

# **Estimation of Low-Flow Statistics at Ungaged Sites on Streams in the Lower Hudson River Basin, New York, From Data in Geographic Information Systems**



Scientific Investigations Report 2017–5019

**Cover.** Horse Heaven Brook at Oak Hill Road, Sand Lake, Rensselaer County, New York, station 01359060, drainage area 1.4 square miles, latitude 42°38'20" N, longitude 73°32'17" W, on September 26, 2010, streamflow 0.013 cubic feet per second. Photograph by Allan Randall.

# **Estimation of Low-Flow Statistics at Ungaged Sites on Streams in the Lower Hudson River Basin, New York, From Data in Geographic Information Systems**

By Allan D. Randall and Douglas A. Freehafer

Scientific Investigations Report 2017–5019

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

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

RYAN K. ZINKE, Secretary

**U.S. Geological Survey**

William H. Werkheiser, Acting Director

U.S. Geological Survey, Reston, Virginia: 2017

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit <https://www.usgs.gov> or call 1–888–ASK–USGS (1–888–275–8747).

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

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

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

Suggested citation:

Randall, A.D., and Freehafer, D.A., 2017, Estimation of low-flow statistics at ungaged sites on streams in the Lower Hudson River Basin, New York, from data in geographic information systems: U.S. Geological Survey Scientific Investigations Report 2017–5019, 42 p., <https://doi.org/10.3133/sir20175019>.

ISSN 2328-0328 (online)

## Contents

Abstract.....	1
Introduction.....	1
Calculation of Low-Flow Indices.....	2
Compilation of Watershed Properties.....	7
Drainage Area .....	7
Surficial Geology.....	7
Amount of Water Available for Runoff .....	8
Riparian Evapotranspiration Losses.....	8
Watershed Geometry.....	8
Land Cover .....	11
Bedrock Geology.....	11
Regression Analysis Using Ordinary Least Squares .....	12
Evaluation of Watershed Properties.....	13
Surficial Geology.....	13
Amount of Water Available for Runoff .....	21
Riparian Evapotranspiration Loss .....	22
Watershed Geometry .....	22
Evaluation of Sources of Predictive Error in Regression Equations .....	22
Logarithmic Transformations .....	25
Regression Analysis Using Weighted Least Squares.....	27
Suggested Methods for Estimating Statistical Indices of Low Flow at Ungaged Sites in the Lower Hudson River Basin .....	31
Summary.....	32
References Cited.....	33
Appendix 1. Aggregation of Soil Series Into Surficial Geology Units and Sources of Abundant Evapotranspiration for Use in Regression Analysis.....	37

## Figures

1. Map showing location of Lower Hudson River Basin and watersheds from which data are analyzed in this report.....3
2. Graphs showing low flows calculated from streamflow measurements compared with low flows estimated by four ordinary least squares regression equations and with four watershed properties used in several equations. ....26
3. Graphs showing low flows calculated from streamflow measurements compared with low flows estimated from weighted least squares equations 7 and 9, also comparing residuals from these equations with ranks of residuals.....30

Tables

- 1. Low-flow indices and drainage areas for watersheds in the Lower Hudson River Basin, New York .....4
- 2. Areas of surficial geologic units and poorly drained soil categories for watersheds in the Lower Hudson River Basin, New York.....9
- 3. Ordinary least squares regression equations for estimating low flow of streams in the Lower Hudson River Basin, New York.....14
- 4. Values of watershed properties that represent water available for runoff, stream geometry, and urban land cover for watersheds in the Lower Hudson River Basin, New York .....19
- 5. Comparison of area of Hydrologic Soil Group A, as delineated by the Natural Resources Conservation Service, with areas of two surficial geology variables compiled for this report in the Lower Hudson River Basin, New York.....21
- 6. Comparison of ordinary least squares equations 1 and 2 (table 3) with logarithmic transformations of those equations.....27
- 7. Weighted least squares regression equations for estimating low flow of streams in the Lower Hudson River Basin, New York.....29

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
Length		
inch (in.)	2.54	centimeter (cm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
Area		
acre	4,047	square meter (m <sup>2</sup> )
acre	0.4047	hectare (ha)
acre	0.004047	square kilometer (km <sup>2</sup> )
square mile (mi <sup>2</sup> )	259.0	hectare (ha)
square mile (mi <sup>2</sup> )	2.590	square kilometer (km <sup>2</sup> )
Volume		
gallon (gal)	3.785	liter (L)
gallon (gal)	0.003785	cubic meter (m <sup>3</sup> )
million gallons (Mgal)	3,785	cubic meter (m <sup>3</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 (ft <sup>3</sup> /s)	28.32	liter per second (L/s)
Hydraulic gradient		
foot per mile (ft/mi)	0.1894	meter per kilometer (m/km)

## Datum

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

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

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

## Abbreviations

7Q2	7-day, 2-year low flow
7Q10	7-day, 10-year low flow
GIRAS	Geographic Information Retrieval and Analysis System
GIS	geographic information systems
NLCD	National Land Cover Database
NLCDe 92	Enhanced National Land Cover Data 1992
NRCS	Natural Resources Conservation Service
OLS	ordinary least squares
SSURGO	Soil Survey Geographic Database
USGS	U.S. Geological Survey
WLS	weighted least squares





# Estimation of Low-Flow Statistics at Ungaged Sites on Streams in the Lower Hudson River Basin, New York, From Data in Geographic Information Systems

By Allan D. Randall and Douglas A. Freehafer

## Abstract

A variety of watershed properties available in 2015 from geographic information systems were tested in regression equations to estimate two commonly used statistical indices of the low flow of streams, namely the lowest flows averaged over 7 consecutive days that have a 1 in 10 and a 1 in 2 chance of not being exceeded in any given year (7-day, 10-year and 7-day, 2-year low flows). The equations were based on streamflow measurements in 51 watersheds in the Lower Hudson River Basin of New York during the years 1958–1978, when the number of streamflow measurement sites on unregulated streams was substantially greater than in subsequent years. These low-flow indices are chiefly a function of the area of surficial sand and gravel in the watershed; more precisely, 7-day, 10-year and 7-day, 2-year low flows both increase in proportion to the area of sand and gravel deposited by glacial meltwater, whereas 7-day, 2-year low flows also increase in proportion to the area of postglacial alluvium. Both low-flow statistics are also functions of mean annual runoff (a measure of net water input to the watershed from precipitation) and area of swamps and poorly drained soils in or adjacent to surficial sand and gravel (where groundwater recharge is unlikely and riparian water loss to evapotranspiration is substantial). Small but significant refinements in estimation accuracy resulted from the inclusion of two indices of stream geometry, channel slope and length, in the regression equations. Most of the regression analysis was undertaken with the ordinary least squares method, but four equations were replicated by using weighted least squares to provide a more realistic appraisal of the precision of low-flow estimates. The most accurate estimation equations tested in this study explain nearly 84 and 87 percent of the variation in 7-day, 10-year and 7-day, 2-year low flows, respectively, with standard errors of 0.032 and 0.050 cubic feet per second per square mile. The equations use natural values of streamflow and watershed properties; logarithmic transformations yielded less accurate equations inconsistent with some conceptualized relationships.

## Introduction

Estimates of the low flow of streams are useful for many purposes, such as management of water quality, assessment of wastewater dilution, evaluation of aquatic ecosystems, and river navigation planning. Regression equations that estimate selected statistical indices of low flow at sites where streamflow has not been measured from properties of the watershed had previously been developed for two regions of New York, the Susquehanna River Basin (Ku and others, 1975, revised by Randall, 2010) and the Lower Hudson River Basin (Barnes, 1986). These studies were consistent with comparable studies in New England in demonstrating that the areal extent of surficial sand and gravel in a watershed has a powerful effect on low flow, because precipitation readily infiltrates sand and gravel and is gradually released to streams, thereby providing the principal source of streamflow during periods of low flow (see Wandle and Randall, 1994, or Kontis and others, 2004, and references in each). Each of these studies relied in part upon unpublished maps of surficial geology and other data that may not exactly correspond to recent products of geographic information systems (GIS) technology.

