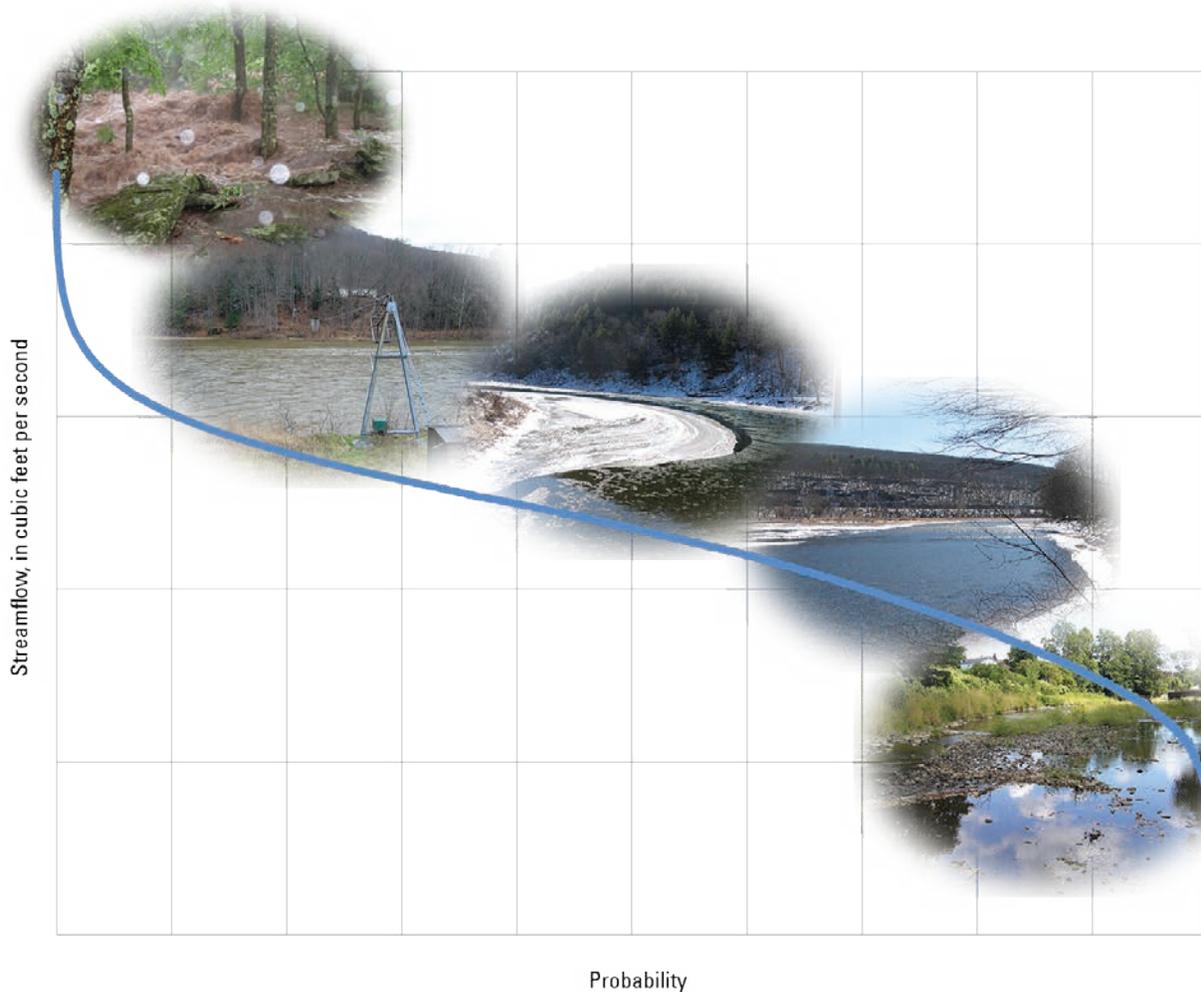
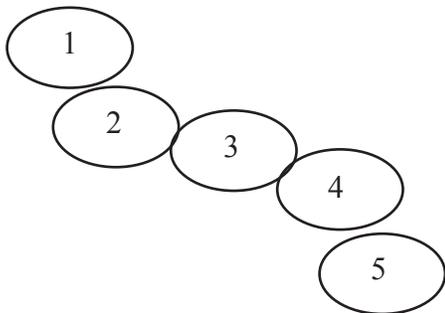


Prepared in cooperation with The Nature Conservancy and the New York State Energy Research and Development Authority

# Estimation of Unaltered Daily Mean Streamflow at Ungaged Streams of New York, Excluding Long Island, Water Years 1961–2010



Scientific Investigations Report 2014–5220



- Cover.**
1. Photograph looking upstream from the left bank at station 01349541 – Sugarloaf Brook south of Tannersville, N.Y.; photograph by Tim Hoffman, U.S. Geological Survey.
  2. Photograph looking from the right bank at station 01421000 – East Branch Delaware River at Fishs Eddy, N.Y.
  3. Photograph looking downstream from the New York State Route 8 bridge, downstream from station 01425000 – West Branch Delaware River at Stilesville, N.Y.
  4. Photograph looking downstream from station 01427510 – Delaware River at Callicoon, N.Y.
  5. Photograph looking downstream from station 01426000 – Oquaga Creek at Deposit, N.Y.
- The blue line represents the flow-duration curve of the U.S. Geological Survey New York Streamflow Estimation Tool.

# **Estimation of Unaltered Daily Mean Streamflow at Ungaged Streams of New York, Excluding Long Island, Water Years 1961–2010**

By Christopher L. Gazoorian

Prepared in cooperation with The Nature Conservancy and the New York State Energy Research and Development Authority

Scientific Investigations Report 2014–5220

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

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

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

U.S. Geological Survey, Reston, Virginia: 2015

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment, visit <http://www.usgs.gov/> or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit <http://www.usgs.gov/pubprod/>.

To order this and other USGS information products, visit <http://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:

Gazoorian, C.L., 2015, Estimation of unaltered daily mean streamflow at ungaged streams of New York, excluding Long Island, water years 1961–2010: U.S. Geological Survey Scientific Investigations Report 2014–5220, 29 p., <http://dx.doi.org/10.3133/sir20145220>.

ISSN 2328-0328 (online)

## Acknowledgements

The author thanks Stacey Archfield of the U.S. Geological Survey for providing technical expertise and support on the use of the QPPQ and map correlation method and her technical review of the report. The author also thanks Marla Stuckey, Gary Firda, John Williams, and Ken Pearsall of the U.S. Geological Survey for their technical review of the report and the New York Streamflow Estimation Tool and Martyn Smith, Doug Freehafer, Elizabeth Nystrom, James Ulrich, and Jeremy Newson for their data analysis and geographic information system and programming support for the New York Streamflow Estimation Tool.



# Contents

Abstract.....	1
Introduction.....	1
Background and Problem.....	2
Physical Setting.....	2
Previous Studies .....	2
Purpose and Scope .....	2
Estimation of Unaltered Daily Mean Streamflow .....	4
Reference Streamgages in New York and Surrounding States.....	5
Selection Criteria .....	5
Record Extension Techniques .....	5
Regression Analysis.....	8
Flow-Duration Curve Statistics.....	9
Basin Characteristics.....	9
Regression Against Basin Characteristics.....	9
Estimation of Streamflow Time Series from a Reference Streamgauge .....	15
Map Correlation .....	15
Using the New York Streamflow Estimation Tool for Estimating Unaltered Daily Mean Streamflow for Ungaged Locations .....	15
Accuracy and Limitations of Estimated Unaltered Streamflows .....	17
Summary.....	25
References Cited.....	26
Appendix 1. Flow-Duration Exceedance Probabilities, Observed and Computed From Streamflow Data, and Regression Equations for Streamgages Used in Regression Analysis .....	29
Appendix 2. Description of Reference Streamgages Used in the Development of the New York Streamflow Estimation Tool With Period of Record and Use of Data .....	29
Appendix 3. Reference Streamgages With Index Streamgauge Used for MOVE.3 Record Extension.....	29
Appendix 4. Basin Characteristics Used in the Development of Flow-Duration Regression Equations.....	29
Appendix 5. User's Guide for the New York Streamflow Estimation Tool (NYSET) .....	29

## Figures

1. Map showing physiographic provinces of New York.....	3
2. Graphs showing the QPPQ method (Fennessey, 1994) used in the New York Streamflow Estimation Tool (NYSET) showing A, observed daily mean streamflow at a reference streamgauge, B, flow-duration curve at the reference streamgauge, C, constructed flow-duration curve at the ungaged location, and D, estimated daily mean streamflow at the ungaged location; modified from Stuckey and others (2012).....	4
3. Map showing location of U.S. Geological Survey reference streamgages in and near New York.....	6
4. Map showing location of U.S. Geological Survey reference streamgages and National Weather Service climate divisions.....	7

5. Example correlation map for U.S. Geological Survey streamgage 01312000, Hudson River near Newcomb, N.Y. ....	16
6. Screen capture of summary report generated by the New York Streamflow Estimation Tool showing flow-duration curves and a hydrograph.....	18
7. Graphs showing comparison of observed and New York Streamflow Estimation Tool estimated daily mean streamflows for U.S. Geological Survey streamgages A, 01414500 Mill Brook near Dunraven, N.Y. and B, 04213500 Cattaraugus Creek at Gowanda, N.Y.....	21
8. Boxplots showing distribution of, A, Nash-Sutcliffe efficiency values and B, root-mean-square error by major river basin obtained from comparison between observed and estimated daily mean streamflows .....	23
9. Graphs showing flow-duration curves and hydrographs showing estimated and observed daily mean flows for U.S. Geological Survey streamgages, A, 04221500, Genesee River at Scio, N.Y. and, B, 04232100, Sterling Creek at Sterling, N.Y. %, percent .....	24

## Tables

1. Period and duration of drought when the lowest Palmer drought severity index occurred during the 1960s drought in streams of New York, excluding Long Island .....	8
2. Basin characteristics used in the development of regression equations for flow-duration exceedance probabilities for New York streams .....	10
2. Basin characteristics used in the development of regression equations for flow-duration exceedance probabilities for New York streams .....	11
2. Basin characteristics used in the development of regression equations for flow-duration exceedance probabilities for New York streams .....	12
3. Regression coefficients for use with flow-duration exceedance probability regression equations for New York streams, excluding Long Island .....	14
4. Comparison of low-flow statistics calculated by empirical ranking method and log-Pearson type III curve fitting at reference stations with observed flows for the period April 1, 1961, to March 31, 2010 .....	19
5. Nash-Sutcliffe efficiencies from comparison of the New York Streamflow Estimation Tool estimated daily-mean values to observed values at reference gages ..	22
6. Normalized root-mean-square errors as percentage from comparison of the New York Streamflow Estimation Tool estimated daily-mean values to observed values at reference gages .....	22
7. Basin characteristics used in development of regression equations to estimate flow-duration exceedance probabilities with the New York Streamflow Estimation Tool for basins in New York .....	25

## Conversion Factors

Multiply	By	To obtain
Length		
inch (in.)	2.54	centimeter (cm)
inch (in.)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
Area		
square mile (mi <sup>2</sup> )	259.0	hectare (ha)
square mile (mi <sup>2</sup> )	2.590	square kilometer (km <sup>2</sup> )
Flow rate		
cubic foot per second (ft <sup>3</sup> /s)	0.02832	cubic meter per second (m <sup>3</sup> /s)
cubic foot per second per square mile [(ft <sup>3</sup> /s)/mi <sup>2</sup> ]	0.01093	cubic meter per second per square kilometer [(m <sup>3</sup> /s)/km <sup>2</sup> ]

Temperature in degrees Fahrenheit (°F) may be converted to degrees Celsius (°C) as  
 $^{\circ}\text{C} = (^{\circ}\text{F} - 32) / 1.8$ .

## Datum

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

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

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

## Abbreviations

BaSE	Baseline Streamflow Estimator
CY	climate year
FDC	flow-duration curve
GIS	geographic information system
GNWISQ	get NWIS Web streamflow data files
LOC	line of organic correlation
LP3	log-Pearson type III
MA SYE	Massachusetts Sustainable-Yield Estimator
MOVE	maintenance of variance
NLCD	National Land Cover Database 2006
NS	Nash-Sutcliffe efficiency
NWIS	National Water Information System
OLS	ordinary least squares
P <sub>n</sub>	nth percentile from flow-duration curve

PRISM	Parameter-elevation Regressions on Independent Slopes Model
QPPQ	Methodology that equates streamflow as a percentile from a flow-duration curve for a particular day at a hydrologically similar reference streamgage with streamflow as a percentile from the flow-duration curve for the same day at an ungaged location
RMSE	root-mean-square error
SREF	Streamflow Record Extension Facilitator
WY	water year

# Estimation of Unaltered Daily Mean Streamflow at Ungaged Streams of New York, Excluding Long Island, Water Years 1961–2010

By Christopher L. Gazoorian

## Abstract

The lakes, rivers, and streams of New York State provide an essential water resource for the State. The information provided by time series hydrologic data is essential to understanding ways to promote healthy instream ecology and to strengthen the scientific basis for sound water management decision making in New York. The U.S. Geological Survey, in cooperation with The Nature Conservancy and the New York State Energy Research and Development Authority, has developed the New York Streamflow Estimation Tool to estimate a daily mean hydrograph for the period from October 1, 1960, to September 30, 2010, at ungaged locations across the State. The New York Streamflow Estimation Tool produces a complete estimated daily mean time series from which daily flow statistics can be estimated. In addition, the New York Streamflow Estimation Tool provides a means for quantitative flow assessments at ungaged locations that can be used to address the objectives of the Clean Water Act—to restore and maintain the chemical, physical, and biological integrity of the Nation's waters.

The New York Streamflow Estimation Tool uses data from the U.S. Geological Survey streamflow network for selected streamgages in New York (excluding Long Island) and surrounding States with shared hydrologic boundaries, and physical and climate basin characteristics to estimate the natural unaltered streamflow at ungaged stream locations. The unaltered streamflow is representative of flows that are minimally altered by regulation, diversion, or mining, and other anthropogenic activities. With the streamflow network data, flow-duration exceedance probability equations were developed to estimate unaltered streamflow exceedance probabilities at an ungaged location using a methodology that equates streamflow as a percentile from a flow-duration curve for a particular day at a hydrologically similar reference streamgage with streamflow as a percentile from the flow-duration curve for the same day at an ungaged location. The reference streamgage is selected using map correlation, a geostatistical method in which variogram models are developed that correlate streamflow at one streamgage with streamflows at all other locations in the study area. Regression

equations used to predict 17 flow-duration exceedance probabilities were developed to estimate the flow-duration curves at ungaged locations for New York using geographic information system-derived basin characteristics.

A graphical user interface, with an integrated spreadsheet summary report, has been developed to estimate and display the daily mean streamflows and statistics and to evaluate different water management or water withdrawal scenarios with the estimated monthly data. This package of regression equations, U.S. Geological Survey streamgage data, and spreadsheet application produces an interactive tool to estimate an unaltered daily streamflow hydrograph and streamflow statistics at ungaged sites in New York. Among other uses, the New York Streamflow Estimation Tool can assist water managers with permitting water withdrawals, implementing habitat protection, estimating contaminant loads, or determining the potential affect from chemical spills.

## Introduction

The lakes, rivers, and streams of New York provide an essential water resource for the State. The U.S. Geological Survey (USGS) operates more than 240 streamgages in New York, but these streamgages only monitor a fraction of the thousands of rivers and streams that carry water throughout the State. This information quantifies the daily mean flows, allows for estimating streamflow statistics such as flow-duration exceedances, and helps water managers to understand the natural flow regime of a stream, which is critical to the sustainability and health of aquatic freshwater ecosystems (Vogel and others, 2007; Poff and Zimmerman, 2010). Water management agencies require an understanding of natural and low streamflow characteristics for planning and management of waste-loads to streams, permitting streamflow alterations, water-quality evaluations, water-supply design, groundwater management, and aquatic-habitat protection.

