

Prepared in cooperation with the U.S. Environmental Protection Agency

SWToolbox: A Surface-Water Toolbox for Statistical Analysis of Streamflow Time Series

Chapter 11 of
Section A, Statistical Analysis
Book 4, Hydrologic Analysis and Interpretation

Techniques and Methods 4–A11

SWToolbox: A Surface-Water Toolbox for Statistical Analysis of Streamflow Time Series

By Julie E. Kiang, Kathleen M. Flynn, Tong Zhai, Paul Hummel, and
Gregory Granato

Chapter 11 of
Section A, Statistical Analysis
Book 4, Hydrologic Analysis and Interpretation

Prepared in cooperation with the U.S. Environmental Protection Agency

Techniques and Methods 4–A11

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

U.S. Department of the Interior

RYAN K. ZINKE, Secretary

U.S. Geological Survey

William H. Werkheiser, Deputy Director

exercising the authority of the Director

U.S. Geological Survey, Reston, Virginia: 2018

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>.

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

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

Suggested citation:

Kiang, J.E., Flynn, K.M., Zhai, Tong, Hummel, Paul, and Granato, Gregory, 2018, SWToolbox: A surface-water toolbox for statistical analysis of streamflow time series: U.S. Geological Survey Techniques and Methods, book 4, chap. A–11, 33 p., <https://doi.org/10.3133/tm4A11>.

Acknowledgments

This software was developed in cooperation with the U.S. Environmental Protection Agency (EPA). The EPA support of this effort is much appreciated. The software was built using computational code originally written for the EPA DFLOW program and the USGS SWSTAT program. The interface is built on the EPA BASINS environment. The EPA helped to set up testing sessions and to identify software reviewers, whose resulting comments helped to improve the user interface.

This report was also much improved as the result of suggestions from USGS colleague reviewers, Paul Barlow and Kernell Ries.

Preface

The software program described in this report, SWToolbox, can be downloaded from the link provided at <https://doi.org/10.3133/tm4-A11>.

The software runs on the Windows operating system and the R statistics program must also be installed in order to use a few specific functions.

Although this computer program has been written and used by the U.S. Geological Survey (USGS), no warranty, expressed or implied, is made by the USGS or the U.S. Government as to the accuracy and functionality of the program and related program material, nor shall the fact of distribution constitute any such warranty, and no responsibility is assumed by the USGS in connection therewith. Users are requested to notify the USGS of any errors found in this report or with the computer program.

Contents

Acknowledgments	iii
Preface	iii
Abstract	1
Overview.....	1
Statistical Analysis Methods	2
Frequency Analysis	2
Computing an n-day Annual Time Series.....	2
Year, Season, or Month.....	2
Year Associated With an n-day Annual Value	2
Missing Record	2
Days Used to Compute an Annual n-day Value	2
N-day Frequency Analysis	3
Screening Tests for Trends, Distribution Fit, and Outliers.....	4
Trend Tests.....	4
Distribution Fit Tests	4
Outlier Test.....	4
Biologically Based and Nonbiologically Based Design Flows.....	5
Duration Analysis.....	5
Duration Option	5
Compare Option	6
Hydrograph Option	6
Using SWToolbox: Getting Started.....	7
Building, Saving, and Reopening a Project	7
Time-Series Data Discovery and Retrieval.....	12
Bypassing the Map Interface	15
Using SWToolbox: Analysis Menu Tools.....	16
Selecting Time-Series Data.....	16
Summarizing, Listing, and Graphing Time-Series Data	18
Data Tree	18
List	18
Graph.....	18
Graph Menus.....	18
Integrated Design Flow.....	22
Select Dates Tab	22
N-Day, Trend, Frequency Tab.....	22
N-Day Low Annual Time Series and Ranking List	22
Trend of Low Annual Time Series and Statistics	22
Screening Tests.....	22
Frequency Statistics Grid	26
Frequency Report	26
Frequency Graph	26
Design Flow Tab	26
Outliers Tab.....	29
Duration/Compare Analysis	29
Duration Hydrograph Analysis	30
Summary.....	33
References Cited.....	33

Figures

1–7.	Screen captures showing the—	
1.	“Welcome to the USGS SWToolbox 1.0” and “USGS SWToolbox 1.0” dialog boxes that appear on first use of the Surface Water Toolbox (SWToolbox)	7
2.	Initial map with data layers active that show States and hydrologic cataloging units in the Surface Water Toolbox	8
3.	Method for adding labels to a map layer in the Surface Water Toolbox. In this example, cataloging-unit numbers are added to the Cataloging Units layer. Options available by right-clicking on the selected layer, “Cataloging Units.” Label-style dialog box that appears after selecting the Label Setup option. The resulting map, with cataloging-unit labels added.....	9
4.	Map interface showing the river network and the streamflow-discharge stations in the basin of interest and surrounding area in the Surface Water Toolbox	12
5.	Steps used to select one or more data sites for which data can be retrieved from the U.S. Geological Survey National Water Information System for use in the Surface Water Toolbox. In this example, streamgages on the Pawcatuck River, Rhode Island, are selected	13
6.	Dialog boxes used to retrieve data from the U.S. Geological Survey National Water Information System and other sources for use in the Surface Water Toolbox.....	14
7.	Data can be downloaded from a list of station identifiers by selecting this menu option from the Download Data window in the Surface Water Toolbox	15
8–28.	Screen capture of—	
8.	Data selection window in the Surface Water Toolbox.....	17
9.	Filter Data form in the Surface Water Toolbox	17
10.	Sample Data Tree form in the Surface Water Toolbox.....	19
11.	Sample Time-Series List form in the SWToolbox	20
12.	Graph selection form in the SWToolbox	20
13.	Example of time-series plot in the Surface Water Toolbox	21
14.	Sample Edit Time-Series Graph form in the Surface Water Toolbox.....	21
15.	Select Dates tab in Integrated Design Flow window in the Surface Water Toolbox	23
16.	N-day, Trend, Frequency tab in Integrated Design Flow window in the Surface Water Toolbox	24
17.	Example of n-day time-series list in the Surface Water Toolbox. In this example, the 7-day annual low flows have been generated.....	25
18.	Example of trend list in the Surface Water Toolbox.....	26
19.	Example of screening-test results in the Surface Water Toolbox.....	26
20.	Example of Frequency Statistics Grid in the Surface Water Toolbox.....	27
21.	Example of graph in Frequency Graph in the Surface Water Toolbox.....	27
22.	Design Flow tab in Integrated Design Flow window in the Surface Water Toolbox	28
23.	Integrated Design Flow results form in the Surface Water Toolbox.....	28
24.	Outlier tab in Integrated Design Flow window in the Surface Water Toolbox	30

25. Duration/Compare Analysis window in the Surface Water Toolbox.....	31
26. Example of a Compare graph in the Surface Water Toolbox	31
27. Window used to select percent exceedances for graphing or reporting in the Surface Water Toolbox	32
28. Example of duration hydrograph in the Surface Water Toolbox.....	32

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
	Flow rate	
cubic foot per second (ft ³ /s)	0.02832	cubic meter per second (m ³ /s)

Abbreviations

7Q10	7-day minimum flow that occurs on average only once every 10 years
BASINS	Better Assessment Science Integrating point & Non-point Sources software (EPA)
EMA	Expected Moments Algorithm
EPA	U.S. Environmental Protection Agency
HUC	hydrologic unit code
K	frequency factor
LP3	log-Pearson Type III
NWIS	National Water Information System database (USGS)
SWToolbox	Surface-Water Toolbox computer program (USGS)
USGS	U.S. Geological Survey

SWToolbox: A Surface-Water Toolbox for Statistical Analysis of Streamflow Time Series

By Julie Kiang,¹ Kate Flynn,¹ Tong Zhai,² Paul Hummel,² and Greg Granato¹

Abstract

This report is a user guide for the low-flow analysis methods provided with version 1.0 of the Surface Water Toolbox (SWToolbox) computer program. The software combines functionality from two software programs—U.S. Geological Survey (USGS) SWSTAT and U.S. Environmental Protection Agency (EPA) DFLOW. Both of these programs have been used primarily for computation of critical low-flow statistics. The main analysis methods are the computation of hydrologic frequency statistics such as the 7-day minimum flow that occurs on average only once every 10 years (7Q10), computation of design flows including biologically based flows, and computation of flow-duration curves and duration hydrographs. Other annual, monthly, and seasonal statistics can also be computed. The interface facilitates retrieval of streamflow discharge data from the USGS National Water Information System and outputs text reports for a record of the analysis. Tools for graphing data and screening tests are available to assist the analyst in conducting the analysis.

Overview

The Surface Water Toolbox (SWToolbox) is a Windows-based desktop computer application that provides a set of procedures for statistical analysis of streamflow time series. In particular, it computes n-day frequency statistics, such as the 7Q10—the 7-day minimum flow that occurs on average only once every 10 years. It also computes biologically based design flows, flow-duration statistics, and a variety of annual, monthly, and seasonal statistics. As an aid to analysis, the program also facilitates download and plotting of streamflow time series from the U.S. Geological Survey (USGS) National Water Information System (NWIS), and performs basic tests on the data such as trend tests.

The program is the result of a joint effort between the USGS and the U.S. Environmental Protection Agency (EPA) and replaces the USGS SWSTAT program (<https://water.usgs.gov/software/SWSTAT/>) and the EPA DFLOW program (<https://www.epa.gov/waterdata/dflow>)

(U.S. Environmental Protection Agency [EPA], 1986). SWToolbox was built using the EPA BASINS platform (<https://www.epa.gov/exposure-assessment-models/basins>) (U.S. Environmental Protection Agency, 2015) to facilitate data retrieval. The BASINS platform itself is built upon the open-source MapWindow spatial data viewer and geographic information system (<http://www.mapwindow.org/>). This manual includes basic information on setting up a project within SWToolbox; however, for additional help with using the interface, the user is referred to the online help system, which provides additional help as developed for BASINS. Users familiar with the USGS Groundwater (GW) Toolbox (Barlow and others, 2014) will find that the processes for setting up a project and downloading data are identical. This manual focuses on those functions that are specific to the USGS SWToolbox “Analysis” menu. These functions include the following statistical capabilities to support water-quantity and water-quality modeling:

- List time-series data.
- Graph time-series data, flow-duration curves, flow-frequency curves, running sums, as well as residuals, cumulative differences, and scatter plots of two time series.
- Compute an n-day high or low annual time series from a daily time series,
- Perform frequency analysis of any annual time series using the log-Pearson Type III (LP3) distribution.
- Compute biologically based design flows using annual time-series data.
- Test for monotonic trend in annual time series.
- Perform flow-duration analysis.
- Compute duration hydrograph tables and curves.
- Compute a variety of annual, seasonal, and monthly statistics, such as the mean, minimum, and maximum value, for a selected time series.

Installation instructions are available with the download package. The R statistical program (R) must also be installed to utilize the R screening tests. Other aspects of the SWToolbox will function properly without R.

¹U.S. Geological Survey.

²RESPEC Consulting and Services.

Statistical Analysis Methods

This section describes the statistical analysis methods that are included in the SWToolbox. The methods were developed for the analysis of mean daily discharges and annual low and high flows derived from mean daily discharges. Other types of daily and annual time series can be analyzed, but the methods may not support the desired results. For example:

- For flood-frequency analysis, the USGS PeakFQ program (<https://water.usgs.gov/software/PeakFQ/>) (Veilleux and others, 2014) based on Bulletin 17C (England and others, 2018) is recommended.
- Streamflow records with (valid) negative values cannot be analyzed with the SWToolbox.
- Tidal records and annual time series derived from tidal records should not be analyzed with the SWToolbox.

Frequency Analysis

The 7Q10 is a low-flow statistic that is commonly used to regulate wastewater discharges to surface waters. The 7Q10 represents the minimum consecutive 7-day average flow that can be expected to occur no more than once every 10 years. In other words, the expected recurrence interval of that flow is 10 years. Variations of this statistic include the 7Q2, the 7-day minimum average flow with a recurrence interval of 2 years, or the 30Q2, the 30-day minimum average flow with a recurrence interval of 2 years. In fact, this can be generalized to any n-day average flow and any recurrence interval. Specific low-flow frequency statistics are denoted using the xQy nomenclature, where x is the number of days over which flows are averaged, and y is the recurrence interval in years.

The SWToolbox can compute maximum n-day annual average flows as well as minimum n-day flows. To calculate these statistics, the n-day time series is computed, it is fit to a statistical distribution, and the recurrence interval is estimated from that statistical distribution. Because a poor fit to the statistical distribution can result in a poor estimate of the n-day low-flow statistic, it is important that the analyst check each individual fit by examining a probability plot of calculated n-day flow values.

Computing an n-day Annual Time Series

An annual n-day time series is derived from the daily record of streamflow. Each n-day value is based on the average values over successive periods of n-consecutive days within the year or season being considered. An annual low-flow series consists of each of the lowest average n-day periods in a year or season for the period of record or period of analysis. An annual high-flow time series consists of each of the highest average n-day periods in a year or season over the full daily record.

Year, Season, or Month

Although full years of daily data are commonly used to determine each annual n-day value, annual time series may also be constructed from n-day values for seasons or months. Time series for full years generally are computed on the basis of the water year of October 1 through September 30 used when analyzing high flows and the climate year of April 1 through March 31 used when analyzing low flows. Generally, in the conterminous United States, streamflow values are the lowest during late summer and early fall and highest during the spring. So, using October 1 for the start of the high-flow year and April 1 for the start of the low-flow year will generally produce annual time series with each annual value independent of the adjacent years. It is good practice to confirm that this is the case. The user is not restricted to the water and calendar years for computations; a year may start on any calendar day and will be 365 days in length (366 in a leap year.)

A season may start on any calendar day and is any time period equal to or shorter than a full year. A month is a season that begins on the first day of a calendar month and ends on the last day of that month. Leap day is included for the month of February and for any season defined to include February 28 and March 1.

Year and month are specific cases of a season. For simplicity, the term season will be used in describing the n-day option.

Year Associated With an n-day Annual Value

To be consistent over all possible seasonal variations, the year associated with a computed n-day annual value is the calendar year associated with the last day of the season over which the n-day value is determined. This is the case for both low-flow and high-flow n-day values. It is independent of the number of days that actually fall in a calendar year.

For example, if the season November 1–January 31 is used, the n-day value calculated from the period November 1, 1983–January 31, 1984, would be associated with calendar year 1984, even though 61 days from 1983 and 31 days from 1984 were used.

Missing Record

No annual n-day value will be computed in a season that contains any missing record. The number of individual missing time steps does not matter. It does not matter if low flows are being computed and the missing time step occurs during a period of high flow. For example, if annual 7-day low flows are being computed and there is a missing value in May, no n-day value will be computed for that year.

Days Used to Compute an Annual n-day Value

In all cases, any n-day length may be specified that is equal to or less than the length of the season being considered. Each n-day average is computed independently of the preceding and following n-day averages. The first n daily values are summed

and divided by n . Then the second through $n+1$ daily values are summed and divided by n . Then the third through $n+2$ daily values are summed and divided by n and so on to the end of the season. The lowest of all the average values is the n -day low-flow value and the highest of all the average values is the n -day high flow.

The SWToolbox is consistent in how it computes both high and low n -day values for time periods of year, season, or month. Only the days in the season specified are used to compute an annual n -day value. The program will not backtrack in time to include days from before the start of the season nor will it forward track in time to include days beyond the end of the season. For example, consider the 365-day low-flow value for the season January 1–December 31. For nonleap years, the 365-day low-flow value is the average of all 365 values. For leap years, the 365-day low-flow value is the lesser of the average of days 1 through 365 and the average of days 2–366. If the program were to compute a 365-day average for every day in the current year by backtracking into the previous year, the resulting 365-day low-flow value would not be statistically independent of the previous year.

