USGS
science for a changing world

Prepared in cooperation with the Puerto Rico Environmental Quality Board

Estimating Flow-Duration Statistics and Low-Flow Frequencies for Selected Streams and the Implementation of a StreamStats Web-Based Tool in Puerto Rico

Scientific Investigations Report 2021–5054

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

Cover. Upstream view of hanging bridge 300 yards downstream from U.S. Geological Survey streamgage 50113800 (Rio Cerrillos above Lago Cerrillos near Ponce, Puerto Rico). Photograph taken by David M. Hernandez, Hydrologic Data Chief, Puerto Rico.

Estimating Flow-Duration Statistics and Low-Flow Frequencies for Selected Streams and the Implementation of a StreamStats Web-Based Tool in Puerto Rico

By Tara Williams-Sether

Prepared in cooperation with the Puerto Rico Environmental Quality Board

Scientific Investigations Report 2021–5054

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

U.S. Geological Survey, Reston, Virginia: 2021

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 https://www.usgs.gov or call 1–888–ASK–USGS.

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

Suggested citation:
Williams-Sether, T., 2021, Estimating flow-duration statistics and low-flow frequencies for selected streams and the implementation of a StreamStats web-based tool in Puerto Rico: U.S. Geological Survey Scientific Investigations Report 2021–5054, 18 p., https://doi.org/10.3133/sir20215054.

Associated data for this publication:
U.S. Geological Survey, 2019, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, https://doi.org/10.5066/F7P55KJN.

Williams-Sether, T., 2021, Data files for the development of regression equations for flow-duration statistics and n-day low-flow frequencies for ungaged streams in Puerto Rico through water year 2018: U.S. Geological Survey data release, https://doi.org/10.5066/P9Y2QVJ6.

ISSN 2328-0328 (online)

Contents

Figures

Tables

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
Length		
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		
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)
inch per year (in/yr)	25.4	millimeter per year (mm/yr)

Temperature in degrees Fahrenheit (°F) may be converted to degrees Celsius (°C) as follows:
°C = (°F – 32) / 1.8.

Datum

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

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

Altitude, as used in this report, refers to distance above the vertical datum.

Supplemental Information

A water year is the 12-month period from October 1 through September 30 and is designated by the calendar year in which it ends.

Abbreviations

Adj-R^2	adjusted coefficient of determination
ALOS/PALSAR	advanced land observing satellite/phased array type L-band synthetic aperture radar
ASEP	average standard error of prediction
BIC	Bayesian information criterion
D50	flow duration with 50-percent exceedance probability
D60	flow duration with 60-percent exceedance probability
D70	flow duration with 70-percent exceedance probability
D80	flow duration with 80-percent exceedance probability
D90	flow duration with 90-percent exceedance probability
D95	flow duration with 95-percent exceedance probability
D98	flow duration with 98-percent exceedance probability
D99	flow duration with 99-percent exceedance probability
DEM	digital elevation model
GIS	geographic information system
GLS	generalized-least-squares
M1D2Y	1-day mean low flow that occurs on average once in 2 years
M1D5Y	1-day mean low flow that occurs on average once in 5 years
M1D10Y	1-day mean low flow that occurs on average once in 10 years
M7D2Y	7-day mean low flow that occurs on average once in 2 years
M7D5Y	7-day mean low flow that occurs on average once in 5 years
M7D10Y	7-day mean low flow that occurs on average once in 10 years
M14D2Y	14-day mean low flow that occurs on average once in 2 years
M14D5Y	14-day mean low flow that occurs on average once in 5 years
M14D10Y	14-day mean low flow that occurs on average once in 10 years
M30D2Y	30-day mean low flow that occurs on average once in 2 years
M30D5Y	30-day mean low flow that occurs on average once in 5 years
M30D10Y	30-day mean low flow that occurs on average once in 10 years
NRCS	Natural Resources Conservation Service
NWIS	National Water Information System
OLS	ordinary-least-squares
PREQB	Puerto Rico Environmental Quality Board
Pseudo R^2	pseudo coefficient of determination
RSE	residual standard error
SMEV	standard model error of variance
USFS	U.S. Forest Service
USGS	U.S. Geological Survey
WLS	weighted-least-squares

Estimating Flow-Duration Statistics and Low-Flow Frequencies for Selected Streams and the Implementation of a StreamStats Web-Based Tool in Puerto Rico

By Tara Williams-Sether

Abstract

Daily mean streamflow data from 28 U.S. Geological Survey streamflow-gaging stations in Puerto Rico with 10 or more years of unregulated or minimally affected flow record through water year 2018 were used to develop regression equations for flow duration and annual *n*-day low-flow statistics. Ordinary least-squares and generalized least-squares regression techniques were used to develop regional regression equations for flow-duration statistics at the 99th, 98th, 95th, 90th, 80th, 70th, 60th, and 50th percent exceedance probabilities and annual *n*-day low-flow frequency statistics for the 1-, 7-, 14-, and 30-day mean low flows with the 2-year (0.5 nonexceedance probability), 5-year (0.2 nonexceedance probability), and 10-year (0.1 nonexceedance probability) recurrence intervals. A StreamStats web application was developed to estimate basin and climatic characteristics for the regional regression equation analysis. Basin and climatic characteristics determined to be significant explanatory variables in one or more regression equations included drainage area, mean total annual reference evapotranspiration, and minimum basin elevation. The adjusted coefficient of determination for the flow-duration regression equations ranged from 57.7 to 81.4 percent. The pseudo coefficient of determination for the annual *n*-day low-flow regression equations ranged from 64.6 to 70.7 percent. The StreamStats web application incorporates the flow duration, and annual *n*-day low-flow regression equations and can provide streamflow estimates for most ungaged sites in the island.

Introduction

Streamflow statistics, such as flow duration and low-flow frequency, are used to characterize flow of a certain magnitude at a location of interest on a stream that is important to Federal, State, Tribal, and local agencies. These flow statistics are commonly used in water-supply planning and design, for estimates of point-source waste loading into rivers, for determining discharge thresholds for aquatic life, for reservoir inflows, and for water-quality analyses.

An analysis of flow duration and low-flow statistics at streamflow-gaging stations in Puerto Rico is necessary because of the increasing demand on surface-water resources. As water demands from streams increase, knowledge of streamflow characteristics becomes increasingly important. Streamflow statistics can be used as an index for water management regulations, to assess the water-supply potential, to adequately evaluate the capacity of the stream to receive waste loads, and to preserve aquatic and wildlife habitats. Although streamflow statistics can be calculated at streamflow-gaging stations, predictive statistical techniques, such as regression equations, can be used to make estimates of streamflow statistics at locations where streamflow-gaging stations do not exist. If an ungaged stream location has 0.5 to 1.5 times the drainage area of a streamflow-gaging station on the same stream, then streamflow information usually can be extrapolated from the streamflow-gaging station record using a drainage-area ratio method (Sauer, 1974; Perry and others, 2004). For ungaged stream locations with greater than 1.5 times the drainage area of a streamflow-gaging station, or where no streamflow-gaging station data are available, regression equations that relate streamflow statistics with physical and climatic characteristics of drainage basins can be used (Perry and others, 2004).

The U.S. Geological Survey (USGS), in cooperation with the Puerto Rico Environmental Quality Board (PREQB), developed regression equations for estimating flow duration and *n*-day low-flow frequency statistics for ungaged sites in Puerto Rico. In addition, the regression equations developed from this study are included in the developed Puerto Rico StreamStats web-based tool that can be accessed through the USGS StreamStats interface (U.S. Geological Survey, 2021). StreamStats allows users to obtain flow statistics, basin characteristics, and other information for user-selected locations on a stream (Ries and others, 2017). Using a geographic information system (GIS)-based interactive map of Puerto Rico, the user can "point and click" at a location on a stream, and StreamStats will rapidly delineate the basin upstream from the selected location, calculate basin characteristics, and provide estimated streamflow statistics computed from available regression equations. The user also can "point and click" on identified USGS streamflow-gaging stations and receive additional flow statistics and information about those stations through provided links to the National Water Information System (NWIS) and the StreamStats Data-Collection Station Report web page. NWIS provides access to data collected at streamflow-gaging stations, and statistics computed and published on the data collected are reported in StreamStats Data-Collection Station Report web pages.

Purpose and Scope

The purpose of this report is to present methods for estimating flow-duration statistics for the 99th, 98th, 95th, 90th, 80th, 70th, 60th, and 50th percent exceedance probabilities (referred to as D99, D98, D95, D90, D80, D70, D60, and D50, respectively), and for annual *n*-day low-flow frequency statistics for the 1-, 7-, 14-, and 30-day mean low flows with the 2-, 5-, and 10-year recurrence intervals (referred to as M1D2Y, M1D5Y, M1D10Y, M7D2Y, M7D5Y, M7D10Y, M14D2Y, M14D5Y, M14D10Y, M30D2Y, M30D5Y, and M30D10Y, respectively) for ungaged stream locations in the main island of Puerto Rico. Recurrence intervals for low flows of 2, 5, and 10 years are alternately referred to as the 0.5, 0.2, and 0.1 nonexceedance probabilities, respectively. A flow duration represents the percentage of time that a given flow, measured during a specific time interval, has been equaled or exceeded during that time interval; for example, a 90-percent flow exceedance, or D90, represents a streamflow that is equaled or exceeded 90 percent of the time. An *n*-day low-flow frequency represents the probability that flow will not exceed different flow levels in any given year. For example, with *n* equal to 7, the minimum 7-day mean low flow with the 10-year recurrence interval (M7D10Y) is the minimum mean flow for 7 consecutive days, which has a 0.1 probability (10-percent chance) of not being exceeded in a given year. The probability of recurrence of low flows will be referred to as nonexceedance probability and is computed as the reciprocal of the recurrence interval. The nonexceedance probabilities of the 2-, 5-, and 10-year recurrence intervals are 0.5, 0.2, and 0.1, respectively.