The U.S. Geological Survey cooperated in a study with The Nature Conservancy and the New York State Energy Research and Development Authority to develop a tool for estimating continuous daily mean streamflow and streamflow

## 2 Estimation of Unaltered Daily Mean Streamflow at Ungaged Streams, Excluding Long Island, Water Years 1961–2010

statistics at ungaged locations for the period from October 1, 1960, to September 30, 2010. The New York Streamflow Estimation Tool (NYSET) estimates daily streamflow hydrographs using data from existing USGS streamgages coupled with explanatory basin characteristics. A graphical user interface, with an integrated spreadsheet summary report, has been developed to display the estimated daily streamflow statistics and evaluate different water management or water withdrawal scenarios with the estimated monthly data. Among other uses, the NYSET can assist water managers with permitting water withdrawals, implementing habitat protection, estimating contaminant loads, or determining the potential affect from chemical spills.

### Background and Problem

The seasonal variability of streamflows in New York makes it more challenging for water managers to protect the natural ecosystem during months with lower flows. Streamflow in many streams in New York are primarily sustained by the base flows from adjacent aquifers and meltwater runoff from upland parts of the basin. Typically, the period from about late June through early October (corresponding to the growing season) is when streamflows are reduced to their lowest levels of the year and recharge to aquifers is reduced. Similarly, the annual minimum streamflow at many USGS streamgages in New York is often observed during late summer and early fall (U.S. Geological Survey, 2012). This naturally occurring low flow period, combined with permitted water withdrawals, effluent discharges and increased demand for water can create a shortage of available water and damage to the natural ecosystem. Additionally, changes in climate patterns, such as droughts or floods, and changes in land use, such as increased industrial and suburban growth, may increase alteration to streamflows and aquatic biota natural to the lakes and streams of New York.

### Physical Setting

The State of New York occupies all or part of eight physiographic provinces that differ widely, from the high-relief areas of the Adirondack and Catskill Mountains to the low-relief areas along the Great Lakes and in the valleys of the Hudson, Mohawk, and St. Lawrence Rivers (fig. 1). Long Island, which is excluded from the study area of this report, is located in the Coastal Plain province (Fenneman, 1938; Lumia and others, 2006). The provinces included in this study represent basins with diverse physical characteristics found throughout the State.

The St. Lawrence Valley province lies at the extreme northern border of New York and is primarily smooth glacial plain. The Adirondack province lies in the northern part of New York and covers an area of about 10,000 square miles (mi<sup>2</sup>). The western half of the province is a plateau and has an abundance of lakes and ponds, whereas the eastern half is

mostly mountainous. The northern part of western New York lies in the Central Lowlands, whereas most of the central and southern areas of western New York are in the Appalachian Plateau province. The Appalachian Plateau province is mostly hilly terrain, with the highest elevations being in the Allegheny and Catskill Mountains. Southeastern New York lies in the Valley and Ridge, New England, and Piedmont provinces (Fenneman, 1938; Lumia and others, 2006).

### Previous Studies

Before this study, regression equations to estimate flow-duration exceedance probabilities for ungaged streams in all of New York had not been developed nor were there any statewide tools available for estimating low- and natural-flow characteristics for ungaged streams in New York. Low-flow statistics were presented by Eissler (1979) at gaged sites across New York State for the minimum 7-day, 2-year discharge (7Q2) and the minimum 7-day, 10-year discharge (7Q10) which are the annual lowest mean discharges for 7 consecutive days with a 2-year and 10-year recurrence interval, respectively. Previous regression studies include regression equations to estimate 7Q10 and 7Q2 low flow values for ungaged drainages in the lower Hudson River Basin (Barnes, 1986), and another, more recent study focused on low flow statistics in the Susquehanna River Basin (not shown; Randall, 2010).

Similar tools for estimating daily streamflow and assessing water availability were developed by the USGS in Massachusetts (Massachusetts Sustainable Yield Estimator [MA SYE]) and Pennsylvania (Baseline Streamflow Estimator [BaSE]; Archfield and others, 2010; Stuckey and others, 2012). The MA SYE, BaSE, and NYSET tools all apply the QPPQ method introduced by Fennessey (1994) and used by Hughes and Smakhtin (1996), Smakhtin (1999), Smakhtin and Masse (2000), Mohamoud (2008), Shu and Ourda (2012), and Linhart and others (2013). The MA SYE and BaSE tools have been effective at estimating the natural unaltered daily streamflow hydrograph at ungaged locations in their respective study areas and provide user-friendly ways of computing streamflow statistics commonly used for water resource management and habitat protection.

### Purpose and Scope

This report describes the New York Streamflow Estimation Tool (NYSET, version 1.0) to estimate streamflow at ungaged sites in New York for water years<sup>1</sup> (WYs) 1961 to 2010 and documents the methods used to develop the tool. Parameter-based regression equations used to predict 17 exceedance probabilities from the flow-duration curve (FDC) for ungaged streams in New York and flow-duration

<sup>1</sup>A water year runs from October 1 through September 30 and is designated by the year in which it ends.

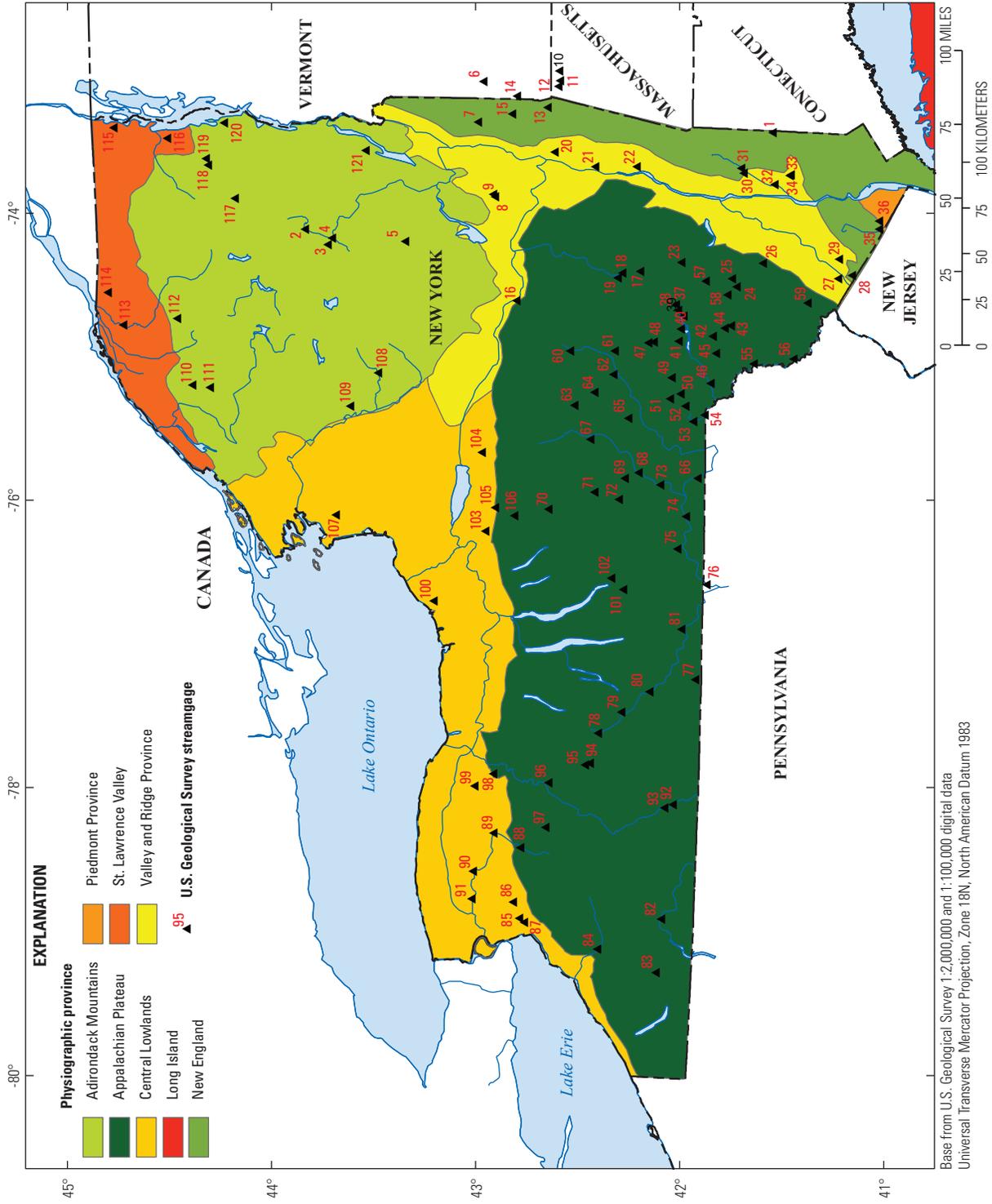


Figure 1. Physiographic provinces of New York; modified from Lumia and others (2006).

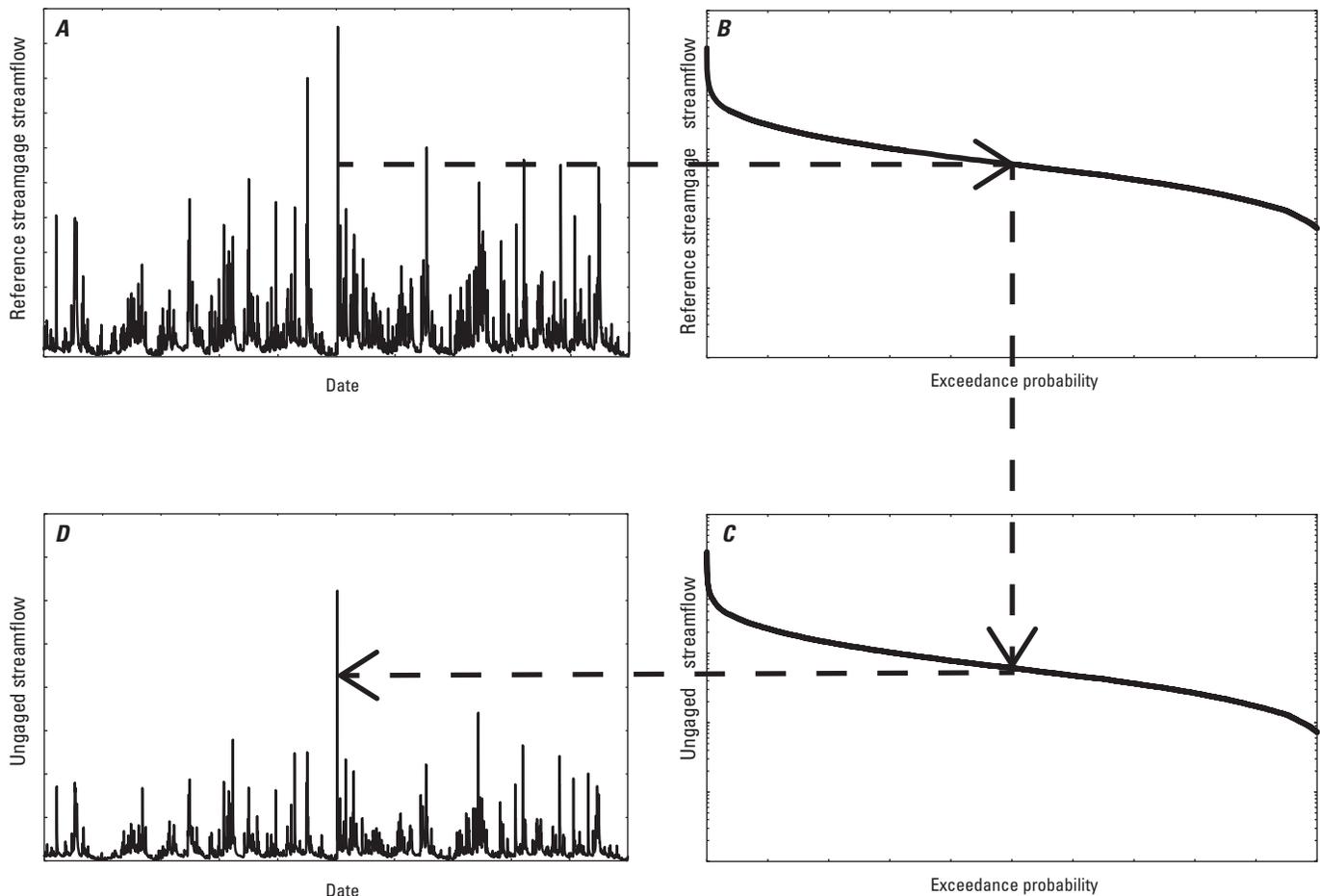
exceedance probabilities for 90 streamgages are presented (appendix 1). Streamflow data from continuous-record streamgages were used to develop correlation maps of the estimated correlation of streamflow between a reference streamgage and an ungaged location. Documentation on the proper use, functionality, and limitations of the tool also are presented in this report. A user's manual for NYSET is included (appendix 5).

## Estimation of Unaltered Daily Mean Streamflow

Daily mean streamflow for unaltered ungaged locations on streams in New York is estimated for WYs 1961 to 2010). The QPPQ method is used to estimate unaltered daily mean streamflow (fig. 2), which assumes equivalence of streamflow, as a percentile from the FDC for a particular day at a reference streamgage, where streamflow is measured,

to the streamflow as a percentile from the FDC for the same day at an ungaged location (Fennessey, 1994; Hughes and Smakhtin, 1996; Smakhtin, 1999; Smakhtin and Masse, 2000; Mohamoud, 2008; Archfield and others, 2010; Shu and Ourda, 2012; Stuckey and others, 2012; Linhart and others, 2013). The reference streamgage must be unaltered by regulation. Stuckey and others (2012) define a reference streamgage as a composite of the upstream land cover, geology, and hydrologic characteristics that can be used to represent ungaged basins with similar characteristics. A set of reference streamgages from the streamgage network in New York and surrounding States was identified to estimate correlations with ungaged locations. Where necessary, each reference streamgage record was extended by streamflow record extension methods to establish a common period of streamflow record. For this study the period from WYs 1961 to 2010 is used.

The FDC, a cumulative frequency curve that shows the percentage of time that specified streamflows are equaled or exceeded (Searcy, 1959), was constructed using the complete records for each of the reference streamgages. An FDC is built by sorting streamflows observed during a given



**Figure 2.** The QPPQ method (Fennessey, 1994) used in the New York Streamflow Estimation Tool (NYSET) showing *A*, observed daily mean streamflow at a reference streamgage, *B*, flow-duration curve at the reference streamgage, *C*, constructed flow-duration curve at the ungaged location, and *D*, estimated daily mean streamflow at the ungaged location; modified from Stuckey and others (2012).

period of time by magnitude and calculating the probability a specified streamflow value will be equaled or exceeded. For this report, the percentage of time that the streamflow is equaled or exceeded is termed exceedance probability, and an individual exceedance probability is termed percentile. Exceedance probability is used for discussing statistics and percentile is used for discussing methodology associated with a generic FDC.

At an ungaged location, the FDC is constructed from point estimates of streamflow for 17 exceedance probabilities. Regression equations, developed using physical and climatic basin characteristics and streamflow data from a subset of reference streamgages, are applied to estimate each of the 17 exceedance probabilities. To construct a continuous daily FDC, streamflow at all other exceedance probabilities was determined by log-log interpolation, completing the FDC between the 17 exceedances. Thus, a FDC consisting of 18,262 streamflow exceedances (one value for each day in WYs 1961–2010) is produced to represent the complete range of calculable percentile flows at the ungaged location.

When estimating unaltered daily mean streamflow, the selection of the reference streamgage is very important in attaining the best estimate of daily streamflow at the ungaged location. A methodology developed by Archfield and Vogel (2010), spatial map correlation, is used to select a reference streamgage. Spatial map correlation is a geostatistical procedure whereby a reference streamgage is selected whose streamflow demonstrates the strongest estimated correlation with an ungaged location. Though a nearby streamgage commonly exhibits strong estimated correlations, this method is unlike the frequently used method of selecting only from nearby sites, which assumes the nearest streamgages are the best correlated to the ungaged location (Archfield and Vogel, 2010).

## Reference Streamgages in New York and Surrounding States

A total of 121 reference streamgages were selected for this study (fig. 3); 115 in New York, 3 in Massachusetts, 2 in Vermont, and 1 in Pennsylvania (site identifier 76 near Waverly, N.Y.). The streamgages used in the study have a contributing drainage area that lies in all or parts of all physiographic regions that occur in New York, excluding the Coastal Plain of Long Island.

## Selection Criteria

The criteria for selecting appropriate streamgages include streamflow minimally altered by regulation, diversion, or mining, and other anthropogenic activities, no days with zero flow, and at least 20 years of continuous record; however, two streamgages with fewer than 20 years of record were included to increase spatial coverage and basin characteristic diversity. The average number of years of record is 49.8 years.

