2020) with the 174 streamgages available for inclusion in the regional regression analysis (appendix 2, and from Kolb and others, 2020). Various combinations of the potential explanatory variables selected from the APS analysis were tested. The response variables and some of the explanatory variables were transformed to logarithms (base 10) prior to the regression analysis. The purpose of the log transformations was to improve linear relations between the dependent and independent variables and to improve the variance of the residuals about the regression line. If the log transformation was shown not to substantially improve either of those characteristics, the independent variables were used in their measured units. Once a smaller subset of potential basin characteristics was determined, stepwise regression also was used to assess the strongest potential explanatory variables.

Along with the APS analysis, correlation matrix plots were used for a visual assessment of potential correlation among the independent variables as well as to assess which independent variables could likely be strong explanatory variables in the regression analysis (fig. 6). Tabular output from the matrix analysis included the Pearson correlation coefficient along with the p-value to assess statistical significance of the correlation. It should be noted that the matrix analysis and plots are useful for assessing the relations between single variables. In the case of a multiple linear regression where several independent variables are included, the importance of individual independent variables may change; therefore, the scatter plots should be considered only as a tool for initial assessments (Montgomery and others, 2012).

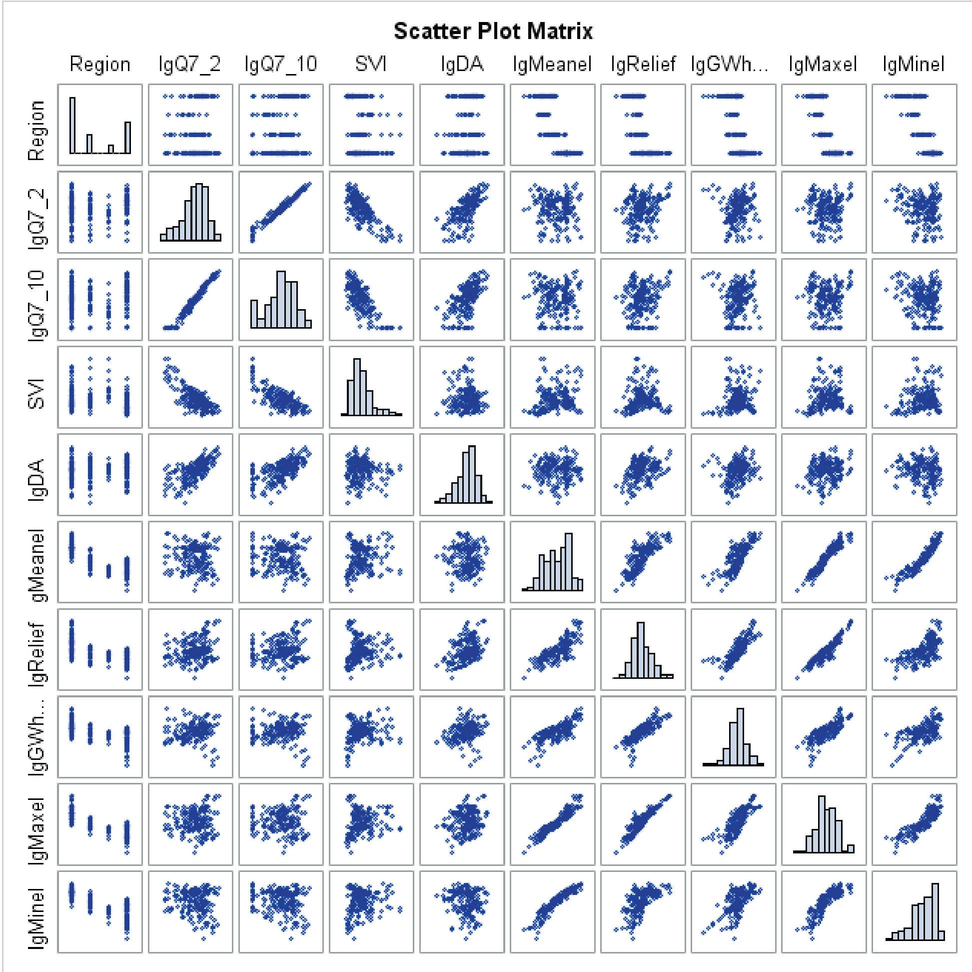


Figure 6. Example of a correlation matrix plot.

Final Regression Equations

At 9 of the 174 streamgages included in the regression analysis, 1 or more of the 3 at-site low-flow frequency statistics were equal to zero (appendix 2). As such, weighted left-censored regression techniques were used for the final regression analyses, with the weights being based on the number of years of record (Lorenz, 2014; Ziegeweid and others, 2015; Gotvald, 2017). For left-censored regression, a threshold value is imposed on the data that censors low (left) values below the threshold, with both the censored and uncensored values used in the regression. Censoring and coding data as "less than" a threshold value allows the use of a log transformation on the data and, therefore, allows all the data (uncensored and censored) to be used in the regression analysis to develop the regional equations (Watson and McHugh, 2014). For datasets that do not contain censored values, weighted left-censored regression provides the same results as weighted least squares regression (Helsel and Hirsch, 2002). Because of the uncertainty in measuring low flows and estimating low-flow frequency statistics less than 0.1 cubic foot per second (ft^3/s), the censoring threshold for the regression analyses in this report was set at 0.1 ft^3/s. Based on this censoring level, five streamgages had one or more of the three at-site low-flow frequency statistics that were less than 0.1 ft^3/s but greater than zero (appendix 2).