The report describes (1) the statistical methodology, sources of basin and climatic characteristics, and methods used to develop regression equations that relate selected basin characteristics to flow duration and annual *n*-day low-flow statistics; and (2) the StreamStats web application for automatically measuring the required basin characteristics data and solving the regression equations to estimate flow duration and annual *n*-day low-flow frequency statistics at ungaged sites. Selected streamflow-gaging stations used in this report were mostly unaffected by regulation and had a minimum of 10 years of continuous daily mean streamflow record through water year 2018. A water year is the 12-month period from October 1 through September 30 and is designated by the calendar year in which it ends. For example, water year 2018 begins October 1, 2017, and ends September 30, 2018. Partial-record streamflow-gaging stations were not used in this study.

Previous Studies

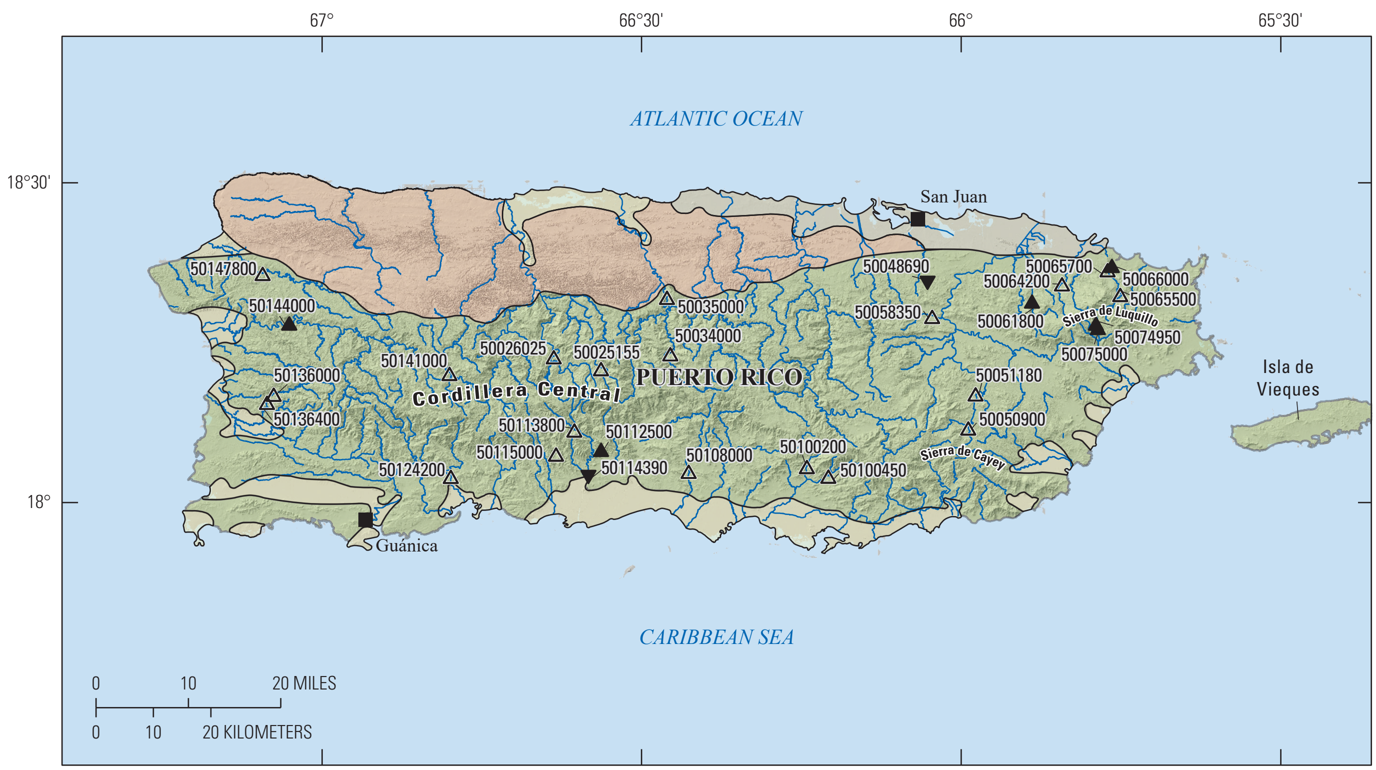
Selected flow duration and *n*-day low-flow frequencies have been previously published in Puerto Rico. Quinones and others (1984) presented monthly flow durations for 78 streamflow-gaging stations in Puerto Rico, with 35 of the 78 streamflow-gaging stations affected by regulation. The period of record of the 78 streamflow-gaging stations ranged from 3 to 25 years. Cobb (1978) presented 7-day, 10-year (0.1 nonexceedance probability) frequency estimates for 31 streamflow-gaging stations in Puerto Rico. These frequency estimates were made for an additional 15 streamflow-gaging stations with inadequate record lengths using nearby streamflow-gaging stations and drainage area comparisons. Colón-Dieppa and Quinones-Aponte (1985; as mentioned in Santiago-Rivera, 1992, 1995, 1998) present 7-day, 10-year (0.1 nonexceedance probability) frequency estimates for selected streamflow-gaging stations. Santiago-Rivera (1992) presented 7-, 14-, 30-, 60-, and 90-day low flows with 2-year and 10-year (0.5 and 0.1 nonexceedance probabilities) frequency estimates for 12 continuous streamflow-gaging stations with at least 10 years of record and 7-, 14-, and 30-day low-flows with 2-year and 10-year (0.5 and 0.1 nonexceedance probabilities) frequency estimates for 81 partial record streamflow-gaging stations. The 93 streamflow gaging stations selected by Santiago-Rivera (1992) were in eastern Puerto Rico and not affected by regulation. Santiago-Rivera (1995) presented 7-, 14-, 30-, 60-, and 90-day low flows with 2-year and 10-year (0.5 and 0.1 nonexceedance probabilities) frequency estimates for 9 continuous streamflow-gaging stations with at least 10 years of record and 7-, 14-, and 30-day low flows with 2-year and 10-year (0.5 and 0.1 nonexceedance probabilities) frequency estimates for 105 partial record streamflow-gaging stations. The 114 streamflow-gaging stations selected by Santiago-Rivera (1995) were in southern and western Puerto Rico. One continuous and 11 partial-record streamflow-gaging stations were affected by regulation. Santiago-Rivera (1998) presented 7-, 14-, 30-, 60-, and 90-day low flows with 2-year and 10-year (0.5 and 0.1 nonexceedance probabilities) frequency estimates for 15 continuous streamflow-gaging stations with at least 10 years of record (1 continuous streamflow-gaging station with less than 10 years of record was used) and 7-, 14-, and 30-day low-flows with 2-year and 10-year (0.5 and 0.1 nonexceedance probabilities) frequency estimates for 94 partial record streamflow-gaging stations. The 109 streamflow-gaging stations selected by Santiago-Rivera (1998) were in northern and central Puerto Rico and not affected by regulation. Regression equations for estimating flow duration and *n*-day low-flow frequency statistics at streamflow-gaging stations with little or no regulation have not been developed in Puerto Rico.

Description of Study Area

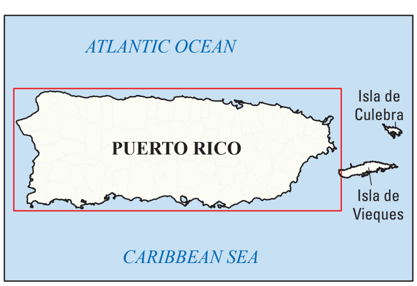
Puerto Rico is the smallest and most eastern island of the Greater Antilles. It is bounded by the Atlantic Ocean to the north and the Caribbean Sea to the south (Ramos-Ginés, 1999).

There are three major physiographic regions in Puerto Rico: the mountainous area, the discontinuous coastal plain, and the area of karst topography (fig. 1). The Cordillera Central and Sierra de Cayey, the central mountain chains in the interior, run east to west and have peak elevations that range from 1,970 to more than 3,940 feet above sea level. The Cordillera Central and Sierra de Cayey divide the island into a northern two-thirds and a southern one-third and form the principal drainage divides of the island's larger streams. The stream valleys in the mountainous areas are rough and deeply incised with steep slopes (Ramos-Ginés, 1999). The coastal lowlands are located mainly along the north and south coastlines with some smaller lowlands near the east and west coasts. These lowland areas include alluvial valleys, which are the result of erosion of the interior mountains. The karst area, located in the northwestern part of the island, consists of rugged limestone and karst features such as caves, sinkholes, haystack hills, and limestone cliffs (Gómez-Gómez and others, 2014).

The Cordillera Central and Sierra de Cayey mountains form an insular hydrologic divide that separates the island of Puerto Rico into two climatologically distinct regions. The northern two-thirds of the island has a somewhat humid climate, whereas the southern one-third of the island is semi-arid (Gómez-Gómez and others, 2014). In coastal areas, air temperatures range from 75 to 81 degrees Fahrenheit, and in the interior mountainous areas, air temperatures range from 72 to 77 degrees Fahrenheit (Gómez-Gómez and others, 2014). There is a general rainy season from April to November and a dry season from December to March. Owing to the topography of the island, rainfall varies greatly across the island.



Base from U.S. Geological Survey digital data
Hillshade from digital elevation model (U.S. Geological Survey, 2020)
North American Datum of 1983



EXPLANATION

Physiography modified from Miller and others, 1999

- Area of karst topography
- Mountainous area
- Discontinuous coastal plain

U.S. Geological Survey streamflow-gaging station and identifier

- 50112500 ▲ Streamflow-gaging stations with positive *n*-day low-flow trend
- 50108000 △ Streamflow-gaging stations with no *n*-day low-flow trend
- 50114390 ▼ Streamflow-gaging stations with negative *n*-day low-flow trend

Figure 1. Streamflow-gaging stations used in the development of flow duration and annual *n*-day low-flow frequency regression equations for Puerto Rico (modified from Miller and others, 1999).