Minimally regulated streamgages for this analysis are those with upstream reservoir impoundments that control less than 15 percent of the contributing drainage area at the streamgage and alter less than half the exceedance flows along the FDC. For streamgages with at least 20 years of unregulated flow data recorded before the impoundment was constructed; only the period before the start of regulation was used in the analysis. Information on diversions was obtained from USGS annual water data reports (data for 2006 and later only, <http://wdr.water.usgs.gov/>; data before 2006, in paper format on file at the USGS New York Water Science Center).

To avoid using reference streamgage basins with potential anthropogenic effects on streamflow, the percentage of impervious area within a reference streamgage basin was limited to less than 10 percent. The average percentage of area classified as impervious from the National Land Cover Database (NLCD) 2006 (Fry and others, 2011) for the streamgages selected for regression analysis is 0.8 percent; with a maximum of 6.7 percent. Additionally, the average percent of area classified as urban from NLCD 2006 for the streamgages selected for the analysis is 5.4; with the two basins with the greatest percent area classified as urban are 29.5 and 29.1 percent urban.

An additional criterion for reference streamgages to also be included in the regression analysis for estimating flow-duration exceedance probabilities is that the observed period of record must include the drought of record for the National Weather Service climate division in which the basin is located (fig. 4, table 1). This was determined based on the monthly Palmer drought severity index (PDSI) as computed by the National Climatic Data Center for the period of record, January 1895 through April 2013 (Northeast Regional Climate Center, 2011). All reference streamgages where the period of greatest PDSI did not occur within the observed period of daily record were excluded from the regression analysis. This ensures that the observed FDCs for all streamgages include the same period of extreme drought that occurred in the 1960s defining the low-flow part of the flow-duration curve. Of the 121 reference streamgages, 90 met this additional criterion. A complete listing of streamgages used in the analysis and if they were used in regression analysis and (or) a time-series reference streamgage, is presented in [appendix 2](#).

## Record Extension Techniques

A complete daily streamflow record at the reference streamgage for the period of interest (WYs 1961 to 2010) is required to estimate daily mean streamflow using the QPPQ method for any ungaged stream location. Of the 121 selected reference streamgages, 26 had a complete record with unregulated flow for the entire period of interest. The remaining streamgages had record lengths of 9 to 88 years that included none or part of the period of interest. Streamflow records that did not include the complete period of interest were extended to complete the WY 1961 to 2010 period using the Streamflow Record Extension Facilitator (SREF; Granato,

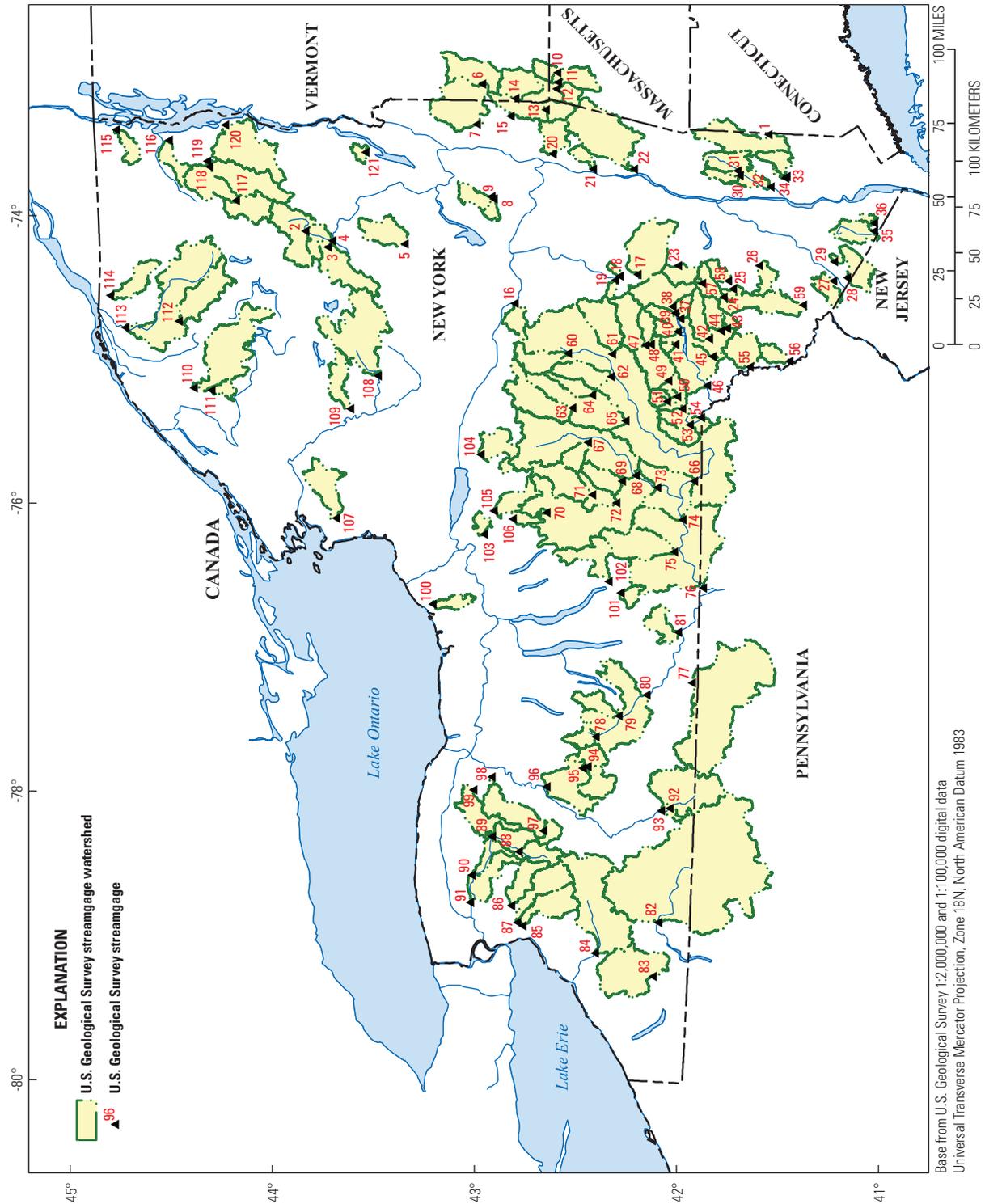


Figure 3. Location of U.S. Geological Survey reference streamgages in and near New York.

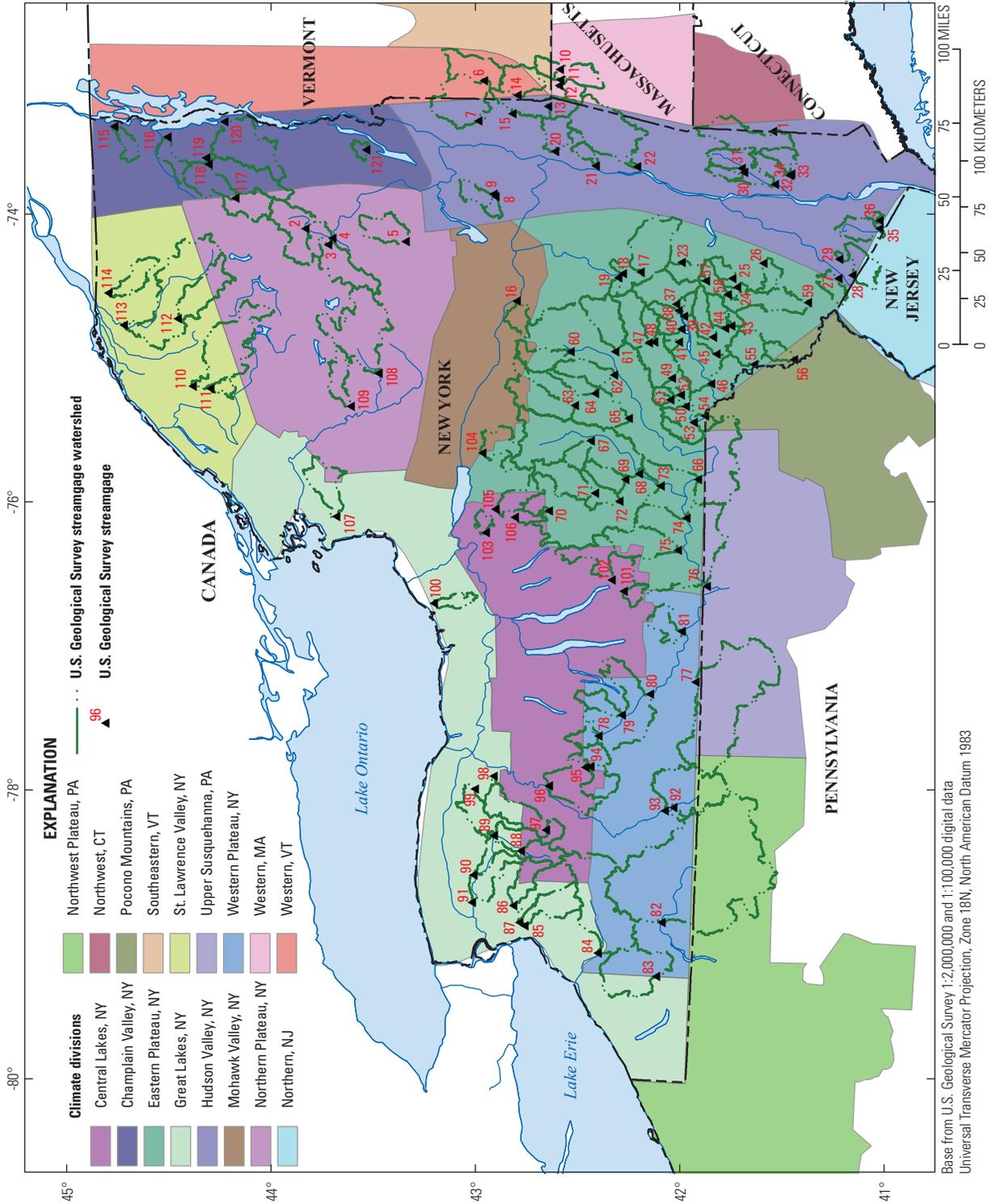


Figure 4. Location of U.S. Geological Survey reference streamgages and National Weather Service climate divisions; from Northeast Regional Climate Center (2011).

## 8 Estimation of Unaltered Daily Mean Streamflow at Ungaged Streams, Excluding Long Island, Water Years 1961–2010

**Table 1.** Period and duration of drought when the lowest Palmer drought severity index occurred during the 1960s drought in streams of New York, excluding Long Island.

[Data are from Northeast Regional Climate Center (2011). PDSI, Palmer drought severity index]

Climate division	Drought period	Duration	Lowest PDSI
Central Lakes, N.Y.	12/1960–1/1961	2 months	-4.18 in 1/1961
Champlain Valley, N.Y.	11/1964–7/1965	9 months	-4.61 in 7/1965
Eastern Plateau, N.Y.	8/1964–2/1966	19 months	-5.99 in 11/1964
Great Lakes, N.Y.	11/1960–1/1961	3 months	-5.09 in 1/1961
Hudson Valley, N.Y.	5/1964–9/1966	29 months	-6.66 in 11/1964
Mohawk Valley, N.Y.	9/1964–1/1965	5 months	-4.62 in 11/1964
Northern Plateau, N.Y.	10/1964–7/1965	10 months	-4.18 in 7/1965
St. Lawrence Valley, N.Y.	11/1960–1/1961	3 months	-3.89 in 1/1961
Upper Susquehanna, Pa.	8/1964–1/1966	18 months	-5.87 in 12/1964
Western, Mass.	6/1964–9/1966	28 months	-5.48 in 11/1964
Western, Vt.	9/1964–9/1965	13 months	-5.22 in 7/1965
Western Plateau, N.Y.	11/1960–1/1961	3 months	-4.64 in 1/1961

2009). The number of incomplete years of record during WYs 1961 to 2010 that required years of record extension range from 1 to 50 years with an average of 21 years. Streamflow data estimated by the years of record extension analysis were not used in the development of regression equations or correlation maps.

The SREF methodology relies on the assumption that long-term (greater than 10 years) streamflow records from hydrologically similar streamgages can be used to estimate missing records at a streamgage of interest (Granato, 2009). The SREF program produces estimated daily mean streamflow at streamgages with limited data for extending or augmenting the streamflow record (Granato, 2009). The line of organic correlation (LOC) regression as part of a maintenance of variance (MOVE) method is applied for record extension in SREF. The LOC for streamflow record extension can predict flows with variance and probability distribution that closely estimate those of the observed record (Helsel and Hirsch, 1992). The MOVE.3 method (Vogel and Stedinger, 1985) was used for this analysis.

For this study, streamgages used to extend the record of a streamgage with an incomplete period of record during WYs 1961 to 2010 are referred to as “index streamgages.” To ensure a reasonable application of MOVE.3, an index streamgage was considered only if there is at least 10 years of concurrent record with the incomplete streamgage record. The 10 years of concurrent record do not need to occur within the historical period of interest (WYs 1961–2010). The period-of-record at the index streamgages was analyzed for correlation and record extension procedures. The records outside of WYs 1961 to 2010 were removed after the streamflow record extensions were completed. A maximum of two streamgages was used for record extension (appendix 3). The period of available concurrent record, strength of correlation, and distribution

of LOC residuals were all considered during selection of the index streamgages. To ensure a good fit (similar hydrologic conditions) between the index streamgage and a streamgage with an incomplete period of record, the concurrent records were evaluated graphically and statistically using correlation and the coefficient of determination ( $R^2$ ) statistic. Record extension correlations ranged from 0.83 to 0.99, with a mean of 0.94. Streamflows were log transformed before LOC regression. A listing of streamgages that required record extension techniques and the index streamgages used to extend the records is provided in appendix 3.

### Regression Analysis

To estimate points along a FDC at ungaged locations, observed streamflow data through WY 2010 from 90 of the 121 reference streamgages in New York and surrounding States were used to develop regression equations for 17 percentiles along the FDC (appendix 1). Physical and climate basin characteristics known to potentially affect a range of streamflow, such as basin slope, soil properties and precipitation were computed for the basins draining to the streamgage locations and flow-duration exceedance probabilities were computed for the streamgages using the entire period of observed unregulated flow. The flow-duration exceedance probabilities (dependent variable) were related to the basin characteristics (independent or explanatory variables) using regression analysis for all 17 exceedance probabilities. To complete the FDC the remaining streamflow exceedance probabilities are determined by log-log interpolation, building a continuous daily FDC.

## Flow-Duration Curve Statistics

Using the program Get NWIS WEB Streamflow Files (GNWISQ; Granato, 2009), daily streamflow values for the selected streamgages were compiled from the National Water Information System (NWIS) Web application (<http://waterdata.usgs.gov/nwis>). GNWISQ provides an easy-to-use interface for batch downloads from NWIS and formats the retrieved files for further analysis. The downloaded observed daily streamflows were reviewed for completeness and accuracy then ranked, using the Weibull plotting position, to compute an exceedance probability for each corresponding ranked streamflow (Stedinger and others, 1993). The observed period of record for unregulated flow at a streamgage was used for computing the flow-duration exceedance probabilities used in the regression analysis. No estimated data from the record extension analysis were used to develop regression equations.

## Basin Characteristics

A total list of 85 climate and physical basin characteristics, which potentially may affect a range of streamflow exceedances, was compiled from various GIS sources (table 2). Basin characteristics derived with Arc Hydro for Esri ArcGIS 10.0 were used exclusively to improve the consistency, reproducibility, and ease-of-use of the resulting regression equations. Some basin characteristics evaluated were used in previous regression analysis in New York (Lumia and others, 2006).

Many of the physical basin characteristics including drainage area, slope and geographic locations of basin centroids and outlets were derived using digital elevation maps (DEM; U.S. Geological Survey, 2000). Most climate basin characteristics were raster-based datasets retrieved from the PRISM in 800-meter resolution for the normal (30-year) period from 1971 to 2000 (Daly, 1996). The two climate variable exceptions were mean annual runoff based on gridded contour maps produced by the USGS (Randall, 1996) and the median seasonal maximum snow depth (Cember and Wilks, 1993). Randall (1996) produced mean annual runoff maps incorporating 1951 to 1980 climate normal, streamflow and local surficial geology data. A similar analysis of mean annual runoff for New York has not been completed using more recent data; therefore, the data from Randall (1996) were considered the most reliable for this study.