As an example, when computing monthly low flows, 28, 29, 30, or 31 daily values will be considered. If the 20-day low flow is being calculated for February, the average values for the following periods would be calculated:

- February 1–20
- February 2–21
- February 3–22
- February 4–23
- February 5–24
- February 6–25
- February 7–26
- February 8–27
- February 9–28
- February 10–29 (leap year only)

Within the SWToolbox, requesting an n -day greater than the length of the time period results in that n -day computation being dropped. For example, if the user specifies that the 1-, 7-, 14-, and 30-day low flows are to be calculated for all months, the 30-day low flow will be dropped from the calculations for February and all subsequent months. In this case, one would need to perform two analyses—January–February and March–December. Similar changes would be required if the user specified 31-day values. Unfortunately, there is no way to get the 29 days for February.

N-day Frequency Analysis

The computed n -day flows are fit to a log-Pearson Type III (LP3) distribution. This distribution is most commonly used to fit an n -day minimum time series and is described in more detail in Riggs (1972). For peak-flow frequency, the LP3 distribution is also commonly used in the United States, but the analysis typically employs instantaneous maxima rather than

daily maxima and follows guidelines specified in Bulletin 17C (England and others, 2018).

When used for low flows, the n -day frequency analysis is used to estimate the probability that flow will not exceed different flow levels in any given year. For example, the 7Q10 is the 7-day minimum flow that on average is not exceeded only once every 10 years. The remainder of this section discusses computation of an n -day minimum series and nonexceedance probabilities, but analogous computations also can be made using n -day maximum series to compute exceedance probabilities.

To fit the LP3 distribution, the first three moments are computed from the log-transformed n -day time series. It is conventional to use the base-10 logarithm, so that

$$X = \log_{10}(Q)$$

where X is the log-transformed time series that will be used to compute the mean (\bar{X}), standard deviation (S), and skew (G), and Q is the n -day time series of flow. These three moments define the LP3 distribution. Individual points along the frequency curve can be computed as

$$\hat{X} = \bar{X} + KS$$

where \hat{X} is the computed estimate and K is a frequency factor, calculated as a function of the skew, G , and the nonexceedance probability of interest.

When there are zeros in the n -day series, it is impossible to make this log transformation. In these cases, the LP3 distribution is fit to the nonzero members of the time series, resulting in nonexceedance probabilities conditioned on flow being nonzero. A conditional-probability adjustment is made to account for the zeros in the record and compute an unconditional nonexceedance probability that includes the possibility of zero flows. The conditional-probability adjustment uses the following computation to determine the unconditional probability that the n -day flow, Q , is less than any specified flow, Q_T , in any particular year:

$$P(Q < Q_T) = P(Q < Q_T | Q = 0) P(Q = 0) + P(Q < Q_T | Q > 0) P(Q > 0)$$

Because $P(Q < Q_T | Q = 0) = 1$, this reduces to

$$P(Q < Q_T) = P(Q = 0) + P(Q < Q_T | Q > 0) P(Q > 0)$$

These unconditional probabilities are then used to define the frequency curve.

While the SWToolbox allows an n -day frequency curve to be computed without first applying a logarithmic transformation to the n -day time series, this is not considered standard practice because streamflow time series typically will not adequately fit the Pearson Type 3 distribution unless the transformation is applied.

4 SWToolbox: A Surface-Water Toolbox for Statistical Analysis of Streamflow Time Series

Confidence intervals for the frequency curve are computed by borrowing from the peak-flow literature. Confidence intervals in the SWToolbox are computed using the confidence intervals for the Expected Moments Algorithm (EMA) for fitting the LP3 distribution (Cohn and others, 2001). As the EMA results are exactly the same as those for the simple method of moments outlined above when time series do not include unusual data types (for example, values of zero flow or flexible representations of data allowed by EMA), the confidence intervals are equivalent and appropriate for this use. However, confidence intervals are only available when the conditional-probability adjustment is **not** employed.

Screening Tests for Trends, Distribution Fit, and Outliers

Computation of accurate n-day frequency statistics requires that the data are stationary and that the LP3 distribution is a reasonable fit to the data. The following tests are implemented in the SWToolbox to assist the analyst in making these assessments. They are meant as screening tools only and should not be considered as comprehensive assessments of adherence to underlying assumptions. In addition to evaluating the screening-test results, each time series and the fitted frequency curve should be visually inspected to check for obvious problems. Additional inspection and checking of a time series is warranted whenever possible issues are identified.

If the skew of the logarithms of the n-day flows are equal to zero, then the n-day values will plot on a straight line on a graph with logarithmic and probability scales and the LP3 will simplify to the lognormal. If the sample of values has a positive skew, it may be fit with an LP3 distribution and the n-day values will plot in a concave-up pattern. Similarly, if the sample of values has a negative skew, it may be fit with a LP3 distribution and the n-day values will plot in a concave-down pattern on such a graph (Riggs, 1968). Obvious and substantial departures of the data from the fitted distribution may indicate a problem with the fit.

Trend Tests

The Mann-Kendall test is used to test for monotonic trends in data series (for additional information on this test, see, for example, Helsel and Hirsch, 2002). In the SWToolbox, Kendall's tau statistic is computed for the n-day time series, and it is used to test for monotonic trends in time. The Kendall tau test statistic is reported along with the p-value and the Sen slope of the trend line. A site should be flagged for further screening when the test results in a p-value of 0.05 or less.

The Spearman Rank Correlation test is used to test whether or not two ordered data series are highly correlated by computing the correlation coefficient of the ranks of values of the data series. (The algorithms are based on methods described by Fisher, 1924; Haan, 1977; Caruso and Cliff,

1997; and Granato, 2014; additional information is available in statistical texts such as Helsel and Hirsch, 2002). In the SWToolbox, the test is used as a second screening test for monotonic trend by testing the correlation between the n-day series and year. A significance level of 0.05 or less indicates that the site should be flagged for further screening.

Distribution Fit Tests

The probability plot correlation coefficient (PPCC) test estimates the probability that input statistics fit a specified distribution (see, for example, Helsel and Hirsch, 2002). In the SWToolbox, the PPCC is used to test whether the n-day series conforms to an LP3 distribution (or Pearson Type III if the logarithmic transformation is not specified). Test statistics for the PPCC tests were derived by using information from Kirby (1972), Vogel and Kroll (1989), and Vogel and McMartin (1991). The SWToolbox applies the test only to the nonzero values that are fit to the distribution. A frequency curve is flagged if the input data do not fit within a given probability (95 percent).

The Kolmogorov-Smirnov test evaluates the probability that input data fit a specified distribution (see, for example, Haan, 1977; Press and others, 1992; and Helsel and Hirsch, 2002). In the SWToolbox, this test is applied to nonzero n-day values, and tests whether the data fit an LP3 distribution (or the Pearson Type III distribution, if the logarithmic transform was not selected) by using methods developed by Kirby (1972) for skewed distributions. A frequency curve fit will be flagged if the input data do not fit within a given probability (95 percent).

Outlier Test

The K-Ratio Outliers test identifies n-day values that may not fit the selected distribution. This test is conducted in iterative fashion, with one n-day value left out of each iteration to test whether it appears to be an outlier as compared to the statistics for the remaining n-day values. When the K-Ratio test statistic is greater than one, the data point is flagged as one that may be an outlier with respect to the distribution fit.

The K-Ratio Outliers test is also applied to identify n-day frequency statistics for a given streamgage that are substantially different from the same statistics calculated for nearby streamgages. This test uses the ratios of Pearson of LP3 frequency factors for nearby streamgages. If the frequency factor (K) of the n-day statistics for the selected streamgage is greater than the maximum expected frequency factor, given the average, standard deviation, skew of the n-day statistics, and number of nearby streamgages, then the n-day value for the streamgage being tested is flagged as a potential outlier. Nearby streamgages are selected by using the great-circle distance among the latitude-longitude coordinates of all selected streamgages (Sinnott, 1984). When the K-Ratio test statistic is greater than one, this indicates that the n-day statistic is outside the expected maximum outliers, given the sample statistics of the n-day values for the nearby

streamgages; in this case, the n-day value for the selected streamgage is flagged as one that may be an outlier, with respect to the distribution fit of the other streamgages.

Biologically Based and Nonbiologically Based Design Flows

The EPA uses biologically and nonbiologically based design flows as water-quality criteria to protect aquatic life (EPA, 1986). The biologically based flows are determined by computing the frequency of excursions below the design-flow threshold. A user may select specific averaging periods and frequency of excursions. For computation of biologically based flows, the averages are computed as harmonic means (see below). The frequency of excursions is specified in terms of the average number of years between excursions. Specific biologically based flows are designated as xBy flows. For example, the 1B3 is computed such that the 1-day average flow dips below the design flow on average no more than once every 3 years. The most commonly used biologically based design flows are the 1B3 (criterion maximum concentration, acute), the 4B3 (criterion continuous concentration, chronic), and the 30B3 (used for ammonia).

Because multiple excursions may happen within a year or season, the low-flow period used for computation of excursions must also be specified. This “excursion clustering period” is typically set at 120 days. Computation of the biologically based design flows is an iterative procedure that continues until the solution converges. For additional information on how biologically based flows are computed, see U.S. Environmental Protection Agency (1986).

Hydrologically based design flows include n-day frequency statistics. Other possible design flows include the harmonic mean—the reciprocal of the mean of the reciprocals of the individual values. The harmonic mean is calculated for short-duration averages during the computation of biologically based design flows, but the harmonic mean can also be computed for the entire time series. For this computation of harmonic mean, only nonzero values are used. Other statistics that may be of interest include a particular percentile of flow or a specific flow value. The average frequency of excursions below these flow values can also be computed.

Duration Analysis

The flow-duration curve shows that the percent of time-specific discharges are equaled or exceeded during an analysis period (Searcy, 1959). In other words, it is a curve showing different percentiles of flow.

Three variations on duration analysis are available in the SWToolbox. The “**Duration**” option provides the traditional flow duration described by Searcy (1959). The “**Compare**” option builds on duration analysis to compare two time series. The “**Hydrograph**” option uses duration analysis to look at each day over a period of analysis.

While developed for analysis of mean daily discharges, the “**Duration**” options may be useful for analyzing other types of time-series data. For convenience, the time-series data may be referred to as flow or discharge in these discussions.

Duration Option

The “**Duration**” option is used to determine the percentage of time streamflow values within a time series are exceeded. For example, the smallest flow in a time series of 100 days was exceeded 99 percent of the time during the period, as 99 of the 100 values are larger. The following features should be considered in deciding whether or not this option is appropriate for the data being evaluated; a daily time step is used for the analysis:

- Data at a time step shorter than 1 day are aggregated to a daily time step. An aggregated daily value may be the average, total, minimum, or maximum of all the values in the day. The day is midnight to midnight, with the first time step beginning at midnight and the last time step ending at midnight. If data are missing for one or more time steps within a day, an aggregated value will not be determined and the daily value will be considered missing.
- Data at a time step greater than 1 day will not be disaggregated and cannot be analyzed.
- Values less than zero are not included in the analysis and are included in the count of missing values. The count for missing daily time steps is included in the output report. For this reason, this program should not be used to calculate n-day statistics for streamgages, such as those in tidal areas, that have legitimate negative flow values (for reversing flows).

To compute the duration curve for the graph, the time-series values are sorted from highest (largest) value to lowest (smallest) value. The exceedance probability, P_i , is computed for each data value using Weibull plotting positions:

$$P_i = (i/(n + 1))$$

where

- n = the total number of nonnegative values;
- i = order index of the flow values, with $i = 1$ for the highest value and $i = n$ for the lowest nonnegative value; and
- P_i = probability that the i th flow value will be equaled or exceeded.

If the probability is to be expressed as a percent rather than a fraction, then the plotting position is multiplied by 100. In the resulting graph, exceedance probabilities are plotted on the x-axis and flow is plotted on the y-axis.

Discussions of n-day flows are commonly expressed in terms of the return period. For example, the 7Q10 is nominally the annual 7-day 10-year low flow. However, this flow does

6 SWToolbox: A Surface-Water Toolbox for Statistical Analysis of Streamflow Time Series

not necessarily occur once every 10 years. Instead, the 10-year determination is based on the probability of occurrence, which is the reciprocal of the annual exceedance probability calculated by using the Weibull plotting position formula on the annual series ($1/P$). Therefore, the 10-year flow will not occur in every decade; on average, the probability of having a flow that is less than or equal to the 7Q10 is 10 percent in any given year. For the purposes of analysis, however, the exceedance probabilities are estimated from the fitted distribution rather than being interpolated between plotting-position values.

Compare Option

In the “**Compare**” option, a duration analysis is done on two time series. Mean absolute error, root mean square error, and bias are computed for each of the duration class intervals and a standard error of estimate is computed. This option produces a graph of the flow-duration curves and a tabular report of computed values. Common uses of the option include comparing the following:

- an observed or measured time series and a simulated time series,
- two time series to evaluate likeness, such as flow at adjacent basins, and
- split time periods for a single location, such as before and after urbanization.

The following features should be considered in deciding whether or not this option is appropriate for your data:

- Data may be analyzed at an annual, monthly, or daily time step. The two time series must be of equal length of time and at the same time step. If the time steps are not the same, one or both can be aggregated or disaggregated to the same time step.
- An aggregated value may be the average, total, minimum, or maximum of all the shorter time-step values that make up the analysis time step. When aggregating to a daily time step, the day is midnight to midnight, with the first time step beginning at midnight and the last time step ending at midnight. When aggregating to a time step shorter than daily, the midnight-midnight bound must be observed. If data are missing for one or more time steps within an aggregation period, no value will be determined and the aggregated value will be considered missing.
- The “same” and “divide” transformations are used to disaggregate to a smaller time step. With “same,” the smaller time-step values are the same as the greater time-step values. With “divide,” the greater time-step value is equally divided across the smaller time steps.

- If a time step is missing for one time series, the corresponding time step in the other time series is not included in the analysis.
- Values less than zero are not included in the analysis and are included in the count of missing values. The count for missing values is included in the output report.

To compute the duration curves for the graph, both time series are checked for missing values; if a value is missing in one time series, the corresponding time step in the other time series is not included in the analysis. For each time series, the values are sorted from highest (largest) value to lowest (smallest) value. As before, the exceedance probability, P , is computed for each data value using the Weibull plotting positions. The resulting graph has the duration curves for both time series.

Hydrograph Option

The “**Hydrograph**” option is used to determine the percentage of time that it is likely that some value will be exceeded on any given day over the period of analysis. This is similar to the “Duration” option, but, instead, duration curves and exceedance values are computed for each day of the season. The following features should be considered in deciding whether or not this option is appropriate for the data being evaluated:

- A daily time step is used for the analysis.
- Data at a time step shorter than 1 day are aggregated to a daily time step. An aggregated daily value may be the average, total, minimum, or maximum of all the values in the day. The day is midnight to midnight, with the first time step beginning at midnight and the last time step ending at midnight. If data are missing for one or more time steps within a day, an aggregated value will not be determined and the daily value will be considered missing.
- Data at a time step greater than 1 day will not be disaggregated and cannot be analyzed.
- Values less than zero are not included in the analysis and are included in the count of missing values. The count for missing daily time steps is included in the output report.
- Leap day is not included in the analysis.

For each day of the season, all values are retrieved for the period of record and then ranked. The ranked values are assigned using the Weibull plotting positions. The percentiles are then determined by linear interpolation. The default probabilities for the percentiles are 0.0 (maximum value), 0.1, 0.2, 0.3, 0.5, 0.7, 0.8, 0.9, and 1.0 (minimum value); these can be modified, with a maximum of up to 12 percentiles specified.