The mean annual total rainfall in Sierra de Luquillo in eastern Puerto Rico is 169 inches per year (Gómez-Gómez and others, 2014). The least amount of rain falls in the vicinity of Guánica (fig. 1) in southwestern Puerto Rico. In this area, the mean annual total rainfall is 30 inches per year (Gómez-Gómez and others, 2014). Puerto Rico does experience the Atlantic hurricane season, which is similar to the remainder of the Caribbean Sea and North Atlantic Ocean. On average, one-fourth of its annual rainfall is contributed by tropical cyclones. Drought conditions do happen yearly within many geographic areas of Puerto Rico. Extended, extreme drought conditions affecting the entire region of Puerto Rico have occurred three times (1920s–1930s, mid-1960s–1968, and 1994) during the 20th century (Gómez-Gómez and others, 2014). Extreme drought in 2015 resulted in water rationing for residents and businesses (Alvarez, 2015).

Selection of Streamflow-Gaging Stations

The USGS retrieved and reviewed 227 streamflow-gaging station records for Puerto Rico from the USGS NWIS database (U.S. Geological Survey, 2019) with daily mean streamflow data for inclusion in the study. The screening criteria for selected streamflow-gaging stations required them to be unaffected or have very minimal effects from regulation or activities that may affect low flows, such as diversions for domestic supply or irrigation. In addition, the selected streamflow-gaging stations were required to have at least 10 years of daily flow record through water year 2018. Through the review process, 103 streamflow-gaging stations were identified with 10 or more years of daily mean streamflow data. Of the 103 streamflow-gaging stations with 10 or more years of daily mean streamflow data, only 28 streamflow-gaging stations were identified with no to minimal affects from regulation or activities that may affect low flows and thus were selected for this study. These 28 streamflow-gaging stations are distributed across Puerto Rico, except for the northwestern part of the island and an area east of the Cordillera Central where no streamflow-gaging stations were identified for use in this study (fig. 1, table 1). Record lengths for the 28 streamflow-gaging stations varied from 11 to 65 water years of data, with 9 streamflow-gaging stations having less than 25 water years of data, 15 streamflow-gaging stations having 25 to 50 water years of data, and 4 streamflow-gaging stations having greater than 50 water years of data. Daily mean streamflow data for the 28 selected streamflow-gaging stations were retrieved from NWIS (U.S. Geological Survey, 2019) for use in computing the flow duration and the annual *n*-day low-flow statistics (tables 2 and 3).

Statistical Methods

This section describes the statistical methods used in the publicly available USGS Surface-Water Toolbox computer program (SWToolbox) (Kiang and others, 2018) that implements the USGS Surface-Water Statistics computer program functionality (Flynn and others, 1995) within a Windows™ interface. The “Duration/Compare” and the “Integrated Frequency Analysis” procedures within SWToolbox were used to compute the flow duration and annual *n*-day low-flow statistics for this study. The statistical methods in SWToolbox were developed for analysis of daily mean streamflow and annual low flows derived from daily mean streamflow.

Flow-Duration Statistics

Flow durations represent the percentage of time that a given flow, measured during a specific 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 used 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):

$$P = 100 \times \left(\frac{M}{n+1}\right), \tag{1}$$

where

- P is the percent probability that a given streamflow will be equaled or exceeded;
- M is the ranked position, from highest to lowest, of all daily mean streamflow values for period of record used; and
- n is the number of daily mean streamflow values for the period of record used.

Flow-duration statistics for this study were computed using daily mean streamflow record through water year 2018 and using the SWToolbox “SWSTAT default” option for class limits generation. The computed flow-duration statistics for the 28 streamflow-gaging stations used in this study are listed in table 2 and are available as a USGS data release (Williams-Sether, 2021).

Annual *n*-Day Low-Flow Series

The *n*-day low-flow frequencies are computed for stations by use of an annual series of selected low flows based on the lowest mean flow for a number of consecutive days. Any combination of number of days of mean minimum flow and years of recurrence may be used to determine *n*-day low-flow frequencies (Kiang and others, 2018). Typically, the annual low-flow series for the determination of *n*-day low-flow frequencies are computed using a climatic year, which is from April 1 of a given year to March 31 of the following year.

Table 1. Streamflow-gaging stations and basin characteristics used in the regression analyses for Puerto Rico.

[HUC, hydrologic unit code; DRNAREA, area that drains to a point on a stream, in square miles; ETPENMON, mean total annual reference evapotranspiration (Penman-Monteith Method; Howell and Evett, 2004), in millimeters; GWHEAD, mean minus minimum basin elevation, in feet (North American Vertical Datum of 1988); MINBELEV, minimum basin elevation, in feet (North American Vertical Datum of 1988); SSURGOB, percentage of area of Hydrologic Soil Type B from SSURGO (Kolb and Ryan, 2019), in percent; PR, Puerto Rico; nr, near; blw, below; Hwy, highway; abv, above]

Station number	Station name	Latitude (decimal degrees)	Longitude (decimal degrees)	HUC	Basin characteristics				
					DRNAREA	ETPENMON	GWHEAD	MINBELEV	SSURGOB
50025155	Rio Saliente at Coabey nr Jayuya, PR	18.2108	−66.5628	21010002	8.87	1,230	1,340	1,670	80.9
50026025	Rio Caonillas at Paso Palma, PR	18.2294	−66.6367	21010002	37.9	1,320	1,360	944	66.6
50034000	Rio Bauta nr Orocovis, PR	18.2340	−66.4544	21010002	16.7	1,290	1,630	781	2.28
50035000	Rio Grande de Manati at Ciales, PR	18.3224	−66.4596	21010002	134	1,340	1,760	136	10.6
50048690	Quebrada Las Curias blw Las Curias outflow, PR	18.3430	−66.0518	21010005	1.13	1,530	195	242	0.466
50050900	Rio Grande de Loiza at Quebrada Arenas, PR	18.1180	−65.9882	21010005	6.02	1,390	614	640	10.8
50051180	Quebrada Salvatierra nr San Lorenzo, PR	18.1713	−65.9768	21010005	3.76	1,500	436	296	1.17
50058350	Rio Cana at Rio Canas, PR	18.2925	−66.0447	21010005	7.56	1,500	462	146	35.7
50061800	Rio Canovanas nr Campo Rico, PR	18.3165	−65.8885	21010005	10.3	1,460	1,290	231	6.25
50064200	Rio Grande nr El Verde, PR	18.3434	−65.8416	21010005	7.29	1,390	1,580	127	15.8
50065500	Rio Mameyes nr Sabana, PR	18.3268	−65.7504	21010005	6.78	1,270	1,380	266	32.5
50065700	Rio Mameyes at Hwy 191 at Mameyes, PR	18.3655	−65.7702	21010005	11.9	1,340	1,160	30.9	34.6
50066000	Rio Mameyes at Mameyes, PR	18.3722	−65.7635	21010005	13.4	1,370	1,060	11.7	30.6
50074950	Quebrada Guaba nr Naguabo, PR	18.2819	−65.7885	21010005	0.059	1,280	210	2,080	0
50075000	Rio Icacos nr Naguabo, PR	18.2752	−65.7854	21010005	1.24	1,270	224	2,020	0
50100200	Rio Lapa nr Rabo Del Buey, PR	18.0577	−66.2407	21010004	9.98	1,490	1,210	376	0.343
50100450	Rio Majada at La Plena, PR	18.0427	−66.2071	21010004	16.4	1,520	1,110	415	0.687
50108000	Rio Descalabrado nr Los Llanos, PR	18.0502	−66.4257	21010004	12.9	1,560	638	214	7.26
50112500	Rio Inabon at Real Abajo, PR	18.0845	−66.5627	21010004	9.69	1,430	1,410	397	35.2
50113800	Rio Cerrillos abv Lago Cerrillos nr Ponce, PR	18.1148	−66.6045	21010004	11.9	1,370	1,360	706	17.5
50114390	Rio Bucana at Hwy 14 bridge nr Ponce, PR	18.0394	−66.5824	21010004	24.5	1,460	1,330	116	13.0
50115000	Rio Portugues nr Ponce, PR	18.0772	−66.6332	21010004	8.81	1,380	1,390	469	5.74
50124200	Rio Guayanilla nr Guayanilla, PR	18.0423	−66.7978	21010004	18.8	1,520	1,180	72.1	8.96
50136000	Rio Rosario at Rosario, PR	18.1708	−67.0749	21010003	17.6	1,440	1,040	216	15.6
50136400	Rio Rosario nr Hormigueros, PR	18.1581	−67.0854	21010003	18.6	1,440	1,060	154	15.0
50141000	Rio Blanco nr Adjuntas, PR	18.2033	−66.7999	21010003	15.2	1,310	789	1,550	18.8
50144000	Rio Grande de Anasco nr San Sebastian, PR	18.2822	−67.0506	21010003	134	1,400	1,430	106	14.2
50147800	Rio Culebrinas at Hwy 404 nr Moca, PR	18.3597	−67.0922	21010003	70.6	1,520	451	45.1	35.1

Table 2. Streamflow associated with flow-duration exceedance probabilities for streamflow-gaging stations used in the regression analyses for Puerto Rico.