The SSURGO (U.S. Department of Agriculture, Natural Resources Conservation Service, 2011) provided nearly complete raster coverages of soil properties for New York State. However, there were gaps in the data, primarily in the Adirondack region. To fill in these gaps and extend the area of soils coverage statewide, the Digital General Soil Map of the United States (STATSGO2) dataset was used in areas of incomplete coverage by SSURGO. The STATSGO2 dataset has the same data structure as SSURGO, but with a 1:250,000 resolution compared with SSURGO with a 1:24,000 resolution (U.S. Department of Agriculture, Natural

Resources Conservation Service, 2013). No STATSGO2 data were used in regression equation development, but are needed to apply the regression equations to ungaged stream locations whose basins lie in part, or completely, within areas where SSURGO data are unavailable.

## Regression Against Basin Characteristics

To relate basin characteristics to the observed flow-duration exceedance probabilities, exploratory ordinary least squares (OLS) and weighted least squares (WLS) regression techniques were applied (Tasker, 1980; Helsel and Hirsch, 1992). Flow-duration exceedance probabilities were weighted for the WLS regression technique using (number of days of record at streamgage divided by sum of days of record of all streamgages) to apply weights accounting for differing periods of record (Stuckey and others, 2012). This method results in longer periods of record being assigned greater weight in the regression with the assumption that the daily flow statistics, particularly variance, are better defined with a longer record. Regression iterations were completed using the statistical software package R (R Core Team, 2013) with additional functionality from R packages MASS (Venables and Ripley, 2002), leaps (Lumley, 2009), and MPV (Braun, 2013). The regression diagnostics used to assess the resulting regressions include graphical relations, multicollinearity, prediction error sum of squares (PRESS) statistics, variance inflation factor (VIF) statistics, standard errors, and coefficients of determination ( $R^2$ ; Helsel and Hirsch, 1992).

Data from the 90 streamgages were used to develop regression equations to estimate the 1-, 5-, 10-, 15-, 20-, 25-, 35-, 50-, 65-, 75-, 80-, 85-, 90-, 95-, and 99-percent flow-duration exceedance probabilities (P1, P5, P10, P15, P20, P25, P35, P50, P65, P75, P80, P85, P90, P95, and P99, respectively). To represent the extreme ends of the FDC from WYs 1961 to 2010, two additional regression equations were developed for the 0.0055- and 99.9945-percent flow-duration exceedance probabilities (P0.0055 and P99.9945, respectively). The P0.0055 and P99.9945-percent exceedance probabilities are the largest and smallest possible daily streamflow quantiles calculated using the Weibull plotting position for a record containing 18,262 daily streamflow values. Limited to observed data for the regression analysis, the number of streamgages used to develop the regression equations for the lower and upper ends of the FDC was restricted to those streamgages with a period of record including at least 18,262 daily streamflow values. Regression equations to estimate the P0.0055 and P99.9945 were developed using data from 46 streamgages (appendix 1) in and near New York. Additional streamgages, originally considered for inclusion in the regression analysis were removed from the analysis before the final equations were developed. Outliers and streamgages with high leverage or influence were only removed from regression analysis if sufficient data or information were found to support the removal of the streamgages, such as questionable levels of regulation

## 10 Estimation of Unaltered Daily Mean Streamflow at Ungaged Streams, Excluding Long Island, Water Years 1961–2010

**Table 2.** Basin characteristics used in the development of regression equations for flow-duration exceedance probabilities for New York streams.

[ft, feet; NLCD, National Land Cover Database; DEM, digital elevation model; mi<sup>2</sup>, square miles; mi, miles; PRISM, Parameter-Elevation Regressions on Independent Slopes Model; in., inches; °F, degrees Fahrenheit; ft/mi, feet per mile; m, meters; SSURGO, Soil Survey Geographic database; UTM, Universal Transverse Mercator]

Basin characteristic	Source	Reference
Agricultural, in percent of basin area	NLCD 2006	Fry and others (2011)
Basin at elevation greater than 1,000 ft, in percent of basin area	DEM	U.S. Geological Survey (2000)
Basin at elevation greater than 1,200 ft, in percent of basin area	DEM	U.S. Geological Survey (2000)
Basin at elevation greater than 2,000 ft, in percent of basin area	DEM	U.S. Geological Survey (2000)
Basin at elevation greater than 500 ft, in percent of basin area	DEM	U.S. Geological Survey (2000)
Drainage area, in mi <sup>2</sup>	DEM	U.S. Geological Survey (2000)
Elevation of basin centroid, in ft	DEM	U.S. Geological Survey (2000)
Elevation of basin outlet, in ft	DEM	U.S. Geological Survey (2000)
Forested, in percent of basin area	NLCD 2006	Fry and others (2011)
Impervious surface, in percent of basin area	NLCD 2006	Fry and others (2011)
Lag factor	DEM	U.S. Geological Survey (2000)
Main channel length, in mi	DEM	U.S. Geological Survey (2000)
Maximum basin elevation, in ft	DEM	U.S. Geological Survey (2000)
Mean annual maximum temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean annual minimum temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean annual precipitation, 1971–2000, 800-m normals, in in.	PRISM	Daly (1996)
Mean annual runoff, 1951–1980, in in.	Gridded contour map	Randall (1996)
Mean April precipitation, 1971–2000, 800-m normals, in in.	PRISM	Daly (1996)
Mean August precipitation, 1971–2000, 800-m normals, in in.	PRISM	Daly (1996)
Mean autumn, in September through November precipitation, 1971–2000, 800-m normals, in in.	PRISM	Daly (1996)
Mean basin elevation, in ft	DEM	U.S. Geological Survey (2000)
Mean basin slope, in degrees	DEM	U.S. Geological Survey (2000)
Mean December precipitation, 1971–2000, 800-m normals, in in.	PRISM	Daly (1996)
Mean February precipitation, 1971–2000, 800-m normals, in in.	PRISM	Daly (1996)
Mean January precipitation, 1971–2000, 800-m normals, in in.	PRISM	Daly (1996)
Mean July precipitation, 1971–2000, 800-m normals, in in.	PRISM	Daly (1996)
Mean June precipitation, 1971–2000, 800-m normals, in in.	PRISM	Daly (1996)
Mean March precipitation, 1971–2000, 800-m normals, in in.	PRISM	Daly (1996)
Mean maximum April temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean maximum August temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean maximum December temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)

**Table 2.** Basin characteristics used in the development of regression equations for flow-duration exceedance probabilities for New York streams.

[ft, feet; NLCD, National Land Cover Database; DEM, digital elevation model; mi<sup>2</sup>, square miles; mi, miles; PRISM, Parameter-Elevation Regressions on Independent Slopes Model ; in., inches; °F, degrees Fahrenheit; ft/mi, feet per mile; m, meters; SSURGO, Soil Survey Geographic database; UTM, Universal Transverse Mercator]

Basin characteristic	Source	Reference
Mean maximum February temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean maximum January temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean maximum July temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean maximum June temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean maximum March temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean maximum May temperature , 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean maximum November temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean maximum October temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean maximum September temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean May precipitation, 1971–2000, 800-m normals, in in.	PRISM	Daly (1996)
Mean minimum April temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean minimum August temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean minimum December temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean minimum February temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean minimum January temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean minimum July temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean minimum June temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean minimum March temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean minimum May temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean minimum November temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean minimum October temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean minimum September temperature, 1971–2000, 800-m normals, in °F	PRISM	Daly (1996)
Mean November precipitation, 1971–2000, 800-m normals, in in.	PRISM	Daly (1996)
Mean October precipitation, 1971–2000, 800-m normals, in in.	PRISM	Daly (1996)
Mean September precipitation, 1971–2000, 800-m normals, in in.	PRISM	Daly (1996)
Mean spring, in March through May precipitation, 1971–2000, 800-m normals, in in.	PRISM	Daly (1996)
Mean summer, in June through August precipitation, 1971–2000, 800-m normals, in in.	PRISM	Daly (1996)
Mean winter, in December through February precipitation, 1971–2000, 800-m normals, in in.	PRISM	Daly (1996)
Minimum basin elevation, in ft	DEM	U.S. Geological Survey (2000)
Overland flow distance, in m	DEM	U.S. Geological Survey (2000)
Riparian zone, in mi <sup>2</sup>	DEM	U.S. Geological Survey (2000)

## 12 Estimation of Unaltered Daily Mean Streamflow at Ungaged Streams, Excluding Long Island, Water Years 1961–2010

**Table 2.** Basin characteristics used in the development of regression equations for flow-duration exceedance probabilities for New York streams.

[ft, feet; NLCD, National Land Cover Database; DEM, digital elevation model; mi<sup>2</sup>, square miles; mi, miles; PRISM, Parameter-Elevation Regressions on Independent Slopes Model ; in., inches; °F, degrees Fahrenheit; ft/mi, feet per mile; m, meters; SSURGO, Soil Survey Geographic database; UTM, Universal Transverse Mercator]

Basin characteristic	Source	Reference
Seasonal maximum snow depth, 50th percentile	Gridded contour map	Cember and Wilks (1993)
Slope 10-85 of channel, in ft/mi	DEM	U.S. Geological Survey (2000)
Slope ratio, unitless	DEM	U.S. Geological Survey (2000)
Slope, lower half of channel, in ft/mi	DEM	U.S. Geological Survey (2000)
Slope, upper half of channel, in ft/mi	DEM	U.S. Geological Survey (2000)
Underlain by clay, in percent of basin area	SSURGO	U.S. Department of Agriculture, Natural Resources Conservation Service, 2011
Underlain by hydrologic soils group A, in percent of basin area	SSURGO	U.S. Department of Agriculture, Natural Resources Conservation Service, 2011
Underlain by hydrologic soils group AD, in percent of basin area	SSURGO	U.S. Department of Agriculture, Natural Resources Conservation Service, 2011
Underlain by hydrologic soils group B, in percent of basin area	SSURGO	U.S. Department of Agriculture, Natural Resources Conservation Service, 2011
Underlain by hydrologic soils group BD, in percent of basin area	SSURGO	U.S. Department of Agriculture, Natural Resources Conservation Service, 2011
Underlain by hydrologic soils group C, in percent of basin area	SSURGO	U.S. Department of Agriculture, Natural Resources Conservation Service, 2011
Underlain by hydrologic soils group CD, in percent of basin area	SSURGO	U.S. Department of Agriculture, Natural Resources Conservation Service, 2011
Underlain by hydrologic soils group D, in percent of basin area	SSURGO	U.S. Department of Agriculture, Natural Resources Conservation Service, 2011
Underlain by sand, in percent of basin area	SSURGO	U.S. Department of Agriculture, Natural Resources Conservation Service, 2011
Underlain by silt, in percent of basin area	SSURGO	U.S. Department of Agriculture, Natural Resources Conservation Service, 2011
Urban, in percent of basin area	NLCD 2006	Fry and others (2011)
Water, in percent of basin area	NLCD 2006	Fry and others (2011)
Wetland, in percent of basin area	NLCD 2006	Fry and others (2011)
Wetness Index, in unitless	DEM	U.S. Geological Survey (2000)
X location of basin centroid, in UTM zone 18N m	DEM	U.S. Geological Survey (2000)
X location of basin outlet, in UTM zone 18N m	DEM	U.S. Geological Survey (2000)
Y location of basin centroid, in UTM zone 18N m	DEM	U.S. Geological Survey (2000)
Y location of basin outlet, in UTM zone 18N m	DEM	U.S. Geological Survey (2000)

or unusual basin characteristics or streamflow. All outliers or streamgages with high leverage that were identified and determined to influence the regression analyses for several flow-duration exceedance probabilities were removed from all 17 regression analyses to maintain consistency. Initially, statewide regression equations were developed and residual standard errors were mapped to determine if regionalization was appropriate. To improve the regression equations, the possibility of regionalizing the State was evaluated during exploratory regression analysis using OLS and WLS. The creation of regions based on the statewide residuals, 8-digit hydrologic unit code boundaries, physiographic provinces, and (or) major basins was explored, but no overall improvement was noted. The final regression equations for the 17 flow-duration exceedance probabilities were developed using WLS on a statewide scale. An additional equation was developed as an alternate for predicting the P99.9945-percent exceedance probability in instances where the primary equation results in a higher flow estimate for the P99.9945 flow than the P99 flow, which is a physical impossibility. The alternate equation uses only the P99 flow as a predictor, forcing the P99.9945 flow to be less than the P99. Of the 46 streamgages used to develop the primary regression equation to estimate the P99.9945-percent exceedance probability flow, only 1 required the implementation of the alternate equation (appendix 1). The alternate equation only affects the flow estimates between the P99 and P99.9945-percent exceedance probability flows (the lowest 0.0055-percent of flows) for the 50-year estimated time series hydrograph. A summary of the regression equation coefficients is available in table 3.

The basin characteristics determined to be significant at the 95-percent confidence level for one or more regression equations are drainage area; mean annual runoff; percent hydrologic soils group A; percent hydrologic soils group B; mean monthly precipitation for the months of May, June and July; mean summer precipitation; maximum June temperature; X location of the basin centroid; Y location of the basin centroid; slope of the lower half of the main channel; and the percentage of basin above 1,200-foot (ft) elevation above North American Vertical Datum of 1988 (NAVD 88; table 2). Basin characteristics of the streamgage basins used in the analysis are listed in appendix 4. To form a near-linear relation between the flow-duration exceedance probabilities and basin characteristics, all independent and dependent variables were log-transformed (base 10) before regression analysis. For the basin characteristics measured as a percent, a constant of 0.0004 was added to the decimal form of the percentage for all zero values. The regression model took the following form in log units:

$$\log \hat{Q}p = A + b \log Da + c \log Mar + d \log HA + e \log HB + f \log P5 + g \log P6 + h \log P7 + i \log PS + j \log T6 + k \log Ycent + l \log Xcent + m \log Sl + n \log E12 \quad (1)$$

or in arithmetic space:

$$\hat{Q}p = 10^A (Da^b)(Mar^c)(HA^d)(HB^e)(P5^f)(P6^g)(P7^h)(PS^i)(T6^j)(Ycent^k)(Xcent^l)(Sl^m)(E12^n) \quad (2)$$

where

log	= log to base 10;
$\hat{Q}p$	= flow-duration exceedance probability, in cubic feet per second;
$A$	= the intercept;
$Da$	= drainage area, in square miles;
$Mar$	= mean annual runoff, in inches per square mile;
$HA$	= basin underlain by hydrologic soils group A, in percent;
$HB$	= basin underlain by hydrologic soils group B, in percent;
$P5$	= mean precipitation for the month of May, in inches;
$P6$	= mean precipitation for the month of June, in inches;
$P7$	= mean precipitation for the month of July, in inches;
$PS$	= mean precipitation during summer (June, July, and August), in inches;
$T6$	= maximum temperature in June temperature, in degrees Fahrenheit;
$Ycent$	= Y location of the basin centroid, in Universal Transverse Mercator meters;
$Xcent$	= X location of the basin centroid, in Universal Transverse Mercator meters;
$Sl$	= slope of the lower half of the main channel, in feet per mile;
$E12$	= basin above 1,200 ft elevation above NAVD 88, in percent; and
$b-n$	= coefficients of regression estimated by WLS.

Standard errors of prediction provide an estimate of reliability of the predicted flow-duration exceedance probabilities and are shown for the developed regression equations in table 3 (Stuckey and others, 2012). The range of standard error of prediction for the flow-duration exceedance probability regression equations is from 0.03 to 0.27 in log units (6.6 to 93 percent), with the average standard error of all 17 flow-duration equations equal to 27 percent (table 3). The highest errors occur in the regression equations estimating the low-flow portions of the FDC; the extreme low-flow exceedance probability of P99.9945 has an error of 93 percent and P99 and P95 have errors of 92 and 49 percent, respectively. The uncertainty associated with the regression equations is estimated by the  $R^2$ . In comparing the  $R^2$  for P65 (0.990) and P99.9945 (0.853), the higher  $R^2$  value indicates that 99.0 percent of the P65 value is explained by the effect of basin characteristics selected for use in the P65 regression

**Table 3.** Regression coefficients for use with flow-duration exceedance probability regression equations for New York streams, excluding Long Island.