Final selection of the independent variables to include in the regression models was based on several factors such as statistical significance, standard error, coefficient of determination, and ease of measurement of the explanatory variables. Correlation among the explanatory variables (multicollinearity) was assessed by using the variance inflation factor (VIF). If one or more VIF values exceed 5, the regression coefficients are likely poorly estimated due to multicollinearity (Montgomery and others, 2012). In such a case, one of the explanatory variables should be removed and the regressions reanalyzed to assess the remaining variables.

For both the OLS and left-censored regression analyses, multiple regression diagnostics were generated and used to identify possible problems with the streamgage data or basin characteristics. Along with reviews to ensure that the regression residuals were randomly distributed around zero and that they were reasonably distributed geographically, other regression diagnostics were reviewed to determine streamgages that had high leverage and (or) high influence. The leverage metric is used to measure how unusual the values of independent variables at one streamgage are compared to the values of the same variables at all other streamgages. The influence metric indicates whether the data at a streamgage had a high influence on the estimated regression metric values (Eng and others, 2009). A streamgage may have a high leverage metric indicating that its independent variables are substantially different from those at all other streamgages, but the same streamgage may not have a high influence on the regression metrics. Conversely, a streamgage with a high influence may not have a high leverage metric. Sometimes, measurement or transposing errors in reported values of some independent variables can produce high leverage or influence metrics. Streamgages with high influence or leverage were given additional review to determine if such errors had been made or if the streamgage should be excluded for other reasons. For the final left-censored equations for the 1Q10, 7Q2, and 7Q10 low-flow frequency statistics, none of the streamgages indicated high leverage. Three streamgages indicated high influence: USGS station 02410000, 02422000, and 02448900 (appendix 2). Additional reviews of the streamgages did not indicate any issues that warranted removing them from the analysis. All three streamgages had at least one of the low-flow frequency statistics being regionalized that was less than 0.1 ft^3/s; therefore, the low-flow frequency statistics for those streamgages were censored in the analysis (appendix 2). Also, the periods of record analyzed for USGS stations 02410000 (1954 to 1987) and 02422000 (1941 to 1971) included data from 1954, which was the driest year of record for the period from 1895 to 2015 with respect to mean annual precipitation in Alabama (Feaster and Lee, 2017). In addition, the period of record analyzed for USGS station 02448900 was 1991 to 2014, for which 2007 and 2000 were the second and fourth driest years, respectively, in Alabama for the period from 1895 to 2015.

For the 1Q10, 7Q2, and 7Q10 low-flow frequency statistics, the combination of drainage area, SVI, mean annual precipitation, percentage of drainage basin in the Piedmont ecoregion, and percentage of drainage basin in the Southeastern Plains ecoregion were included in the final left-censored regression equations (table 2). All independent variables were statistically significant at the probability value (p-value) of ≤0.05, and the VIF for all independent variables was less than 1.60, indicating no issues of multicollinearity among the variables. The final regressions were done on a statewide basis with the mean annual precipitation, SVI, and percentage of drainage basin in the Piedmont and Southeastern Plains ecoregions as the independent variables that would incorporate regional differences in the low-flow characteristics. Drainage area was the only independent variable that was log transformed for the regression analysis, and for the transformation, the natural log was used.

For the mean annual flow, the combination of drainage area, mean annual precipitation, and percentage of the drainage basin in the Southeastern Plains ecoregion was included in the final regression equations. The mean annual flow data did not include any zero flows. All independent variables were statistically significant at a p-value of ≤0.05, and the VIF for all independent variables was less than 1.15. Drainage area was the only independent variable that was log transformed for the regression of the mean annual flow with the base 10 logarithm being used.

Figure 7 shows plots of the observed and predicted 1Q10, 7Q2, 7Q10, and mean annual flow, along with the line of equality and for the low-flow frequency results, the censoring level of 0.1 ft^3/s. Because the plots are shown using a log-log scale, zero flows are not shown for the low-flow frequency plots. The plots indicate a reasonable scatter distribution around the line of equality throughout the range of flows.

Accuracy and Limitations

Users of regional regression equations should be aware of the uncertainty in the results. One measure of uncertainty is the average standard error of estimate (SEE). In the case of the low-flow frequency equations, the average SEE is a measure of the average uncertainty in a prediction from the regression equations based on all streamgage data throughout the applicable region. However, users typically will be interested

Table 2. Selected low-flow frequency and mean annual flow regression equations for ungaged streams in Alabama.