The objective of the present study was to test a variety of watershed properties now (2016) available from GIS databases as a means of estimating the statistical indices of low flow that had been calculated by Barnes (1986) for streamflow measurement sites in the Lower Hudson River Basin, and particularly to ascertain whether the extent of soils derived from sand and gravel adequately represents the effect of surficial sand and gravel on low flow. This report first explains the calculation by Barnes (1986) of low-flow indices, and a few modifications thereof during this study. Next, it analyzes in general terms several categories of watershed properties that might be expected to affect the low flow of streams, based on literature and on tests of particular properties in trial regression equations. Then, the four types of watershed properties that proved to be significantly influential are evaluated in more detail. Several ordinary least squares regression equations are presented to document the improvement in estimation of low flow that results from including each property. Finally, weighted least squares versions of four equations are presented for potential use in estimating low flows at ungaged sites.

The Lower Hudson River Basin consists of the entire area tributary to the Hudson River downstream from the junction of the Hudson and Mohawk Rivers at Troy, New York (fig. 1; streams and cities not shown).

## Calculation of Low-Flow Indices

Barnes (1986) compiled, for each of 53 watersheds in the Lower Hudson River Basin, two widely used statistical indices of low flow:

- *7-day, 10-year low flow (7Q10)*: the annual lowest streamflow, averaged over 7 consecutive days, that has a 1 in 10 chance of not being exceeded in any given year
- *7-day, 2-year low flow (7Q2)*: the annual lowest streamflow, averaged over 7 consecutive days, that has a 1 in 2 chance of not being exceeded in any given year

These indices were computed for a 20-year reference period of April 1, 1958, through March 31, 1978, and are presented in table 1. A common reference period was used because of the variability of climate from year to year. For example, the early 1960s was a period of low groundwater levels and extreme low flows, whereas the mid-1970s had high groundwater levels and higher than normal low flows. If low-flow indices for some streams included one of these periods but not the other, the low-flow indices would be in part a function of timewise variability in climate, which would distort or complicate the regression analysis. Barnes excluded watersheds affected by streamflow regulation or diversion, or by urban area that exceeded 10 percent of watershed area. For 16 of the 53 selected watersheds in the Lower Hudson River Basin, the streamflow measurement stations were streamgaging stations with 15 or more years of continuous streamflow record, for which low-flow indices were calculated by the log-Pearson type III technique (Riggs, 1972). Results were checked by visual inspection of low-flow frequency curves. Low-flow indices for the remaining 37 stations were calculated as described in the following excerpt from Barnes (1986, p. 5–8).

If a gaging station having from 8 to 14 years of discharge records in the reference period also had climatological and geologic characteristics similar to those at a nearby long-term gaging station, its 7-day, 10-year low flows were obtained by a comparison of flow-duration curves. The duration curve for the station of shorter record and the curve for the corresponding period at the 20-year station were both plotted to verify that their slopes were similar. The 7-day, 10-year low flows for the shorter periods, based on log-Pearson analysis, were located on each curve, and if they represented similar flow durations, the 7-day, 10-year low flow for the reference period was located on the duration curve for the abbreviated period for the station of longer record. The percent duration was noted, and the flow corresponding

to that percent duration at the site of shorter record was taken as the 7-day, 10-year low-flow value.

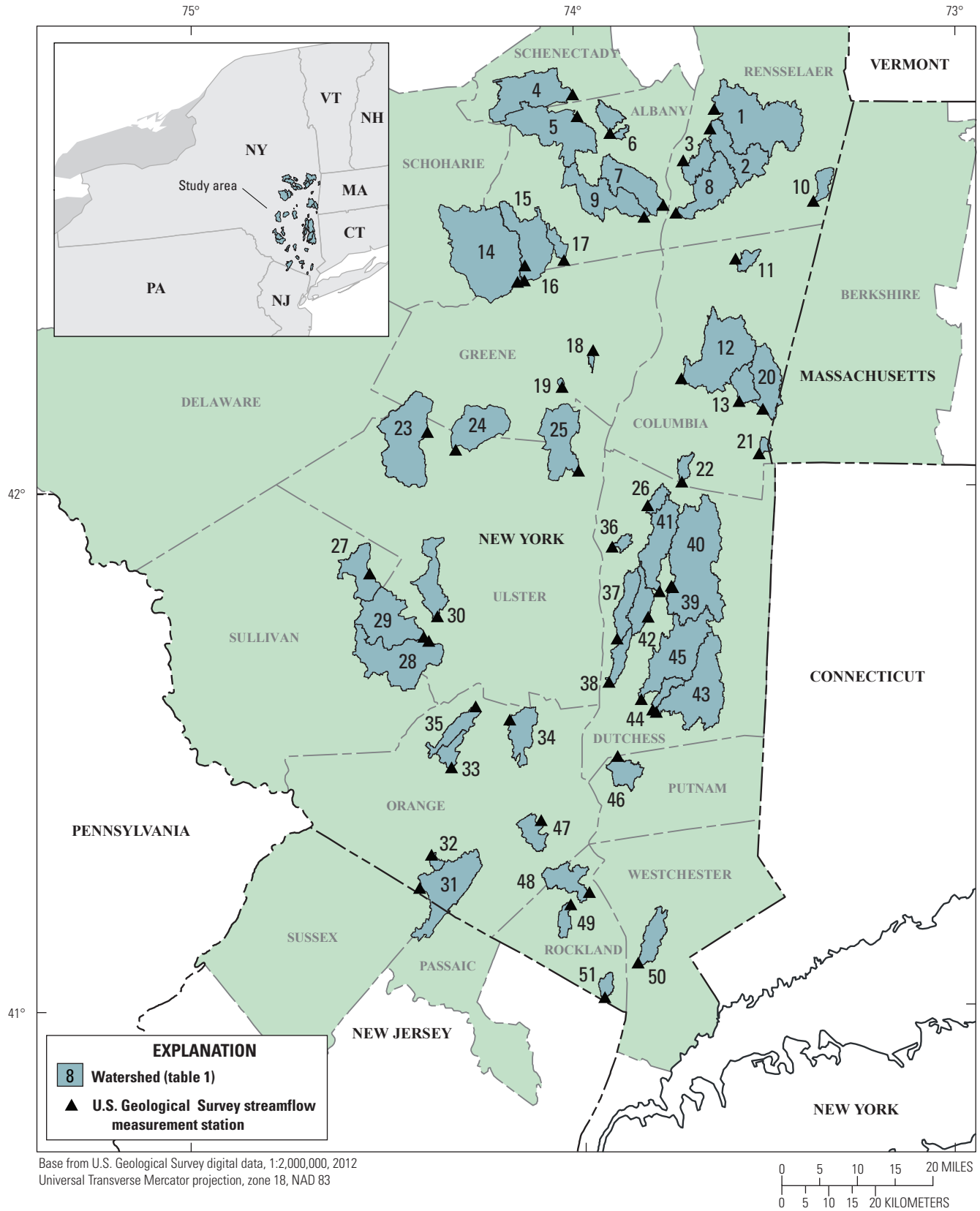
If a suitable index station for determination of low-flow statistics through duration curves could not be found, a simple ratio of low-flow statistics was used. The 7-day, 10-year low flows were calculated for the station with 8 to 14 years of record and for a concurrent period at several long-term sites. The ratio of short-period to long-period flow statistics for the long-term sites was then used to adjust the 7-day, 10-year low flow at the site with the short record.

If a site was a low-flow partial-record station or a gaged station with fewer than 8 years of data, measured flows were correlated with concurrent daily mean flows at index stations, and the 7-day, 10-year statistics were obtained graphically as described by Riggs [1972]. The 7-day, 2-year low flows were determined in a manner similar to that described above for 7-day, 10-year flows.