[D99, flow duration with 99-percent exceedance probability, in cubic feet per second; D98, flow duration with 98-percent exceedance probability, in cubic feet per second; D95, flow duration with 95-percent exceedance probability, in cubic feet per second; D90, flow duration with 90-percent exceedance probability, in cubic feet per second; D80, flow duration with 80-percent exceedance probability, in cubic feet per second; D70, flow duration with 70-percent exceedance probability, in cubic feet per second; D60, flow duration with 60-percent exceedance probability, in cubic feet per second; D50, flow duration with 50-percent exceedance probability, in cubic feet per second]

Station number	Streamflow associated with flow-duration exceedance probabilities							
	D99	D98	D95	D90	D80	D70	D60	D50
50025155	3.45	3.80	4.90	6.28	8.60	10.9	13.0	16.2
50026025	14.3	16.0	18.9	23.1	31.0	38.0	45.3	54.1
50034000	3.50	4.00	5.10	6.50	8.70	11.0	13.0	15.6
50035000	24.0	32.0	40.2	51.6	67.0	83.0	101	122
50048690	0.040	0.060	0.120	0.210	0.320	0.450	0.630	0.910
50050900	4.00	4.84	5.80	6.86	8.80	10.7	12.2	14.8
50051180	0.470	0.540	0.720	0.920	1.20	1.40	1.70	2.00
50058350	1.60	1.90	2.74	3.30	4.20	5.00	6.05	7.50
50061800	2.70	3.20	4.30	5.56	7.26	9.00	11.0	13.3
50064200	3.50	4.30	5.58	7.10	9.67	12.0	15.0	19.0
50065500	10.0	11.3	14.0	17.0	21.0	25.0	29.3	35.0
50065700	8.70	10.0	13.0	17.0	23.0	28.0	34.0	40.0
50066000	9.58	11.0	14.0	17.9	23.4	29.0	35.9	43.4
50074950	0.100	0.110	0.140	0.170	0.220	0.260	0.310	0.370
50075000	3.09	3.60	4.20	5.00	6.00	6.78	7.80	9.00
50100200	0.010	0.040	0.130	0.250	0.490	0.740	0.980	1.30
50100450	[1]0.005	[1]0.020	0.090	0.340	0.650	1.00	1.55	2.18
50108000	[1]0.005	[1]0.020	0.030	0.140	0.450	0.770	1.14	1.80
50112500	1.80	2.10	2.80	3.60	5.00	6.30	8.00	10.0
50113800	3.80	4.30	5.31	6.80	8.94	11.0	13.7	16.6
50114390	2.35	3.02	3.80	4.30	4.97	5.60	6.60	8.00
50115000	1.70	1.90	2.30	3.00	4.00	5.00	6.20	8.00
50124200	1.90	2.20	3.10	4.00	5.40	6.83	8.60	11.0
50136000	5.40	6.90	9.00	11.0	14.0	19.0	24.7	31.0
50136400	7.80	9.10	11.0	13.0	17.0	20.4	25.6	32.5
50141000	7.00	7.50	8.80	10.0	12.8	15.0	17.0	22.0
50144000	52.0	57.0	67.0	79.0	103	131	165	205
50147800	26.0	29.0	35.2	44.0	61.0	81.0	108	142

[1]Numbers have been adjusted from zero.

Table 3. Streamflow associated with *n*-day mean low-flow frequencies for streamflow-gaging stations used in the regression analyses for Puerto Rico.

[M1D2Y, 1-day mean low flow that occurs on average once in 2 years, in cubic feet per second; M1D5Y, 1-day mean low flow that occurs on average once in 5 years, in cubic feet per second; M1D10Y, 1-day mean low flow that occurs on average once in 10 years, in cubic feet per second; M7D2Y, 7-day mean low flow that occurs on average once in 2 years, in cubic feet per second; M7D5Y, 7-day mean low flow that occurs on average once in 5 years, in cubic feet per second; M7D10Y, 7-day mean low flow that occurs on average once in 10 years, in cubic feet per second; M14D2Y, 14-day mean low flow that occurs on average once in 2 years, in cubic feet per second; M14D5Y, 14-day mean low flow that occurs on average once in 5 years, in cubic feet per second; M14D10Y, 14-day mean low flow that occurs on average once in 10 years, in cubic feet per second; M30D2Y, 30-day mean low flow that occurs on average once in 2 years, in cubic feet per second; M30D5Y, 30-day mean low flow that occurs on average once in 5 years, in cubic feet per second; M30D10Y, 30-day mean low flow that occurs on average once in 10 years, in cubic feet per second]

Station number	Streamflow associated with *n*-day mean low-flow frequencies											
	M1D2Y	M1D5Y	M1D10Y	M7D2Y	M7D5Y	M7D10Y	M14D2Y	M14D5Y	M14D10Y	M30D2Y	M30D5Y	M30D10Y
50025155	5.17	3.74	3.11	5.73	4.19	3.50	6.29	4.51	3.72	7.51	5.24	4.24
50026025	19.7	15.3	13.2	21.0	16.3	14.1	22.5	17.4	15.1	26.9	20.5	17.5
50034000	6.38	4.58	3.79	6.85	4.92	4.06	7.27	5.16	4.25	8.20	5.70	4.64
50035000	46.9	30.1	23.1	51.9	33.7	26.0	56.8	37.8	29.6	64.8	42.7	33.2
50048690	0.103	0.050	0.033	0.156	0.077	0.050	0.243	0.122	0.079	0.399	0.179	0.107
50050900	5.39	4.14	3.59	5.95	4.58	3.97	6.36	4.91	4.29	7.15	5.46	4.75
50051180	0.645	0.422	0.331	0.770	0.526	0.421	0.853	0.605	0.496	1.02	0.759	0.656
50058350	2.89	1.91	1.51	3.35	2.23	1.78	3.79	2.57	2.07	4.39	2.89	2.28
50061800	4.47	2.85	2.16	4.89	3.23	2.56	5.35	3.52	2.78	6.52	4.26	3.32
50064200	4.91	3.40	2.73	5.80	4.05	3.27	6.88	4.76	3.84	9.21	6.34	5.12
50065500	11.4	8.97	7.78	12.9	10.5	9.39	15.0	12.0	10.6	19.1	14.9	13.0
50065700	10.9	8.09	6.80	12.9	9.91	8.69	15.4	11.7	10.1	21.1	15.2	12.7
50066000	9.48	7.34	6.55	11.4	8.87	7.92	14.2	11.4	10.2	18.8	14.5	12.8
50074950	0.109	0.082	0.071	0.130	0.096	0.082	0.151	0.114	0.098	0.194	0.146	0.126
50075000	3.70	2.82	2.38	4.28	3.45	3.03	4.76	3.82	3.36	5.80	4.66	4.11
50100200	[1]0.051	[1]0.025	[1]0.016	[1]0.065	[1]0.038	[1]0.025	[1]0.075	[1]0.057	[1]0.039	0.512	0.127	0.038
50100450	[1]0.051	[1]0.025	[1]0.016	[1]0.065	[1]0.038	[1]0.025	[1]0.075	[1]0.057	[1]0.039	[1]0.097	[1]0.064	[1]0.019
50108000	[1]0.051	[1]0.025	[1]0.016	[1]0.065	[1]0.038	[1]0.025	[1]0.075	[1]0.057	[1]0.039	[1]0.097	[1]0.064	[1]0.019
50112500	2.40	1.59	1.25	2.85	1.97	1.60	3.25	2.33	1.93	3.91	2.86	2.40
50113800	5.91	4.38	3.65	6.39	4.72	3.93	6.80	5.10	4.31	7.58	5.63	4.73
50114390	3.34	2.49	2.12	3.98	2.94	2.46	4.69	3.49	2.89	5.19	3.68	3.05
50115000	2.38	1.64	1.29	2.57	1.87	1.56	2.77	2.05	1.73	3.22	2.37	1.98
50124200	3.10	1.97	1.52	3.38	2.30	1.87	3.72	2.61	2.16	4.30	3.07	2.57
50136000	9.28	6.35	4.83	10.1	7.06	5.39	10.9	7.59	5.80	11.4	8.31	6.80
50136400	10.5	7.30	5.61	11.0	8.08	6.61	11.6	8.87	7.55	12.9	10.5	9.49
50141000	8.05	6.57	5.92	8.52	6.96	6.27	9.03	7.42	6.75	10.0	8.20	7.46
50144000	59.9	48.2	42.9	65.2	52.7	46.9	69.8	56.4	50.4	77.9	63.7	57.6
50147800	30.1	24.4	22.1	32.6	26.5	24.1	35.2	28.2	25.5	39.3	31.5	28.4

[1]Numbers have been adjusted from zero.

However, unlike the conterminous United States, the low-flow period in Puerto Rico occurs within the water year. To verify using a water year instead of a climatic year to compute *n*-day low-flow frequencies, boxplots of monthly mean streamflow were examined for all 28 streamflow-gaging stations. The season when low flows typically occur should not span years, whether defined as a water or climatic year. Lowest monthly mean streamflows at streamflow-gaging station 50034000 (fig. 2) occur in December–March and June–August, which are within the span of a water year. This boxplot example (fig. 2) is representative of all 28 streamflow-gaging stations and verifies that water year can be used to determine an independent annual *n*-day low-flow series in Puerto Rico; thus, the annual *n*-day low-flow time series computed for the *n*-day low-flow frequency estimations for this study were based on water year. An annual *n*-day low-flow time series consists of each of the lowest mean *n*-day periods in a year for the period of record; for example, if *n* equals 7, the annual 7-day low-flow time series consists of the lowest mean streamflow during any 7-day consecutive period within each water year of record. An annual *n*-day low-flow value was not computed for a water year that contained any missing record.

N-Day Low-Flow Frequency

The log-Pearson Type III probability distribution (Interagency Advisory Committee on Water Data, 1982; Kiang and others, 2018) was used for determining the *n*-day low-flow frequencies for this study. An overview of techniques used to compute low-flow frequency statistics is provided by Riggs (1972) and Kiang and others (2018). A *n*-day low-flow frequency analysis is used to estimate the probability that flow will not exceed different flow levels in any given year. The SW Toolbox SWSTAT program automatically adjusted the annual nonexceedance probability Log-Pearson Type III low-flow curves when *n*-day low-flow values of zero were present. The computed *n*-day low-flow frequencies for the 28 streamflow-gaging stations used in this study are listed in table 3 and are available as a USGS data release (Williams-Sether, 2021).