Note: It may be that there is an insufficient number of values to compute the percentiles for some or all days for some probabilities near 0.0 and (or) 1.00. Minimum and maximum probabilities are $1/(n+1)$ and $n/(n+1)$, respectively, where n is number of years with a value for a given day.

Using SWToolbox: Getting Started

The first steps in the use of the SWToolbox for any analyses are to build a SWToolbox project for the geographic area of interest, retrieve spatial and time-series datasets to populate the project, and review the time-series data. This section of the report provides instructions that can be used to complete these tasks. It is assumed that the SWToolbox has already been installed on the user's computer; information on how to download and install the program can be found at the website provided in the Preface to this report. Recall that the R screening tests will only provide results when R is installed on the user's computer. Other SWToolbox functions are not affected by whether or not R has been installed.

A number of text-formatting conventions are used throughout the report. Components of the SWToolbox interface, such as dialog boxes, panels, or buttons, are referred to in quotes, such as the "Preview Map" panel. Specific menu options and tools are referred to by bold text, such as the "**Open Project**" option in the "**File**" menu, which is available on the main toolbar.

Building, Saving, and Reopening a Project

A SWToolbox project is a collection of spatial and time-series data and analysis results for a specific geographic area. The project area can consist of a State, county, or HUC-8 hydrologic cataloging unit. A cataloging unit is one of the divisions of the hydrologic unit code (HUC) classification system of river basins in the United States (Seaber and others, 1987). Each cataloging unit is given an 8-digit code; for example, the Pawcatuck River at Wood River Junction, Rhode Island, USGS streamgage (station number 01117500), is within the Pawcatuck-Wood cataloging unit (01090005). (Cataloging units also are referred to as basins in this report.)

After installation, double-clicking on the "USGS SWToolbox" icon on the Windows desktop results in the two dialog boxes shown in figure 1: the "USGS SW Toolbox 1.0" and "Welcome to USGS SW Toolbox 1.0" dialog boxes. If the SWToolbox has not previously been used, the user should select the "**Build New Project**" option. Selecting this option produces a map of the United States showing the boundaries of States and cataloging units (fig. 2). The user can then zoom into an area of interest with the "Pan" and "Zoom" tools.

Labels can be added to identify the cataloging units in a zoomed-in view. This is done by right-clicking on the Cataloging Units layer, which results in several menu options (fig. 3A); in this example, the user selects the "**Label Setup**" option, which opens the "Label style" dialog box (fig. 3B).

The user double-clicks on the "CU" field and then selects the "Apply" check box. A label-generation dialog box then appears (not shown on fig. 3) that allows the user to select the position and orientation of the labels. Once these are selected, the cataloging-unit labels are added to the map (fig. 3C).

The "**Select**" tool can then be used to select the cataloging unit of interest, which will be highlighted in yellow. The user selects the "Build New Project" check box from the "Build New USGS SW Toolbox 1.0 Project" dialog box (fig. 1). At this point, the SW Toolbox will prompt the user with a "Save new project as..." dialog box to select a folder name and location in which to save the project. The default location in which to save the project is the C:\USGS-SWToolbox\data folder, and a subfolder is created for the project with the HUC-8 code as the default folder name (for example, 01090005). Project information is stored in a MapWindow project file in the project folder (.mwprj file) with the HUC-8 code in the default file name (01090005.mwprj). If multiple projects have been created for the HUC-8 area, a sequence number will be appended to the subfolder and project names (for example, 01090005-2). A "Projection Properties" dialog box will appear and the user can select the desired projection for the project data layers [the default "Albers Equal-Area (Conterminous U.S.)" option was selected for this example]. The project then will be populated with a set of standard data layers. As additional layers are added, it may be helpful to close the "Preview Map" panel by clicking on the "X" on the top right of the "Preview Map" panel or by toggling off the checkmark in the "**Preview Map**" section from the "**Panels**" option in the "**View**" menu. The resulting map and data layers are shown in figure 4.

The "Welcome to USGS SW Toolbox 1.0" dialog box can be reopened at any time using the "**Welcome Screen**" option from the "**Help**" menu. Also, as an alternative to the approach described in the previous paragraphs, a new project can be built by selecting the "**New**" option from the "**File**" menu. An existing project can be reopened from the "Welcome to USGS SW Toolbox 1.0" dialog box or with the "**Open Project**" or "**Open SWToolbox Project**" options from the "**File**" menu.

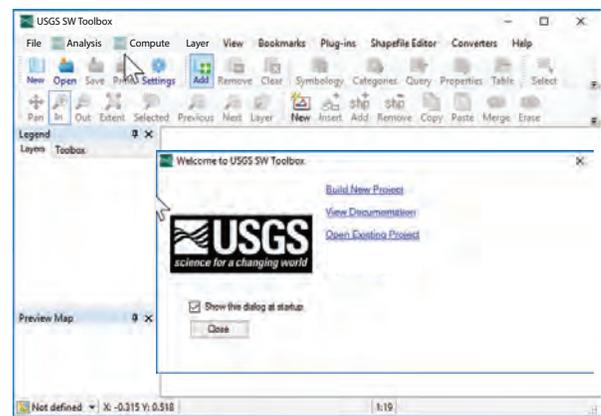


Figure 1. Screen captures showing the "Welcome to the USGS SWToolbox 1.0" and "USGS SWToolbox 1.0" dialog boxes that appear on first use of the Surface Water Toolbox (SWToolbox).

8 SWToolbox: A Surface-Water Toolbox for Statistical Analysis of Streamflow Time Series

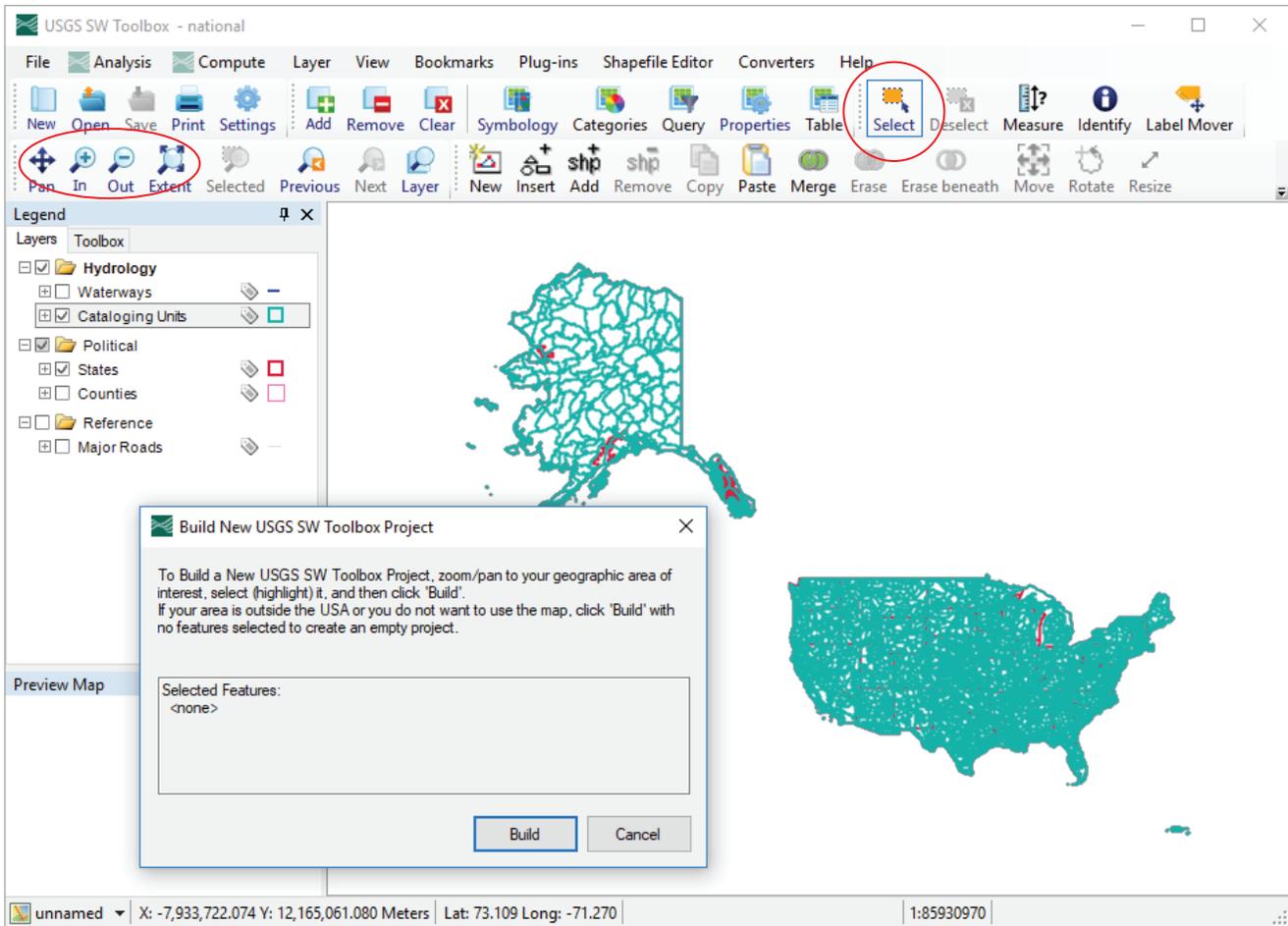


Figure 2. Screen captures showing the initial map with data layers active that show States and hydrologic cataloging units in the Surface Water Toolbox (SWToolbox). Pan, Zoom In, Zoom Out, and Select tools are identified with red circles.

A

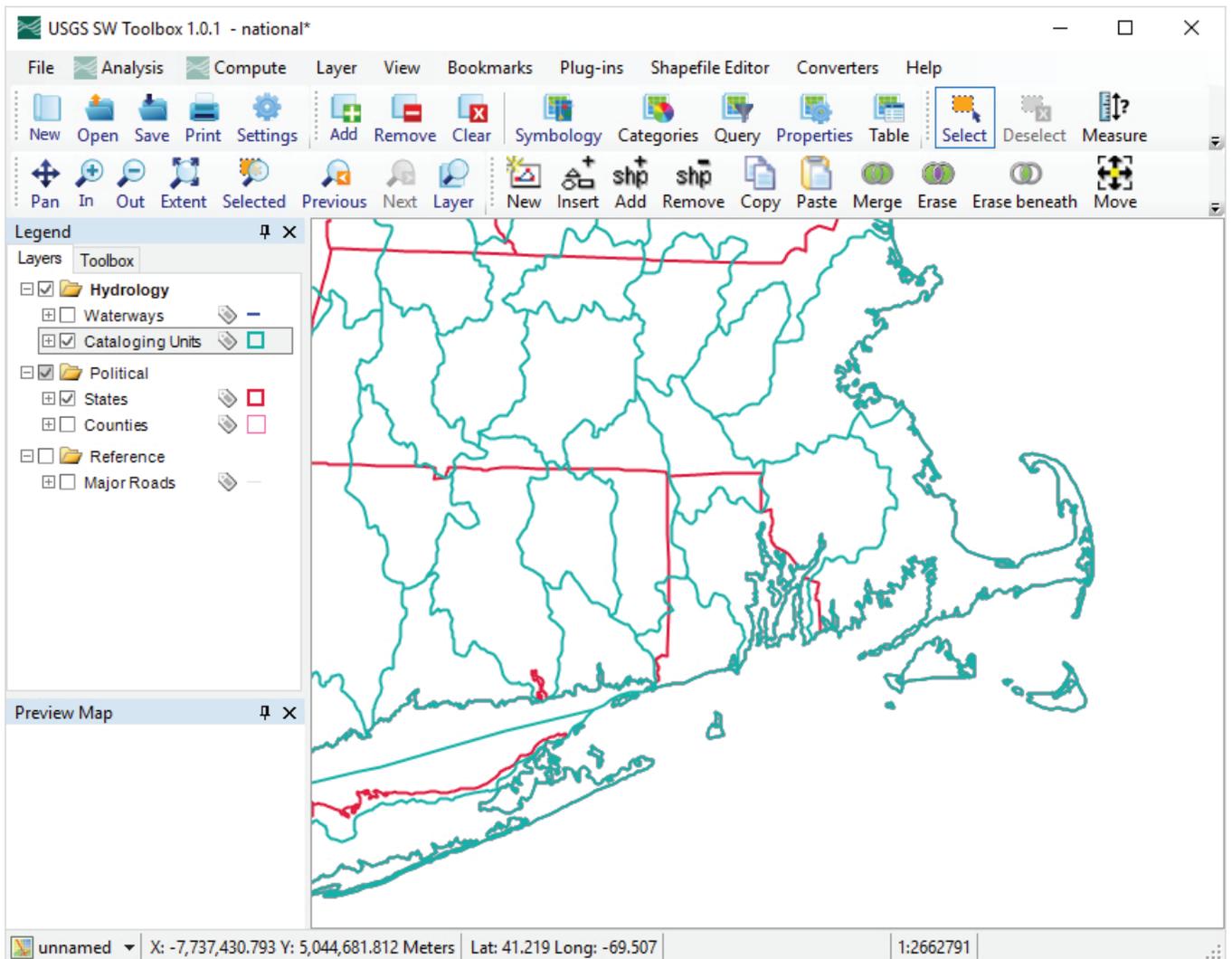


Figure 3. Screen captures showing the method for adding labels to a map layer in the Surface Water Toolbox (SWToolbox). In this example, cataloging-unit numbers are added to the Cataloging Units layer. *A*, Options available by right-clicking on the selected layer, "Cataloging Units." *B*, Label-style dialog box that appears after selecting the Label Setup option. *C*, The resulting map, with cataloging-unit labels added.