The low-flow indices calculated by Barnes (1986) were adopted for the present study, with three exceptions described below:

*Peekskill Hollow Creek (station 01374300)*.—This stream has been the source of public water supply for the city of Peekskill since at least 1875. A pump station, along with a low dam sufficient only to submerge the pump intake but not to store water, was constructed near the hamlet of Van Cortlandtville. In about 1912 and 1940, dams with a combined storage capacity of 1,260 million gallons were constructed on Wicopee Brook, a tributary of Peekskill Hollow Creek 9 miles upstream from the pump station. Thereafter, water was released from these reservoirs during periods of low flow as needed to maintain sufficient flow at the pump station to sustain the municipal withdrawals. From 1953 through 1962, the U.S. Geological Survey (USGS) measured the flow of Peekskill Hollow Creek occasionally during periods of low flow about 1,200 feet upstream from the pump station. Because some or all of the measurements likely included water released from reservoir storage in addition to natural streamflow, this measurement station was deleted from the study.

*Coeymans Creek (station 01359902)*.—When this streamgaging station was in operation, from August 1967 through September 1977, filter-backwash water was discharged to a tributary of Coeymans Creek from the city of Albany water-filtration plant 4 miles upstream (U.S. Geological Survey, 1978). For 8 of these years, the volume of backwash water in each month was recorded; the number of backwash events in a month typically averaged 2.7 to 4.6 events per day, and the monthly average rate of backwash discharge was typically  $0.65 \pm 0.1$  cubic feet per second in the months of July through October when all annual low flows took place (R. Borfritz, assistant chief operator, written commun., 2007). The low-flow indices calculated by Barnes (1986) did not exclude this wastewater discharge.



**Figure 1.** Location of Lower Hudson River Basin and watersheds from which data are analyzed in this report.

#### 4 Low-Flow Statistics at Ungaged Sites on Streams in the Lower Hudson River Basin, New York

**Table 1.** Low-flow indices and drainage areas for watersheds in the Lower Hudson River Basin, New York.

[Lowest annual 7-day flows are from Barnes (1986). Where a range of years of record is given, the period of record begins April 1 in the first year listed and ends September 30 or later in the second year listed. "P" indicates a partial-record station and is followed by the number of years in which at least one measurement was made]

Watershed number (fig. 1)	U.S. Geological Survey streamflow measurement station		Lowest annual 7-day mean flow, April 1958 through March 1978, in cubic feet per second, with recurrence interval		Watershed area, in square miles ( <i>DRNAREA</i> )	Years of record for station, from 1958 through 1978
	Number	Name	10 years (7Q10)	2 years (7Q2)		
1 <sup>a</sup>	01358500	Poesten Kill	2.8	7.6	72.348	1958–68
2	01359100	Wynants Kill	3.5	5.1	29.138	P 7
3 <sup>b</sup>	01359155	Mill Creek	0.6	1.1	13.241	P 3
4	01359200	Normans Kill	0.25	0.7	42.263	P 6
5 <sup>c</sup>	01359340	Bozen Kill	0.3	0.8	50.493	P 8
6	01359513	Hunger Kill (plus Blockhouse Creek)	3.9	7	10.144	1967–77
7 <sup>b</sup>	01359600	Vloman Kill	0.01	0.1	29.872	P 8
8 <sup>bd</sup>	01359750	Moordener Kill	2	3.8	31.591	1958–78
9 <sup>c</sup>	01359902	Coeymans Creek	0.2	1.15	34.825	1967–77
10	01359990	East Brook	0.8	1.6	7.196	P 5
11	01360530	Trout Brook	0.05	0.13	5.322	P 7
12	01361200	Claverack Creek	3.5	7.2	60.962	1960–68
13	01361250	Taghkanic Creek	0.6	1.2	12.579	P 7
14	01361500	Catskill Creek	0.21	1.3	95.708	1958–78
15	01361550	Ten Mile Creek 1	0.7	1.7	19.134	P 8
16	01361570	Ten Mile Creek 2	0.92	2.1	37.142	1969–78
17	01361760	Wolf Fly Creek	0.004 <sup>g</sup>	0.03 <sup>g</sup>	6.314	P 8
18 <sup>b</sup>	01362005	Bell Brook	0.02	0.1	1.304	P 4
19 <sup>b</sup>	01362040	Marys Glen	0.004 <sup>g</sup>	0.04	0.810	P 1
20	01362100	Roeliff Jansen Kill	2.2	5	27.339	1958–59
21 <sup>b</sup>	01362155	Preechey Hollow Brook	0.06	0.19	2.529	P 6
22	01362168	Fall Kill	0.004 <sup>g</sup>	0.03 <sup>g</sup>	4.952	P 6
23	01362198	Esopus Creek	4.6	8.7	59.394	1964–78
24	01362400	Stony Clove Creek	0.7	2.8	32.381	P 5
25	01364400	Platte Kill	0.2	0.8	36.727	P 6
26	01364700	Saw Kill	0.15	0.4	6.309	P 7
27	01365500	Chestnut Creek	3	5.4	21.031	1958–78
28	01366650	Sandburg Creek	6.8	12	52.772	1958–77
29 <sup>b</sup>	01366750	Beer Kill	0.9	1.9	41.975	P 6
30 <sup>b</sup>	01366800	Vernooy Kill	2	3.5	22.752	P 8
31	01368810	Wawayanda Creek	3.3	7.4	44.997	P 6
32 <sup>b</sup>	01369650	Stony Creek	0.02	0.08	2.617	P 6
33 <sup>b</sup>	01370600	Crystal Brook	0.06	0.2	8.311	1964–68
34 <sup>b</sup>	01370800	Tin Brook	0.05	0.2	18.960	P 5
35 <sup>b</sup>	01370836	Dwaar Kill	0.07	0.25	12.747	P 3
36 <sup>b</sup>	01372030	Fallsburg Creek	0.004 <sup>g</sup>	0.03	3.409	P 8
37 <sup>b</sup>	01372050	Fall Kill	0.8	1	16.543	P 7
38	01372065	Casper Creek	0.9	2	10.311	1969–75

**Table 1.** Low-flow indices and drainage areas for watersheds in the Lower Hudson River Basin, New York.—Continued

[Lowest annual 7-day flows are from Barnes (1986). Where a range of years of record is given, the period of record begins April 1 in the first year listed and ends September 30 or later in the second year listed. “P” indicates a partial-record station and is followed by the number of years in which at least one measurement was made]

Watershed number (fig. 1)	U.S. Geological Survey streamflow measurement station		Lowest annual 7-day mean flow, April 1958 through March 1978, in cubic feet per second, with recurrence interval		Watershed area, in square miles ( <i>DRNAREA</i> )	Years of record for station, from 1958 through 1978
	Number	Name	10 years (7Q10)	2 years (7Q2)		
39 <sup>bf</sup>	01372100	East Branch Wappinger Creek	2	7	33.568	1958–63
40 <sup>b</sup>	01372200	Wappinger Creek	4.7	15	92.633	1958–75
41	01372300	Little Wappinger Creek	0.35	1.7	32.906	1958–75
42	01372400	Great Spring Creek	0.1	0.35	15.475	1960–65
43 <sup>b</sup>	01372800	Fishkill Creek	1.8	6.7	57.269	1958–75
44	01372850	Whortlekill Creek	1.6	2.7	6.719	1959–68
45 <sup>f</sup>	01372900	Sprout Creek	0.8	4.2	51.068	P 9
46	01372950	Clove Creek	0.3	0.9	12.469	P 7
47	01373690	Woodbury Creek	0.4	1.1	11.063	1966–68
48	01374440	Cedar Pond Brook	0.4	0.7	17.430	1960–62
49	01374460	South Branch Minisceongo Creek	0.4	0.8	5.882	P 6
50 <sup>b</sup>	01376100	Pocantico River	0.2	0.5	15.612	P 4
51	01376270	Sparkill Creek	0.7	1	4.664	1960–63 & 1965–66

<sup>a</sup>*Poesten Kill*. Topographic drainage area is 89.4 square miles (U.S. Geological Survey, 2006a). However, water from the Quacken Kill tributary is diverted for municipal use to Tomhannock Reservoir in an adjacent watershed. The diversion “amounts to the entire flow during periods of low water” at the point of diversion, 1 mile downstream from the hamlet of Quackenkil, New York (Robert Beall, U.S. Geological Survey, written commun., 1962). Accordingly, the area upstream from this diversion was excluded for low-flow analysis.