[mi<sup>2</sup>, square miles; °F, degrees Fahrenheit; m, meters; ft/mi, feet per mile; P[n], [n] probability exceedance; --, basin characteristic not significant]

Flow duration exceedance	Basin characteristics coefficients													Coefficient of determination (R <sup>2</sup> ) adjusted for the number of predictor variables	Observed flow statistic coefficient, P99	Standard error of model
	Intercept	Drainage area, in mi <sup>2</sup>	Mean annual runoff, in inches 1951–1980	Percentage of hydrologic soils		Mean precipitation, in inches, 1971–2000			Maximum temperature in °F 1971–2000	Location of basin centroid, in m		Slope, lower half of main channel, in ft/mi	Percentage of basin above 1,200 feet elevation			
				Group A	Group B	May	June	July		Summer	X					
P0.0055	65.2397	0.9866	--	--	--	1.8825	--	--	0.2133	-9.8940	0.1172	--	0.951	0.094	23	
P1	0.3206	0.9604	1.1831	--	-0.0604	--	-1.1457	--	--	--	--	--	0.993	0.052	14	
P5	-0.3018	1.0031	1.0315	--	-0.0273	--	-0.5273	--	--	--	--	--	0.997	0.036	8.8	
P10	-0.6798	1.0118	0.9187	--	-0.0148	--	--	--	--	--	--	--	0.997	0.030	7.3	
P15	-0.6530	1.0156	0.9594	--	0.2581	--	-0.3477	--	--	--	--	--	0.998	0.028	6.6	
P20	-0.7731	1.0181	0.9581	--	0.3841	--	-0.3976	--	--	--	--	--	0.997	0.029	7.0	
P25	-0.8510	1.0163	0.9921	--	0.0210	0.5214	--	-0.5355	--	--	--	--	0.997	0.031	7.2	
P35	-1.1180	1.0165	1.0417	--	0.0385	0.6195	--	-0.5551	--	--	--	--	0.996	0.036	8.7	
P50	-2.0038	1.0166	1.0929	--	0.0533	--	0.6097	--	--	--	--	--	0.993	0.045	11	
P65	3.8155	1.0310	0.7557	--	0.1123	--	1.4385	--	-3.3025	--	--	--	0.990	0.061	14	
P75	13.1196	1.0396	--	0.0570	0.1781	--	2.8016	--	-8.3346	--	--	--	0.979	0.087	22	
P80	14.9630	1.0481	--	0.0779	0.2081	--	2.8620	--	-9.4231	--	--	--	0.973	0.102	26	
P85	16.7651	1.0554	--	0.1047	0.2402	--	2.8999	--	-10.4886	--	--	--	0.964	0.122	31	
P90	18.0498	1.0562	--	0.1342	0.2740	--	2.8915	--	-11.2622	--	--	--	0.952	0.143	38	
P95	20.5953	1.0654	--	0.1695	0.3218	--	2.8778	--	-12.7425	--	--	--	0.937	0.170	49	
P99	25.9723	1.0728	--	0.2362	0.4022	--	2.9287	--	-15.8240	--	--	--	0.908	0.229	92	
P99.9945	-7.7526	1.4689	--	--	0.5999	--	5.4351	--	--	0.5897	0.3840	--	0.853	0.266	93	
Alternate equation for 0.999945 flow duration exceedance probability																
P99.9945	-0.545800	--	--	--	--	--	--	--	--	--	--	--	1.160400	0.931	0.329	167.9

equation. Whereas only 85.3 percent of the P99.9945 value is described by its regression equation, indicating that additional variables may be affecting this extreme low-flow (table 3). The flow-duration exceedance probabilities computed from observed streamflow data and regression equations for streamgages used in the regression analysis are displayed in [appendix 1](#).

It is important to use the same data source as was used to develop the regression equations for determination of the basin characteristics to maintain consistency in the accuracy of the regression estimates. This is not possible for some areas where SSURGO data were unavailable; STATSGO2 data have been included in StreamStats to complete areas where SSURGO data are missing. Accuracy may still differ from what was found in the regression analysis.

## Estimation of Streamflow Time Series from a Reference Streamgage

Selection of an appropriate reference streamgage is crucial to achieving the most accurate estimates when applying the QPPQ method. Selecting the closest reference streamgage to the ungaged location as the best reference streamgage is intuitive, assuming the closer the ungaged location is to a reference streamgage, and the more hydrologically similar they are (Stuckey and others, 2012). However, the closest streamgage may not have the most physically or hydrologically similar basin to those at the ungaged location. The spatial map correlation method described by Archfield and Vogel (2010) determined that the closest reference streamgage is not always the best choice.

## Map Correlation

Map correlation is a geostatistical approach used to select a reference streamgage with the highest correlated streamflow with an ungaged location. When applying the map correlation method, a variogram model for each reference streamgage is developed describing a map of continuous streamflow correlation estimates between it and all other available reference streamgages in New York and surrounding States (fig. 5; Archfield and Vogel, 2010). The expected correlation at the ungaged location is estimated by applying ordinary kriging (Isaaks and Srivastava, 1989). The streamgage whose map has the highest estimated correlation coefficient at the ungaged location is selected as the most appropriate reference streamgage.

Stuckey and others (2012) determined that using the Spearman's rho correlation coefficient, a nonparametric measure of correlation, was preferable to alternate correlation metrics. Stuckey and others (2012) evaluated measures of distance and determined that using distance between basin centroids may be more appropriate than distances between basin outlets. For this study, Spearman's rho correlation centroid distances were used to apply the map correlation

method. The spherical variogram model is a broadly applied variogram function and was selected to describe the spatial structure within the correlation of daily streamflow between the reference streamgages.

Variogram models were developed for 117 reference streamgages in and near New York with minimally altered streamflow and at least 20 years of continuous record during WYs 1961 to 2010 using the Geostatistical Analyst extension of Esri ArcMAP 10.0. Variogram models for four reference streamgages with 9 to 18 years of record also were included to improve spatial coverage. The reference streamgages listed in [appendix 2](#) were used to apply the map correlation method.

It is ideal for the common period of streamflow record between reference streamgages to represent a broad range of streamflow values. One potential source of error from insufficient common periods of record is poor selection of a reference streamgage based on misleading correlation coefficients. Streamflow records of shorter duration often exhibit a more limited range of streamflows than those of longer duration, which may indicate they should be removed from the dataset due to not adequately estimating a representative historical hydrograph. No streamgage records with short common periods of record were removed in this study. A potential benefit may be gained from retaining short period of record correlation pairs because the performance of map correlation (kriging) is contingent on data density (Skoien and Blöschl, 2007; Archfield and Vogel, 2010), and data obtained from short common periods of record have value for defining the spatial-correlation structure (Stuckey and others, 2012).

## Using the New York Streamflow Estimation Tool for Estimating Unaltered Daily Mean Streamflow for Ungaged Locations

The NYSET is used to estimate daily streamflow hydrographs for WYs 1961 to 2010 at rural ungaged sites in New York using data from existing USGS streamgages coupled with explanatory basin characteristics. A graphical user interface, with an integrated spreadsheet summary report, displays the estimated daily streamflow statistics and allows water management to evaluate different scenarios or water withdrawal scenarios with estimated monthly data. This package of regression equations, USGS streamgage data, and spreadsheet application produce an interactive tool to estimate an unaltered daily streamflow hydrograph and streamflow statistics at ungaged sites in New York. The New York Streamflow Estimation Tool uses a similar model as the Massachusetts Sustainable-Yield Estimator (MA SYE; Archfield and others, 2010) and BaSE (Stuckey, 2012) tools; however, parts of both models and the original programming for BaSE have been modified for use in this application

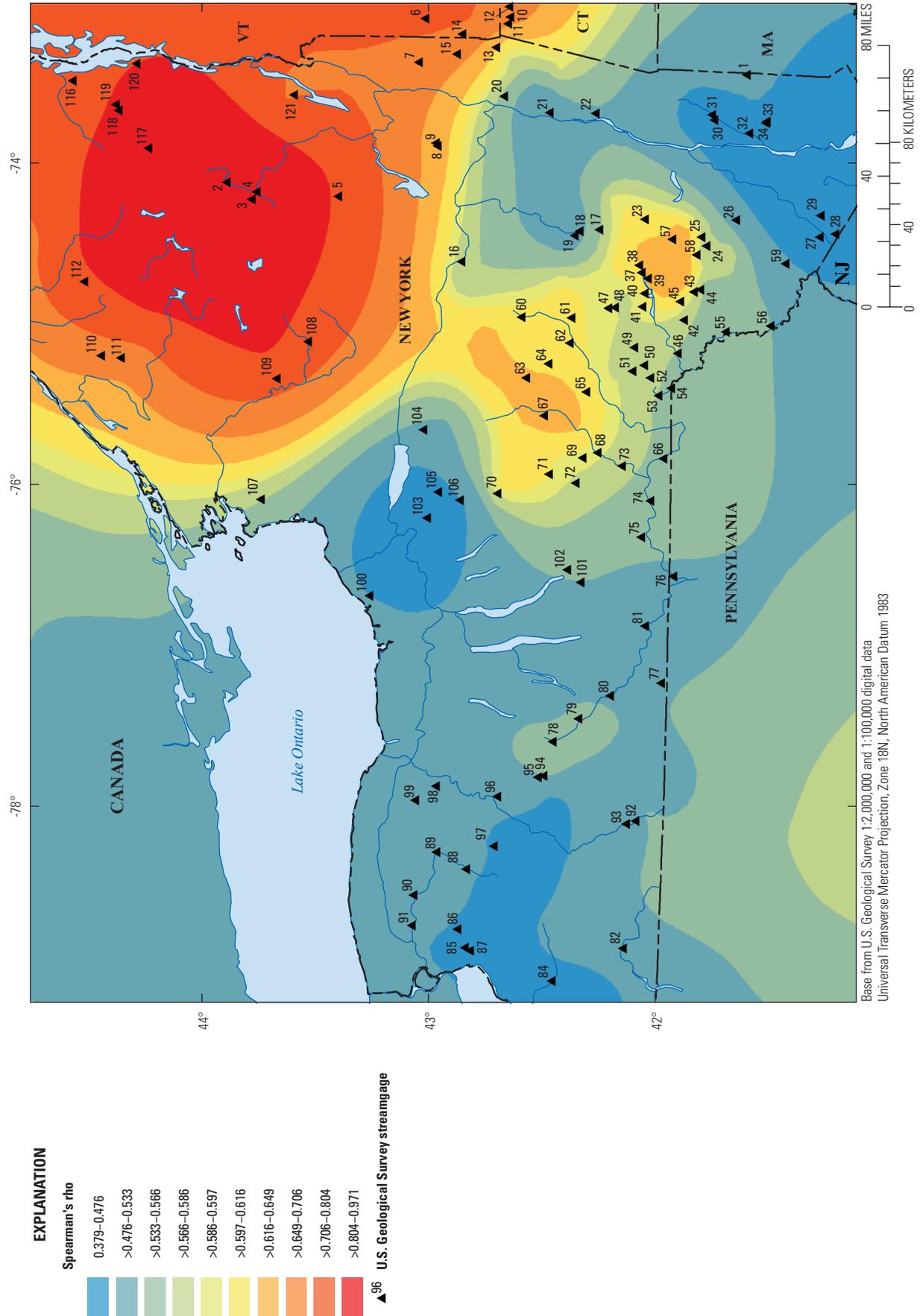


Figure 5. Example correlation map for U.S. Geological Survey streamgage 01312000, Hudson River near Newcomb, N.Y. (fig. 1, streamgage 2)

and written as a standalone application on a Visual Basic .NET (VB.NET) platform with Microsoft Excel output. The NYSET output files include reference streamgauge information, unaltered daily mean streamflow, daily and monthly flow statistics, FDCs, and hydrographs.

Basin characteristic information for the ungaged location may be calculated using USGS StreamStats for New York on the opening page of NYSET. The calculated basin characteristics may then be entered in the NYSET Basin Characteristics tab manually or imported from an ArcGIS 10.1 geodatabase file downloaded from StreamStats (U.S. Geological Survey, 2014). NYSET can import the geodatabase file downloaded from StreamStats and complete the basin characteristics screen form with the required information. By default, NYSET selects an appropriate reference streamgauge for a user-entered ungaged location by maximizing the estimated streamflow correlation. A user-specified option of manually selecting a different reference streamgauge is also available. With complete information for the ungaged location of interest entered into NYSET, the Compute Unaltered Streamflows function computes the unaltered daily mean streamflows and associated streamflow statistics for the ungaged location for WYs 1961 to 2010.

A report output file in the form of a Microsoft Excel spreadsheet is generated summarizing the reference streamgauge and ungaged location information. This summary includes basin characteristic information for the ungaged location and reference streamgages, percent difference in basin characteristics between the two locations, and any warnings associated with the basin characteristics (fig. 6). FDCs and hydrographs are presented for the ungaged location in cubic feet per second and cubic feet per second per square mile. The estimated daily flows for the ungaged location can be easily exported to a text file, which can be used in statistical software packages to determine additional daily streamflow statistics. An additional spreadsheet displaying select monthly exceedance probability flows also is created.

The user also can manipulate additional monthly withdrawals and discharges, designate monthly instream-flow targets and determine the remaining instream-flow after the user defined alterations. The Unaltered streamflow column is automatically filled in with the mean monthly flows for each month and can be edited by the user. The instream-flow target represents the user-defined minimum acceptable flows. The “Difference from target” column represents the surplus or deficit from the instream-flow target designated by the user for the exceedance or mean flows used in the “Unaltered streamflow” column. For example, if the Q50 (50-percent exceedance flow) flow in January is 100 ft<sup>3</sup>/s (unaltered streamflow) and the instream-flow target was 80 ft<sup>3</sup>/s, then there would be a surplus (difference from target) of 20 ft<sup>3</sup>/s, 50 percent of the time. If additional withdrawals were specified, the difference from target value would be reduced by those withdrawals. More detailed information, instructions for use, and all related links to files for the NYSET tool are in [appendix 5](#).

In addition to the estimated daily mean streamflow, the NYSET computes a suite of daily mean flow statistics. The statistics computed are mean, median, 7Q2, 7Q10, and monthly probability flows for the 1-, 2-, 5-, 10-, 15-, 20-, 25-, 30-, 35-, 40-, 45-, 50-, 55-, 60-, 65-, 70-, 75-, 80-, 85-, 90-, 95-, 98- and 99-percent exceedances. These statistics, including the 7Q2 and 7Q10, are computed by empirically ranking the flows using the Weibull plotting position for the period of interest.

The 7-day low-flow statistics are computed in NYSET by identifying the lowest mean consecutive 7-day period in each climate year<sup>2</sup> (CY) in the estimated streamflow dataset. The 7-day flow values calculated for each CY are then ranked to identify the 50- and 90-percent exceedances flow values (7Q2 and 7Q10, respectively). A log-Pearson type III (LP3) fit to the data is not used, though typically the USGS would use an LP3 type analysis to estimate 7-day low flow statistics (Riggs, 1972; Suro and Gazoorian, 2011). The complexity involved in fitting LP3 computations to the estimated dataset would substantially slow down the processing speed of the NYSET and add an additional layer of uncertainty by transfer of at-site skew values from a reference streamgauge. To improve performance and not introduce additional uncertainty, the *n*-day statistics computed by the NYSET are rank-based flow estimates. Based on an analysis of all reference streamgages with observed periods of record including CYs 1961 to 2009, a 49-year daily mean flow record, the empirical ranking method produces 7Q2 and 7Q10 estimates similar to those from an LP3 curve fitting (table 4). The monthly probability flows are computed by extracting each daily mean flow value that occurs in the given month for the estimated period and ranking them to identify the value of each of the probability flows estimated.

## Accuracy and Limitations of Estimated Unaltered Streamflows

Accuracy of unaltered daily streamflow estimates for ungaged locations depends on the uncertainties associated with the multiple steps, as defined by Stuckey and others (2012). The steps that may each produce individual uncertainties include (1) streamflow measurements at reference streamgages, (2) streamflow record extensions at reference streamgages, (3) selection of a reference streamgauge using the map correlation method, (4) transfer of exceedance probabilities from a reference streamgauge to the ungaged location, and (5) estimation of the FDC for an ungaged location based on regression equations and basin characteristics.

Streamflow at each reference streamgauge is determined daily during the period of record based on a stage-discharge relation unique to each individual streamgauge, which is

<sup>2</sup>A climate year runs from April 1 through March 31 and is designated by the year in which it starts.