[SEE, standard error of estimate; 1Q10, annual minimum 1-day flow with a 10-year recurrence interval, in cubic feet per second; DA, drainage area, in square miles; SVI, streamflow-variability index, dimensionless; Precip, mean annual precipitation, in inches; Pied, percentage of drainage basin in the Piedmont ecoregion; SEP, percentage of drainage basin in the Southeastern Plains ecoregion; 7Q2, annual minimum 7-day average flow with a 2-year recurrence interval, in cubic feet per second; 7Q10, annual minimum 7-day average flow with a 10-year recurrence interval, in cubic feet per second; Mean annual, mean annual flow, in cubic feet per second]

Statistic	Regression equation	SEE (percent)	Number of left-censored streamgages
1Q10	$0.478*DA^{1.149}*e^{-9.67SVI}*e^{0.0352Precip}*e^{-0.0128Pied}*e^{-0.00558SEP}$	38.8	14
7Q2	$0.722*DA^{1.066}*e^{-7.30SVI}*e^{0.0263Precip}*e^{-0.00356Pied}*e^{-0.00228SEP}$	18.7	2
7Q10	$0.572*DA^{1.126}*e^{-9.39SVI}*e^{0.0332Precip}*e^{-0.0115Pied}*e^{-0.00521SEP}$	35.4	13
Mean annual flow	$0.159*DA^{0.997}*10^{0.0184*Precip}*10^{-0.000677SEP}$	12.3	0

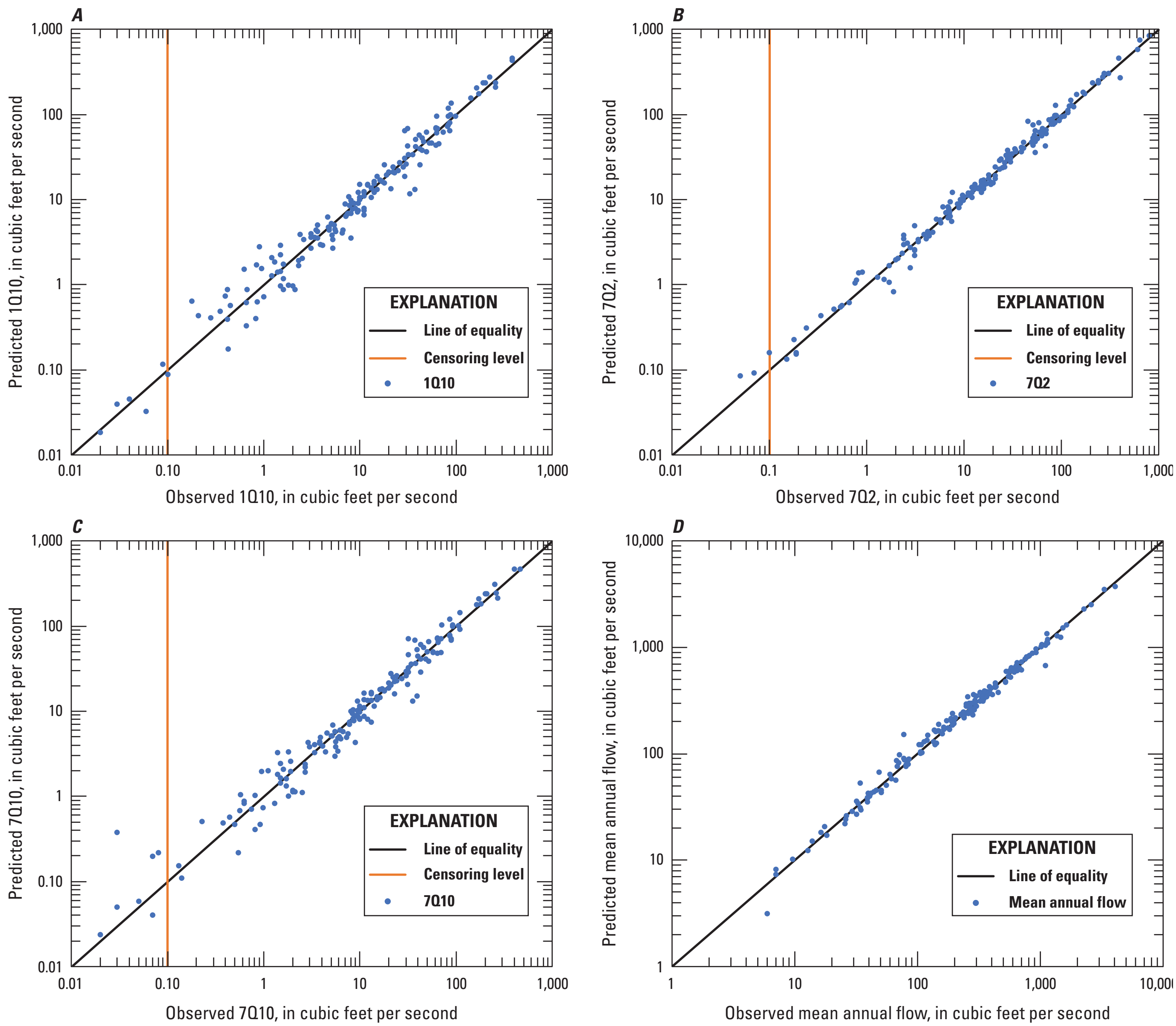


Figure 7. The relation between the observed and predicted *A*, annual minimum 1-day flow with a 10-year recurrence interval (1Q10); *B*, annual minimum 7-day flow with a 2-year recurrence interval (7Q2); *C*, annual minimum 7-day flow with a 10-year recurrence interval (7Q10); and *D*, mean annual flow.