<sup>b</sup>Stream channel appears to be incised in lacustrine clay, till, or bedrock at this measurement site; therefore, underflow is probably not significant even at flows as small as 7Q10.

<sup>c</sup>*Bozen Kill*. Thompsons Lake is topographically part of the Bozen Kill watershed and briefly overflows into a Bozen Kill tributary during periods of high runoff. Ordinarily, however, Thompsons Lake drains through conduits in limestone bedrock to Pitcher Farm Spring in Fox Creek watershed, as does a closed topographic depression at 1,300 feet altitude south of Thompsons Lake (Baker, 1976; Palmer and others, 1991a, b). The area tributary to Thompsons Lake has not been considered by the U.S. Geological Survey to be part of the Bozen Kill watershed, and the area tributary to the nearby closed depression was also excluded for low-flow analysis.

<sup>d</sup>*Moordener Kill*. The drainage divide south of the hamlet of East Schodack, New York, was revised based on field observation of streamflow in a small, poorly drained area, December 7, 2007.

<sup>e</sup>*Coeymans Creek*. Topographic drainage area is 35.1 square miles (U.S. Geological Survey, 2006b). Vly Creek Reservoir, completed in 1958, captures and diverts for municipal use all runoff during periods of low flow from a small part of the watershed. Accordingly, the area upstream from this reservoir was excluded for low-flow analysis.

<sup>f</sup>*Sprout Creek and East Branch Wappinger Creek*. A small area north of Tyrrel Lake is topographically part of Sprout Creek watershed, but contains three ponds in ice-block depressions whose altitudes are about 10 feet below Tyrrel Lake. Therefore, this locality was inferred to drain northward to East Branch Wappinger Creek through permeable gravel beneath the topographic divide. The drainage area previously compiled for Sprout Creek was reduced slightly, and that for East Branch Wappinger Creek increased correspondingly, to reflect this interpretation.

<sup>g</sup>Zero values of 7Q10 or 7Q2 calculated by Barnes (1986) were replaced by 0.004 (7Q10) and 0.03 (7Q2) to avoid zero values in some applications.



To remove the wastewater discharge from the recorded low flows of Coeymans Creek, the procedures specified by Barnes (1986, p. 5) for calculating 7-day low flows at streamgaging stations with 8 to 14 years of record were replicated. Flow duration curves for two nearby long-term streamgaging stations for 1968–77 (the years of low-flow record at Coeymans Creek) were plotted. Then, 7Q10 and 7Q2 for 1968–77 at the two long-term stations were computed graphically (Riggs, 1972, p. 1), plotted on their flow-duration curves, and found to agree in that 7Q2 plotted at 95-percent flow duration and 7Q10 at 99.8-percent flow duration for each station. Therefore, the 7Q2 and 7Q10 values for the 1958–78 reference period at these two stations were also plotted on their 1968–77 flow duration curves, and the same flow duration percentages on the 1968–77 flow duration curve for Coeymans Creek were assumed to represent 1958–78 7Q2 and 7Q10 at Coeymans Creek. Finally, the average rate of backwash discharge (0.65 cubic feet per second) was subtracted, yielding an estimated natural 1958–78 7Q2 and 7Q10 for Coeymans Creek of 1.15 and 0.21 cubic feet per second, respectively. The transfer process was approximate because plotting the 1958–78 7Q2 and 7Q10 values at the long-term stations required extrapolation of their 1958–68 duration curves, but the values of 7Q2 and 7Q10 for Coeymans Creek thus estimated are substantially smaller than values estimated by Barnes (1986) and are judged to be a closer approximation to natural conditions.

*Blockhouse Creek (station 01359517) and Hunger Kill (station 01359513).*—These two watersheds are almost adjacent, and their hydrogeology is identical: 85-percent surficial outwash sand exceeding 8 feet in thickness, overlying lacustrine silt and clay; downstream reaches of the principal stream channels are incised through the sand into silt and clay. The large values of 7Q10 and 7Q2 per square mile from both watersheds demonstrate the powerful effect of surficial sand and gravel on low flows. In preliminary regression analyses, however, low flow of Hunger Kill was consistently underestimated, whereas low flow of Blockhouse Creek was consistently overestimated. One likely explanation for this discrepancy is that watersheds in a sand plain are bounded by groundwater divides, which could not be delineated in this locality from the scant data available on depth to the water table, and which do not necessarily coincide with the topographic divides from which drainage areas were computed for this study. Accordingly, in order to reduce the statistical scatter within the dataset, these two watersheds were merged, each watershed property being summed or combined as a weighted average.

Regional appraisals of low flow by the USGS are commonly based on streamflow data from multiyear reference periods that continue nearly up to the year in which the study was launched, in order to capture possible trends in climate or land use over time that may affect low flows. The 1958–78 reference period adopted by Barnes (1986) was not updated for the present study for three reasons:

1. The number of streamflow measurement sites on unregulated streams—long-term streamgaging stations as well as partial-record stations where occasional streamflow measurements had been made—was substantially greater in 1958–78 than in subsequent years. Inasmuch as the goal of this study was to evaluate the spatial variation in watershed properties compiled from GIS databases as influences on low flow, the greater number and variety of watersheds monitored in 1958–78 was an advantage.
2. Limitations on personnel and funding precluded the substantial effort required to regenerate a new set of low-flow statistics.
3. Recent studies of climatic influences on low flows in this region have not consistently documented timewise trends that would warrant adjustment of low-flow statistics from 1958–78 to represent current conditions. Unregulated streams in the northeastern United States experienced a small (not statistically significant) increase in annual runoff during the 20th century but no significant trend in magnitude or timing of summer or fall low flows (Hodgkins and others, 2005; Hayhoe and others, 2007, p. 392). Changes in magnitude of annual 7-day low flows of 51 regulated streams in the Northeast during the 21st century, as projected by an array of climate models, range from a decrease of more than 10 percent to little or no change (Hayhoe and others, 2007, p. 398). By contrast, at 5 streamgaging stations on unregulated streams in New York that have operated continuously since at least 1944, 7Q10 calculated for 1976–2006 ranged from 0 to 18.6 percent higher than 7Q10 calculated for periods of record through 1975, and annual 7-day low flows over periods of record through 2006 showed an upward trend at 4 of the 5 stations (Suro and Gazoorian, 2011). Burns and others (2007) calculated an increasing trend from 1953 to 2005 in monthly precipitation and monthly mean runoff in the Catskill region, immediately west of the Lower Hudson River Basin, for each month from June through October. However, the magnitudes of the increases calculated for most months were deemed not statistically significant, and the calculated increases in monthly mean runoff are not pertinent to annual 7-day low flows because the method of computation is entirely different. Saft and others (2015) present evidence that in many watersheds in southeastern Australia a multiyear dry period resulted in a downward shift in the rainfall-runoff relationship compared with the historical norm, such that any particular annual rainfall generated less annual streamflow than would have been expected during the historical period of record. However, they observed that such a downward shift was most likely in watersheds of low relief, relatively dry climate, and a low proportion of forest cover (characteristics that are not typical of the Lower Hudson River Basin).

## Compilation of Watershed Properties

More than 50 watershed properties were evaluated during this study for statistically significant correlation with low flows. These properties could be grouped into seven categories: drainage area, surficial geology, amount of water that is available for runoff (that is, precipitation minus evapotranspiration), areas of intense riparian evapotranspiration (that is, wetlands and poorly drained soils), watershed geometry (slope, relief, and stream-channel gradient), bedrock geology, and land cover. Each of these categories is discussed in turn in the following sections.

### Drainage Area

The area of each watershed had been delineated on 1:24,000-scale topographic maps and measured by planimeter by the USGS prior to the study by Barnes (1986). Subsequently, watershed divides were scanned into a GIS database, and drainage areas were computed by the GIS program. These values appear in table 1 and are stored in the USGS National Water Information System database. Many of these values differ slightly from the older values reported by Barnes (1986), even though the delineations of the divides in GIS replicate the original source maps faithfully. Manmade diversions for municipal water supply or natural diversions through cavernous limestone or surficial gravel capture all runoff during periods of low flow from parts of five watersheds. For each of these watersheds, the remaining drainage area that contributes to low flow at the measurement site was delineated during this study and is reported in table 1 with a footnote that explains the basis for delineation.