18 Estimation of Unaltered Daily Mean Streamflow at Ungaged Streams, Excluding Long Island, Water Years 1961–2010

USGS NY WSC Streamflow Estimation Tool (MYSET)						
Summary Report by: userID			Date: 12/4/14			
Project Description:						
Reference streamgage:	01350000	Schoharie Creek at Prattsville, NY				
Correlation coefficient:	0.98					
Distance from ungaged site:	12.45					
<a href="#">Export Daily Values</a>						
<b>Streamflow statistics for ungaged site, 1961 - 2010, in cubic feet per second</b>						
Mean daily streamflow					10.40	
Median daily streamflow					5.60	
7Q2					0.99	
7Q10					0.36	
<b>Monthly streamflow statistics for ungaged site, 1961 - 2010 in cubic feet per second</b>						
	Mean	Q10	Q50	Q75	Q95	Q99
January	11.13	21.56	5.83	4.32	3.25	2.91
February	10.31	20.76	6.34	4.22	2.78	2.21
March	18.31	37.20	11.40	6.53	3.72	2.53
April	21.88	40.36	15.65	9.56	6.12	5.12
May	12.25	23.90	8.42	5.68	3.98	3.40
June	7.83	14.23	4.37	3.40	1.70	1.09
July	4.28	6.84	3.09	1.83	0.77	0.63
August	3.26	5.32	2.04	1.20	0.65	0.22
September	4.74	8.78	1.84	1.12	0.48	0.12
October	7.81	16.43	3.80	1.81	0.75	0.33
November	10.94	21.07	6.89	4.14	1.66	0.20
December	12.16	24.03	7.21	4.96	3.54	2.56
<b>Basin characteristics</b>						
	Ungaged site	Reference streamgage	Range*	Units	different from reference	
Drainage area	4.80	237.00	3.14-4,780	miles squared	-97.97	
Mean annual runoff, 1951 - 1980	30.20	27.60	11.6-37.4	inches	8.42	
Percent of basin underlain by hydrologic soils group A	1.24	2.54	0.52-51.9	percent	-1.30	
Percent of basin underlain by hydrologic soils group B	1.46	3.25	1.16-69.6	percent	-1.79	
Slope of lower half of channel	53.20	21.50	1.56-152	feet/mile	147.44	
Percent of basin above 1200 ft sea level	100.00	99.70	0.00-100	percent	0.30	
Mean May precipitation, 1971 - 2000	5.27	4.42	3.16-5.68	inches	19.23	
Mean June precipitation, 1971 - 2000	5.07	4.23	3.59-5.33	inches	19.86	
Mean July precipitation, 1971 - 2000	5.13	4.00	3.20-5.26	inches	28.25	
Mean Summer precipitation, 1971 - 2000	15.10	12.10	10.48-15.51	inches	24.79	
Maximum June temperature, 1971 - 2000	68.90	71.80	68.8-78.8	degrees Fahrenheit	-4.04	
X location of basin centroid	561544.00	560452.76	-	UTM meters	-	
Y location of basin centroid	4670100.00	4678527.28	-	UTM meters	-	
X location of basin outlet	559845.00	546444.42	-	UTM meters	-	
Y location of basin outlet	4670505.00	4685407.78	-	UTM meters	-	
<b>*Streamflow estimates may not be valid if value for ungaged site is outside of this range</b>						

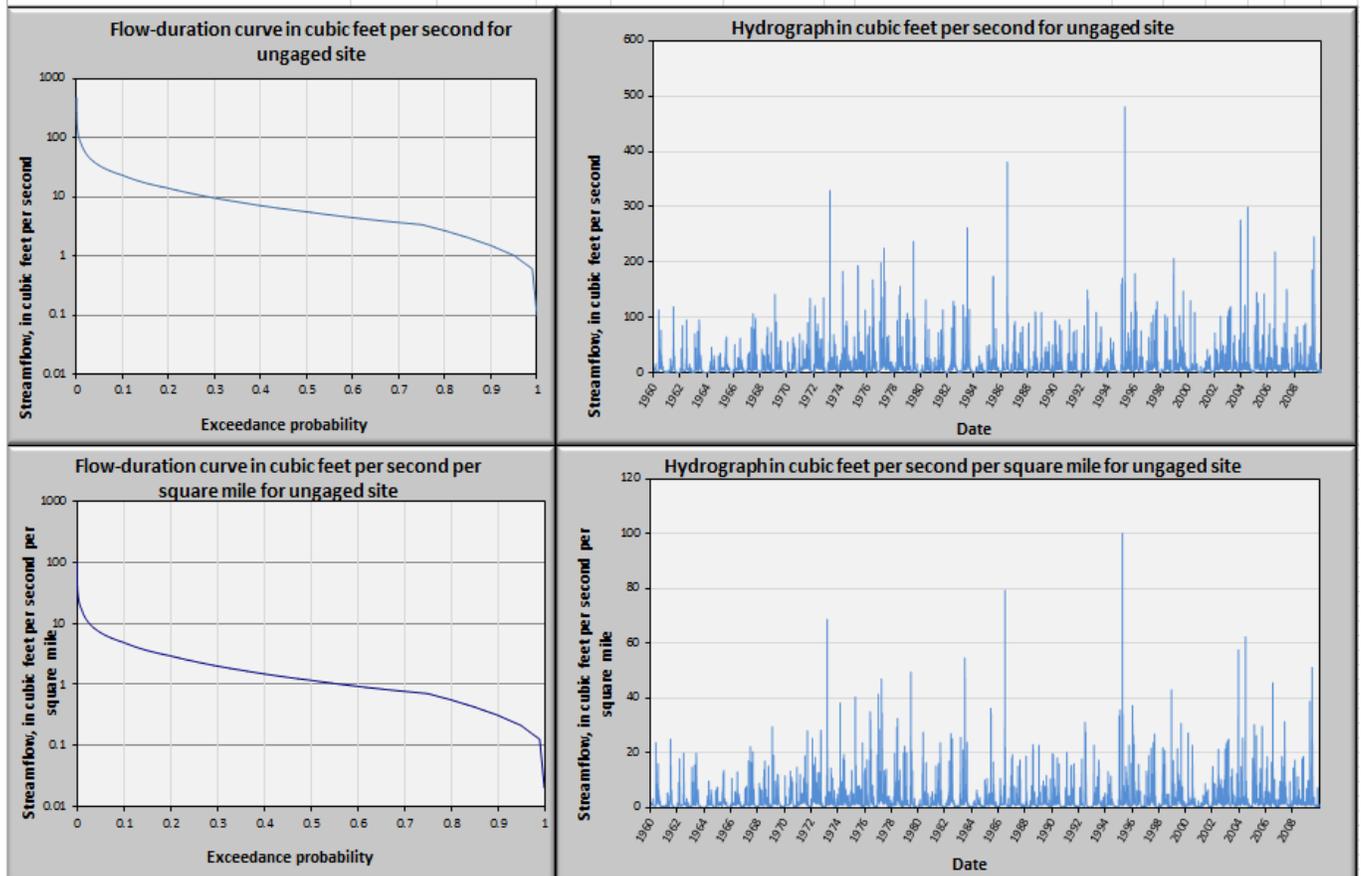


Figure 6. Screen capture of summary report generated by the New York Streamflow Estimation Tool showing flow-duration curves and a hydrograph.

**Table 4.** Comparison of low-flow statistics calculated by empirical ranking method and log-Pearson type III curve fitting at reference stations with observed flows for the period April 1, 1961, to March 31, 2010.

[Data are for climate years 1961 to 2009. ID, Identification number; mi<sup>2</sup>, square miles; 7Q2, 7-day, 2-year discharge ; ft<sup>3</sup>/s, cubic feet per second; LP3, log-Pearson type III; %, percent; 7Q10, 7-day, 10-year discharge]

U.S. Geological Survey streamgage			Drainage area, in mi <sup>2</sup>	7Q2, in ft <sup>3</sup> /s			7Q10m in ft <sup>3</sup> /s		
ID	Map number (fig. 1)	Name		LP3	Empirical ranking <sup>1</sup>	Difference, in %	LP3	Empirical ranking <sup>1</sup>	Difference, in %
01334500	15	Hoosic River near Eagle Bridge, NY	510	151	146	-3.2	94.7	95.3	0.6
01350000	17	Schoharie Creek at Prattsville, NY	237	20.8	19.1	-7.9	8.78	8.14	-7.3
01365000	24	Rondout Creek near Lowes Corners, NY	38.3	11.7	11.1	-4.5	6.12	6.29	2.7
01372500	32	Wappinger Creek near Wappingers Falls, NY	181	17.7	18.6	4.8	5.80	4.71	-18.7
01413500	37	East Branch Delaware River at Margaretville, NY	163	22.2	22.0	-0.9	10.8	10.4	-3.1
01414500	39	Mill Brook at Dunraven, NY	25.2	4.52	4.57	1.2	2.24	2.29	2.1
01420500	45	Beaver Kill at Cooks Falls, NY	241	66.5	65.1	-2.0	37.2	38.7	4.0
01423000	49	West Branch Delaware River at Walton, NY	332	43.9	41.3	-5.9	22.1	23.3	5.4
01435000	58	Neversink River near Claryville, NY	66.6	27.1	27.6	1.6	13.8	13.7	-0.7
01503000	66	Susquehanna River at Conklin, NY	2,232	311	289	-7.0	178	187	4.7
01512500	73	Chenango River near Chenango Forks, NY	1,483	242	235	-2.8	136	145	6.5
01529500	80	Cohocton River near Campbell, NY	470	46.5	46.7	0.5	26.4	26.7	1.2
03011020	82	Allegheny River at Salamanca, NY	1,608	242	238	-1.6	141	129	-8.3
04214500	85	Buffalo Creek at Gardenville, NY	142	12.8	13.0	1.4	5.93	5.57	-6.0
04215500	87	Cazenovia Creek at Ebenezer, NY	135	13.5	14.1	4.6	7.19	6.57	-8.5
04217000	89	Tonawanda Creek at Batavia, NY	171	12.7	11.7	-7.9	6.60	6.00	-9.1
04230500	98	Oatka Creek at Garbutt, NY	200	33.3	33.0	-1.0	20.4	19.0	-6.8
04231000	99	Black Creek at Churchville, NY	130	5.23	6.29	20.1	1.41	1.00	-29.2
04233000	101	Cayuga Inlet near Ithaca, NY	35.2	4.70	4.57	-2.7	2.83	2.57	-9.2
04234000	102	Fall Creek near Ithaca, NY	126	18.8	19.3	2.4	8.82	9.29	5.3
04243500	104	Oneida Creek at Oneida, NY	113	26.3	25.4	-3.2	17.3	18.3	5.8
04256000	109	Independence River at Donnatsburg, NY	88.7	33.0	33.9	2.6	22.7	21.6	-4.8

<sup>1</sup>Weibull plotting position.

constructed from measured discharge and observed stage data. The associated accuracy for each reference streamgauge is documented for each year of record in annual data reports. Most published streamflow records for stations in New York are rated as good to fair (U.S. Geological Survey, 2011). These ratings reflect that 95 percent of the data are within 10 (good) to 15 (fair) percent of their true values.

Additional uncertainty to unaltered streamflow estimates is introduced by streamflow record extension and is difficult to quantify. Because the length of record requiring extension at the reference streamgages varied from none to 100 percent of the record during the period of interest, the uncertainty introduced to the estimated record is inconsistent. A general sense of the accuracy of these estimates can be estimated by observing the correlation coefficients. For this study, the correlation coefficients range from 0.75 to 0.99 (a perfect correlation would be equal to 1). Streamgages requiring a smaller part of the record be extended generally will have less overall uncertainty than those requiring a long part of the record be extended, assuming similar quality correlations to the index streamgauge used for record extension. Correlations, and associated uncertainties, are less likely to agree when there is a nonlinear relation between the logarithm of streamflow at a streamgauge requiring record extension and the logarithm of streamflow at the index streamgauge used to apply the extension. Therefore, the overall uncertainty introduced by record extension is determined by the length of extension period, strength of index streamgauge correlation, and quality of the relation between daily streamflow and daily exceedance probabilities when streamflow is used as a surrogate for exceedance probability.

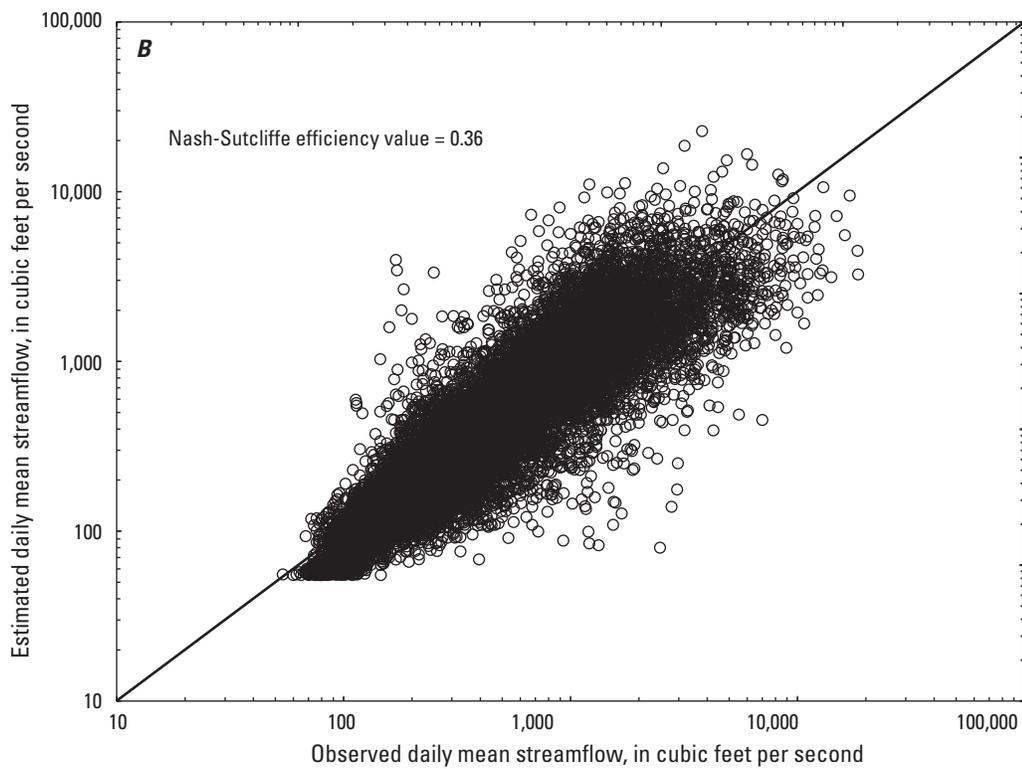
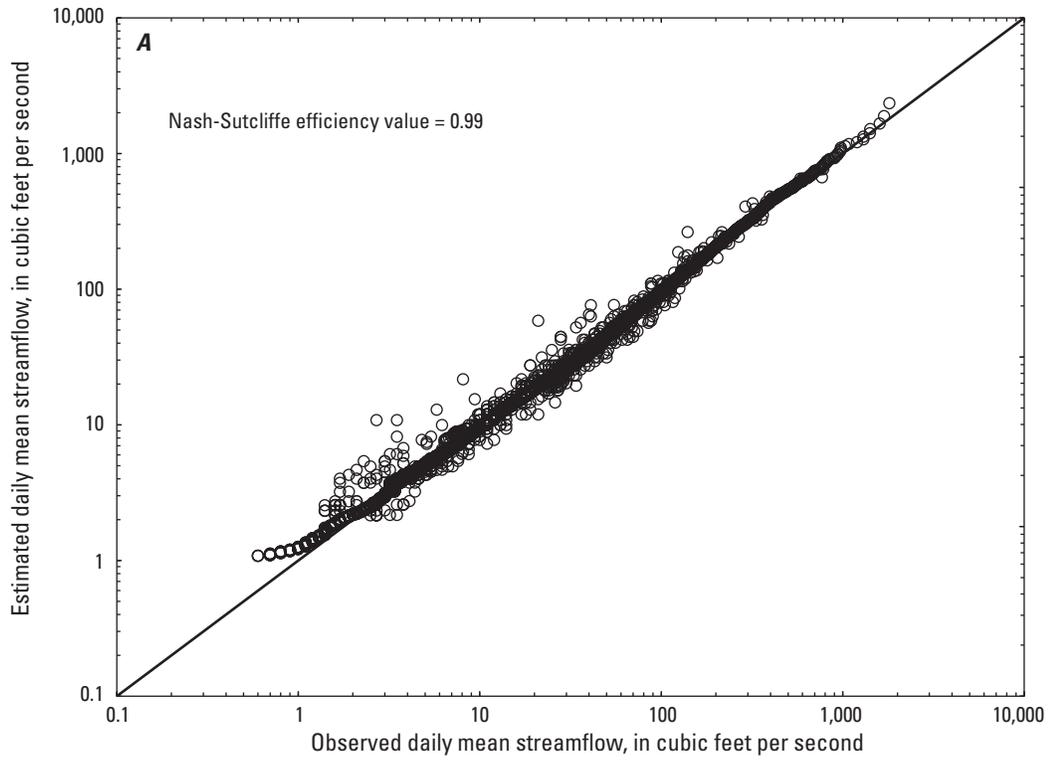
The accuracy with which the best correlated reference streamgauge is selected by map correlation affects the uncertainty in the unaltered streamflow estimates. Uncertainty is introduced into the daily exceedance probabilities for the ungaged location because even the best correlated reference streamgauge would not have a perfect correlation. As the correlation between an ungaged location and the reference streamgauge decrease, the uncertainty in unaltered streamflow estimates increase. If map correlation does not select the best correlated reference streamgauge, the uncertainty also would increase. The accuracy of map correlation to ungaged locations across New York remains unknown, but could be assessed by measuring streamflows at ungaged locations, computing correlations to reference streamgages and comparing them with the estimated correlations from the map correlation.