in a measure of uncertainty of the low-flow estimate at a specific ungaged location for which the regional regression equation is being applied. One such measure of uncertainty is the confidence interval of the prediction, also known as the prediction interval. The prediction interval is a range in values of an estimated response variable, such as the 1Q10, 7Q2, and 7Q10 statistics, over which the true value of that response variable occurs within some stated probability. For example, the 90-percent prediction interval for an estimated flow value means that there is a 90-percent probability that the true value lies with that interval. The USGS StreamStats application (https://streamstats.usgs.gov/ss/; USGS, 2019c) uses the 90-percent prediction interval estimates as part of the computation of low-flow frequency estimates for ungaged stream locations. Tasker and Driver (1988) determined that a 100 $(1-\alpha)$ prediction interval for a flow statistic estimated at an ungaged location from a regression equation can be computed as follows:

$$Q/C < Q < CQ, \tag{5}$$

where

Q is the streamflow characteristic for the ungaged site; and

C is computed as:

$$C = 10^{t_{(\alpha/2, n-p)} SE_{p,i}} \quad (6)$$

where

$t_{(\alpha/2,\, n-p)}$ is the critical value from the Student's *t*-distribution at a particular alpha-level divided by 2 ($\alpha/2$) and degrees of freedom ($n-p$) and is equal to 1.68 for an α of 0.10, which corresponds to a prediction interval of 90 percent; and

$SE_{p,i}$ is the standard error of prediction for site *i* and is computed for a weighted left-censored regression as

$$SE_{p,i} = \left[MSE + X_i U X_i^T \right]^{0.5} \quad (7)$$

where

MSE is the mean square error;

X_i is a row vector of the explanatory variables for site *i*, augmented by a 1 as the first element;

U is the covariance matrix for the regression coefficients; and

X_i^T is the transpose of X_i (Ludwig and Tasker, 1993).

The values for *MSE* and *U* are presented in table 3. The prediction intervals are provided for users when the statistics are computed by using the USGS StreamStats application (USGS, 2019c).

Limitations for Applying the Regional Regression Equations

The Alabama low-flow regional regression equations should be applied considering the following limitations.

1. The equations are not applicable to tidally influenced streams.

2. The range of explanatory variables used to develop the Alabama low-flow frequency and mean annual flow regional regression equations are listed in table 4. Applying the equations at stream locations having explanatory variables outside the range of those used in this investigation may result in prediction errors that are considerably greater than those suggested by the SEEs listed in table 2.

3. The low-flow regression equations are applicable to streams with minimal anthropogenic influences and, therefore, should not be applied for streams that are known to be substantially affected by regulation, diversions, or urbanization.

4. Computation of basin characteristics needed to apply the regression equations at ungaged locations should be computed by using the same GIS datasets and computation methods as those used in this investigation (Kolb and others, 2020). The USGS StreamStats application includes the same GIS data layers and computation methods used in this investigation (Ries and others, 2008; USGS, 2019c; Kolb and others, 2020).

5. Because of the uncertainty in measuring and estimating streamflows less than 0.1 ft^3/s and the relatively small percentage of flows below that threshold limit at streamgages included in this investigation, a censoring threshold of 0.1 ft^3/s was used to develop left-censored regression equations for Alabama. Consequently, a regression analysis at an ungaged location that results in a predicted value of less than 0.1 ft^3/s should be reported as less than 0.1 ft^3/s.

6. Caution should be used when applying the regression equations in areas known to have karst topography. Low flows in karst topography can be substantially affected by gains from large springs and losses from sinkholes (Eash and Barnes, 2017).

Table 3. Values used to determine prediction intervals for the Alabama low-flow regression equations.