Drainage area does not appear as an independent variable in regression equations presented in this report. Instead, areas of particular terranes that significantly contribute to or detract from low flow appear in equations that estimate low flow in cubic feet per second, and all watershed properties are divided by drainage area in equations that estimate low flow in cubic feet per second per square mile. Drainage area was tested as an independent variable in some trial equations (along with areas of the particular significant terranes, each of which was expressed as a fraction of drainage area), but these trial equations had less favorable diagnostic statistics than equations presented in this report.

### Surficial Geology

Bedrock is overlain by unconsolidated glacial and postglacial sediments nearly everywhere in the Lower Hudson River Basin. Many previous studies (see Wandle and Randall, 1994, and references cited therein) reported that low flow of streams is strongly affected by surficial geology, particularly the areal extent of sand and gravel relative to till and lacustrine silty clay. Detailed (1:24,000-scale) maps of surficial geology

are not available for most localities in New York. Therefore, for this study the distribution of unconsolidated glacial and postglacial sediments overlying bedrock was inferred from county soil surveys, as compiled in the Soil Survey Geographic Database (SSURGO) by the Natural Resources Conservation Service (2007).

Each individual soil series in each county was assigned to 1 of 12 geologic units, as detailed in appendix 1 (table 1–1), largely on the basis of soil description in the latest soil survey report for that county, particularly the description of the parent material or C horizon. (For a few counties, SSURGO incorporates revised soil series nomenclature that postdates the county soil survey report, in which case soil series descriptions can be obtained online at <https://websoilsurvey.sc.egov.usda.gov/>.) The geologic units distinguish several lithologies that were not expected to be individually significant as influences on low flow but could be compiled in various ways in regression equations to ascertain which combination best predicted low flow. Many small areas of excavated, backfilled, or urbanized land were not classified to a soil series by SSURGO; each of these was assigned to a geologic unit on the basis of topography, adjacent soils, and any clues in the soil survey descriptions of these areas. Subsequently, an attempt was made to automate the assignment of soils in a few counties to geologic units on the basis of the “Parent Material Group Name” in SSURGO. Most of the assignments that had been made individually could be replicated in this manner, but not all.

The Natural Resources Conservation Service (NRCS) has classified soil series into four “hydrologic soil groups,” or combinations thereof, on the basis of estimated rate of infiltration during long-term storms when the soils are thoroughly wet:

- Group A: high infiltration, chiefly well-drained to excessively drained sand and gravel
- Group B: moderate infiltration, chiefly well- to moderately well-drained soils of moderate texture
- Group C: slow infiltration, chiefly soils with a layer that impedes vertical flow of water or that have a fine to moderately fine texture
- Group D: very slow infiltration, chiefly soils that are composed of clay, that have a high water table, or that are shallow over nearly impervious material

This classification has been used in conjunction with base-flow separation from streamflow hydrographs to develop a procedure for estimating average rates of groundwater recharge at sites in Dutchess County, N.Y. (Chazen Companies, 2006). The hydrologic soil groups constitute an alternative to the reclassification of soil series developed in the present study and were tested as independent variables in the regression analysis.

## Amount of Water Available for Runoff

Water input to the Lower Hudson River Basin is exclusively from precipitation. A substantial part of that input is lost to evapotranspiration; the remainder becomes available for immediate storm runoff or for groundwater recharge that will eventually be discharged to streams. Mean annual precipitation is strongly correlated with mean annual runoff and with altitude, and several studies (cited in Randall, 2010, and Randall and Johnson, 1988) have shown that low flow of streams in the glaciated northeastern United States is in part a function of one or another of these three indices of water availability.

The regression equations presented by Barnes (1986) for estimating low flow in the Lower Hudson River Basin included as independent variables mean annual precipitation for 1920–49 (Knox and Nordensen, 1955) and a simple approximation of mean watershed altitude (34 percent of the range from minimum to maximum altitude, added to the minimum altitude). The datasets for the present study included mean watershed altitude as calculated by the GIS from USGS 30-meter digital elevation models. The datasets also included 1951–80 mean annual precipitation and mean annual runoff, as interpolated by the GIS from contour maps by Randall (1996) that were drawn by using a protocol that forced the precipitation contours to parallel the runoff contours and ensured that each set of contours was constrained both by precipitation at precipitation stations and runoff from gaged watersheds.

## Riparian Evapotranspiration Losses

During periods of low flow, streamflow consists almost entirely of groundwater discharge from earth materials bordering the stream channel. Where the water table near the stream is at or close to land surface, however, evaporation from open water or saturated soil and transpiration by plants diverts to the atmosphere some groundwater discharge before it can reach the stream channel, thereby reducing low flows. Where streams are ponded, the ratio of water area to flow increases and evaporation losses increase. Studies of low flow in the Susquehanna River Basin of New York (Randall, 2010), in central New England (Wandle and Randall, 1994), and in Connecticut (L. Weiss and R. Melvin, USGS, written commun., 1983; equation published in Randall and Johnson, 1988) all concluded that the area of lakes and swamps bordered by sand and gravel is negatively correlated with low flow, and that incorporation of this term in regression equations resulted in a modest but statistically significant improvement in the predictive accuracy of the equations. Evapotranspiration from riparian swamps and lakes bordered by till also diverts to the atmosphere some potential groundwater discharge to streams. In mountainous regions of central New England, the total area of swamps and lakes proved to be more significant than the area of swamps and lakes bordered by sand and gravel (Wandle and Randall, 1994); elsewhere, however, including

area of lakes and swamps bordered by till as part of the wetland variable in regression equations degraded predictive accuracy, probably because the till-mantled uplands contribute such a small fraction of total low flow in most watersheds that evapotranspiration from upland wetlands had little impact on total watershed low flow.

At the time of this study (2007–13), several GIS datasets were available that delineated total wetland areas (lakes, plus swamps variously defined), including the U.S. Fish and Wildlife Service (USFWS) National Wetlands Inventory, the National Land Cover Database (NLCD) (Homer and others, 2007), the USGS Geographic Information Retrieval and Analysis System (GIRAS) (Price and others, 2006), and the NRCS SSURGO dataset (Natural Resources Conservation Service, 2007). None of these datasets distinguished upland wetlands bordered by till from wetlands bordered by sand and gravel in valleys or lowlands. Accordingly, an algorithm was devised to scan the perimeter of each polygon designated by SSURGO as water, and each polygon designated as one of several soil series described as organic muck or peat, and then to sum areas of water (lakes) that are included in each in three categories and also areas of swamps in each of these categories:

- Bordered entirely or partly by outwash sand and gravel and (or) by alluvium
- Not bordered by outwash sand and gravel or by alluvium, but bordered entirely or partly by clay, silt, or fine sand of lacustrine origin
- All others (generally bordered only by till)

This algorithm generally produced reasonable results, although a few large lakes and swamps were reclassified as bordered by sand and gravel even though only a tiny segment of their perimeter was sand and gravel.

SSURGO also classified all soil series into seven categories as to drainage potential. During periods of low flow, evapotranspiration losses from very poorly, poorly, or somewhat poorly drained soils could be expected to exceed losses from moderately well to excessively drained soils. Accordingly, somewhat poorly to very poorly drained soils that are derived from outwash sand and gravel, from alluvium, and from lacustrine fine sand to clay were identified (appendix 1, table 1–2), and the areas in these three categories in each watershed were compiled (table 2).

## Watershed Geometry

Topographic relief, land-surface slope, and (or) stream-channel gradient have commonly been included among the watershed properties tested as explanatory variables in regression analyses designed to estimate low or high streamflow. These geomorphic indices have been included because they could be measured conveniently on topographic maps and because steep slopes were perceived by the investigators to



**Table 2.** Areas of surficial geologic units and poorly drained soil categories for watersheds in the Lower Hudson River Basin, New York.

[Acronyms used in this report are in parentheses; acronyms bracketed over 2 or 3 columns represent the sum of areas in those columns for each streamflow measurement station. Acronyms are defined in the text section “Regression analysis using ordinary least squares.”]