The Kendall tau test (Lins and Cohn, 2011) was used to test for nonstationarity (trends) in the annual *n*-day low-flow time series used for the low-flow frequency estimates. A *p*-value threshold of 5 percent (α=0.05) was used in this study for the Kendall's tau test, and *p*-values less than or equal

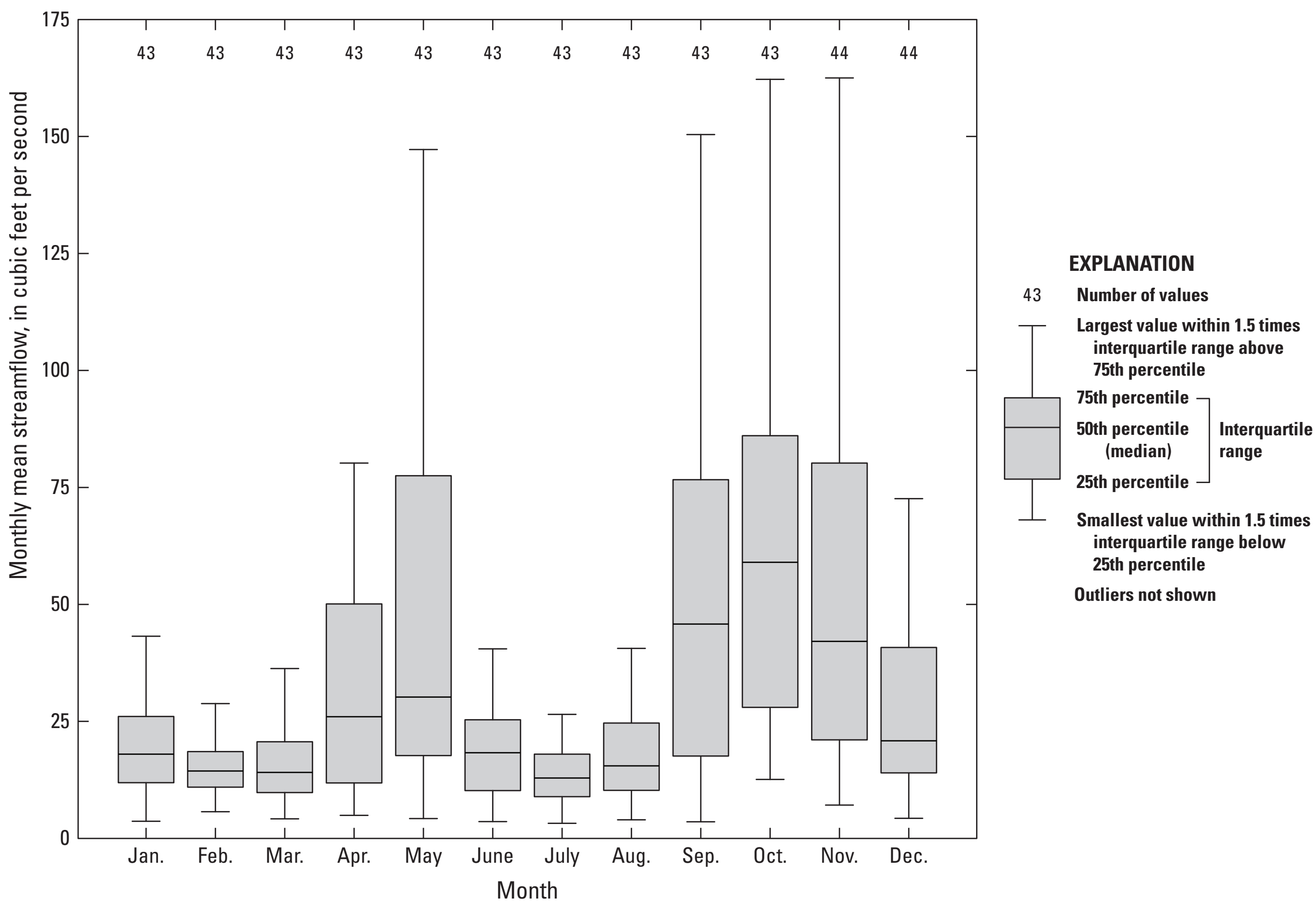


Figure 2. Boxplot of monthly mean streamflow for streamflow-gaging station 50034000 (Rio Bauta near Orocovis, Puerto Rico), 1969–2018.

to 5 percent were flagged as having statistically significant trends (positive or negative). Eight streamflow-gaging stations were found to have statistically significant trends in *n*-day low flows. Five stations were in northeastern Puerto Rico, 2 in southern Puerto Rico, and 1 in western Puerto Rico (fig. 1). Positive trends were identified for streamflow-gaging stations 50061800 (7- and 30-day low flows), 50066000 (1- and 7-day low flows), 50074950 (1- and 7-day low flows), 50075000 (1-, 7-, and 14-day low flows), 50112500 (1-, 7-, 14-, and 30-day low flows), and 50144000 (1-, 7-, 14-, and 30-day low flows). Negative trends were identified for streamflow-gaging stations 50048690 (1-day low flow) and 50114390 (1-, 7-, 14-, and 30-day low flows). Results of the Kendall tau test are available as a USGS data release (Williams-Sether, 2021). The *n*-day low-flow time series where trends were indicated were not adjusted owing to uncertainties in the cause of the trends. For example, short-term records may show a trend, but the trend could be temporary based on the record time period and does not mean that there is an actual change in the system or conditions long term.

Basin and Climatic Characteristics

The USGS StreamStats Program was created by the USGS in cooperation with Esri to make GIS-based estimation of streamflow statistics easier, faster, and more consistent than previously used manual techniques. The USGS StreamStats Program is a map-based internet interface designed for national application, with each State, Territory, or group of States responsible for creating unique geospatial datasets and regression equations to compute streamflow statistics. Further information about StreamStats usage and limitations can be accessed at U.S. Geological Survey (2021). A Puerto Rico StreamStats web application was developed to provide the estimates of basin and climatic characteristics at the 28 streamflow-gaging stations used in this study (table 1). The Puerto Rico StreamStats web application may be used to provide estimates of streamflow frequency statistics for user-selected ungaged sites on Puerto Rico streams. The Puerto Rico web application is only applicable to the main island and does not include the islands Vieques and Culebra (not shown in fig. 1). Forty-seven basin and climatic characteristic layers were developed for the Puerto Rico StreamStats study, and the GIS raster layers are available in Kolb and Ryan (2019). The 47 characteristics that were considered for potential use in the regression analysis are listed in table 4.

In regional regression studies, independent and dependent variables often are assumed to have a normal distribution (and can therefore be described using the parameters of mean and standard deviation) and need to be transformed into log space (log-normal distribution) before the regression equation is created to ensure a linear relation among the independent and dependent variables. However, some independent (basin and climatic characteristics) and dependent (duration flows and *n*-day low flows) variables for some stations had a value of zero that could not be logarithmically transformed. A logarithmic transformation is undefined when a value is zero, so a constant of +1 was added to the independent variable (basin and climatic characteristic) for all station data used if the independent variable had a zero value and a logarithmic transformation was needed. Zero values in the dependent flow statistics were adjusted using a constant value of one-half of the smallest value of all station data used for a given flow statistic. Log transformations had a base unit of 10 ($\log_{10}$).

The 47 basin and climatic characteristics (tables 1 and 4) were tested as explanatory variables for regression analysis. Correlation matrix charts of the log-transformed flow statistics (D70 and M14D10Y), log-transformed explanatory variables, and untransformed explanatory variables were generated in R (version 1.2.1335) (R Core Team, 2019) to evaluate if log transformation of the explanatory variables was needed and to check for correlation of the explanatory variables with flow statistics. Explanatory variables that indicated a weak correlation (values between −0.3 and 0.3) with D70 and M14D10Y were removed from further consideration, leaving a subset of 14 potential explanatory variables (identified in table 4). Cross-correlation was identified visually and noted in table 4 for 4 of the 14 explanatory variables (BASINPERIM, DRNAREA, LFPLENGTH, and STRMTOT).

Development of Regional Regression Equations

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

Table 4. Basin and climatic characteristics considered for use in the flow duration and annual *n*-day low-flow frequency regression analysis for Puerto Rico.

[ALOS/PALSAR, advanced land observing satellite/phased array type L-band synthetic aperture radar; DEM, digital elevation model; NLCD 2001, 2001 national land cover database (Kolb and Ryan, 2019); NLCD 2011, 2011 national land cover database (Multi-Resolution Land Characteristics Consortium, 2018); USFS, U.S. Forest Service; NRCS, Natural Resources Conservation Service; SSURGO, soil survey geographic database (Kolb and Ryan, 2019)]

Label	Definition	Units	Minimum	Maximum
ALPA10BARE	Percent bare ground land-use/land-cover category for Puerto Rico based on 2010 ALOS/PALSAR	percent	0	0.170
ALPA10FLWD	Percent flat woodland land-use/land-cover category for Puerto Rico based on 2010 ALOS/PALSAR	percent	0	6.17
ALPA10FRST	Percent forests land-use/land-cover category for Puerto Rico based on 2010 ALOS/PALSAR	percent	37.8	99.2
ALPA10FRWT	Percent forested wetland land-use/land-cover category for Puerto Rico based on 2010 ALOS/PALSAR	percent	0	1.76
ALPA10HERB	Percent herbaceous land-use/land-cover category for Puerto Rico based on 2010 ALOS/PALSAR	percent	0	42.7
ALPA10MTWD	Percent mountain woodland land-use/land-cover category for Puerto Rico based on 2010 ALOS/PALSAR	percent	0.175	30.1
ALPA10SHSC	Percent shrubs/scrub land-use/land-cover category for Puerto Rico based on 2010 ALOS/PALSAR	percent	0	12.5
ALPA10URBA	Percent urban land-use/land-cover category for Puerto Rico based on 2010 ALOS/PALSAR	percent	0	16.0
ALPA10WATR	Percent water land-use/land-cover category for Puerto Rico based on 2010 ALOS/PALSAR	percent	0	3.82
BASINPERIM[1,2]	Perimeter of the drainage basin as defined in SIR 2004–5262 (Gingerich, 2005)	miles	1.30	89.7
BSLDEM10M	Mean basin slope computed from 10-meter DEM	feet per mile	24.9	60.8
CSL10_85fm	Change in elevation between points 10 and 85 percent of length along main channel to basin divide divided by length between points	feet per mile	31.4	912
DRNAREA[1,2]	Area that drains to a point on a stream	square miles	0.059	134
ELEV	Mean basin elevation	feet	437	3,020
ELEVMAX[1]	Maximum basin elevation	feet	790	4,380
ETPENMON[1]	Mean total annual reference evapotranspiration (Penman-Monteith Method; Howell and Evett, 2004)	inches	1,230	1,560
GWHEAD[1]	Mean minus minimum basin elevation	feet	195	1,760
KARST	Presence/absence of karst in Puerto Rico	percent	0	20.7
LAKESNWI	Percentage of lakes and ponds as determined from the National Wetlands Inventory (2001) (U.S. Fish and Wildlife Service, 2021)	percent	0	5.89
LC01BARE	Percentage of area barren land, NLCD 2001 category 31	percent	0	2.44
LC01CROP	Percentage of area cultivated crops, NLCD 2001 category 82	percent	0	0.438
LC01CRPHAY	Percentage of cultivated crops and hay, classes 81 and 82, from NLCD 2001	percent	0	3.30
LC01DEV	Percentage of land-use from NLCD 2001 classes 21–24	percent	0	21.9
LC01EVERG	Percentage of area evergreen forest, NLCD 2001 category 42	percent	31.5	100
LC01FOREST	Percentage of forest from NLCD 2001 classes 41–43	percent	31.5	100
LC01HERB	Percentage of herbaceous upland from NLCD 2001 class 71	percent	0	53.5
LC01IMP	Average percentage of impervious land cover from NLCD 2001	percent	0	6.62
LC01SHRUB	Percentage of shrub scrub from NLCD 2001 class 52	percent	0	16.2