[MSE, the mean square error in equation 7; U, the covariance matrix used in equation 7; Intercept, y-axis intercept of regression equation; E, scientific notation indicating 10 to the power of; DA, drainage area, in square miles; SVI, streamflow-variability index, dimensionless; Precip, mean annual precipitation, in inches; Pied, percentage of drainage basin in the Piedmont ecoregion; SEP, percentage of drainage basin in the Southeastern Plains ecoregion; 1Q10, annual minimum 1-day flow with a 10-year recurrence interval, in cubic feet per second; 7Q2, annual minimum 7-day average flow with a 2-year recurrence interval, in cubic feet per second; 7Q10, annual minimum 7-day average flow with a 10-year recurrence interval, in cubic feet per second; Mean annual, mean annual flow, in cubic feet per second]

Statistic	MSE	U						
1Q10	**0.1405**		**Intercept**	**DA**	**SVI**	**Precip**	**Pied**	**SEP**
		Intercept	3.14E−01	−2.30E−03	−6.49E−02	−4.72E−03	−2.68E−04	−1.91E−05
		DA	−2.30E−03	5.93E−04	−1.14E−03	−3.27E−07	−4.42E−06	−3.46E−06
		SVI	−6.49E−02	−1.14E−03	4.90E−02	7.88E−04	1.14E−04	5.04E−05
		Precip	−4.72E−03	−3.27E−07	7.88E−04	7.63E−05	3.54E−06	−3.55E−07
		Pied	−2.68E−04	−4.42E−06	1.14E−04	3.54E−06	1.30E−06	3.95E−07
		SEP	−1.91E−05	−3.46E−06	5.04E−05	−3.55E−07	3.95E−07	5.37E−07
7Q2	**0.0343**		**Intercept**	**DA**	**SVI**	**Precip**	**Pied**	**SEP**
		Intercept	7.22E−02	−6.83E−04	−1.23E02	−1.09E−03	−5.99E−05	−3.20E−06
		DA	−6.83E−04	1.33E−04	−1.25E04	1.85E−06	−7.60E−07	−7.25E−07
		SVI	−1.22E−02	−1.25E−04	8.36E−03	1.46E−04	2.13E−05	8.91E−06
		Precip	−1.09E−03	1.85E−06	1.47E−04	1.77E−05	7.96E−07	−8.78E−08
		Pied	−5.99E−05	−7.60E−07	2.13E−05	7.96E−07	3.01E−07	8.78E−08
		SEP	−3.20E−06	−7.25E−07	8.91E−06	−8.78E−08	8.78E−08	1.22E−07
7Q10	**0.1183**		**Intercept**	**DA**	**SVI**	**Precip**	**Pied**	**SEP**
		Intercept	2.64E−01	−1.95E−03	−5.38E−02	−3.96E−03	−2.25E−04	−1.63E−05
		DA	−1.95E−03	4.88E−04	−8.98E−04	2.78E−07	−3.33E−06	−2.82E−06
		SVI	−5.38E−02	−8.98E−04	4.02E−02	6.51E−04	9.33E−05	4.16E−05
		Precip	−3.96E−03	2.78E−07	6.51E−04	6.40E−05	2.96E−06	−2.94E−07
		Pied	−2.25E−04	−3.33E−06	9.33E−05	2.96E−06	1.08E−06	3.29E−07
		SEP	−1.63E−05	−2.82E−06	4.16E−05	−2.94E−07	3.29E−07	4.49E−07
Mean annual	**0.0029**		**Intercept**	**DA**	**Precip**	**SEP**		
		Intercept	4.21E−03	−1.76E−04	−6.82E−05	1.43E−06		
		DA	−1.76E−04	5.72E−05	9.31E−07	−8.82E−08		
		Precip	−6.82E−05	9.31E−07	1.19E−06	−2.92E−08		
		SEP	1.43E−06	−8.82E−08	−2.92E−08	7.88E−09		

Table 4. Ranges of explanatory variables used to develop the low-flow frequency and mean annual flow regression equations for Alabama.

Basin characteristic	Minimum	Maximum
Drainage area (square miles)	2.01	2,469
Streamflow-variability index (dimensionless)	0.169	1.20
Mean annual precipitation (inches)	48.71	67.45
Percentage of drainage basin from the Piedmont ecoregion	0	100
Percentage of drainage basin from the Southeastern Plains ecoregion	0	100

StreamStats

USGS StreamStats is a web-based GIS application that provides a range of analytical tools useful for water-resource managers, planners, and engineers (Ries and others, 2008). The StreamStats application can be used to delineate drainage areas for user-selected sites on streams and then generate basin characteristics and estimates of streamflow statistics for the selected site where the functionality is available. StreamStats users also can select USGS data-collection streamgages and, where available, get streamflow statistics and other information about the streamgage. StreamStats is currently (2020) available for Alabama and can be used to estimate rural and urban flood frequency (Hedgecock, 2004; Hedgecock and Feaster, 2007; Hedgecock and Lee, 2010; USGS, 2019c).

The regression equations developed in this investigation to estimate the 1Q10, 7Q2, and 7Q10 low-flow frequency statistics and the mean annual flow have been incorporated into the USGS Alabama StreamStats application. Complete instructions for using StreamStats are provided on the StreamStats website (USGS, 2019c). The website provides links to (1) information about general limitations of the application, (2) other State applications, (3) user instructions, (4) definitions of terms, (5) answers to frequently asked questions, (6) downloadable presentations and other technical information about the application, and (7) contact information.