Water-shed number (fig. 1)		U.S. Geological Survey streamflow measurement station		Area, in square miles															
				Surficial glacial meltwater deposits						Postglacial alluvium (Q)			Organic sediments in swamps, bordering (SW/sqft)			Somewhat to very poorly drained soils, formed on			Hydro-logic Group A (HSG_A)
				(S)		(FY)		(SY)		(QAL)		(QALL)	(TTH)	(TSR)	(V)				
Sand and gravel (SG)	Sand (SD)	Fine sand and silt (FL)	Silt and clay (LY)	Thin sand over silt and clay (SY)	All or partly sand and gravel (QAL)	Silt loam or silt to deeper than 6 feet (QALL)	Thickness greater than 6 feet (TTH)	Overlies bedrock at less than 6 feet depth (TSR)	Commonly overlies silt, sand or gravel (V)	Some gravel and sand (SW/sq)	Some fine sand to clay, no gravel (SWfy)	Fine sand, silt and clay (FYp)	Post-glacial alluvium (Qp)	Sand and gravel (Sp)					
1	01358500	Poesten Kill	4.883	0.026	0.076	0.018	0	1.352	1.458	48.276	13.222	0	0.784	0.049	0.094	1.921	0.4	3.884	
2	01359100	Wynants Kill	7.129	0.118	0.082	0	0	0.198	0.662	15.225	3.684	0	0.588	0	0.082	0.788	0.36	6.163	
3	01359155	Mill Creek	1.478	0.048	0.07	1.088	0	0.117	0.344	7.706	2.094	0	0.064	0.032	0.156	0.385	0	1.17	
4	01359200	Normans Kill	1.365	0.327	0.007	0.175	0.031	1.166	0.218	26.392	11.861	0	0.174	0.054	0.158	0.17	0.209	0.923	
5	01359340	Bozen Kill	2.188	0.002	1.364	2.276	0.009	1.697	0.074	25.808	15.681	0.818	0.218	0.143	2.799	0.437	0	1.833	
6	01359513	Hunger Kill (plus Blockhouse Creek)	0.007	8.489	0.119	1.067	0.242	0.12	0	0.031	0	0	0	0.033	0.113	0.04	0.734	5.865	
7	01359600	Vloman Kill	0.828	3.84	3.999	10.973	1.88	1.117	0.055	4.396	0.853	0.362	0	0	6.076	0.393	0.603	0.88	
8	01359750	Moordener Kill	5.822	0.132	0.044	0.675	0.502	0.492	1.114	13.857	7.764	0	0.633	0.025	0.46	1.496	0.207	5.028	
9	01359902	Coeymans Creek	1.222	0.085	1.01	4.318	0.087	1.06	0.299	8.798	17.233	0.346	0.002	0.022	1.983	0.15	0.008	1.011	
10	01359990	East Brook	1.053	0.018	0	0	0	0.383	0.006	2.275	3.396	0	0.035	0	0	0.2	0.028	0.991	
11	01360530	Trout Brook	0.622	0	0.065	0	0	0.027	0	3.443	1.115	0	0.012	0	0.065	0	0.057	0.549	
12	01361200	Claverack Creek	8.814	0.973	1.313	0.295	0	1.728	0	17.416	28.789	0	0.586	0.164	1.391	0	0.643	8.826	
13	01361250	Taghkanic Creek	0.476	0	0	0	0	1.063	0	3.512	7.327	0	0.161	0	0	0	0.006	0.46	
14	01361500	Catskill Creek	2.513	0	0.145	0.926	0.011	2.103	0.024	53.372	35.354	0.135	0.183	0.131	0.791	0.308	0	2.158	
15	01361550	Ten Mile Creek 1	1.095	0	0.061	0	0	0.545	0	11.735	5.115	0.182	0.112	0	0.061	0.011	0	1.028	
16	01361570	Ten Mile Creek 2	1.804	0	0.142	0.037	0	1.005	0	24.797	8.085	0.282	0.125	0.151	0.179	0.092	0	1.677	
17	01361760	Wolf Fly Creek	0.319	0	0.235	0	0	0.024	0	3.367	2.043	0.152	0	0.008	0.224	0	0	0.284	
18	01362005	Bell Brook	0.211	0	0	0	0	0	0	0.051	0.968	0.029	0.008	0	0	0	0	0.211	
19	01362040	Marys Glen	0.008	0	0	0	0	0.015	0	0.133	0.654	0	0	0	0	0	0	0.008	
20	01362100	Roeliff Jansen Kill	2.831	0.02	0	0	0	1.454	0	9.07	13.691	0	0.076	0	0	0	0.061	2.649	
21	01362155	Preechey Hollow	0.332	0	0	0	0	0.104	0	0.248	1.845	0	0	0	0	0	0	0.325	
22	01362168	Fall Kill	0.273	0	0	0	0	0.044	0	0.494	4.026	0	0.069	0	0	0	0.13	0.136	
23	01362198	Esopus Creek	1.673	0	0	0.654	0.188	1.513	0.009	11.89	42.887	0.536	0	0	0	0	0.008	2.142	
24	01362400	Stony Clove Creek	1.005	0	0	0.15	0.189	1.135	0	12.397	17.235	0.24	0	0	0	0	0.021	1.503	
25	01364400	Platte Kill	2.479	0.109	1.025	0.004	0	0.852	0.293	2.891	28.717	0.053	0.017	0.039	0.698	0	0.118	2.251	
26	01364700	Saw Kill	0.314	0	0	0.001	0	0	0.271	2.244	3.448	0	0	0	0.001	0.271	0.07	0.225	

**Table 2.** Areas of surficial geologic units and poorly drained soil categories for watersheds in the Lower Hudson River Basin, New York.—Continued

[Acronyms used in this report are in parentheses; acronyms bracketed over 2 or 3 columns represent the sum of areas in those columns for each streamflow measurement station. Acronyms are defined in the text section “Regression analysis using ordinary least squares.”]