Table 4. Basin and climatic characteristics considered for use in the flow duration and annual *n*-day low-flow frequency regression analysis for Puerto Rico.—Continued

[ALOS/PALSAR, advanced land observing satellite/phased array type L-band synthetic aperture radar; DEM, digital elevation model; NLCD 2001, 2001 national land cover database (Kolb and Ryan, 2019); NLCD 2011, 2011 national land cover database (Multi-Resolution Land Characteristics Consortium, 2018); USFS, U.S. Forest Service; NRCS, Natural Resources Conservation Service; SSURGO, soil survey geographic database (Kolb and Ryan, 2019)]

Label	Definition	Units	Minimum	Maximum
LC01WETLND	Percentage of wetlands, classes 90 and 95, from NLCD 2001	percent	0	0.232
LC11CANOPY	Percent tree canopy cover for Puerto Rico from the NLCD 2011 USFS Percent Tree Canopy Cover (Analytical Version)	percent	61.9	96.7
LFPLENGTH[1,2]	Length of longest flow path	miles	0.379	43.2
MINBELEV[1]	Minimum basin elevation	feet	11.7	2,080
PRECIP	Basinwide mean annual precipitation	inches	57.2	155
RCN[1]	Runoff-curve number as defined by NRCS (U.S. Department of Agriculture, 1986)	dimensionless	0	77.8
RELIEF[1]	Maximum minus minimum elevation	feet	462	4,250
ROCKDEP	Depth to rock	feet	0.010	4.69
RUGGED[1]	Ruggedness number computed as stream density times basin relief	feet per mile	0	11,100
SLOPERAT[1]	Slope ratio computed as main channel slope divided by basin slope	dimensionless	1.26	21.2
SSURGOA	Percentage of area of Hydrologic Soil Type A from SSURGO	percent	0	2.97
SSURGOB[1]	Percentage of area of Hydrologic Soil Type B from SSURGO	percent	0	80.9
SSURGOC	Percentage of area of Hydrologic Soil Type C from SSURGO	percent	0	83.8
SSURGOD	Percentage of area of Hydrologic Soil Type D from SSURGO	percent	0	96.0
SSURGODEP	Area-weighted average soil depth from NRCS SSURGO database	inches	0.711	3.60
SSURGOKSAT	Saturated hydraulic conductivity in micrometers per second from NRCS SSURGO database	micrometers per second	3.01	33.3
SSURGWATCP[1]	Available water capacity of the top 60 inches of soil—Determined from SSURGO data	inch per inch	0.114	0.183
STRDEN	Stream density—Total length of streams divided by drainage area	miles per square mile	0	3.31
STRMTOT[1,2]	Total length of mapped streams in basin	miles	0	372

[1]Basin characteristics used in the flow duration and annual *n*-day low-flow frequency regression analysis.

[2]Cross-correlated basin characteristics.

To determine which of the 14 explanatory variables were significant in estimating flow statistics, potential models were selected using OLS regression. An exhaustive search was conducted in R (R Core Team, 2019) using the R function “regsubsets” to identify the best two models of all sizes up to the maximum of 14 variables for flow-duration statistics and annual *n*-day low-flow frequency statistics. Selection criteria was based on the Bayesian information criterion (BIC) (Schwarz, 1978), Mallows' Cp (Mallows, 1973), and adjusted coefficient of determination (Adj-R^2) (Theil, 1958). The BIC is a criterion for model selection among a finite set of models. BIC favors more parsimonious models over more complex models; thus, it adds a penalty based on the number of parameters being estimated in the model. The model with the lowest BIC is preferred. Mallows’ Cp is used to assess the fit of a regression model that has been estimated using OLS. It estimates the size of the bias that is introduced into the predicted responses by having an underspecified model. A small value of Cp means that the model is somewhat precise. The Adj-R^2 compares the explanatory power of regression models that contain different numbers of predictors, and will increase if added variables to a model explain the variability and decrease if added variables to a model offer no benefit.

Farmer and others (2019) suggest one variable for every 10 streamflow-gaging stations used in final regression models. Because this study uses data from 28 streamflow-gaging stations, the explanatory variable selection for the final regression models was limited to no more than three variables. Of the four cross-correlated explanatory variables (BASINPERIM, DRNAREA, LFPLENGTH, and STRMTOT; table 4), DRNAREA was used as one of the final variables considered because it is more commonly used. Flow-duration regression variables for final analysis were DRNAREA, ETPENMON, MINBELEV, RCN, and SSURGOB. Annual *n*-day low-flow regression variables for final analysis were DRNAREA, ETPENMON, GWHEAD, MINBELEV, and SSURGOB. The island was not subdivided into subregions owing to the paucity of the number of streamflow-gaging stations used.

Flow-Duration Regressions

The final flow-duration regression equations were developed in R (R Core Team, 2019) using the “lm” function, which is an OLS technique. The GLS or WLS technique was not applicable for flow duration because a time series for flow duration was not available. Final model selections for each flow duration were based on the largest Adj-R^2 and minimum residual standard error (RSE). The final selected variables (DRNAREA, ETPENMON, and MINBELEV) resulted in the largest Adj-R^2 and minimum RSE. The RSE is a measure of how well the regression model fits the data. The final flow-duration regression equations are listed in table 5. Flow-duration statistics generated during this study are available as a USGS data release (Williams-Sether, 2021).

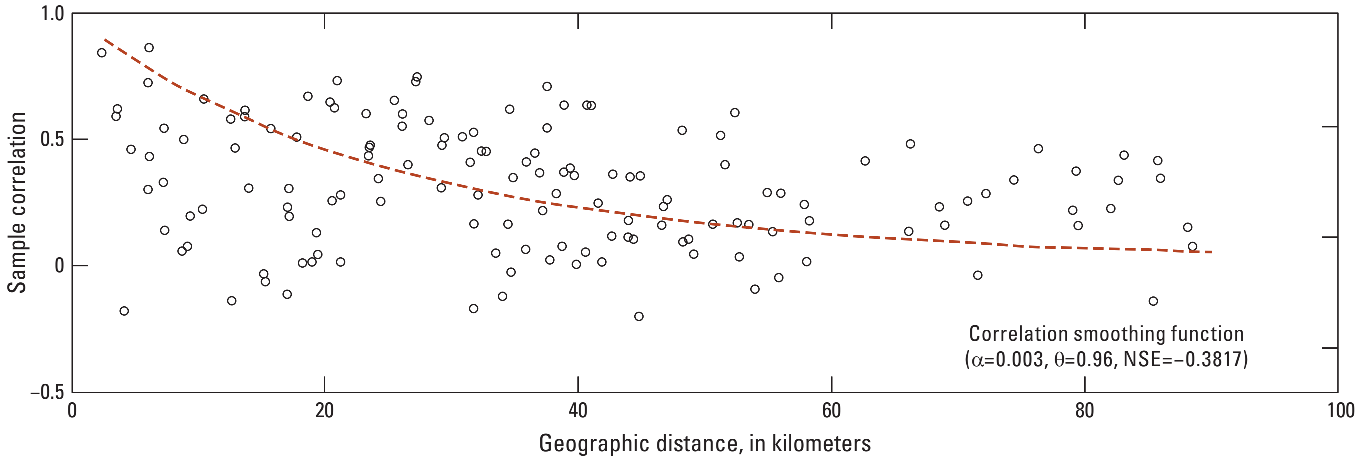
N-Day Low-Flow Regressions

Final annual *n*-day low-flow regression equations were developed using the OLS and GLS techniques in the WREG program. The benefit of using GLS instead of OLS for the *n*-day low-flow frequency statistics was evaluated by comparing resulting statistics of both techniques. The differences in mean square error and Adj-R^2 between the OLS and GLS techniques were less than 0.5 percent for all *n*-day regression models except for the M14D10Y regression model where the mean square error difference was 1.3 percent and the Adj-R^2 difference was 0.7 percent. The GLS technique was used for the *n*-day low-flow frequency statistics because GLS computes streamflow-gaging station weights while accounting for cross-correlation between the stations, varying flow-record lengths, and variances in the annual *n*-day low flows. The correlation smoothing function used by WREG to compute a weighting matrix for the data from the 28 streamflow-gaging stations included in the development of the GLS regression equations is shown in figure 3. The smoothing function relates the correlation between *n*-day low flows at two streamflow-gaging stations to the geographic distance between the

Table 5. Regression equations for flow-duration exceedance probability estimates in Puerto Rico.