B

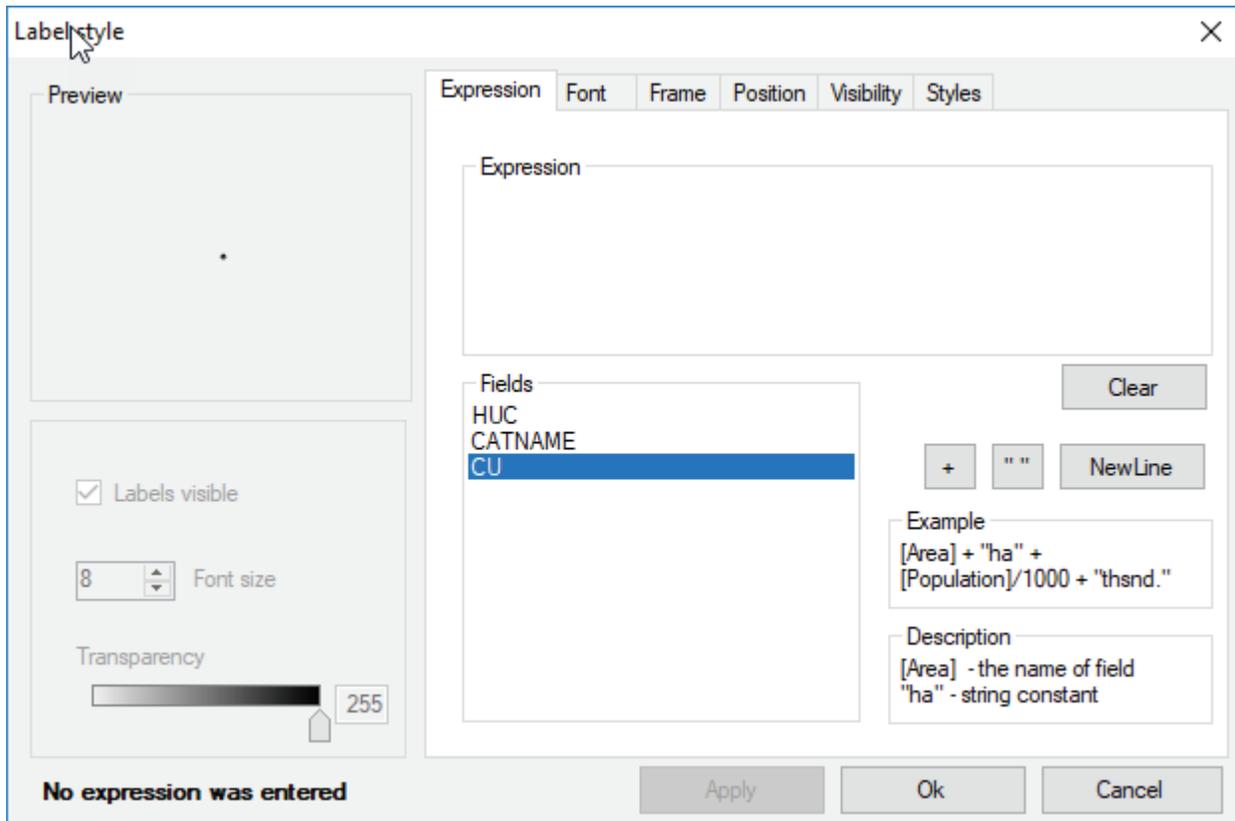


Figure 3. Screen captures showing the method for adding labels to a map layer in the Surface Water Toolbox (SWToolbox). In this example, cataloging-unit numbers are added to the Cataloging Units layer. *A*, Options available by right-clicking on the selected layer, “Cataloging Units.” *B*, Label-style dialog box that appears after selecting the Label Setup option. *C*, The resulting map, with cataloging-unit labels added.—Continued

C

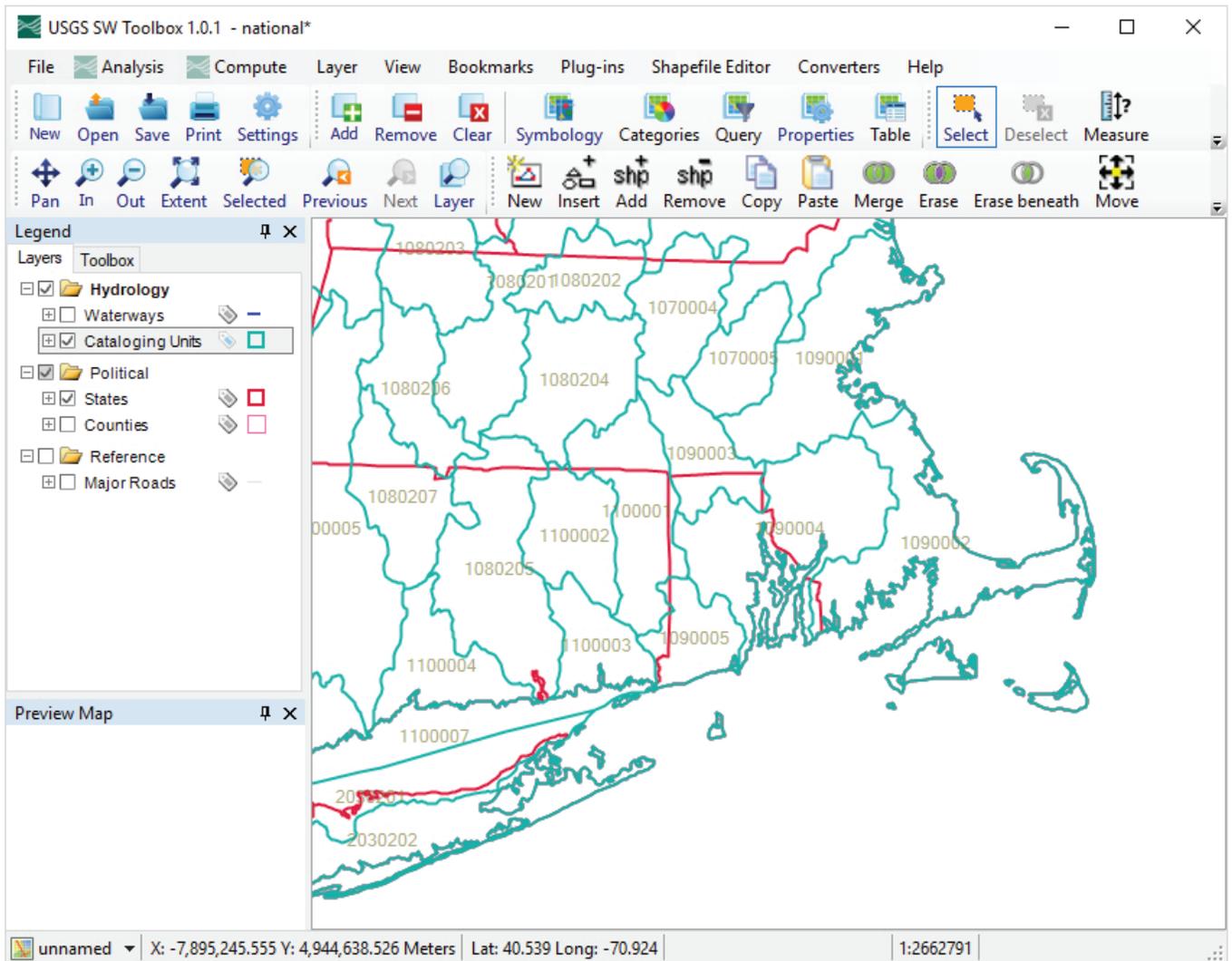


Figure 3. Screen captures showing the method for adding labels to a map layer in the Surface Water Toolbox (SWToolbox). In this example, cataloging-unit numbers are added to the Cataloging Units layer. *A*, Options available by right-clicking on the selected layer, “Cataloging Units.” *B*, Label-style dialog box that appears after selecting the Label Setup option. *C*, The resulting map, with cataloging-unit labels added.—Continued

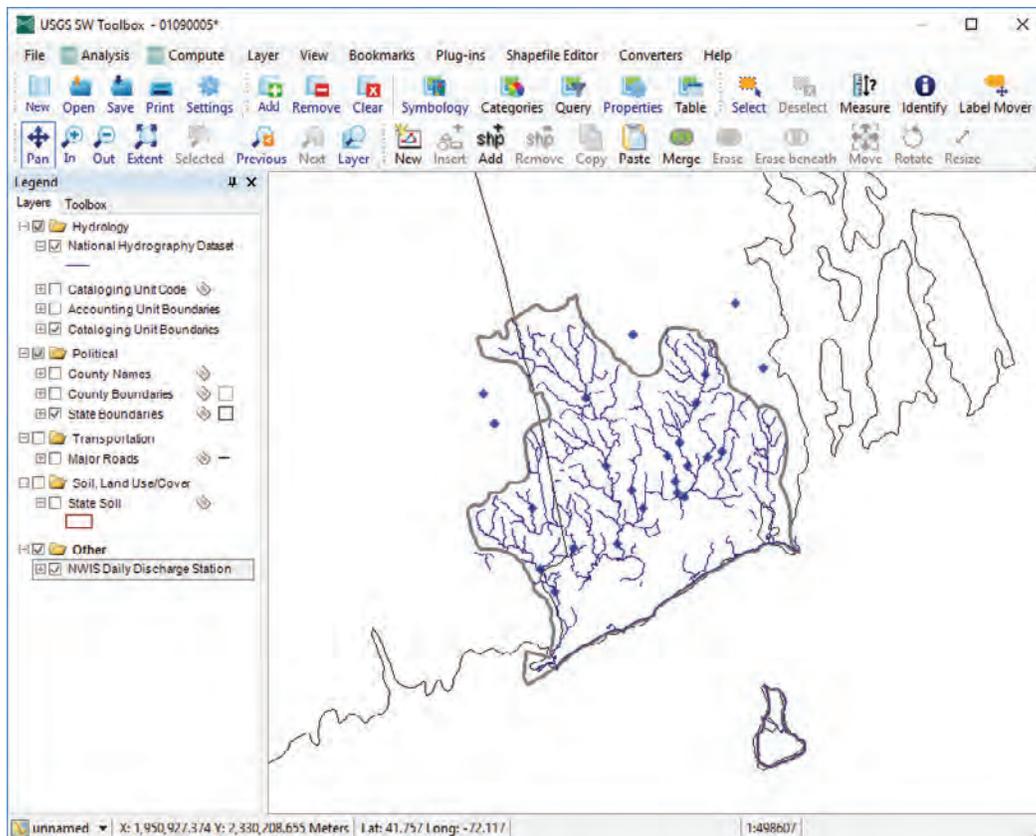


Figure 4. Screen capture of map interface showing the river network and the streamflow-discharge stations in the basin of interest and surrounding area in the Surface Water Toolbox (SWToolbox).

Time-Series Data Discovery and Retrieval

When a SWToolbox project is created, a large cache of map layers is downloaded and stored in the project folder. A subset of these layers, which is shown in the “Legend” panel (fig. 4), is made available as part of the initial build of the project. Data-collection sites for observed data are shown by default on the project map. These are sites at which a minimum of 10 data values of daily streamflow (discharge) are available through the NWIS database.

Time-series data may be discovered and retrieved interactively through use of the map interface, or by specifying station numbers individually or from a user-provided list. The By-Passing the Map Interface section below describes this second option. The first step in use of the map interface for data retrieval is to select specific data sites from the project map for which data will be downloaded from the NWIS. For this example, streamflow data will be retrieved for use in demonstrated analyses. The discharge-locations data layer shown on the map indicates that there are several streamgages within and near the Pawcatuck-Wood Basin (fig. 4). At this point, it is convenient to either add the streamgage numbers using the “**Label Setup**” option, as described previously

(fig. 3), or to import a data layer of the drainage-basin boundary for the streamgage of interest, if one is available. For this example, a shapefile of the drainage basin for the Pawcatuck River at Wood River Junction streamgage (01117500) was imported by use of the “**Add Layer**” option from the “**Layer**” menu. Streamgage identifiers also were added to the map (fig. 5). To select a streamgage, the user clicks on the data layer of interest (“NWIS Daily Discharge Stations”) and then clicks on the “**Select**” tool (fig. 5). The user can then select the streamgage of interest by clicking on it. One or more streamgages of interest can be selected at the same time by drawing a box around them with the “**Select**” tool (fig. 5) or by using the “**Ctrl**” key on the keyboard and “**Select**” tool to select multiple stations for download. The selected stations will then appear as grayish yellow on the map.

After a site is selected, the “**Download Data**” option from the “**File**” menu can be used to retrieve data from NWIS for the station (fig. 6A). As shown on figure 6A, there are several types of data that can be downloaded, but for this example the “**Daily Discharge**” button within the “**Data Values** from US Geological Survey National Water Information System” panel already has been selected because of the steps completed in the previous paragraph. The user then clicks the

“Download” button and data for the selected site are retrieved (there is a pause as the data are retrieved). The user is then informed that data for the selected site have been retrieved and saved in a file that uses the USGS NWIS RDB (relational database) time-series format (fig. 6B). The daily discharge data are reported in units of cubic feet per second (USGS NWIS parameter code 00060). The data are now part of the project and available for use in the SWToolbox.

As part of the data-discovery process, it can be advantageous to add other data layers to complement the default data layers shown on the project map. This can be done by toggling inactive data layers shown on the “Legend” (such as “Major Roads”) to active (by clicking on the check boxes next to each data layer) or by retrieving data layers that have been preprocessed for the SWToolbox and are available for download. These data layers are listed on the “Download Data” dialog box (fig. 6A). After the user selects a particular

data layer and then clicks on the “Download” button, there may be a pause while the requested data files are downloaded and projected, merged, and clipped as needed to the download region (which, in this case, is hydrologic unit 01090005). When the download is complete, the user will be returned to the SWToolbox interface and the added layers will appear on the explanation and map.

The user also can adjust the minimum number of measurement values required at an observation site for the site to be shown on the project map. This is done by adjusting the value specified in “Min Count” in the “Download Data” dialog box. The default value for “Min Count” is 10 measurement values, but this value can be modified. After the value of “Min Count” is adjusted for a selected data type (for example, “Daily Streamflow”), the user clicks on the “Download” button, which causes the map to be updated with sites that meet the minimum specified count.

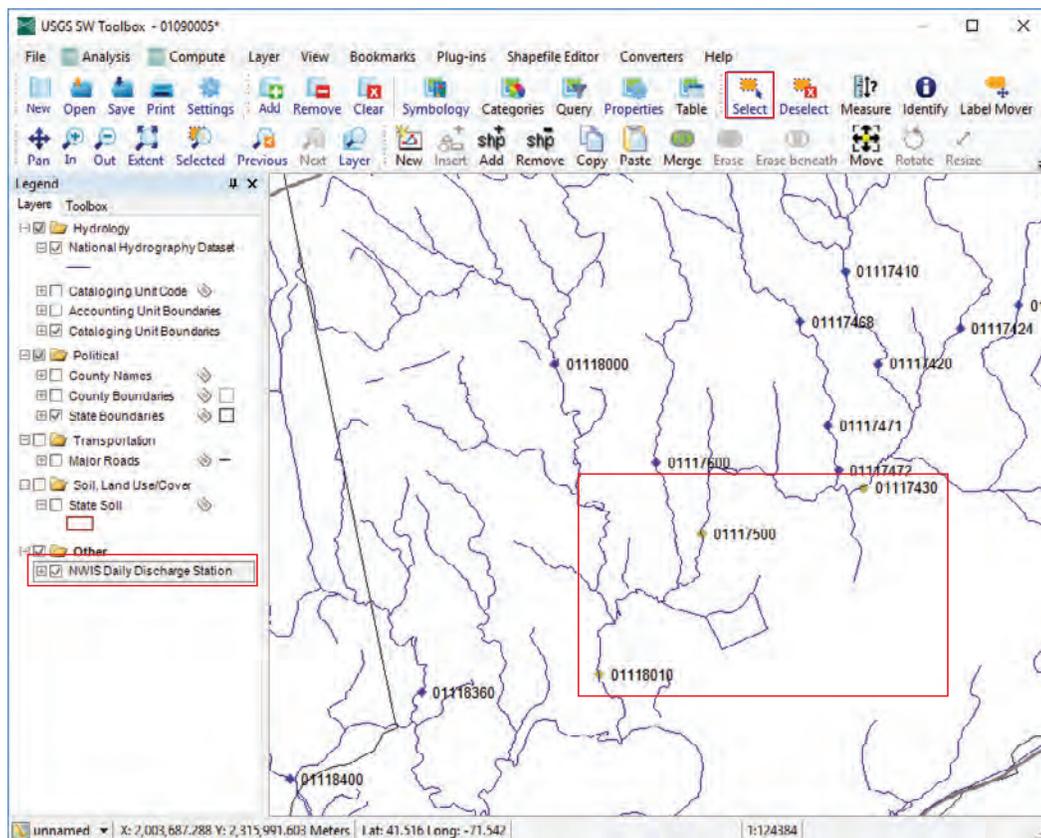


Figure 5. Screen capture showing steps used to select one or more data sites for which data can be retrieved from the U.S. Geological Survey National Water Information System for use in the Surface Water Toolbox (SWToolbox). In this example, streamgages on the Pawcatuck River, Rhode Island, are selected (sites 01117500, 01118010, and 01117430). Steps are discussed in the text. Part of the drainage basin for the streamgage is shown in red.