U.S. Geological Survey streamflow measurement station			Area, in square miles															
			Surficial glacial meltwater deposits						Postglacial alluvium (Q)		Till		Organic sediments in swamps, bordering (SWsq/ft)		Somewhat to very poorly drained soils, formed on			Hydro- logic Soil Group A (HSG_A)
			(S)		(FY)													
Sand and gravel (SG)	Sand (SD)	Fine sand and silt (FL)	Silt and clay (LY)	Thin sand over silt and clay (SY)	All or partly sand and gravel (QAL)	Silt loam or silt to deeper than 6 feet (QALL)	Thickness greater than 6 feet (TTH)	Overlies bedrock at less than 6 feet depth (TSR)	Commonly overlies silt, sand or gravel (V)	Some gravel and sand (SWsq)	Some fine sand to clay, no gravel (SWft)	Fine sand, silt and clay and gravel (FYp)	Post- glacial alluvium (Qp)	Sand and gravel (Sp)				
27	01365500	Chestnut Creek	1.146	0	0.007	0	0	0.215	0.022	14.447	4.885	0.048	0.021	0	0.007	0.022	0.017	0.889
28	01366650	Sandburg Creek	5.297	0.402	0.375	0.291	0	1.705	0.265	19.162	23.269	1.166	0.036	0	0.299	0.014	0.283	4.818
29	01366750	Beer Kill	1.128	0.441	0.246	0	0	0.22	0.303	23.784	11.39	0.861	0.41	0.327	0.246	0	2.147	0.909
30	01366800	Vernooij Kill	1.278	0.171	0.021	0	0	0.31	0	5.821	13.349	1.398	0.039	0.01	0.021	0	0.841	0.617
31	01368810	Wawayanda Creek	3.923	0	2.822	0.587	0	1.62	0	21.069	12.137	0	0.373	0.486	0	0.065	0.431	2.779
32	01369650	Stony Creek	0.13	0	0.065	0	0	0.17	0	1.368	0.864	0	0	0	0	0	0	0.108
33	01370600	Crystal Brook	0.384	0	0.231	0	0	0.266	0	5.56	1.679	0	0	0	0	0	0.019	0.355
34	01370800	Tin Brook	1.871	0.012	1.387	0.818	0	0.661	0	10.156	3.238	0	0.052	0.553	0	0	0.063	1.594
35	01370836	Dwaar Kill	0.619	0	0.811	0.249	0	0.622	0	8.176	1.847	0	0.144	0.148	0	0	0.006	0.535
36	01372030	Fallsburg Creek	0	0	0	0.319	0	0	0.082	0.34	2.22	0	0	0.129	0.279	0.082	0	0
37	01372050	Fall Kill	1.652	0	0	0.27	0	0.064	0.403	2.324	11.114	0	0.3	0.062	0.27	0.407	0.122	1.472
38	01372065	Casper Creek	1.536	1.541	0.03	0.815	0	0.065	0.585	1.365	4.086	0	0	0.108	0.456	0.346	0.345	1.682
39	01372100	East Branch Wappinger Creek	4.321	0	0	0	0	0.763	1.004	9.518	16.952	0	0.415	0	0	0.992	0.186	4.059
40	01372200	Wappinger Creek	13.042	0.041	0.586	0.016	0	1.537	2.361	20.781	50.931	0	1.187	0.131	0.602	2.342	1.373	11.124
41	01372300	Little Wappinger Creek	2.111	0	0	0	0	0.603	0.821	7.212	21.116	0	0.321	0	0	0.852	0.348	1.614
42	01372400	Great Spring Creek	1.122	0	0	0.144	0	0.063	0.284	4.116	9.032	0	0.141	0.144	0.144	0.347	0.396	0.441
43	01372800	Fishkill Creek	5.403	0	0	0.007	0	1.61	2.063	17.815	28.235	0	0.83	0	0.007	2.209	0.614	3.129
44	01372850	Whortlekill Creek	2.402	0	0	0	0	0.168	0.357	1.255	2.412	0	0.084	0	0	0.357	0.151	2.121
45	01372900	Sprout Creek	6.541	0	0	0.378	0	1.918	2.418	10.57	27.36	0	1.149	0.068	0.378	2.596	0.719	5.605
46	01372950	Clove Creek	0.702	0	0.013	0	0	0.313	0	5.615	5.464	0	0.018	0.007	0.013	0	0.001	0.175
47	01373690	Woodbury Creek	1.082	0	0.11	0.059	0	0.674	0	6.282	2.644	0	0.011	0	0	0	0.102	0.891
48	01374440	Cedar Pond Brook	0.23	0	0.01	0	0	0.04	0	6.706	9.69	0	0.015	0	0.01	0.031	0.049	0.168
49	01374460	South Branch Miniscongo Creek	0.768	0	0	0	0	0.081	0	3.981	0.528	0	0.462	0	0	0.081	0	0.327
50	01376100	Pocantico River	0.217	0.005	0.057	0	0	0.394	0	9.313	5.204	0	0.057	0.006	0.057	0.075	0.023	0.065
51	01376270	Sparkill Creek	1.005	0	0	0	0	0.323	0	2.145	1.116	0	0.075	0	0	0.262	0	0.021

favor storm runoff and (or) rapid depletion of groundwater storage and, therefore, small low flows (Paulsen and others, 1940, p. 440; Schumm, 1956; Zecharias and Brutsaert, 1988a, b). Indeed, the valley-fill sand and gravel that is the principal source of low flow generates negligible storm runoff and has much smaller internal relief and slope than the adjacent till-mantled uplands. Average watershed relief, slope, and stream gradient are largely functions of upland topography, however, and increased relief between uplands and valleys is associated with an increase in infiltration of upland storm runoff into the valley-fill sand and gravel, thereby augmenting its subsequent contribution to low flow (Morrissey and others, 1988). Watershed or channel slope and relief have not proven consistently significant in regression equations designed to estimate low flow (see review in Randall, 2010, p. 23–24). Nevertheless, algorithms for GIS computation of several geomorphic indices were readily available at the time of the present study and were tested in regression equations for the Lower Hudson River Basin.

## Land Cover

The type of vegetation, or other land cover, could be expected to affect the proportion of precipitation that becomes groundwater recharge (and thus ultimately becomes groundwater discharge to streams). For example, the amount of precipitation intercepted by and evaporated from tree branches before reaching the ground can substantially exceed that intercepted from grass or low brush (Law, 1957). Furthermore, considerable soil water is tapped and transpired by trees, whose roots penetrated to depths between 10 and 25 feet at about half of 91 sites studied by Stone and Kalisz (1991) and have been observed at depths of 80 feet (Lewis, 1968), whereas grass roots penetrate only to depths of a few feet. However, Hibbert (1969) reported that clear-cutting a 22-acre forested mountain watershed in North Carolina and reseeding with fertilized grass resulted in no change in annual water use by vegetation. Schneider and Ayer (1961) evaluated the effect of reforestation of three small watersheds in central New York over 25 years; total annual runoff and peak discharges were reduced, but base-flow recession rates and annual minimum daily flows were not significantly changed.

Urbanization is generally inferred to reduce groundwater recharge to the extent that precipitation on buildings, and on paved streets and parking areas, is routed to sewers as storm runoff rather than infiltrating into the soil. Also, roadside drainage ditches on upland hillsides can intercept some surface or shallow subsurface runoff that would otherwise have flowed downslope to recharge valley-fill sand and gravel. On the other hand, leaky water mains and sanitary sewers in urban areas can locally augment natural recharge, and replacement of trees by impervious surfaces reduces evapotranspiration. Net recharge to a valley-fill gravel aquifer within Binghamton and Johnson City, N.Y., was estimated by Randall (1977) to have been slightly less under natural conditions than under 1967

urban conditions (mostly freestanding homes on individual lots plus some large commercial buildings, with only 14 percent of the area left as open land).

Land-cover information was available for this study from two GIS databases: GIRAS and the Enhanced National Land Cover Data 1992 (NLCD 92) dataset (Nakagaki and others, 2010). GIRAS was interpreted from aerial photographs taken in the 1970s (which is within the period of streamflow record analyzed for this study) and was compiled at scales of 1:250,000 or 1:100,000. The NLCD 92 dataset was based on satellite imagery collected from 1986 to 1995 and was compiled at a spatial resolution of 30 meters. Each database distinguished forests, grassland, wetlands, residential land cover, and urban land cover. A more precise mapping of urban density in 2001 was provided by the National Land Cover Database (NLCD), expressed as percent impervious surfaces and also available at a 30-meter spatial resolution (Homer and others, 2007).

Despite the evidence cited for greater evapotranspiration from forest areas than from grass, shrubs, or agricultural fields, the percentage of forest cover as compiled in the NLCD 92 and GIRAS datasets was *positively* correlated with low flow in several trial regression equations, was only weakly significant at best, and did not improve the predictive accuracy of the equations. Similar results were obtained in trial regression analyses based on streamgaging stations all across New York (C.L. Gazorian, USGS, written commun., 2013). Therefore, this watershed property was not considered further.

Including area of urban land cover from the NLCD 92 dataset as an independent variable slightly improved predictive accuracy of one trial equation, but the significance of this term was weak and varied with changes in representation of alluvium. Neither urban land cover nor urban plus residential land cover from the NLCD 92 dataset were significant in other trial equations. Urban area compiled in the GIRAS dataset (the sum of categories LU12 through LU17) was also not generally significant, nor was the percentage of impervious area statistic. Urban land use exceeded 6 percent of watershed area in only 3 watersheds (NLCD 92) or 6 watersheds (GIRAS) of the 51 watersheds studied in the Lower Hudson River Basin. These percentages reflect the statement by Barnes (1986, p. 4) that watersheds with more than 10 percent urban area were excluded from this dataset. That exclusion may have precluded regression analysis in the present study from recognizing the possible influence of urban land cover.

## Bedrock Geology

The extent to which low flow of streams may be affected by differences in water-transmitting capacity among various bedrock units was considered during this study. Much of the Lower Hudson River Basin is underlain by Ordovician shales and (in Putnam, Westchester, and Rockland Counties) much older crystalline metamorphic bedrock. These bedrock units are poorly productive aquifers, generally capable of yielding

only a few gallons per minute to wells (although larger yields have been obtained from scattered fracture zones). Three bedrock units that are less widely distributed but somewhat more permeable were evaluated as possible sources of augmented low flow: carbonate bedrock, the Newark Supergroup in Rockland County, and the Catskill facies in Devonian bedrock near the western margin of the Lower Hudson River Basin.