[R^2, coefficient of determination, in percent; Adj-R^2, adjusted coefficient of determination, in percent; RSE, residual standard error, in $\log_{10}$ of cubic feet per second units; D*xx*, duration flow for *xx* exceedance probability; DRNAREA, drainage area, in square miles; ETPENMON, mean total annual reference evapotranspiration (Penman-Monteith Method (Howell and Evett, 2004)), in millimeters; MINBELEV, minimum basin elevation, in feet (North American Vertical Datum of 1988 (NAVD 88)]

Regression equation	R^2	Adj-R^2	RSE
$\log_{10}$ D99 = 13.1 + 0.701 * $\log_{10}$(DRNAREA) − 0.008 * ETPENMON − 0.735 * $\log_{10}$(MINBELEV)	62.4	57.7	0.702
$\log_{10}$ D98 = 11.3 + 0.720 * $\log_{10}$(DRNAREA) − 0.007 * ETPENMON − 0.630 * $\log_{10}$(MINBELEV)	68.0	64.0	0.566
$\log_{10}$ D95 = 10.1 + 0.710 * $\log_{10}$(DRNAREA) − 0.006 * ETPENMON − 0.564 * $\log_{10}$(MINBELEV)	71.3	67.7	0.484
$\log_{10}$ D90 = 8.69 + 0.714 * $\log_{10}$(DRNAREA) − 0.005 * ETPENMON − 0.488 * $\log_{10}$(MINBELEV)	77.6	74.8	0.377
$\log_{10}$ D80 = 7.93 + 0.721 * $\log_{10}$(DRNAREA) − 0.005 * ETPENMON − 0.443 * $\log_{10}$(MINBELEV)	81.3	78.9	0.323
$\log_{10}$ D70 = 7.58 + 0.730 * $\log_{10}$(DRNAREA) − 0.005 * ETPENMON − 0.424 * $\log_{10}$(MINBELEV)	82.7	80.5	0.303
$\log_{10}$ D60 = 7.32 + 0.731 * $\log_{10}$(DRNAREA) − 0.004 * ETPENMON − 0.419 * $\log_{10}$(MINBELEV)	83.2	81.1	0.294
$\log_{10}$ D50 = 7.04 + 0.734 * $\log_{10}$(DRNAREA) − 0.004 * ETPENMON − 0.403 * $\log_{10}$(MINBELEV)	83.5	81.4	0.286



EXPLANATION

Model—Fit of the proposed approximation of cross-correlation

Observation (*Y*=25)—*Y* is the minimum number of years of concurrent record required to estimate cross-correlation

NSE **Nash-Sutcliffe efficiency of the correlation model**

Figure 3. Screen capture of the weighted-multiple-linear regression program (WREG) smoothing function for generalized-least-squares (GLS) correlation of annual *n*-day low flows as a function of the distance between 28 streamflow-gaging stations in Puerto Rico with 25 years of concurrent *n*-day low-flow record.

streamflow-gaging stations for every paired combination of the 28 streamflow-gaging stations with 25 years of concurrent flow. There is some evidence of cross-correlation shown in figure 3 by the paired points (shown as “Observation”) for 25 years of concurrent flow that form the tail of the curve that extends towards the bottom right side of the graph. Final GLS regression models were selected primarily on the basis of minimizing values of the standard model error of variance (SMEV) and the average standard error of prediction (ASEP) and maximizing values of the pseudo coefficient of determination (Pseudo R^2). The final annual *n*-day low-flow regression equations are listed in table 6. Annual *n*-day low-flow data generated during this study are available as a USGS data release (Williams-Sether, 2021).

Accuracy and Limitations of the Regional Regression Equations

The Adj-R^2 for the flow-duration regression equations ranged from 57.7 to 81.4 percent, and the RSE ranged from 0.286 to 0.702 $\log_{10}$ of cubic feet per second units (table 5). The Pseudo R^2 for the annual *n*-day low-flow regression equations ranged from 64.6 to 70.7 percent, the SMEV ranged from 124 to 209 percent, and the ASEP ranged from 138 to 243 percent (table 6). Perhaps owing to the paucity of the dataset used and the presence of some zero flows, the Adj-R^2 and Pseudo R^2 values indicate a moderate to poor model fit. Likewise, the large percentages of the SMEV and the ASEP indicate low accuracy of the *n*-day low-flow regression models. The accuracy of the flow-duration and annual *n*-day low-flow regression models could be improved if more data became available and by including partial-record streamflow-gaging stations using record-extension techniques.

The following limitations warrant consideration when using the regression equations to compute flow durations and *n*-day flow frequencies for Puerto Rico streams: (1) the stream sites should be in rural drainage basins and be unaffected or have very minimal effects from regulation or activities that may affect low flows, such as diversions for domestic supply or irrigation; (2) the explanatory variables should be computed using the same GIS 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 2 and 3.

Table 6. Regression equations for annual *n*-day low-flow frequency estimates in Puerto Rico.

[Adj-R^2, adjusted coefficient of determination, in percent; Pseudo R^2, pseudo coefficient of determination, in percent; *AVP*, average variance of prediction, in $\log_{10}$ of cubic feet per second units; *ASEP*, average standard error of prediction, in percent; *MEV*, mean error variance, in $\log_{10}$ of cubic feet per second units; *SMEV*, standard model error variance, in percent; M*n*D*x*Y, *n* consecutive day mean low flow at the *x*-year return interval; DRNAREA, drainage area, in square miles; ETPENMON, mean total annual reference evapotranspiration (Penman-Monteith Method; Howell and Evett, 2004), in millimeters; MINBELEV, minimum basin elevation, in feet (North American Vertical Datum of 1988)]

Regression equation	Adj-R^2	Pseudo R^2	*AVP*	*ASEP*	*MEV*	*SMEV*
$\log_{10}$ (M1D2Y) = 9.92 + 0.760* $\log_{10}$ (DRNAREA) − 0.010* ETPENMON − 0.520* $\log_{10}$ (MINBELEV)	66.1	66.4	0.297	196	0.259	172
$\log_{10}$ (M1D5Y) = 10.5 + 0.770* $\log_{10}$ (DRNAREA) − 0.010* ETPENMON − 0.540* $\log_{10}$ (MINBELEV)	65.1	66.0	0.338	224	0.295	194
$\log_{10}$ (M1D10Y) = 10.9 + 0.780* $\log_{10}$ (DRNAREA) − 0.010* ETPEN-MON − 0.560* $\log_{10}$ (MINBELEV)	64.5	65.9	0.364	243	0.317	209
$\log_{10}$ (M7D2Y) = 9.81 + 0.730* $\log_{10}$ (DRNAREA) − 0.010* ETPENMON − 0.530* $\log_{10}$ (MINBELEV)	66.4	66.8	0.276	182	0.241	161
$\log_{10}$ (M7D5Y) = 10.2 + 0.750* $\log_{10}$ (DRNAREA) − 0.010* ETPENMON − 0.540* $\log_{10}$ (MINBELEV)	66.3	67.0	0.300	198	0.261	173
$\log_{10}$ (M7D10Y) = 10.6 + 0.750* $\log_{10}$ (DRNAREA) − 0.010* ETPEN-MON − 0.560* $\log_{10}$ (MINBELEV)	65.6	66.8	0.324	214	0.282	186
log10(M14D2Y) = 9.87 + 0.700* $\log_{10}$ (DRNAREA) − 0.010* ETPEN-MON − 0.560* $\log_{10}$ (MINBELEV)	66.3	66.8	0.267	177	0.233	156
$\log_{10}$ (M14D5Y) = 9.66 + 0.720* $\log_{10}$ (DRNAREA) − 0.010* ETPEN-MON − 0.540* $\log_{10}$ (MINBELEV)	67.8	69.5	0.252	167	0.219	148
$\log_{10}$ (M14D10Y) = 9.62 + 0.730* $\log_{10}$ (DRNAREA) − 0.010* ETPEN-MON − 0.530* $\log_{10}$ (MINBELEV)	66.8	70.7	0.258	171	0.223	151
$\log_{10}$ (M30D2Y) = 9.39 + 0.660* $\log_{10}$ (DRNAREA) − 0.010* ETPEN-MON − 0.540* $\log_{10}$ (MINBELEV)	70.1	70.6	0.202	138	0.176	124
$\log_{10}$ (M30D5Y) = 9.78 + 0.680* $\log_{10}$ (DRNAREA) − 0.010* ETPEN-MON − 0.560* $\log_{10}$ (MINBELEV)	68.9	70.3	0.230	155	0.201	138
$\log_{10}$ (M30D10Y) = 11.1 + 0.670* $\log_{10}$ (DRNAREA) − 0.010* ETPEN-MON − 0.640* $\log_{10}$ (MINBELEV)	62.5	64.6	0.355	236	0.309	203

StreamStats Web Application for Regional Regression Equations

The Puerto Rico StreamStats web application (U.S. Geological Survey, 2021) incorporates the flow duration and annual *n*-day low-flow regression equations and provides streamflow estimates for most ungaged sites in the island. Streamflow estimates will not be available for sites on a stream considered to be regulated or affected by anthropogenic activities. The web application currently (2021) includes (1) a mapping tool to specify a location on a stream where flow statistics are needed; (2) a database that includes duration, *n*-day low-flow frequency statistics, hydrologic characteristics, location, and descriptive information for USGS streamflow-gaging stations used in this study; and (3) an automated GIS procedure that measures the required basin and climatic characteristics and solves the regression equations to estimate flow statistics for user-selected sites. For a more in-depth discussion of the capabilities of StreamStats and tutorials for common tasks, see the user manual provided in the help menu of the interface (U.S. Geological Survey, 2021).