Summary

Methods for estimating selected low-flow frequency statistics and mean annual flow at ungaged locations on Alabama streams were developed as a part of this investigation. Because of the uncertainty in measuring and estimating streamflows less than 0.1 cubic foot per second (ft^3/s) and the relatively small percentage of flows below that threshold limit at streamgages included in this investigation, a censoring threshold of 0.1 ft^3/s was used to develop left-censored regression equations for Alabama. The at-site low-flow frequency statistics for the 174 streamgages included in the regression analyses were previously published as part of a U.S. Geological Survey investigation updating low-flow frequency statistics by using streamflow data through March 2014. The majority of the streamgages were from Alabama, but streamgages from Florida, Georgia, Mississippi, and Tennessee that have shared drainage basins with Alabama were also included. The streamflow data used in the mean annual flow regression were computed for the same 174 streamgages by using data through September 2018.

Numerous basin characteristics were tested as explanatory variables for the regression equations. The final regression equations for the 1Q10, 7Q2, and 7Q10 low-flow frequency statistics included drainage area, streamflow-variability index, mean annual precipitation, and percentage of the drainage basin located in the Piedmont and Southeastern Plains ecoregions. The average standard errors of estimate for the 1Q10, 7Q2, and 7Q10 low-flow frequency equations were 38.8, 18.7, and 35.4 percent, respectively. The final regression equation for the mean annual flow included drainage area, mean annual precipitation, and percentage of the drainage basin located in the Southeastern Plains ecoregion as explanatory variables. The average standard error of estimate for the mean annual flow was 12.3 percent.

For the low-flow frequency regressions, the analysis was done by transforming the low-flow frequency statistics and drainage area by using natural logarithms. The remaining explanatory variables were used in their measured units. Mean annual flows computed at the streamgages and drainage area were transformed by using base 10 logarithms. The remaining explanatory variables were used in their measured units.

The streamflow data included in the analyses represent streamgages with minimal anthropogenic influences. Consequently, the regression equations should be applied at locations that are not substantially affected by regulation, diversions, or other anthropogenic influences. The equations also should not be applied on streams that are known to be tidally influenced. The equations should not be applied outside of the range of the basin characteristics used in the regression analysis because doing so can result in prediction errors that are considerably greater than those suggested by the standard error of estimates. The ranges of the basin characteristics used in the regression analysis are as follows: drainage area, 2.01 to 2,469 square miles; streamflow-variability index, 0.169 to 1.20; and mean annual precipitation, 48.71 to 67.45 inches. The ranges of percentage of drainage area in the Piedmont and Southeastern Plains ecoregions are 0 to 100 percent.

The low-flow frequency and mean annual flow regression equations developed in this investigation have been incorporated into the USGS StreamStats application for Alabama. The StreamStats application delineates the drainage basin at a selected location on a stream and then generates the explanatory variables for the delineated basin needed by the regression equations. Along with the streamflow estimates computed by StreamStats, the output from the StreamStats application also will provide 90-percent prediction intervals for those estimates.

References Cited

Alabama Department of Environmental Management, 2012, Water Quality Program, Volume 1, Division 335-6 [revised February 3, 2017]: Alabama Department of Environmental Management, Water Division, accessed March 20, 2013, at http://www.adem.alabama.gov/alEnviroRegLaws/files/Division6Vol1.pdf.

Bingham, R.H., 1982, Low-flow characteristics of Alabama streams: U.S. Geological Survey Water-Supply Paper 2083, 27 p. [Also available at https://doi.org/10.3133/wsp2083.]

Eash, D.A., and Barnes, K.K., 2017, Methods for estimating selected low-flow frequency statistics and harmonic mean flows for streams in Iowa (ver. 1.1, November 2017): U.S. Geological Survey Scientific Investigations Report 2012–5171, 99 p. [Also available at https://doi.org/10.3133/sir20125171.]

Eng, Ken, Chen, Y.-Y., 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 https://pubs.usgs.gov/tm/tm4a8/.]

Esri, 2019, How Natural Neighbor works, accessed May 28, 2019, at https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/how-natural-neighbor-works.htm.

Falcone, J.A., 2011, GAGES-II—Geospatial attributes of gages for evaluating streamflow: U.S. Geological Survey [digital spatial dataset], accessed July 31, 2017, at https://doi.org/10.3133/70046617.

Farmer, W.H., Kiang, J.E., Feaster, T.D., and Eng, K., 2019, Regionalization of surface-water statistics using multiple linear regression: U.S. Geological Survey Techniques and Methods, book 4, chap. A12, 40 p. [Also available at https://doi.org/10.3133/tm4A12.]

Feaster, T.D., Gotvald, A.J., and Weaver, J.C., 2014, Methods for estimating the magnitude and frequency of floods for urban and small, rural streams in Georgia, South Carolina, and North Carolina, 2011 (ver. 1.1, March 2014): U.S. Geological Survey Scientific Investigations Report 2014–5030, 104 p. [Also available at https://dx.doi.org/10.3133/sir20145030.]