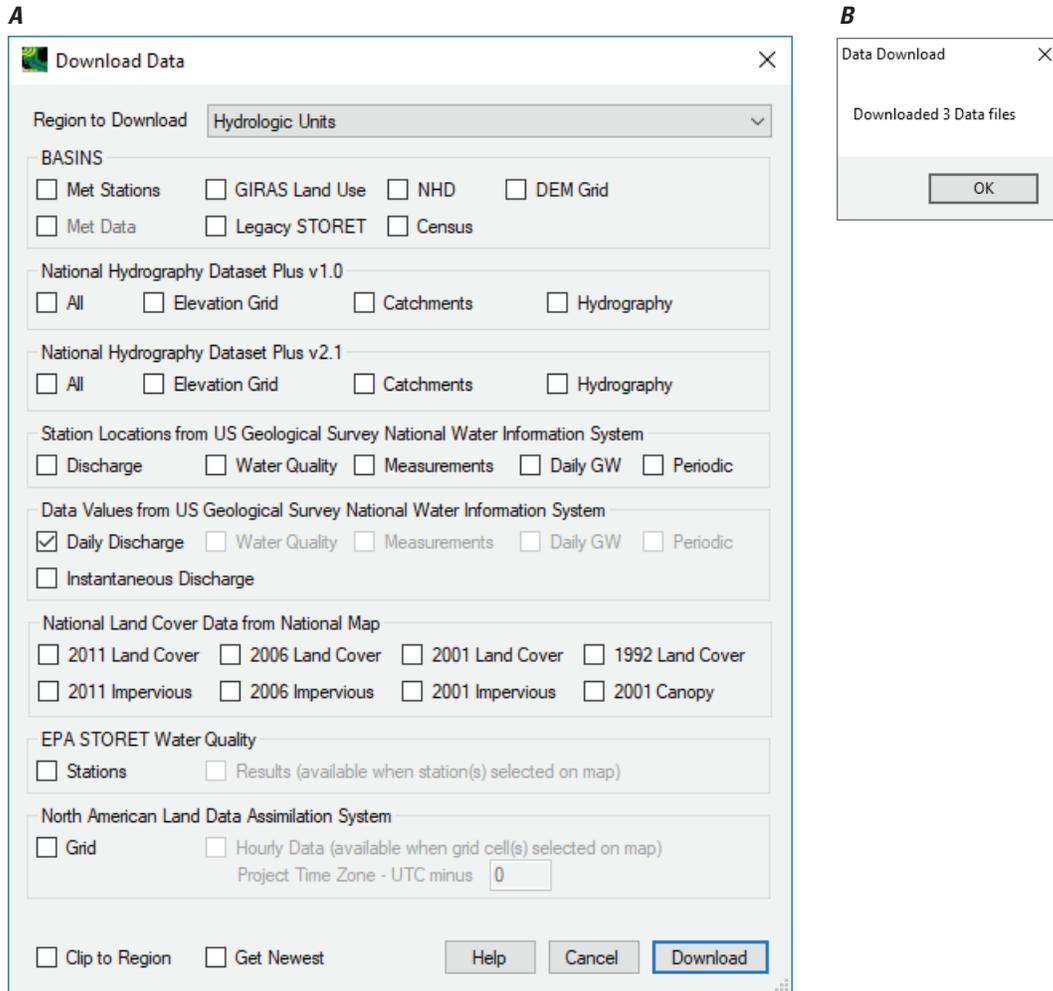


Figure 6. Screen captures showing dialog boxes used to retrieve data from the U.S. Geological Survey National Water Information System and other sources for use in the Surface Water Toolbox (SWToolbox). *A*, The “Download Data” dialog box with the check box for retrieving daily streamflow (discharge) data selected. *B*, The dialog box indicating that the data for the three sites of interest have been retrieved and are available for use.selected (sites 01117500, 01118010, and 01117430). Steps are discussed in the text.

Bypassing the Map Interface

Once a project has been set up, data for any streamgage—regardless of whether or not it is within the project area—can be downloaded by selecting “Download Data” from the “File” menu. When the “Download Data” window appears, select “Station IDs” (fig. 7) from the Region to Download dropdown

menu, and then click on the “Download” button at the bottom of the window. This opens up a dialog box where the USGS streamgage station identifier (ID) can be entered manually for one or more stations, or a list can be cut and pasted from another source. The station locations must be within the selected project area. Discharge data for the listed station IDs will be downloaded.

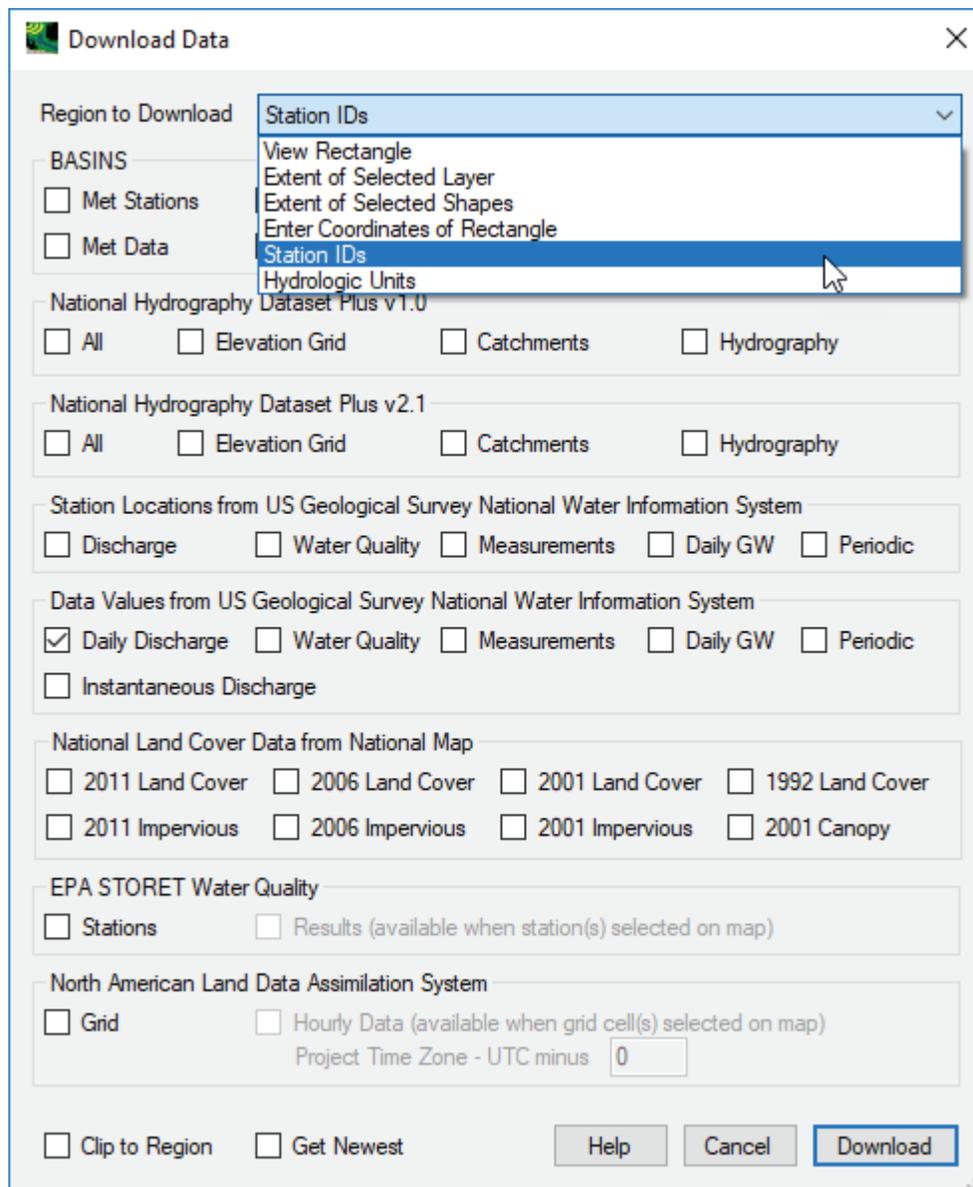


Figure 7. Screen capture showing data can be downloaded from a list of station identifiers (IDs) by selecting this menu option from the Download Data window in the Surface Water Toolbox (SWToolbox).

Using SWToolbox: Analysis Menu Tools

The SWToolbox “**Analysis**” menu contains six main options, which in addition to graphing the data, execute the types of analyses described in previous sections of this report. The following menu options are:

- **List**—Lists time-series data for selected stations.
- **Data Tree**—Lists the attributes and summary statistics for a dataset.
- **Graph**—Basic plots of time series.
- **Integrated Frequency Analysis**—Includes Basic, N-Day generation, Trend (simple), and Frequency capabilities, as well as Design Flows.
- **Duration/Compare**—Includes duration analysis for a single time series or statistical comparison of two time series.
- **Duration Hydrograph**—Performs duration hydrograph analysis.

Selecting Time-Series Data

Selection of most of the “**Analysis**” tools brings up the “Select Data” form (fig. 8), which allows selection of one or more time-series datasets available for the project. The exceptions are the “batch” tools nested under “USGS Integrated Design Flow.” Information on the batch job tools can be found in the online help pages within the SWToolbox.

The top section of the “Select Data to Graph” form is titled “Select Attribute Values to Filter Available Data” and contains drop-down lists with time-series attributes. The number and content of these lists can be managed through the “**Attributes**” menu. The default attributes that are shown with the SWToolbox are “Scenario,” “Location,” and “Constituent,” but these default selections can be changed

by the user. Attributes are shown in figure 8 for the observed datasets that were retrieved in previous steps (that is, the streamflow data for streamgages 01117430, 01117500, and 01118010). When the user selects an attribute from any of the lists, the “Matching Data” section is populated with the datasets that meet the attribute(s) selected. In the example shown in figure 8, streamgages 01117430 and 01117500 have been selected from the “Location” attribute list (step 1), which adds the datasets to the “Matching Data” section. The user then selects these datasets (step 2) and the selected time series show up in the “Selected Data” section of the dialog box (step 3). The “Date Range of Selected Data” in the lower left corner provides information about the date period available for selected time series, including the full (or “All”) extent along with the “Common” range when multiple time series are selected.

Two optional check boxes are available in the lower right corner. The first allows choosing whether or not provisional data should be included with the selected time series. By default, provisional data are **not** included in the SWToolbox analyses. The second check box allows transforming the selected data in a variety of ways such as date subsetting, seasonal splitting, and data value filtering. When this check box is selected and the “Ok” button is clicked, the “Filter Data” form is displayed (fig. 9).

The “Dates to Include” section of the dialog box allows selection of the date range of the data to be shown on the graph. Three columns of dates are shown. The first column indicates the “Start” and “End” dates that are inclusive of all selected time series. The “**All**” button above these dates allows selecting this time period for use with all time series on the graph. The second column indicates the period of record that is common to all selected time series. The “**Common**” button allows selecting the common period for use with all time series. The “Start” and “End” dates that appear in the white fields in the third column initially are the same as those that appear in the first column, but users can change them to specify an alternate range of dates of the graph. Clicking on “Ok” then brings up the next form for the selected “**Analysis**” tool (for example, graph, list, and so forth).

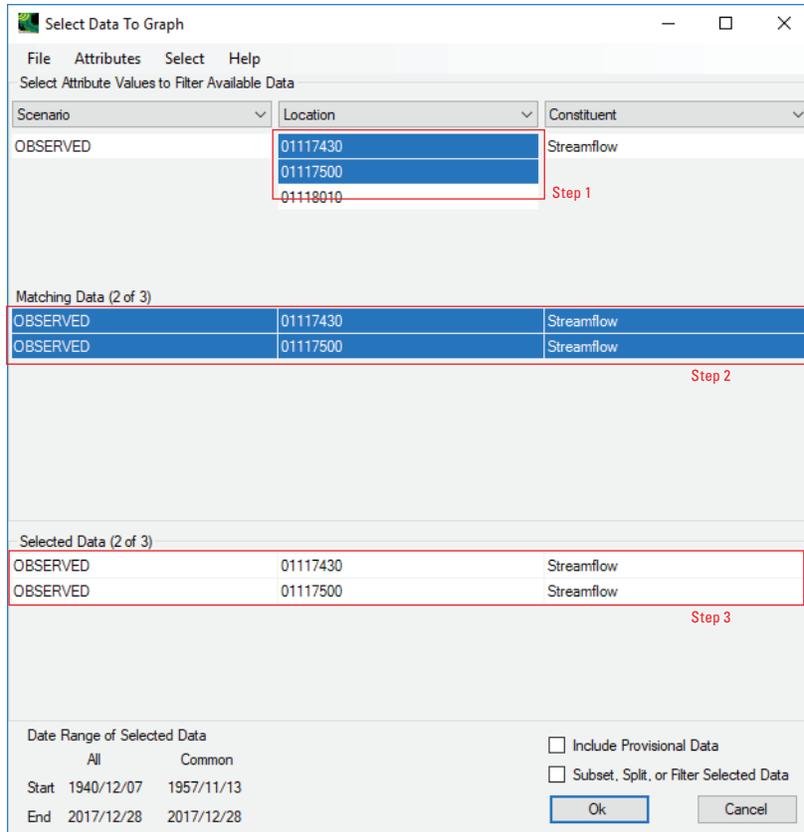


Figure 8. Screen capture of data selection window in the Surface Water Toolbox (SWToolbox).

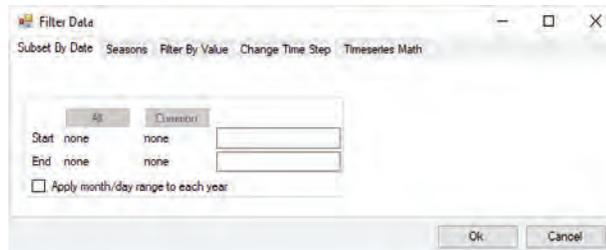


Figure 9. Screen capture of Filter Data form in the Surface Water Toolbox (SWToolbox).

Summarizing, Listing, and Graphing Time-Series Data

The SWToolbox provides the following three tools to help inspect and visualize the data: the “**Data Tree**,” “**List**,” and “**Graph**” tools under the “**Analysis**” menu option. A sample of each tool is shown here using the data selected in the previous section (figs. 10–13).

Data Tree

The “**Data Tree**” tool provides a listing of many of the attributes and computed statistics for the dataset (fig. 10). The “Attributes” branch of the tree listing includes information about the time series that was determined when the source data files were read and the dataset was created. The “Computed” branch of the tree includes statistical values that can be computed once the dataset is created. Additional information about the dataset and a view of its values can be found in the “Internal” and “Data” tree branches, respectively.

List

The “**List**” tool provides a listing of the selected values of the time series (see fig. 11). The entire period is listed at the native time step unless the user specifies otherwise using the “Filter Data” option to subset or transform the values (see Selecting Time-Series Data section above).

If the user opted to include Provisional Data, the List tool can be used to see the exact dates of such values. Within the “Time Series List” form, the user selects the “**Value Attributes**” option under the “**View**” option. This action adds several columns to the “**Time Series List**” form, including the “A” (for approved) and “P” (for provisional) columns. Data identified as “True” in the “P” column are provisional.

Graph

After selecting the time series to be graphed, a form to select the type(s) of graph(s) to produce is displayed (fig. 12). Any number of graphs may be selected, though some types require two datasets to be selected. For this example, a standard time-series graph will be produced.

The time-series graph indicates that data are complete throughout the period of record (fig. 13); however, some data for the period are provisional data that have not yet been approved for publication by the USGS. These provisional data are shown in red on the graph. Exact dates of these data are difficult to identify on the graph because of the very short period of provisional data for this example. The user can determine the exact date range of the provisional data with the “**List**” tool as described in the previous section.

Graph Menus

The “File” menu can be used to print the graph or save it to a file. The time series displayed on the graph can be changed using the ‘Select Data’ menu item from the “File” menu.