The QPPQ method relies on the assumption that both locations have equivalent exceedance probabilities at the same time. Although this assumption is more likely to be true for locations with similar size drainage and in close proximity to the reference streamgauge, its validity for ungaged locations statewide is unknown.

The FDC constructed for the ungaged locations by regression estimates of the 17 specified exceedance probabilities includes errors in the regressions and associated

uncertainties introduced from interpolation of the remaining exceedance probabilities occurring between the 17 exceedance probabilities on the FDC. Part of the regression errors are inherited from uncertainties of individual datasets from which basin characteristics are computed. Additionally, the dataset from 1951 to 1980 for mean annual runoff could be a possible limitation given that it is being applied to estimate streamflow during periods beginning 10 years after 1951 and ending 30 years after 1980. Standard errors of prediction for the 17 regression estimates range from 6.6 to 93 percent, and the standard error of prediction for the alternative equation is 168 percent (table 3). Archfield and others (2010) and Stuckey and others (2012) noted “hook” features when plotting observed and estimated daily mean streamflows at the extreme highest and lowest streamflows. The hook feature also was apparent when plotting the data from this study. Stuckey and others (2012) determined that a use of log-log interpolation smoothed out the hook features in the higher flows; however, a minor hook feature remained, affecting the extreme low streamflows (below P99) observed for more than the 50-year period, typically resulting in under prediction of extreme low flows. The NYSET data also indicated that a log-log interpolation method produced a straighter fit in the higher flow part of the plots, but at low flows the hook persisted. It is unclear why the hook remains in the low flow part even when using a log-log interpolation method. For development of the NYSET, the log-log interpolation method was used.

To evaluate the effectiveness of the overall QPPQ method applied within NYSET, a comparison of observed and estimated daily mean streamflows was done for each of the reference streamgages, except for two that have no unaltered flows concurrent with the historical period of interest estimated by NYSET. Only observed daily mean streamflows for WYs 1961 to 2010 were included for this comparison. Estimated streamflows were produced using the NYSET and the default option of map correlation for reference streamgauge selection. The comparison is a reasonable method to show how the NYSET can be expected to perform over the range of basin characteristics found across the State, though an analysis of streamgages not included in the model was not completed. Nash-Sutcliffe (NS) efficiency values (Nash and Sutcliffe, 1970) and normalized root-mean-square error (RMSE), as a percent, were computed as a measure of goodness of fit. NS efficiency values were evaluated for daily mean flows and the log of daily mean flows. Examples of a comparison of the observed and estimated streamflows for streamgages 01414500, Mill Brook near Dunraven, N.Y. and 04213500, Cattaraugus Creek at Gowanda, N.Y. are shown in figure 7A and B, respectively. These streamgages represent the best (01414500) and worst (04213500) NS efficiency values of the 119 sites evaluated. To estimate how well the model performed, the results were evaluated using the complete set of streamgages and also by major river basin. Median NS efficiencies for daily mean flows and log daily mean flows, statewide, were 0.84 to 0.89, respectively, and the median normalized RMSE was 2.05. The distribution of NS



**Figure 7.** Graphs showing comparison of observed and New York Streamflow Estimation Tool estimated daily mean streamflows for U.S. Geological Survey streamgages *A*, 01414500 Mill Brook near Dunraven, N.Y. and *B*, 04213500 Cattaraugus Creek at Gowanda, N.Y.

## 22 Estimation of Unaltered Daily Mean Streamflow at Ungaged Streams, Excluding Long Island, Water Years 1961–2010

efficiencies and normalized RMSE by major river basin is in tables 5 and 6 and figure 8. Based on these metrics, the NYSET performed best in the Delaware and Susquehanna River basins, indicating results in these basins are generally the most reliable. Although model performance in other areas of New York State was generally good, the diversity of basins tested resulted in a large range of NS and normalized RMSE values.

Flow-duration curves and hydrographs generated in NYSET for streamgages 04221500, Genesee River at Scio, N.Y. and 04232100, Sterling Creek at Sterling, N.Y. are shown in figure 9A and B, respectively. These streamgages represent the best (04221500) and worst (04232100) correlations with reference streamgages with observed streamflow for WY 1961 to 2010. Streamgage 04221500 was associated with reference streamgage 04221000 with a correlation of 0.999, and streamgage 04232100 was associated with reference

streamgage 04245000 with a correlation of 0.888 for estimates of daily mean flow. The period from 2002 to 2006 is shown in the hydrographs in figure 9.

NYSET uses estimates of streamflow derived from regression equations, and is not intended to be used for streams with basin characteristics outside the range used to develop the equations. Streamflow estimates may not be valid for streams with basin characteristics outside this range. The range of basin characteristics used in the development of the regression equations is shown in table 7. Estimated streamflows produced by the NYSET do not include alterations to streamflow by regulation, mining, or other large water uses. Results from the NYSET also may not be valid where groundwater and surface-water divides are not coincident. This can occur in areas with karst topography or mining activity.

**Table 5.** Nash-Sutcliffe efficiencies from comparison of the New York Streamflow Estimation Tool estimated daily-mean values to observed values at reference gages.

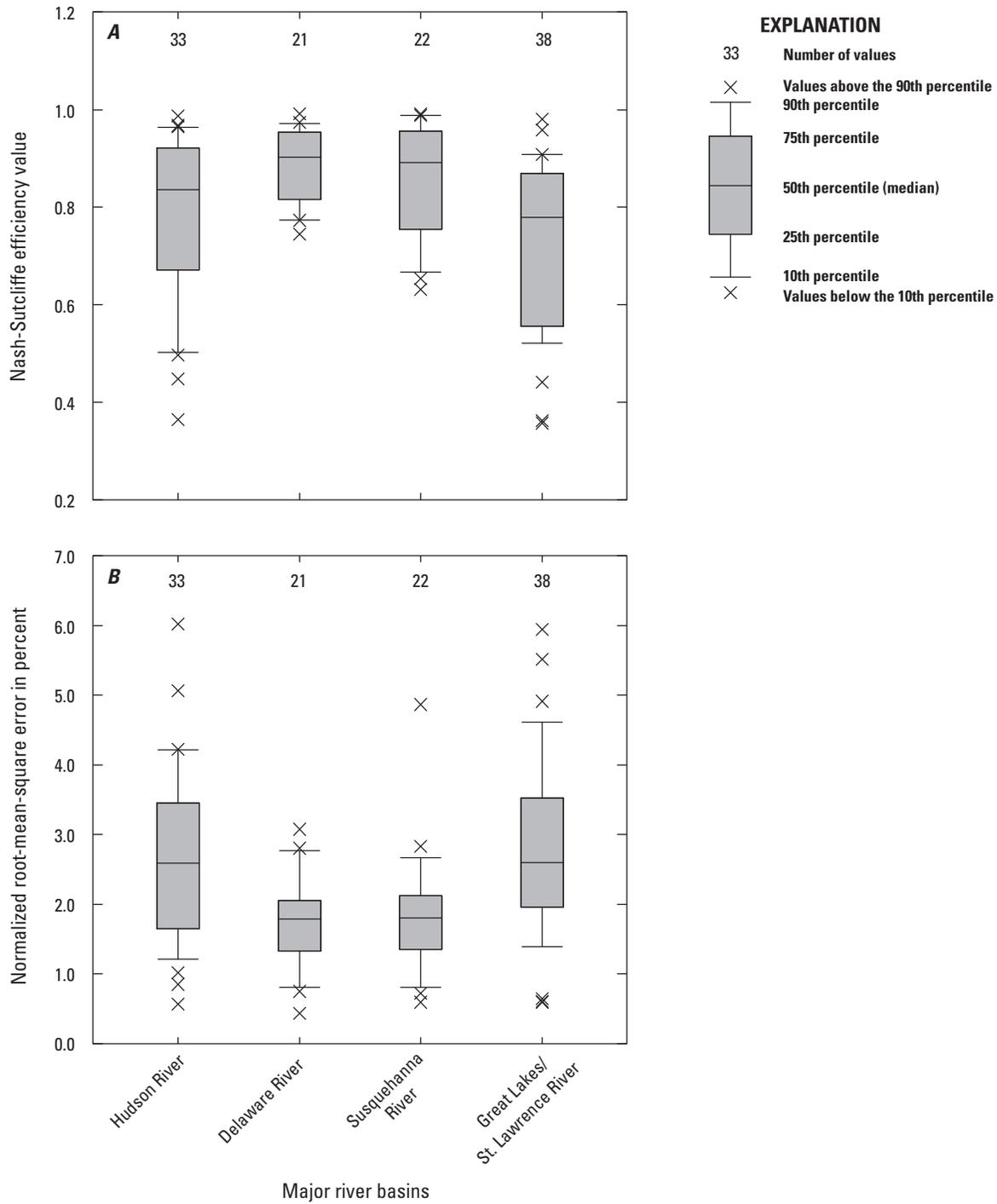
[Excludes two reference gages (01421000, 01437000) with no unregulated, observed flows concurrent with the historical period of interest]

Major river basin	Number of sites	Nash-Sutcliffe efficiency of flow				Nash-Sutcliffe efficiency of log flows			
		Minimum	Maximum	Mean	Median	Minimum	Maximum	Mean	Median
All	119	0.36	0.99	0.79	0.84	-6.98	0.99	0.75	0.89
Hudson	33	0.36	0.99	0.78	0.84	-1.98	0.97	0.76	0.89
Delaware	21	0.74	0.99	0.88	0.90	0.49	0.99	0.88	0.90
Susquehanna	22	0.63	0.99	0.89	0.86	0.76	0.99	0.91	0.93
Great Lakes, Lake Champlain, St. Lawrence	38	0.36	0.98	0.72	0.78	-6.98	0.98	0.72	0.78
Other	5	0.47	0.91	0.71	0.72	0.76	0.94	0.86	0.88

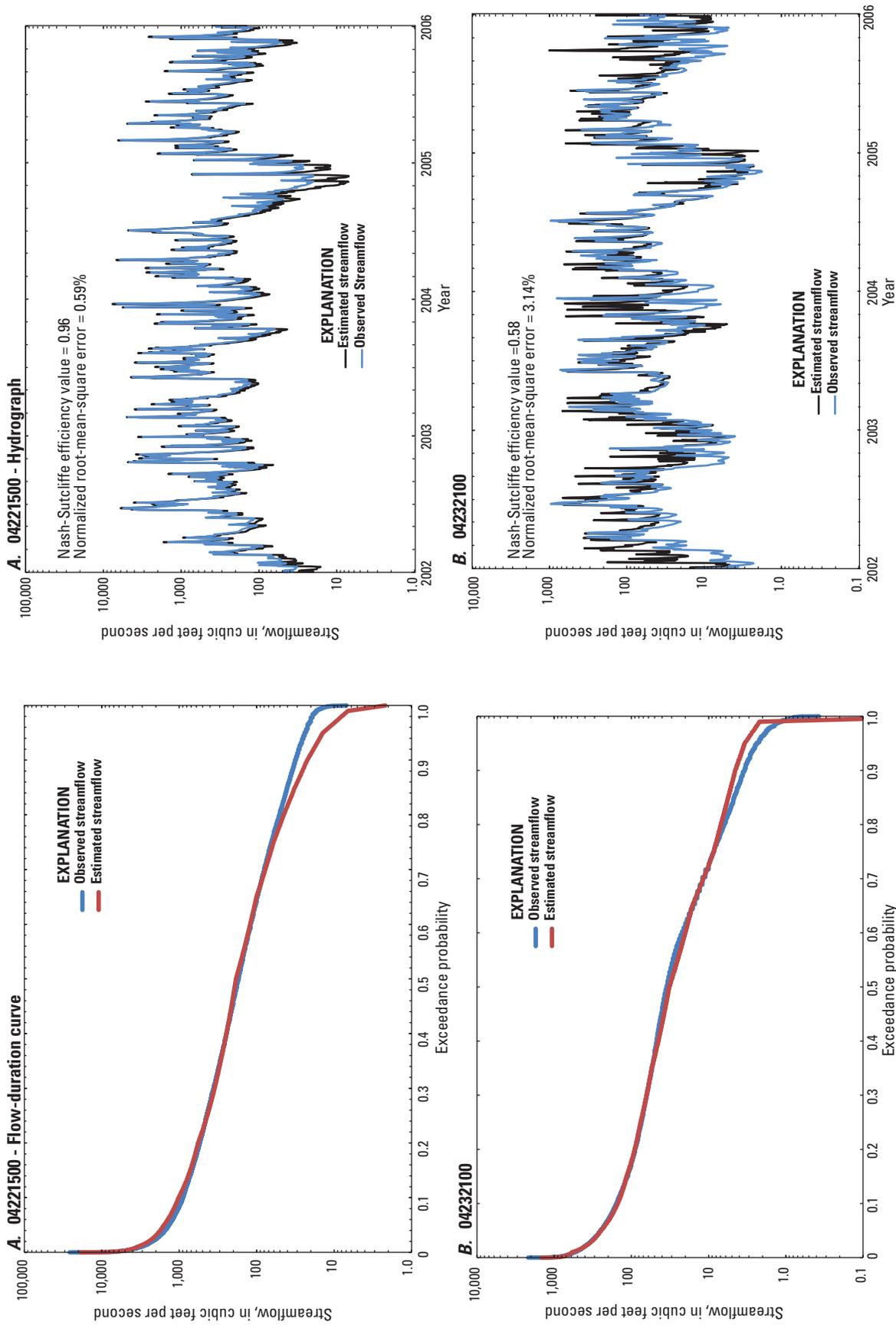
**Table 6.** Normalized root-mean-square errors as percentage from comparison of the New York Streamflow Estimation Tool estimated daily-mean values to observed values at reference gages.

[Excludes two reference gages (01421000, 01437000) with no unregulated, observed flows concurrent with the historical period of interest]

Major river basin	Number of sites	Normalized root-mean-square error			
		Minimum	Maximum	Mean	Median
All	119	0.43	6.34	2.38	2.05
Hudson	33	0.56	6.02	2.64	2.59
Delaware	21	0.43	3.08	0.88	0.90
Susquehanna	22	0.59	4.87	1.84	1.80
Great Lakes, Lake Champlain, St. Lawrence	38	0.59	5.95	0.72	0.78
Other	5	1.40	6.34	2.91	2.32



**Figure 8.** Boxplots showing distribution of, *A*, Nash-Sutcliffe efficiency values and *B*, root-mean-square error by major river basin obtained from comparison between observed and estimated daily mean streamflows.



**Figure 9.** Flow-duration curves and hydrographs showing estimated and observed daily mean flows for U.S. Geological Survey streamgages, *A*, 04221500, Genesee River at Scio, N.Y. and, *B*, 04232100, Sterling Creek at Sterling, N.Y. %, percent.

**Table 7.** Basin characteristics used in development of regression equations to estimate flow-duration exceedance probabilities with the New York Streamflow Estimation Tool for basins in New York.