Feaster, T.D., and Lee, K.G., 2017, Low-flow frequency and flow-duration characteristics of selected streams in Alabama through March 2014: U.S. Geological Survey Scientific Investigations Report 2017–5083, 371 p. [Also available at https://doi.org/10.3133/sir20175083.]

Friel, E.A., Embree, W.N., Jack, A.R., and Atkins, J.T., Jr., 1989, Low-flow characteristics of streams in West Virginia: U.S. Geological Survey Water-Resources Investigations Report 88–4072, 38 p. [Also available at https://doi.org/10.3133/wri884072.]

Gotvald, A.J., 2017, Methods for estimating selected low-flow frequency statistics and mean annual flow for ungaged locations on streams in North Georgia: U.S. Geological Survey Scientific Investigations Report 2017–5001, 25 p. [Also available at https://doi.org/10.3133/sir20175001.]

Hedgecock, T.S., 2004, Magnitude and frequency of floods on small rural streams in Alabama: U.S. Geological Survey Scientific Investigations Report 2004–5135, 10 p. [Also available at https://doi.org/10.3133/sir20045135.]

Hedgecock, T.S., and Feaster, T.D., 2007, Magnitude and frequency of floods in Alabama, 2003: U.S. Geological Survey Scientific Investigations Report 2007–5204, 28 p., + app. [Also available at https://pubs.water.usgs.gov/sir2007-5204.]

Hedgecock, T.S., and Lee, K.G., 2010, Magnitude and frequency of floods for urban streams in Alabama, 2007: U.S. Geological Survey Scientific Investigations Report 2010–5012, 17 p. [Also available at https://doi.org/10.3133/sir20105012.]

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, 522 p. [Also available at https://doi.org/10.3133/twri04A3.]

Hirsch, R.M., 1982, A comparison of four streamflow record extension techniques: Water Resources Research, v. 18, no. 4, p. 1081–1088. [Also available at https://doi.org/10.1029/WR018i004p01081.]

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 https://doi.org/10.3133/sir20065035.]

Kolb, K.R., Clark, J.M., Feaster, T.D., and Painter, J.A., 2020, Supporting data for estimating selected low-flow frequency statistics and mean annual flow for ungaged locations on streams in Alabama: U.S. Geological Survey data release, https://doi.org/10.5066/P994UFS7.

Koltun, G.F., and Kula, S.P., 2013, Methods for estimating selected low-flow statistics and development of annual flow-duration statistics for Ohio: U.S. Geological Survey Scientific Investigations Report 2012–5138, 195 p. [Also available at https://doi.org/10.3133/sir20125138.]

Koltun, G.F., and Schwartz, R.R., 1986, Multiple-regression equations for estimating low flows at ungaged stream sites in Ohio: U.S. Geological Survey Water-Resources Investigations Report 86–4354, 39 p., 6 pls. [Also available at https://doi.org/10.3133/wri864354.]

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, 57 p. [Also available at https://doi.org/10.3133/wri024068.]

Lane, E.W., and Lei, K., 1950, Stream flow variability—Proceedings of American Society of Civil Engineers: Transactions, v. 115, p. 1084–1134.

Lorenz, D.L., 2014, smwrQW—R functions to support water-quality data analysis for statistical methods in water resources: U.S. Geological Survey R Archive Network, accessed April 25, 2019, at https://github.com/USGS-R/smwrQW.

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 https://doi.org/10.3133/wri934013.]

Martin, G.R., and Arihood, L.D., 2010, Methods for estimating selected low-flow frequency statistics for unregulated streams in Kentucky: U.S. Geological Survey Scientific Investigations Report 2010–5217, 83 p. [Also available at https://doi.org/10.3133/sir20105217.]

Martin, G.R., and Ruhl, K.J., 1993, Regionalization of harmonic-mean streamflows in Kentucky: U.S. Geological Survey Water-Resources Investigations Report 92–4173, 47 p., 1 pl. [Also available at https://doi.org/10.3133/wri924173.]

Montgomery, D.C., Peck, E.A., and Vining, G.G., 2012, Introduction to linear regression analysis (5th ed.): Hoboken, New Jersey, John Wiley & Sons, Inc., 645 p.

Omernik, J.M., 1987, Ecoregions of the conterminous United States: Annals of the Association of American Geographers, v. 77, no. 1, p. 118–125, scale 1:7,500,000.

Ries, K.G., III, 1994, Estimation of low-flow duration discharges in Massachusetts: U.S. Geological Survey Water-Supply Paper 2418, 50 p. [Also available at https://doi.org/10.3133/wsp2418.]

Ries, K.G., III, and Friesz, P.J., 2000, Methods for estimating low-flow statistics for Massachusetts streams: U.S. Geological Survey Water-Resources Investigations Report 00–4135, 81 p. [Also available at https://pubs.usgs.gov/wri/wri004135/.]

Ries, K.G., III, Guthrie, J.G., 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 https://doi.org/10.3133/fs20083067.]