The “Edit” menu can be used to modify the appearance (for instance curves, axes, labels, and so forth) of the graph. The form, shown in figure 14, contains five tabs (Axes, Curves, Lines, Legend, and Text). When a graph is first drawn it is given default values for all these parameters.

Menu items on the “View” menu are used to specify zooming/panning on the horizontal or vertical axes. If the graph has been zoomed in, the “Zoom to All” menu option returns the graph to the original extents.

The “Analysis” menu allows the user to add another of the data viewing tools. If, for instance, the current graph includes two time series, choosing “List” from the Analysis menu opens a listing of the same two time series.

The “Coordinates” menu turns on a feature that displays the coordinates of the mouse pointer as it is moved over the graph.



Figure 10. Screen capture of sample Data Tree form in the Surface Water Toolbox (SWToolbox).

History 1	from NWS_discharge_01117500.rdb
Constituent	Streamflow
Id	1
Min	9.14
Max	3,320
Mean	196.51
1940/12/07 24:00	117
1940/12/08 24:00	139
1940/12/09 24:00	153
1940/12/10 24:00	148
1940/12/11 24:00	143
1940/12/12 24:00	137
1940/12/13 24:00	134
1940/12/14 24:00	132
1940/12/15 24:00	130
1940/12/16 24:00	139
1940/12/17 24:00	165
1940/12/18 24:00	182
1940/12/19 24:00	178
1940/12/20 24:00	172
1940/12/21 24:00	155
1940/12/22 24:00	155
1940/12/23 24:00	158
1940/12/24 24:00	157

Figure 11. Screen capture of sample Time-Series List form in the SWToolbox.

Choose Graphs to Create

- Time Series
- Flow/Duration
- Frequency (needs annual time series)
- Running Sum
- Residual (TS2 - TS1) (two datasets needed but 1 datasets)
- Cumulative Difference (two datasets needed but 1 datasets)
- Scatter (TS2 vs TS1) (two datasets needed but 1 datasets)
- Double-Mass Curve
- Box Whisker

Multiple WQ Plots

All None Cancel **Generate**

Figure 12. Screen capture of graph selection form in the SWToolbox.

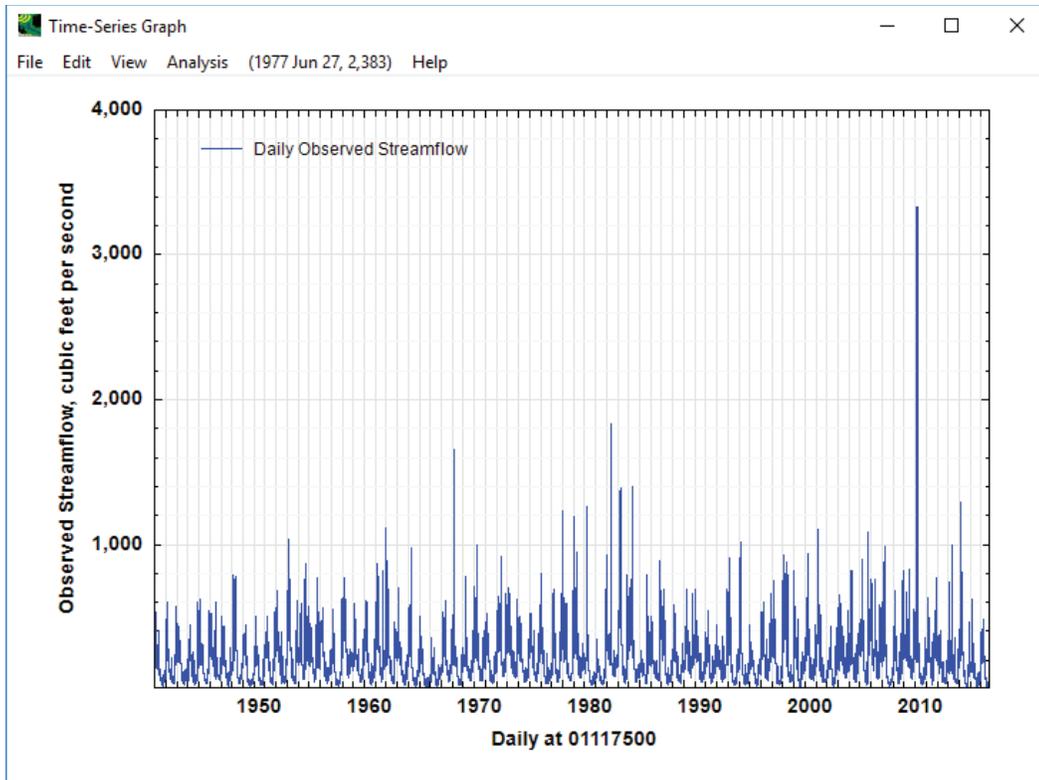


Figure 13. Screen capture of example of time-series plot in the Surface Water Toolbox (SWToolbox).

Figure 14. Screen capture of sample Edit Time-Series Graph form in the Surface Water Toolbox (SWToolbox).

Integrated Design Flow

When choosing interactive mode, the Integrated Design Flow window consists of the following four tabs:

- Select Dates
- N-Day, Trend, and Frequency
- Design Flow
- Outlier Test

These components are discussed in more detail in their respective sections. Each section assumes that the time series to analyze has been selected as described earlier.

Select Dates Tab

When the Integrated Frequency Analysis option is selected, the “**Select Dates**” tab is initially displayed (fig. 15), where the user is able to specify the “**Flow Condition**” (“High” or “Low”) and “**Date/Season**” options for use in subsequent analysis. While the “**Flow Condition**” is used only in n-day, trend, and frequency analysis, the “**Year/Season Boundaries**” and “**Years to Include in Analysis**” provide great flexibility in defining which portions of the data record to analyze and also apply to the Design Flow analysis. In the “**Years to Include in Analysis**” panel, the user has three options: All, Common, or Custom. Once the user chooses “**Custom**,” they are able to change the years and the dates in the appropriate text boxes. After making the other selections on this tab, clicking on the “**Display Basic Statistics**” button at the bottom will result in a pop-up window appearing, which displays the minimum, maximum, standard deviation, count, and count missing for each selected time series. The outlined example below demonstrates the All Dates option for two time series.

In the example shown in Figure 15, the “**Low Flow Condition**” is selected, and the “Year/Season Boundaries panel” is used to specify a year starting on April 1 and extending to March 31 of the following year; these dates should be used for the analysis. This is the climate year that is typically used for low-flow analysis. The “**All**” option is selected from the “Years to Include in Analysis” panel to use all available data.

N-Day, Trend, Frequency Tab

The “**N-Day, Trend, Frequency**” option computes annual n-day high and low flows for a daily time series. On the “**N-Day, Trend, Frequency**” tab (fig. 16), a default list of “Number of Days” to generate is provided for initial use. This list of n-day values can be modified by adding (+ button) or subtracting (– button) values in the text box below the list. The next time the SWToolbox is run, the contents of the list and the items selected will be preserved.

As mentioned earlier, the “**Select Dates**” tab contains the “**Flow Condition**” option, which is used to define whether the annual n-day time series will be for “High” or “Low” values. The following sections describe the buttons on the “**N-Day, Trend, Frequency**” tab.

N-Day Low Annual Time Series and Ranking List

The “**N-Day Timeseries List**” button (fig. 16) brings up a window showing the n-day time series that were computed by the software (fig.17). The set of attributes displayed at the top of the listing may be managed using the “File:Select Attributes” menu item. This may be especially helpful distinguishing between time series when multiple n-days are listed.

These results may be saved by clicking “File:Save Grid as Text” and saving to a text file. This file is in a tab-separated form that is readily imported by other spreadsheet programs. Other tabulated results also can be saved in this way. The “**Analysis**” menu of this window also provides the ability to graph the time series.

Trend of Low Annual Time Series and Statistics

The Trend of Low Annual Time Series and Statistics list is used for annual time series as a preprocessing step to frequency analysis. Figure 18 displays the results, including the Kendall tau statistic (KENTAU), p-value (KENPLV), and Sen median slope (KENSPL) of the time series specified earlier. For the example shown, the Kendall tau statistic (KENTAU=0.11737) and p-value (KENPLV=0.14608) indicate no statistically significant trend. As with the n-day list, the set of attributes displayed as columns in the listing may be managed using the “File:Select Attributes” menu item.

Screening Tests

An additional set of screening tests are run by selecting the “**Screening Tests (R)**” button, and all test results are brought up in a new window (fig. 19). Results will be available only if the R statistical program is installed on the user’s computer. The four tests included here are as follows:

- Spearman—An alternative trend test. Results are given as “PASS” or “FAIL,” with PASS indicating that there is no apparent trend detected by this test.
- PPCC—This test computes a probability plot correlation coefficient for the observed data and the fitted distribution. Results are given as PASS or FAIL, with PASS indicating that the data are not significantly different from the specified distribution with a 95-percent confidence limit.
- KSFit—The Kolmogorov-Smirnov test is used to assess the fit of the distribution. Results are given as PASS or FAIL, with PASS indicating that the data are not significantly different from the specified distribution.
- KRatioOutliers—This function indicates data points that may not fit within the selected distribution. Years for which the data point may be suspect are listed. This information can be used to subject flagged years to additional scrutiny, but having one or more flagged years (as in the example below) is not necessarily a problem. Data should not be removed simply because they appear on this list.

The screenshot shows the 'Integrated Design Flow' window with the 'Select Dates' tab selected. The window has a menu bar with 'File', 'Analysis', and 'Help'. Below the menu bar are three tabs: 'Select Dates', 'N-Day, Trend, Frequency', 'Design Flow', and 'Group Outlier Test'. The 'Select Dates' tab contains the following controls:

- Flow Condition:** Radio buttons for 'High' and 'Low'. 'Low' is selected.
- Year / Season Boundaries:** Two rows of controls. The first row has 'Start' with a dropdown menu set to 'April' and a text box containing '1'. The second row has 'End' with a dropdown menu set to 'March' and a text box containing '31'.
- Years to Include in Analysis:** Radio buttons for 'All: 1942/05/01 to 2017/03/17', 'Common: 1942/05/01 to 2017/03/17', and 'Custom:'. 'Common' is selected. Below these are two rows of controls: 'Start Year' with a text box containing '1942' and 'Data Starts 1942/05/01'; 'End Year' with a text box containing '2017' and 'Data Ends 2017/03/17'.
- Display Basic Statistics:** A button located at the bottom of the tab.

Figure 15. Screen capture of Select Dates tab in Integrated Design Flow window in the Surface Water Toolbox (SWToolbox).

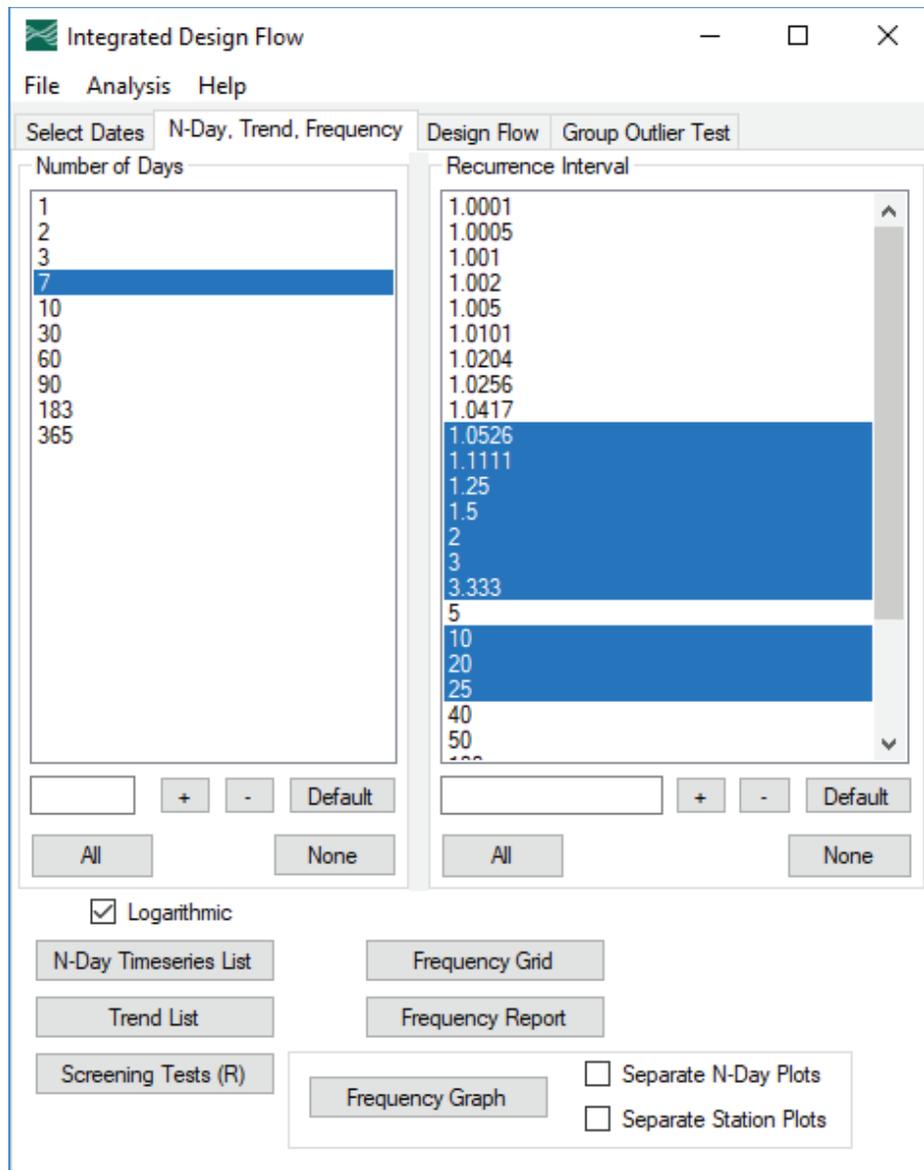


Figure 16. Screen capture of N-day, Trend, Frequency tab in Integrated Design Flow window in the Surface Water Toolbox (SWToolbox).

Year	L007	Date	Rank
1944	0.4	1943/10/04 24:00	3
1945	0.24286	1944/08/28 24:00	2
1946	8.3286	1945/07/14 24:00	56
1947	3.0429	1946/09/16 24:00	21
1948	3.9714	1947/10/17 24:00	33
1949	5.6143	1948/10/02 24:00	42
1950	8.8286	1949/09/13 24:00	59
1951	6.8286	1950/08/19 24:00	49
1952	3.9286	1951/09/29 24:00	31
1953	9.9571	1952/10/24 24:00	64
1954	2.2	1953/09/04 24:00	18
1955	0.75714	1954/08/03 24:00	5
1956	4.2	1955/08/07 24:00	35
1957	9.1143	1956/09/06 24:00	60
1958	1.0286	1957/09/09 24:00	8
1959	5.8571	1958/09/17 24:00	44
1960	2.0429	1959/09/29 24:00	16
1961	8.6714	1960/10/19 24:00	57
1962	3.4429	1961/10/04 24:00	27
1963	1.3143	1962/08/27 24:00	9
1964	1.5571	1963/10/15 24:00	10
1965	1.6286	1964/08/31 24:00	13
1966	1.6	1965/08/22 24:00	11
1967	0.042857	1966/07/28 24:00	1
1968	8.1857	1967/09/21 24:00	55
1969	3.8429	1968/09/03 24:00	30
1970	4.3143	1969/09/02 24:00	36

Figure 17. Screen capture of example of n-day time-series list in the Surface Water Toolbox (SWToolbox). In this example, the 7-day annual low flows have been generated.