Summary

Techniques for estimating flow-duration and low-flow frequency statistics for ungaged streams throughout Puerto Rico were developed in a cooperative study between the U.S. Geological Survey and the Puerto Rico Environmental Quality Board. Components of the study included computing flow statistics at 28 streamflow-gaging stations with 10 or more years of unregulated daily mean streamflow data through water year 2018, development of a StreamStats web application to compute basin and climatic characteristics at these stations, and developing regional regression equations to predict flow statistics at ungaged sites. Flow statistics and regression equations were developed using ordinary-least-squares and generalized-least-squares regression techniques for flow-duration quantiles at the 99th, 98th, 95th, 90th, 80th, 70th, 60th, and 50th percent exceedances and annual *n*-day low-flow frequency statistics for the 1-, 7-, 14-, and 30-day mean low flows with the 2-year (0.5 nonexceedance probability), 5-year (0.2 nonexceedance probability), and 10-year (0.1 nonexceedance probability) recurrence intervals. The final regression equations were chosen based on minimizing values of the residual standard error, the standard model error of variance, and the average standard error of prediction, maximizing values of the adjusted coefficient of determination and pseudo coefficient of determination, and examination of regression residuals. Basin and climatic characteristics determined to be significant explanatory variables in one or more regression equations included drainage area, mean total annual reference evapotranspiration, and minimum basin elevation. The Adj-R^2 for the flow-duration regression equations ranged from 57.7 to 81.4 percent. The Pseudo R^2 for the annual *n*-day low-flow regression equations ranged from 64.6 to 70.7 percent. The StreamStats web application incorporates the flow duration and annual *n*-day low-flow regression equations and can provide streamflow estimates for most ungaged sites on the island. StreamStats will not be available for sites along a stream identified as regulated or affected by anthropogenic activities.

References Cited

Alvarez, L., 2015, Water crisis brings out Puerto Rico's creative side: New York Times, July 14, 2015, accessed April 22, 2021, at https://www.nytimes.com/2015/07/15/us/in-drought-puerto-rico-rations-water-setting-off-a-collection-frenzy.html?_r=2.

Cobb, E.D., 1978, Estimates of 7-day, 10-year minimum flow at selected stream sites in Puerto Rico: U.S. Geological Survey Open-File Report 78–583, 47 p. [Also available at https://doi.org/10.3133/ofr78583.]

Colón-Dieppa, E., and Quinones-Aponte, V., 1985, Estimates of 7-day, 10-year low flow at ungaged streams in Puerto Rico: U.S. Geological Survey Water-Resources Investigations Report 84–4089, 2 p. [Also available at https://doi.org/10.3133/wri844089.]

Eng, K., Chen, 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://doi.org/10.3133/tm4A8.]

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.]

Flynn, K.M., Hummel, P.R., Lumb, A.M., and Kittle, J.L., Jr., 1995, User’s manual for ANNIE, Version 2, a computer program for interactive hydrologic data management: U.S. Geological Survey Water-Resources Investigation Report 95–4085, 216 p., accessed April 1, 2021, at https://water.usgs.gov/water-resources/legacy software/.

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 https://doi.org/10.3133/sir20045262.]

Gómez-Gómez, F., Rodríguez-Martínez, J., and Santiago, M., 2014, Hydrogeology of Puerto Rico and the outlying islands of Vieques, Culebra, and Mona: U.S. Geological Survey Scientific Investigations Map 3296, 40 p. plus 2 pls. [Also available at https://doi.org/10.3133/sim3296.]

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.

Howell, T.A., and Evett, S., 2004, The Penman-Monteith Method. U.S. Department of Agriculture, 17 p. [Also available at https://www.researchgate.net/publication/241492864_The_Penman-Monteith_Method.]

Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood-flow frequency, Bulletin 17B of the Hydrology Subcommittee, Office of Water Data Coordination: Reston, Va., U.S. Geological Survey, 183 p., accessed February 13, 2021, at https://water.usgs.gov/osw/bulletin17b/dl_flow.pdf.

Kiang, J.E., Flynn, K.M., Zhai, T., Hummel, P., and Granato, G., 2018, SWToolbox—A surface-water tool-box for statistical analysis of streamflow time series: U.S. Geological Survey Techniques and Methods, book 4, chap. A11, 33 p. [Also available at https://doi.org/10.3133/tm4A11.]

Kolb, K.R., and Ryan, P.J., 2019, Basin characteristics rasters for Puerto Rico StreamStats: U.S. Geological Survey data release, accessed January 22, 2020, at https://doi.org/10.5066/P9HK9SSQ.

Lins, H.F., and Cohn, T.A., 2011, Stationarity—Wanted dead or alive?: Journal of the American Water Resources Association, v. 47, no. 3, p. 475–480. [Also available at https://doi.org/10.1111/j.1752-1688.2011.00542.x.]

Mallows, C.L., 1973, Some comments on CP: Technometrics, v. 15, no. 4, p. 661–675, accessed May 12, 2020, at https://www.jstor.org/stable/1267380.

Miller, J.A., Whitehead, R.L., Oki, D.S., Gingerich, S.B., and Olcott, P.G., 1999, Ground water atlas of the United States—Segment 13, Alaska, Hawaii, Puerto Rico, and the U.S. Virgin Islands: Hydrologic Atlas 730–N, 36 p. [Also available at https://doi.org/10.3133/ha730N.]

Multi-Resolution Land Characteristics Consortium, 2018, National Land Cover Database 2011 (NLCD 2011). Multi-Resolution Land Characteristics Consortium (MRLC), accessed June 4, 2021, at https://data.nal.usda.gov/dataset/national-land-cover-database-2011-nlcd-2011.

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 https://doi.org/10.3133/sir20045033.]

Quinones, F., Colon-Dieppa, E., and Juarbe, M., 1984, Flow duration at streamflow gaging stations in Puerto Rico: U.S. Geological Survey Open-File Report 84–127, 93 p. [Also available at https://doi.org/10.3133/ofr84127.]

R Core Team, 2019, R—A language and environment for statistical computing: R Foundation for Statistical Computing, Vienna, Austria, accessed February 11, 2021, at https://www.R-project.org/.

Ramos-Ginés, O., 1999, Estimation of magnitude and frequency of floods for streams in Puerto Rico—New empirical models: U.S. Geological Survey Water-Resources Investigation Report 99–4142, 41 p. [Also available at https://doi.org/10.3133/wri994142.]

Ries, K.G., III, Newson, J.K., Smith, M.J., Guthrie, J.D., Steeves, P.A., Haluska, T.L., Kolb, K.R., Thompson, R.F., Santoro, R.D., and Vraga, H.W., 2017, StreamStats, version 4: U.S. Geological Survey Fact 2017–3046, 4 p. [Also available at https://doi.org/10.3133/fs20173046.]

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 https://doi.org/10.3133/twri04B1.]

Santiago-Rivera, L., 1992, Low-flow characteristics at selected sites on stream in eastern Puerto Rico: U.S. Geological Survey Water-Resources Investigations Report 92–4063, 55 p. [Also available at https://doi.org/10.3133/wri924063.]

Santiago-Rivera, L., 1995, Low-flow characteristics at selected sites on stream in southern and western Puerto Rico: U.S. Geological Survey Water-Resources Investigations Report 95–4147, 71 p. [Also available at https://doi.org/10.3133/wri954147.]

Santiago-Rivera, L., 1998, Low-flow characteristics at selected sites on stream in northern and central Puerto Rico: U.S. Geological Survey Water-Resources Investigations Report 98–4200, 62 p. [Also available at https://doi.org/10.3133/wri984200.]

Sauer, V.B., 1974, Techniques for calculating magnitude and frequency of floods in Oklahoma, with compilations of flood data through 1971: U.S. Geological Survey Water-Resources Investigations Report 73–52, 24 p. [Also available at https://doi.org/10.3133/wri7352.]

Schwarz, G., 1978, Estimating the dimension of a model: Annals of Statistics, v. 6, no. 2, p. 461–464, accessed May 12, 2020, at https://doi.org/10.1214/aos/1176344136.

Stedinger, J.R., and Tasker, G.D., 1985, Regional hydrologic analysis 1—Ordinary, weighted, and generalized least squares compared: Water Resources Research, v. 21, no. 9, p. 1421–1432. [Also available at https://doi.org/10.1029/WR021i009p01421.]

Tasker, G.D., and Stedinger, J.R., 1989, An operational GLS model for hydrologic regression: Journal of Hydrology (Amsterdam), v. 111, no. 1–4, p. 361–375. [Also available at https://doi.org/10.1016/0022-1694(89)90268-0.]

Theil, H., 1958, Economic forecasts and policy: Amsterdam, North-Holland Publishing Co., 562 p.

U.S. Department of Agriculture, 1986, Urban hydrology for small watersheds: U.S. Department of Agriculture, Natural Resources Conservation Service, Technical Release 55, 164 p. [Also available at https://www.wcc.nrcs.usda.gov/ftpref/wntsc/H&H/other/TR55documentation.pdf.]

U.S. Fish and Wildlife Service, 2021, National Wetlands Inventory website: Washington, D.C., U.S. Department of the Interior, Fish and Wildlife Service, accessed June 9, 2021, at https://www.fws.gov/wetlands/.

U.S. Geological Survey, 2019, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed June 6, 2019, at https://doi.org/10.5066/F7P55KJN.

U.S. Geological Survey, 2020, 1/3 arc-second digital elevation models (DEMs)—USGS National Map 3DEP downloadable data collection: U.S. Geological Survey digital data, accessed August 12, 2020, at https://apps.nationalmap.gov/downloader/#/.

U.S. Geological Survey, 2021, USGS StreamStats: U.S. Geological Survey digital data, accessed February 10, 2021, at https://streamstats.usgs.gov/ss/.

Williams-Sether, T., 2021, Data files for the development of regression equations for flow-duration statistics and n-day low-flow frequencies for ungaged streams in Puerto Rico through water year 2018: U.S. Geological Survey data release, https://doi.org/10.5066/P9Y2QVJ6.

For more information about this publication, contact:

Director, USGS Dakota Water Science Center
821 East Interstate Avenue, Bismarck, ND 58503
1608 Mountain View Road, Rapid City, SD 57702
605–394–3200

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

Publishing support provided by the
Rolla and Lafayette Publishing Service Centers

Williams-Sether—**Estimating Flow-Duration Statistics and Low-Flow Frequencies and Implementation of a StreamStats Web-Based Tool**—SIR 2021–5054

ISSN 2328-0328 (online)
https://doi.org/10.3133/sir20215054