Riggs, H.C., 1972, Low-flow investigations: U.S. Geological Survey Techniques of Water Resources Investigations, book 4, chap. B1, 18 p.

Ruhl, K.J., and Martin, G.R., 1991, Low-flow characteristics of Kentucky streams: U.S. Geological Survey Water-Resources Investigations Report 91–4097, 50 p. [Also available at https://pubs.usgs.gov/wri/wrir_91-4097/.]

SAS Institute, 2020, SAS 9.4 software website, accessed April 16, 2020, at https://www.sas.com/en_us/software/sas9.html.

Searcy, J.K., 1959, Flow-duration curves, manual of hydrology—Part 2. Low-flow techniques: U.S. Geological Survey Water-Supply Paper 1542–A, 33 p.

Southard, R.E., 2013, Computed statistics at streamgages, and methods for estimating low-flow frequency statistics and development of regional regression equations for estimating low-flow frequency statistics at ungaged location in Missouri: U.S. Geological Survey Scientific Investigations Report 2013–5090, 28 p. [Also available at https://doi.org/10.3133/sir20135090.]

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. [Also available at https://doi.org/10.1111/j.1752-1688.1988.tb03026.x.]

U.S. Department of Agriculture, Natural Resources Conservation Service, 2017, Soil data access website, accessed on October 1, 2017, at https://sdmdataaccess.sc.egov.usda.gov/.

U.S. Environmental Protection Agency [EPA], 2016, Ecoregions of the United States: U.S. Environmental Protection Agency website, accessed June 28, 2016, at https://www.epa.gov/eco-research/ecoregions.

U.S. Geological Survey, [USGS], 2016, How much of your state is wet?: U.S. Geological Survey Water Science School website, accessed June 28, 2016, at https://water.usgs.gov/edu/wetstates.html.

U.S. Geological Survey, [USGS], 2019a, Annual water data reports, accessed November 13, 2019, at https://wdr.water.usgs.gov/.

U.S. Geological Survey, [USGS], 2019b, 3D Elevation Program, accessed November 14, 2019, at https://www.usgs.gov/core-science-systems/ngp/3dep.

U.S. Geological Survey, [USGS], 2019c, StreamStats—Streamflow statistics and spatial analysis tools for water-resources applications, accessed November 14, 2019, at https://www.usgs.gov/mission-areas/water-resources/science/streamstats-streamflow-statistics-and-spatial-analysis-tools?qt-science_center_objects=0#qt-science_center_objects.

U.S. Geological Survey [USGS], 2020, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed April 2020 at https://dx.doi.org/10.5066/F7P55KJN.

Watson, K.M., and McHugh, A.R., 2014, Regional regression equations for the estimation of selected monthly low-flow duration and frequency statistics at ungaged sites on streams in New Jersey: U.S. Geological Survey Scientific Investigations Report 2014–5004, 59 p., accessed October 1, 2017, at https://dx.doi.org/10.3133/sir20145004.

Watson, K.M., Reiser, R.G., Nieswand, S.P., and Schopp, R.D., 2005, Streamflow characteristics and trends in New Jersey, water years 1897–2003: U.S. Geological Survey Scientific Investigations Report 2005–5105, 131 p. [Also available at https://doi.org/10.3133/sir20055105.]

Wolock, D.M., 2003, Base-flow index grid for the conterminous United States, accessed October 1, 2017, at https://water.usgs.gov/GIS/metadata/usgswrd/XML/bfi48grd.xml.

Ziegeweid, J.R., Lorenz, D.L., Sanocki, C.A., and Czuba, C.R., 2015, Methods for estimating flow-duration curve and low-flow frequency statistics for ungaged locations on small streams in Minnesota: U.S. Geological Survey Scientific Investigations Report 2015–5170, 23 p. [Also available at https://doi.org/10.3133/sir20155170.]

Appendix 1. U.S. Geological Survey streamgages that were excluded from the regional regression analysis for estimating selected low-flow frequency statistics and mean annual flow for ungaged locations in Alabama. Available online at https://doi.org/10.3133/sir20205099

Appendix 2. U.S. Geological Survey streamgages and independent and dependent variables used in the low-flow frequency and mean annual flow regression analyses for ungaged locations on streams in Alabama. Available online at https://doi.org/10.3133/sir20205099

For more information about this publication, contact

Director, Lower Mississippi-Gulf Water Science Center
U.S. Geological Survey
640 Grassmere Park, Suite 100
Nashville, TN 37211

For additional information visit
https://www.usgs.gov/centers/lmg-water/

Publishing support provided by
Lafayette Publishing Service Center

USGS

Feaster and others—**Methods for Estimating Low-Flow Frequency Statistics and Mean Annual Flow for Ungaged Locations, Alabama**—SIR 2020–5099, ver. 1.1

ISBN 978-1-4113-4378-8

ISSN 2328-031X (print)
ISSN 2328-0328 (online)
https://doi.org/10.3133/sir20205099