STAID	STANAM	KENTAU	KENPLY	KENSLP	From	To	Count	Not Used	Min	Max	Constituent
01639000	MONOCACY RIVER AT BRIDGEPORT, MD	0.13318	0.096501	0.036814	1944/03/31	2016/03/31	73	0	0.042857	35.286	L007

Figure 18. Screen capture of example of trend list in the Surface Water Toolbox (SWToolbox).

STAID	STANAM	Count	Not Used	Min	Max	NDay	Spearman	PPCC	KSFit	KRatioOutliers
01639000	MONOCACY RIVER AT BRIDGEPORT, MD	72	0	0.042857	35.286	7	PASS	PASS	PASS	1967,2004,1980,1997

Figure 19. Screen capture of example of screening-test results in the Surface Water Toolbox (SWToolbox).

Frequency Statistics Grid

With this option, annual time series may be analyzed with the LP3 distribution. Procedures have been included to handle zero values and periods of missing records. Missing record is indicated using values of -1 or less. Output is provided in a table and, optionally, the computed annual time-series datasets are added to a Watershed Data Management file.

The “**N-Day, Trend, Frequency**” tab, in the “**Integrated Design Flow**” window allows selection of any desired N-Day/Recurrence Interval combination(s) for use in computing frequency statistics. As with the “Number of Days list,” the “Recurrence Interval” list has an initial set of defaults that may be modified using the text box and + and – buttons below the list. Clicking the “**Frequency Grid**” button causes the “**Frequency Statistics**” grid form to appear in a separate window, with the computed statistics shown (fig. 20). The selection combination shown in figure 16 was used to create the output grid shown in figure 20. In this example, the “**Logarithmic**” option is checked, meaning that the base-10 logarithms of each value in the time series were taken before fitting the frequency distribution. In other words, the LP3 distribution was used (leaving it unchecked would use the Pearson Type III distribution).

From this screen, the “**File**” menu contains the following output options:

- **Save Grid**—Produces tab-delimited file that duplicates the results shown on the grid-output form.
- **Save Report**—Produces traditional USGS Frequency Output report.
- **Save/View N-Day**—Enables users to save and (or) view n-day annual time series, including frequency results as attributes.
- **Export Results**—Produces tab-delimited file of resulting attributes and values.

Frequency Report

A formatted text file is produced by clicking on the “**Frequency Report**” button. This report includes both the frequency results and full information on inputs and settings chosen.

Frequency Graph

Clicking on the “**Frequency Graph**” button will produce a graph showing the fitted frequency curve(s) (fig. 21).

The two check boxes next to the “**Frequency Graph**” button can be used to generate frequency plots for different combinations of chosen n-day and flow data stations. If only the “**Separate N-Day Plots**” option is checked, then all stations will be plotted on a single graph for each n-day selected. If only the “**Separate Station Plots**” is checked, then separate graphs will be produced for each station, and each graph will contain frequency plots for all chosen n-days. If both boxes are checked, then there will be a frequency plot created for every combination of chosen n-day(s) and stations.

Design Flow Tab

The “**Design Flow**” tab presents separate panels for calculating biologically based flows and hydrologically based flows used for design purposes (fig. 22). Check boxes in each panel provide options for computing commonly used design flows or custom flow, with the commonly used design flows checked by default. For the custom option, the user must specify the length of the flow averaging period to be used (in days), the average number of years between excursions, the length of the low-flow period, or excursion clustering period, and the maximum number of excursions to be counted within each excursion clustering period. After selecting “**Calculate**”, each biologically based flow and its corresponding flow percentile are provided as output, along with the average number of excursions for each x-day period in 3 years (fig. 23).

Data Set	Probability	Recurrence Interval	7-day
OBSERVED 01639000 Streamflow	0.9500	1.0526	13.942
OBSERVED 01639000 Streamflow	0.9000	1.1111	12.16
OBSERVED 01639000 Streamflow	0.8000	1.25	9.7059
OBSERVED 01639000 Streamflow	0.6667	1.5	7.3401
OBSERVED 01639000 Streamflow	0.5000	2	5.0503
OBSERVED 01639000 Streamflow	0.3333	3	3.1714
OBSERVED 01639000 Streamflow	0.3000	3.333	2.83
OBSERVED 01639000 Streamflow	0.1000	10	0.94105
OBSERVED 01639000 Streamflow	0.0500	20	0.49272
OBSERVED 01639000 Streamflow	0.0400	25	0.40197

Figure 20. Screen capture of example of Frequency Statistics Grid in the Surface Water Toolbox (SWToolbox).

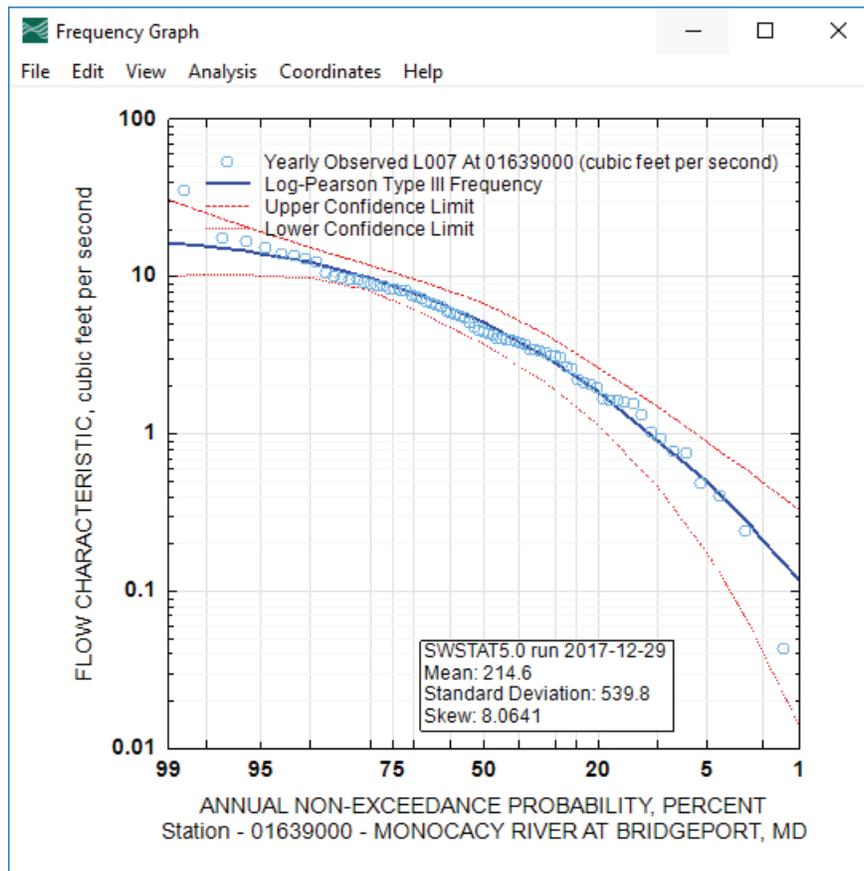


Figure 21. Screen capture of example of graph in Frequency Graph in the Surface Water Toolbox (SWToolbox).

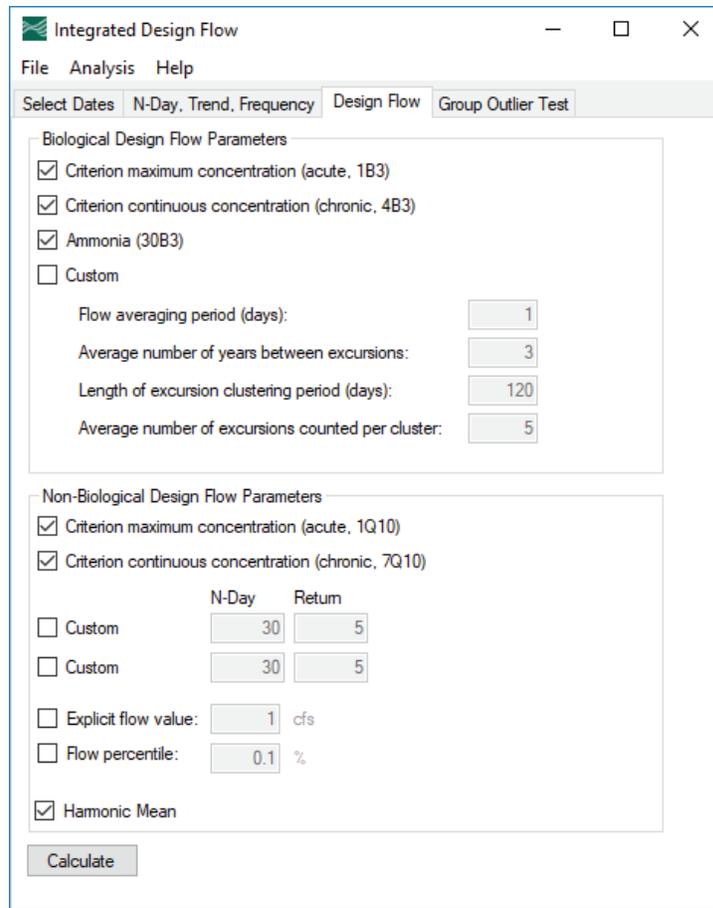


Figure 22. Screen capture of Design Flow tab in Integrated Design Flow window in the Surface Water Toolbox (SWToolbox).

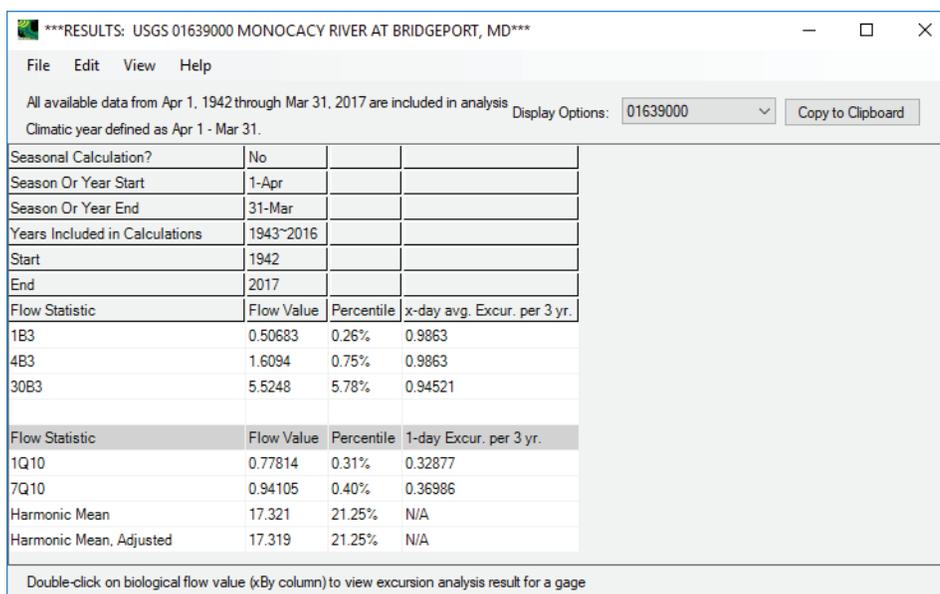


Figure 23. Screen capture of Integrated Design Flow results form in the Surface Water Toolbox (SWToolbox).

Check boxes in the lower panel allow the user to specify parameters for additional design flows. First, hydrologically based design flows may be selected. These are n-day frequency statistics and are equivalent to the statistics computed using the “**N-day, Trend, Frequency**” tab. The “Custom” option allows the user to select a specific n-day average and recurrence interval. The “explicit flow value” option provides the user with the flow percentile corresponding to that flow value. Conversely, when a “Flow percentile” is selected, the output shows the flow value corresponding to that flow percentile. If “harmonic mean” is selected, the harmonic mean of all values in the streamflow time series is computed along with the flow percentile corresponding to that flow value. For all of these options, except harmonic mean, the number of 1-day excursions in a 3-year period is given.

Note that the hydrologically based flows produce the same results as obtained using the “**N-day, Trend, Frequency**” tab. However, the more comprehensive set of analysis and graphing tools available with the “**N-day, Trend, Frequency**” analyses are not available using the “**Design Flow**” tab. For this reason, use of the “**N-day, Trend, Frequency**” tab is recommended for computing these types of flow statistics.

Outliers Tab

The “**Outlier Test**” tab is reserved for another R screening test for 7Q10 values computed from gages that are expected to have similar unit runoff for this low-flow value (fig. 24). The screening test, however, is not yet fully implemented.

This tab will be populated with information from stations that were selected in the “Select Data” window at the beginning of the workflow. Check boxes to the left of each station number allow control of which stations to include in the test. The “**None**” button at the bottom allows selecting none of the stations, and the “**All**” button allows selecting all listed stations. The test can only be run when five or more stations are selected. After the desired stations are selected, click on the “**Calculate Statistics**” button to compute the 7Q10 for the stations. The “**Detect Outliers**” button is used to run the outlier test but is currently disabled.

Duration/Compare Analysis

The “**Duration**” and “**Compare**” features perform traditional flow-duration analysis on time-series data. While the “**Duration**” option performs the analysis on individual time series, the “**Compare**” option performs the analysis

on two related time series, such as simulated and observed streamflows or upstream and downstream streamflows, to compute error and bias. Due to the intrinsic commonality between duration and compare analyses, they are managed together in terms of dataset selection and report generation.

The “**Duration**” analysis can be done on more than one time series whereas the “**Compare Analysis**” needs exactly two time series. If more than two time series are selected for the **Compare** analysis, only the first two of the selected group will be used. **Note:** For the **Compare Analysis**, the two selected time series should have common time duration and time step. This can be accomplished by choosing the “Subset, Split, or Filter Selected Data” check box to bring up the Filter Data form, and then clicking the “**Common**” button on that form.

Once the desired time series are selected, the main entry form for “**Duration/Compare Analysis**” will be shown (fig. 25).

In the above form, the Class Limits Generation panel is used to specify conditions for creating class limits to be used for breaking time-series data values into appropriate classes or “bins” of flow magnitude. The class limits are used in the text report, but not for generation of the flow-duration curve or the flow percentiles. The “**Number**” of class limits must be specified by the user before clicking on the “**Generate**” button. The “**Min**” and “**Max**” limits are optional and allow restricting the range of the class thresholds to be generated. The “**SWSTAT Default**” check box, when selected, will apply a set of 35 class limits (ranging from 0 to 100000), regardless of what the user specifies for the Number, Min, and Max values.

The “**Class Limits**” panel displays all available class thresholds. Users can add or remove class thresholds and select some class limits for use in the analysis by highlighting them. If none is selected (highlighted), then all class limits are used in the analysis. The “**Default**” button will create 35 class limits based on the currently selected overall minimum and maximum values of the time series. Thus, different time series will have their own “default” set of class limits.

The “**Choose Analysis**” panel allows the user to select either the “**Duration**” or “**Compare**” analysis by checking their corresponding radio buttons. Information about the selected time series will be displayed in the area directly under the two radio buttons.

Once the final selection of class limits and the choice of an analysis are made, the user can proceed to click the “**Report**” button to carry out the analysis, and the resulting text report will be displayed in a subsequent popup window.

The user can also click the “**Graph**” button to produce a duration plot for the selected time-series data (fig. 26).

Duration Hydrograph Analysis

This analysis will compute duration-hydrograph tables and curves for the selected time series. The duration hydrograph displays the percentage of time that a particular flow value has been exceeded on a given day over the period of record. The duration hydrograph curves are plotted to show range of exceedance levels for each time step. Data are analyzed at a daily time step. Although it was originally developed for streamflow data, any time series may be analyzed.

Once the desired time series are selected, as discussed in previous sections, the main entry form for “**Duration Hydrograph**” analysis will be shown (fig. 27).