Basin characteristic	Minimum	Maximum	Mean
Drainage area, in square miles	3.14	4,780	288
Mean annual runoff, 1951–1980, in inches	11.6	37.4	22.9
Percent hydrologic soils group A	0.62	51.2	6.52
Percent hydrologic soils group B	1.14	65.7	16.8
Mean May precipitation, 1971–2000, in inches	3.15	5.68	4.10
Mean June precipitation, 1971–2000, in inches	3.59	5.33	4.32
Mean July precipitation, 1971–2000, in inches	3.20	5.26	4.12
Mean summer precipitation, 1971–2000, in inches	10.5	15.5	12.5
Maximum June temperature, in degrees Fahrenheit, 1971–2000	68.8	78.8	74.2
X location of basin centroid, in meters	166,690.76	657,531.03	480,630.94
Y location of basin centroid, in meters	4,557,789.39	4,913,180.43	4,713,574.01
Slope, lower half of main channel, in feet per mile	1.56	152	25.3
Percent of basin above 1,200 feet elevation	0	100	66.6

## Summary

The ability to estimate daily mean streamflow for any location on a stream in New York can aid in managing the water resources of the State. Time series hydrologic data are essential to understanding ways to promote healthy instream ecology and to strengthen the scientific basis for sound water management decision making in New York. Generating daily mean flows at ungaged streams, allows for estimating streamflow statistics such as flow-duration exceedances, and helps water managers to understand the natural flow regime of a stream, which is critical to the sustainability and health of aquatic freshwater ecosystems.

The U.S. Geological Survey, in cooperation with The Nature Conservancy and the New York State Energy Research and Development Authority, has developed the New York Streamflow Estimation Tool (NYSET) to estimate a continuous daily mean streamflow hydrograph for the period from October 1, 1960, to September 30, 2010, at ungaged locations across New York State. The NYSET equates the streamflow as a percentile from the flow-duration curve (FDC) for a particular day at an ungaged location with the streamflow as a percentile for the same day at a reference streamgage by applying a modified QPPQ approach. Estimated streamflow correlations at ungaged locations, by map correlation, is used to select an appropriate reference streamgage for streamflow estimation. Basin characteristics are used to develop regression equations to construct a daily FDC for ungaged locations. The estimated FDC is used to select streamflow percentiles corresponding to those at the reference streamgage.

Regression equations were developed to predict flow-duration exceedance probabilities for 17 percentiles along the FDC using data from 90 streamgages. The standard errors of

prediction for the flow-duration regression equations range from 6.6 to 93 percent, with the average standard error for the suite of equations equal to 27 percent. The map correlation method was used for the selection of a reference streamgage with minimally altered streamflow in and near New York using a spherical variogram model.

The NYSET generates daily mean streamflow estimates for an ungaged location by first selecting a reference streamgage using the map correlation method as a default. An option exists for manual selection by the user. After a reference streamgage is identified, the NYSET then equates the percentiles at the gaged site with percentiles at the ungaged location for each date from October 1, 1960, to September 30, 2010. Regression equations and interpolation are used to convert the percentiles to streamflow at the ungaged location. The NYSET outputs a summary report in the form of a Microsoft Excel spreadsheet. This summary includes basin characteristic information for the ungaged location and reference streamgages, percent difference in basin characteristics between the two locations, and any warnings associated with the basin characteristics. Mean and median streamflows, 7Q2, 7Q10, and select monthly flow statistics are computed for the ungaged location. FDCs and hydrographs are presented for the ungaged location in cubic feet per second and cubic feet per second per square mile. The estimated daily flows for the ungaged location can be easily exported to a text file, which can be used in a statistical software package to determine additional daily streamflow statistics.

Accuracy of estimated unaltered daily mean streamflow for ungaged locations is affected by the uncertainties introduced by the multiple steps involved in the process. Uncertainty is introduced during the selection of the reference streamgage, estimation of exceedance probabilities for the

ungaged location, and the QPPQ process. It is difficult to quantify the overall uncertainty associated with the estimated daily mean flows at the ungaged location because of the number of potential sources. Observed and estimated daily mean flows for a subset of reference streamgages were examined for accuracy. Nash-Sutcliffe (NS) efficiency values were evaluated for daily mean flows and the log of daily mean flows. To estimate how well the model performed, the results were evaluated using the complete set of streamgages and by subsetting them into groups based on which major river basin each site is in. Median NS efficiencies for daily mean flows and log daily mean flows, statewide, were 0.84 to 0.89, respectively, and the median normalized root-mean-square error was 2.05.

The New York Streamflow Estimation Tool NYSET uses estimates of streamflow derived from regression equations, and is not intended to be used for streams with basin characteristics outside the range used to develop the equations. Streamflow estimates may not be valid for streams with basin characteristics outside this range. Estimated streamflows produced by the NYSET do not include alterations to streamflow by regulation, mining, or other large water uses. Results from the NYSET also may not be valid where groundwater and surface-water divides are not coincident. This can occur in areas with karst topography or mining activity.

## References Cited

- Archfield, S.A., and Vogel, R.M., 2010, Map correlation method—Selection of a reference streamgage to estimate daily streamflow at ungaged catchments: *Water Resources Research*, v. 46, no. 10, 15 p. [Also available at <http://dx.doi.org/10.1029/2009WR008481>.]
- Archfield, S.A., Vogel, R.M., Steeves, P.A., Brandt, S.L., Weiskel, P.K., and Garabedian, S.P., 2010, The Massachusetts sustainable-yield estimator—A decision-support tool to assess water availability at ungaged stream locations in Massachusetts: U.S. Geological Survey Scientific Investigations Report 2009–5227, 41 p., plus CD-ROM. [Also available at <http://pubs.usgs.gov/sir/2009/5227/>.]
- Barnes, C.R., 1986, Method for estimating low-flow statistics for ungaged streams in the lower Hudson River basin, New York: U.S. Geological Survey Water-Resources Investigations Report 85–4070, 22 p. [Also available at <http://pubs.er.usgs.gov/publication/wri854070>.]
- Braun, W.J., 2013, MPV—Data sets from Montgomery, Peck and Vining’s book: The Comprehensive R Archive Network Web page, accessed December 5, 2014, at <http://cran.r-project.org/web/packages/MPV/index.html>.
- Cember, R.P., and Wilks, D.S., 1993, Climatological atlas of snowfall and snow depth for the northeastern United States and southeastern Canada: Ithaca, N.Y., Northeast Regional Climate Center Publication RR 93–1, 18 p.
- Daly, Christopher, 1996, Overview of the PRISM model: PRISM Climate Group Web page, accessed December 7, 2011, at <http://www.ocs.orst.edu/prism/overview.html>.
- Eissler, B.B., 1979, Low-flow frequency analysis of streams in New York: New York State Department of Environmental Conservation Bulletin 74, 176 p., 3 pls.
- Fenneman, N.M., 1938, Physiography of eastern United States: New York, McGraw-Hill, 714 p.
- Fennessey, N.M., 1994, A hydro-climatological model of daily streamflow for the northeast United States: Medford, Mass., Tufts University, Ph.D. dissertation, [variously paged].
- Fry, J.A., Xian, George, Jin, Suming, Dewitz, J.A., Homer, C.G., Yang, Limin, Barnes, C.A., Herold, N.D., and Wickham, J.D., 2011, Completion of the 2006 National Land Cover Database for the conterminous United States: *PE&RS*, v. 77, no. 9, p. 858–864.
- Granato, G.E., 2009, Computer programs for obtaining and analyzing daily mean streamflow data from the U.S. Geological Survey National Water Information System web site: U.S. Geological Survey Open-File Report 2008–1362, 123 p., 5 appendixes, CD-ROM. [Also available at <http://pubs.usgs.gov/of/2008/1362/>.]
- Helsel, D.R., and Hirsch, R.M., 1992, Statistical methods in water resources: Amsterdam, Elsevier, *Studies in Environmental Science* 49, 522 p.
- Hughes, D.A., and Smakhtin, V.U., 1996, Daily flow time series patching or extension—A spatial interpolation approach based on flow-duration curves: *Hydrological Sciences Journal*, v. 41, no. 6, p. 851–871.
- Isaaks, E.H., and Srivastava, R.M., 1989, An Introduction to applied geostatistics: New York, Oxford University Press, 561 p.
- Linhart, S.M., Nania, J.F., Christiansen, D.E., Hutchinson, K.J., Sanders, C.L., Jr., and Archfield, S.A., 2013, Comparison between two statistically based methods, and two physically based models developed to compute daily mean streamflow at ungaged locations in the Cedar River basin, Iowa: U.S. Geological Survey Scientific Investigations Report 2013–5111, 7 p. [Also available at <http://pubs.usgs.gov/sir/2013/5111/>.]
- Lumia, Richard, Freehafer, D.A., and Smith, M.J., 2006, Magnitude and frequency of floods in New York: U.S. Geological Survey Scientific Investigations Report 2006–5112, 152 p. [Also available at <http://pubs.usgs.gov/sir/2006/5112/>.]

- Lumley, Thomas, 2009, Leaps—Regression subset selection: The Comprehensive R Archive Network Web page, accessed December 5, 2014, at <http://cran.r-project.org/web/packages/leaps/index.html>.
- Mohamoud, Y.M., 2008, Prediction of daily flow-duration curves and streamflow for ungauged catchments using regional flow-duration curves: *Hydrological Sciences Journal*, v. 53, no. 4, p. 706–724.
- Nash, J.E., and Sutcliffe, J.V., 1970, River flow forecasting through conceptual models; part I—A discussion of principles: *Journal of Hydrology*, v. 10, no. 3, p. 282–290.
- Northeast Regional Climate Center, 2011, U.S. drought monitor—Northeast: Ithaca, N.Y., Northeast Regional Climate Center Northeast Drought database, accessed November 4, 2011, at [http://www.nrcc.cornell.edu/page\\_drought.html](http://www.nrcc.cornell.edu/page_drought.html).
- Poff, N.L., and Zimmerman, J.K.H., 2010, Ecological responses to altered flow regimes—A literature review to inform the science and management of environmental flows: *Freshwater Biology*, v. 55, no. 1, p. 194–205.
- R Core Team, 2013, R—A language and environment for statistical computing: R Foundation for Statistical Computing Web page, accessed December 5, 2014, at <http://www.R-project.org/>.
- Randall, A.D., 1996, Mean annual runoff, precipitation, and evapotranspiration in the glaciated northeastern United States, 1951–80: U.S. Geological Survey Open-File Report 96–395, 2 sheets. [Also available at <http://ny.water.usgs.gov/pubs/of/of96395/OF96-395.html>.]
- Randall, A.D., 2010, Low flow of streams in the Susquehanna River basin of New York: U.S. Geological Survey Scientific Investigations Report 2010–5063, 57 p., accessed December 5, 2014, at <http://pubs.usgs.gov/sir/2010/5063/>.
- Riggs, H.C., 1972, Low-flow investigations: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. B1, 18 p. [Also available at [http://pubs.usgs.gov/twri/twri4b1/pdf/twri\\_4-B1\\_a.pdf](http://pubs.usgs.gov/twri/twri4b1/pdf/twri_4-B1_a.pdf).]
- Searcy, J.K., 1959, Flow-duration curves, part 2 of Manual of hydrology—Low flow techniques: U.S. Geological Survey Water-Supply Paper 1542–A, p. 1–33. [Also available at <http://pubs.usgs.gov/wsp/1542a/report.pdf>.]
- Shu, Chang, and Ourda, T.B.M.J., 2012, Improved methods for daily streamflow estimates at ungauged sites: *Water Resources Research*, v. 48, no. 2, 15 p. [Also available at <http://dx.doi.org/10.1029/2011WR011501>.]
- Skoien, J.O., and Blöschl, Günter, 2007, Spatiotemporal topological kriging of runoff time series: *Water Resources Research*, v. 43, no. 9, 21 p. [Also available at <http://dx.doi.org/10.1029/2006WR005760>.]
- Smakhtin, V.Yu., 1999, Generation of natural daily flow time-series in regulated rivers using non-linear spatial interpolation technique: *Regulated Rivers Research and Management*, v. 15, no. 4, p. 311–323.
- Smakhtin, V.Yu., and Masse, B., 2000, Continuous daily hydrograph simulation using duration curves of a precipitation index: *Hydrological Processes*, v. 14, no. 6, p. 1083–1100.
- Stedinger, J.R., and Tasker, G.D., 1985, Regional hydrologic analysis—1. Ordinary, weighted and generalized least squares compared: *Water Resources Research*, v. 21, no. 9, p. 1421–1432.
- Stedinger, J.R., Vogel, R.M., and Foufoula-Georgiou, Efi, 1993, Frequency analysis of extreme events, chap. 18 of Maidment, D.R., ed., *Handbook of hydrology*: New York, McGraw-Hill, 1,424 p.
- Stuckey, M.H., Koerke, E.H., and Ulrich, J.E., 2012, Estimation of baseline daily mean streamflows for ungauged locations on Pennsylvania streams, water years 1960–2008: U.S. Geological Survey Scientific Investigations Report 2012–5142, 61 p. [Also available at <http://pubs.usgs.gov/sir/2012/5142/>.]
- Suro, T.P., and Gazoorian, C.L., 2011, Changes in low-flow frequency from 1976–2006 at selected streamgages in New York, excluding Long Island: U.S. Geological Survey Scientific Investigations Report 2011–5112, 21 p. [Also available at <http://pubs.usgs.gov/sir/2011/5112/>.]
- Tasker, G.D., 1980, Hydrologic regression with weighted least squares: *Water Resources Research*, v. 16, no. 6, p. 1107–1113.
- U.S. Department of Agriculture, Natural Resources Conservation Service, 2011, Web soil survey [database for New York]: U.S. Department of Agriculture, Natural Resources Conservation Service, accessed August 2011, at <http://websoilsurvey.nrcs.usda.gov/>.
- U.S. Department of Agriculture, Natural Resources Conservation Service, 2013, Web soil survey [database for New York]: U.S. Department of Agriculture, Natural Resources Conservation Service, accessed August 2013, at <http://websoilsurvey.nrcs.usda.gov/>.
- U.S. Geological Survey, 2000, US GeoData digital elevation models: U.S. Geological Survey Fact Sheet 00–040, 2 p. [Also available at <http://pubs.er.usgs.gov/publication/fs04000>.]
- U.S. Geological Survey, 2012, Water data for the nation: U.S. Geological Survey National Water Information System Web page, accessed June 23, 2011, at <http://waterdata.usgs.gov/nwis/>.

U.S. Geological Survey, 2014, New York: U.S. Geological Survey StreamStats Program Web page, accessed December 5, 2014, at [http://water.usgs.gov/osw/streamstats/new\\_york.html](http://water.usgs.gov/osw/streamstats/new_york.html).

Venables, W.N., and Ripley, B.D., 2002, *Modern applied statistics with S* (4th ed.): New York, Springer, 495 p.

Vogel, R.M., and Stedinger, J.R., 1985, Minimum variance streamflow record augmentation procedures: *Water Resources Research*, v. 21, no. 5, p. 715–723.

Vogel, R.M., Sieber, Jack, Archfield, S.A., Smith, M.P., Apse, C.D., and Huber-Lee, Annette, 2007, Relations among storage, yield, and instream flow: *Water Resources Research*, v. 43, no. 5, paper W05403, 12 p., accessed August 2013, at <http://dx.doi.org/10.1029/2006WR005226>.

## Appendixes 1–5

[Available at <http://dx.doi.org/10.3133/sir20145220>]

---

- Appendix 1. Flow-Duration Exceedance Probabilities, Observed and Computed From Stream flow Data, and Regression Equations for Streamgages Used in Regression Analysis.
- Appendix 2. Description of Reference Streamgages Used in the Development of the New York Streamflow Estimation Tool With Period of Record and Use of Data.
- Appendix 3. Reference Streamgages With Index Streamgage Used for MOVE.3 Record Extension.
- Appendix 4. Basin Characteristics Used in the Development of Flow-Duration Regression Equations.
- Appendix 5. User's Guide for the New York Streamflow Estimation Tool (NYSET).



Prepared by the Pembroke Publishing Service Center.

For more information concerning this report, contact:

Director  
New York Water Science Center  
U.S. Geological Survey  
425 Jordan Road  
Troy, NY 12180-8349  
dc\_ny@usgs.gov

or visit our Web site at:  
<http://ny.water.usgs.gov>