The “**Select Percent Exceedance**” list box contains the percentiles for which exceedance values will be extracted from the selected time series for each day of the records. The user can add or remove percent-exceedance levels but they have to be within the range of 0 and 1. If none are selected, then all percentages are used for the analysis. The user clicks on the “**Report**” button to start the analysis. Multiple time-series data can be analyzed and text reports will be displayed in a popup window with one time-series result following another. The user then clicks on the “**Graph**” button to generate a percent-exceedance hydrograph for the selected percentile levels for each of the selected time series that are plotted (fig. 28).

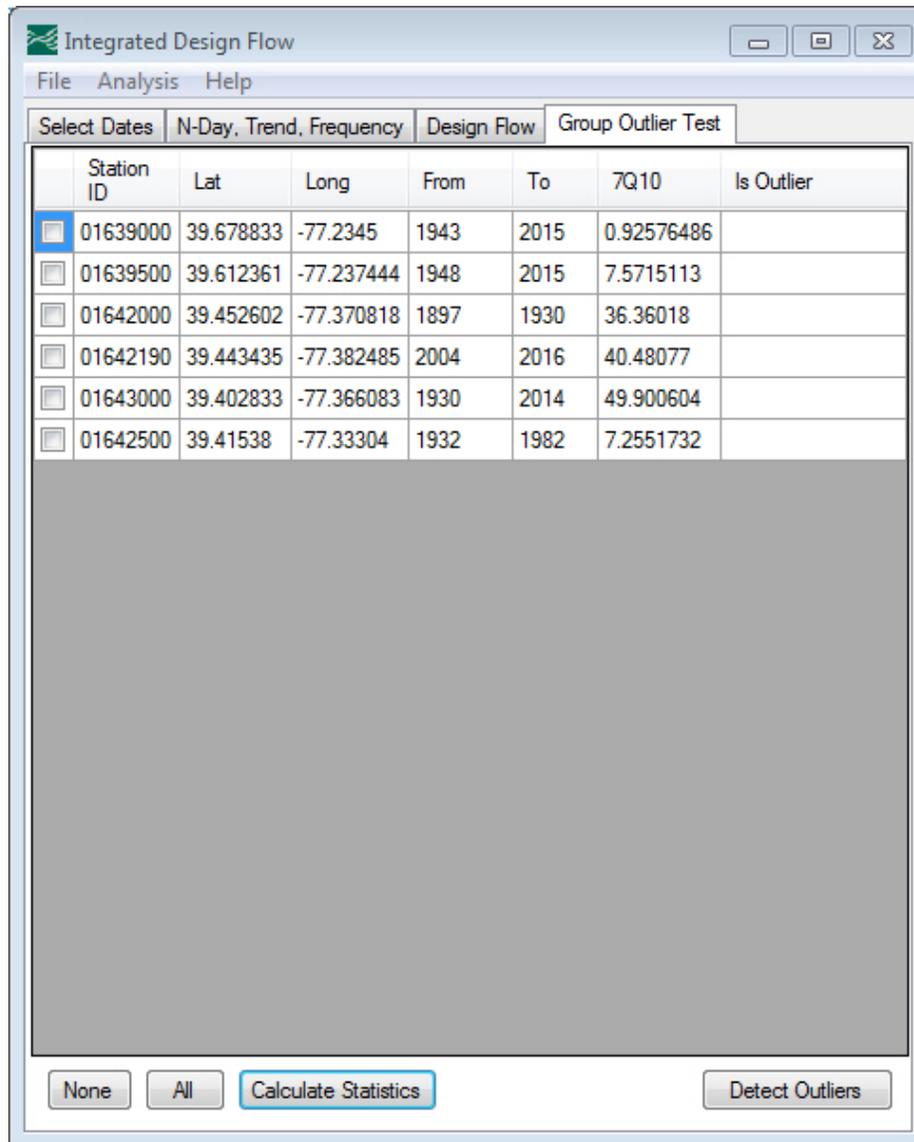


Figure 24. Screen capture of Outlier tab in Integrated Design Flow window in the Surface Water Toolbox (SWToolbox).

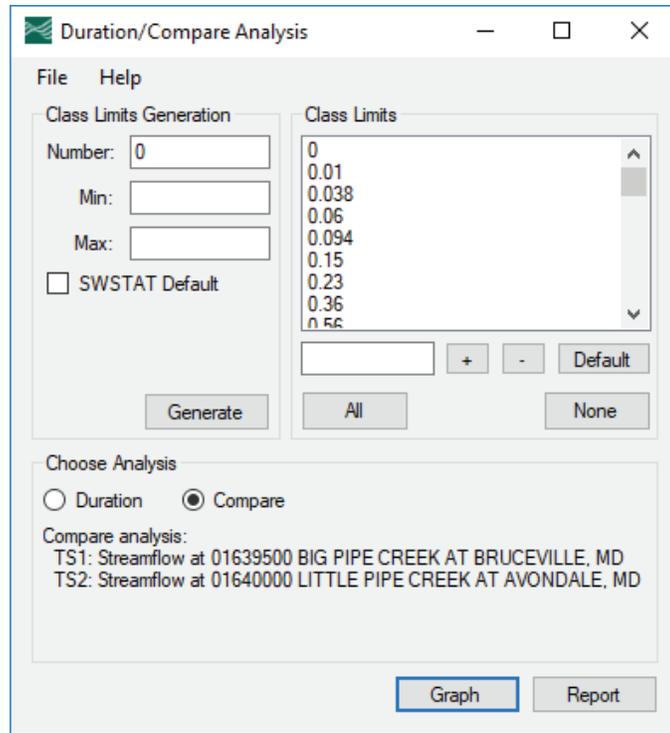


Figure 25. Screen capture of Duration/Compare Analysis window in the Surface Water Toolbox (SWToolbox).

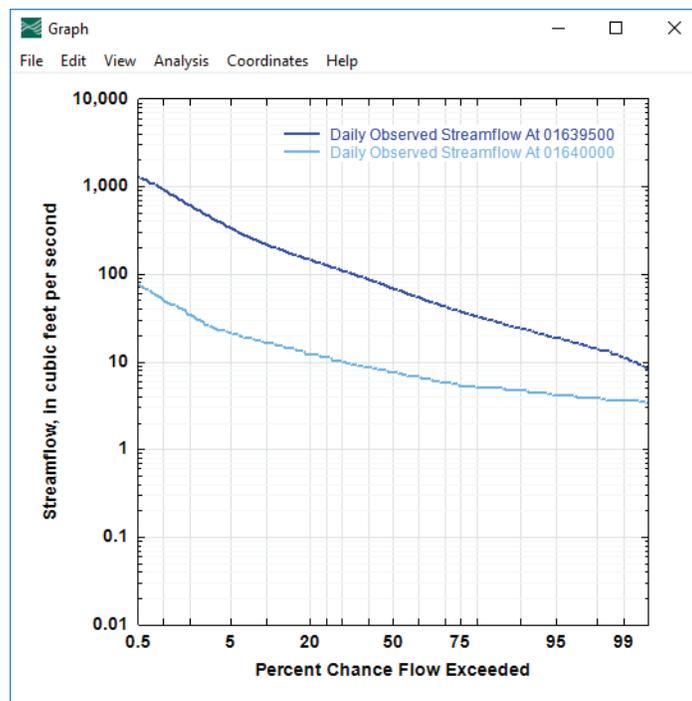


Figure 26. Screen capture of example of a Compare graph in the Surface Water Toolbox (SWToolbox).

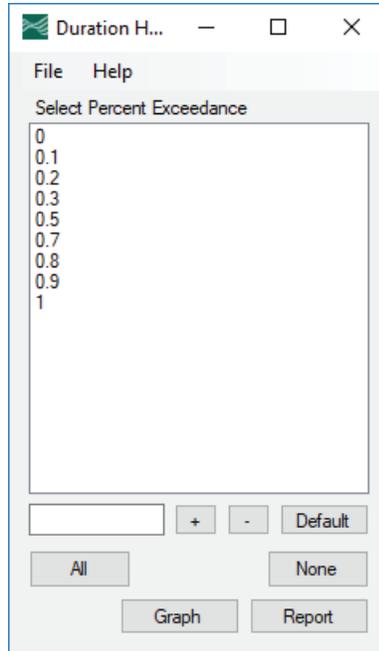


Figure 27. Screen capture of window used to select percent exceedances for graphing or reporting in the Surface Water Toolbox (SWToolbox).

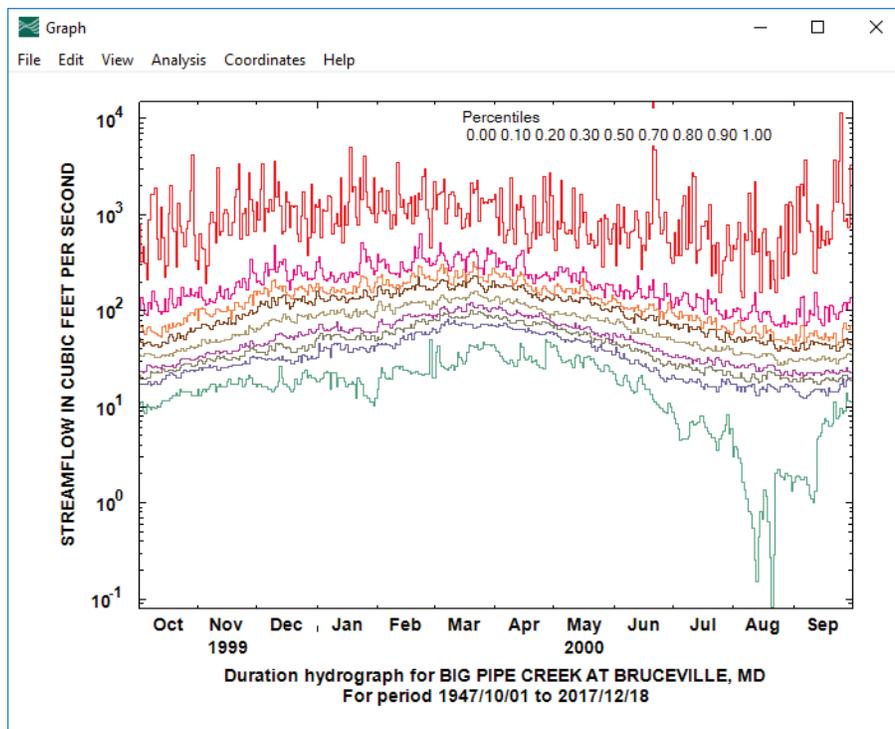


Figure 28. Screen capture of example of duration hydrograph in the Surface Water Toolbox (SWToolbox).

Summary

The SWToolbox provides a point-and-click interface to download and analyze streamflow time series. Two specialized functions are to compute n-day frequency statistics and biologically based design flows. To assist in the analysis of n-day frequency statistics, several additional screening tests are included to test for trends and conformity of the data to the log-Pearson Type III distribution. Noninterpretive statistics such as the flow-duration statistics are also computed by the SWToolbox, as well as basic plotting tools. In all cases, careful analysis and consideration of the handling of missing data are needed.

References Cited

- Barlow, P.M., Cunningham, W.L., Zhai, Tong, and Gray, Mark, 2014, U.S. Geological Survey Groundwater Toolbox, a graphical and mapping interface for analysis of hydrologic data (version 1.0)—User guide for estimation of base flow, runoff, and groundwater recharge from streamflow data: U.S. Geological Survey Techniques and Methods, book 3, chap. B10, 27 p., September 7, 2017, at <https://dx.doi.org/10.3133/tm3B10>.
- Caruso, J.C., and Cliff, Norman, 1997, Empirical size, coverage, and power of the confidence intervals for Spearman's rho: Educational and Psychological Measurement, v. 57, no. 4, p. 637–653.
- Cohn, T.A., Lane, W.L., and Stedinger, J.R., 2001, Confidence intervals for expected moments algorithm flood quantile estimates: Water Resources Research, v. 37, no. 6, p. 1695–1706, September 7, 2017, at <https://dx.doi.org/10.1029/2001WR900016>.
- England, J.F., Jr., Cohn, T.A., Faber, B.A., Stedinger, J.R., Thomas, Jr., W.O., Veilleux, A.G., Kiang, J.E., and Mason, Jr., R.R., 2017, Guidelines for determining flood flow frequency—Bulletin 17C: U.S. Geological Survey Techniques and Methods book 4, chap. B5, XXX p., <https://dx.doi.org/10.3133/tm4-B5>.
- Fisher, R.A., 1924, On a distribution yielding the error functions of several well known statistics: Proceedings of the International Congress of Mathematics, 1924, Toronto, v. 2, p. 805–813.
- Granato, G.E., 2014, Hydrologic Drought Decision Support System (HyDroDSS): U.S. Geological Survey Open-File Report 2014–1003, 91 p., with CD-ROM, September 7, 2017, at <http://dx.doi.org/10.3133/ofr20141003>.
- Haan, C.T., 1977, Statistical methods in hydrology: Ames, Iowa, Iowa State University Press, 378 p.
- 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.
- Kirby, William, 1972, Computer-oriented Wilson-Hilferty transformation that preserves the first three moments and the lower bound of the Pearson Type 3 Distribution: Water Resources Research, v. 8, no. 5, p. 1251–1254.
- Press, W.H., Flannery, B.P., Teukolsky S.A., and Vetterling, W.T., 1992, Numerical recipes in Fortran 77—The art of scientific computing (2d ed.): New York, Cambridge University Press, 992 p.
- Riggs, H.C., 1968, Frequency curves: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. A2, 15 p.
- Riggs, H.C., 1972, Low-flow investigations: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. B1, 18 p.
- Seaber, P.R., Kapinos, F.P., and Knapp, G.L., 1987, Hydrologic units maps: U.S. Geological Survey Water-Supply Paper 2294, 63 p.
- Searcy, J.K., 1959, Manual of hydrology: Part 2, Low-flow techniques: U.S. Geological Survey Water-Supply Paper 1542-A.
- Sinnott, R.W., 1984, Virtues of the haversine: Sky and Telescope, v. 68, no. 2, p. 159.
- U.S. Environmental Protection Agency, 1986, Technical guidance manual for performing wasteload allocation, book VI, Design conditions, chapter 1, Stream design flow for steady-state modeling: U.S. Environmental Protection Agency, EPA Publication 440/4-86-014, 65 p.
- U.S. Environmental Protection Agency (EPA), 2015, BASINS 4.1 (Better Assessment Science Integrating point & Non-point Sources) Modeling Framework: National Exposure Research Laboratory, RTP, North Carolina, September 7, 2017, at <https://www.epa.gov/exposure-assessment-models/basins>.
- Veilleux, A.G., Cohn, T.A., Flynn, K.M., Mason, R.R., Jr., and Hummel, P.R., 2014, Estimating magnitude and frequency of floods using the PeakFQ 7.0 program: U.S. Geological Survey Fact Sheet 2013–3108, 2 p., September 7, 2017, at <https://dx.doi.org/10.3133/fs20133108>.
- Vogel, R.M., and Kroll, C.N., 1989, Low-Flow frequency analysis using probability-plot correlation coefficients: Journal of Water Resources Planning and Management, v. 115, no. 3, p. 338–357.
- Vogel, R.M., and McMartin, D.E., 1991, Probability plot goodness-of-fit and skewness estimation procedures for the Pearson Type III Distribution: Water Resources Research, v. 27, no. 12, p. 3149–3158.

Manuscript was approved September 29, 2017

Prepared by the USGS Science Publishing Network
Reston Publishing Service Center

For more information about this publication, contact:

Chief, Analysis and Prediction Branch

Water Mission Area

U.S. Geological Survey

12201 Sunrise Valley Drive, MS 415

Reston, VA 20192