*Carbonate bedrock.*—In the Lower Hudson River Basin, limestone, dolomite, or marble, composed largely of calcium and magnesium carbonate, are present in two narrow outcrop belts: (1) west of the Hudson River, near the base of the Helderberg escarpment in western Schenectady and Albany Counties, thence southward in valleys immediately east of the Catskill Mountains, and (2) near the eastern border of New York, where marble underlies several prominent narrow valleys (Fisher and others, 1970). In these areas, runoff occurs in part through subsurface conduits that were developed by solution and that locally discharge as large springs (Baker, 1973, 1976; Palmer and others, 1991a, b).

The distribution of carbonate bedrock, and the distribution of a few units that consist of carbonates interbedded with other lithologies are shown on a geological map of New York at a scale of 1:250,000 (Fisher and others, 1970). Areas of these carbonate units were tested in several regression equations during this study but proved not to be significant.

*Newark Supergroup.*—Shale, sandstone, and conglomerate assigned to the Newark Supergroup of Mesozoic age underlies large areas in Connecticut, Massachusetts, and New Jersey, where parts of this unit are recognized as a substantially more productive aquifer than older metamorphic bedrock nearby (Ryder and others, 1981). The Newark Supergroup continues from New Jersey into Rockland County, N.Y., where it sustains larger well yields than nearby metamorphic bedrock, particularly in the western part of its outcrop belt where relatively coarse-grained sandstone and conglomerate predominate (Heisig, 2010).

Sedimentary bedrock of the Newark Supergroup underlies 30 to 100 percent of three watersheds in the Lower Hudson River Basin that were evaluated in this study: Cedar Pond Brook, South Branch Minisceongo Creek, and Sparkill Creek. Low flows in the first two of these were consistently overpredicted by regression equations. Low flow of Sparkill Creek was slightly overpredicted by some trial equations, slightly underpredicted by others; diurnal streamflow fluctuations reported at this site (U.S. Geological Survey, 1978, and other annual data reports) were investigated in 1960 and found to be caused by discharge 6 days per week of industrial wastewater, the water having been derived from wells. Thus, the regression results from these three watersheds suggest that the aquifer properties of the Newark Supergroup are insufficient to significantly enhance low flows, so areas of this bedrock unit were not compiled as a watershed variable for regression analysis.

*Catskill facies.*—The western fringe of the Lower Hudson River Basin (in Schoharie, Sullivan, western Ulster, western Greene, and western Albany Counties) is much higher in altitude and relief than areas to the east and is underlain

by shale, siltstone, sandstone, and conglomerate originally deposited by rivers as the headward nonmarine Catskill facies of a vast marine delta of Devonian age (Rickard, 1975). This facies is coarser grained and more massive than the shale and siltstone facies of the same age that predominates west of the Lower Hudson River Basin, and has larger and more continuous fractures, hence presumably greater water-transmitting capacity (Coates, 1971). Nevertheless, low flows were substantially overpredicted in 3 of the 7 watersheds in this part of the Lower Hudson River Basin dataset, but comparably underpredicted in 3 others, which indicates that inclusion of this bedrock facies in the dataset would not have appreciably improved the regression equations.

It has long been recognized that streams draining productive aquifers have larger low flows per square mile than streams draining poorly permeable earth materials (Tolman, 1937). The strong positive correlation of low flow with extent of surficial sand and gravel in the Lower Hudson River Basin is consistent with this principle. The discussion in the foregoing paragraphs suggests that discharge to streams during periods of low flow from the various bedrock units in the Lower Hudson River Basin is either uniformly small or perhaps heterogeneous but controlled by properties other than bedrock lithology. A study of 49 watersheds in central New England likewise reported no correlation between bedrock lithology and residual variation in low flow not accounted for by surficial geology, water availability, and riparian evapotranspiration (Wandle and Randall, 1994, p. 6). Therefore, no measures of bedrock geology are included among the watershed properties in regression equations presented in this report.

## Regression Analysis Using Ordinary Least Squares

All regression equations were designed to be executed by PROC REG (SAS Institute, 1990, 2008), which was programmed to generate a variety of statistical measures of the reliability of each equation. Most regressions used ordinary least squares (OLS), which assumes that each measurement site is of equal importance. All regressions used natural values of streamflow and watershed properties, as did several earlier studies in the glaciated northeastern United States (Cervione and others, 1982; Male and Ogawa, 1982; Wandle and Randall, 1994; Randall, 2010). Logarithmic transformation of natural values was considered but not applied, as explained in the subsection “Logarithmic Transformations.” Many trial equations were generated to test alternative selections, combinations, or mathematical manipulations of data hypothesized to represent each category of watershed properties described previously, in the section “Compilation of Watershed Properties.” Thirteen regression equations are presented in table 3 to document the extent to which adding or substituting particular watershed properties affects estimation of low flow. Each



equation is accompanied by several statistical measures of accuracy. Subsequently, four of the OLS equations in table 3 were replicated by using weighted least squares (WLS), in which measurement sites are weighted in proportion to length of streamflow measurement records, as described in the section “Regression Analysis Using Weighted Least Squares.”

Two formats were used for regression equations in this study:

- *Standard-value equations*.—Low-flow indices (7Q10, 7Q2) in cubic feet per second, as a function of watershed properties in common units of measure: area in square miles, altitude and relief in feet, and slope in feet per mile. With this format, large watersheds are more influential than small watersheds because the regression process selects an equation that minimizes estimation error, measured in cubic feet per second.
- *Per-square-mile equations* (“specific discharge”).—Low flow indices expressed in cubic feet per second per square mile, as a function of measured watershed properties divided by watershed area. With this format, all watersheds have equal influence regardless of size.

The first two independent variables in each equation in table 3 represent surficial geology and water available for runoff. Five of these equations also include two independent variables that represent riparian evapotranspiration loss and some aspect of watershed geometry. The independent variables in these equations are defined conceptually in the following list; lowercase letters in some variable names represent qualifications applicable to unconsolidated sediments represented by the preceding uppercase letters. The values of individual variables for each watershed in the Lower Hudson River Basin that was studied for this report are given in tables 2 and 4. The variables that represent surficial geology and riparian evapotranspiration were derived from SSURGO as explained in appendix 1 (table 1–3). The sources for variables that represent mean annual runoff and watershed geometry are noted in the following list of definitions. Methods for accessing or calculating these variables to estimate statistical indices of low flow are presented later.

- *SV*: area, in square miles, of surficial sand and gravel deposited by glacial meltwater, plus area of Valois soil
- *SVQ*: same as *SV*, plus area of alluvium
- *SV\_Sp*, *SVQ\_Sp*: same as *SV* and *SVQ*, each minus area of somewhat poorly to very poorly drained surficial sand and gravel deposited by glacial meltwater
- *SQFYpSWsqfy*: area, in square miles, of surficial sand and gravel (*S*), alluvium (*Q*), fine sand (*F*), and silt or clay (*Y*) that are somewhat to very poorly drained; plus area of swamps bordered at least in part by surficial sand and gravel, alluvium, fine sand, silt, or clay
- *SWsqFY*: area, in square miles, of swamps bordered at least in part by surficial sand and gravel and (or)

alluvium, plus area of fine sand, silt, or clay (surficial or mantled by a few feet of sand)

- *TTH*: area, in square miles, of till greater than 6 feet thick
- *HSG\_A*: area of Hydrologic Soil Group A, as delineated by NRCS: mostly deep, well-drained sand or gravel with high infiltration capacity and low runoff potential when thoroughly wet
- *MAQ*: mean annual runoff from the watershed, in cubic feet per second, calculated from a contour map of mean annual runoff for 1951–80 (Randall, 1996)
- *DRNAREA*: drainage area, in square miles, of the watershed upstream from a measurement site, as delineated on 7.5-minute USGS topographic maps, then calculated by digitizing the watershed outline
- *LENGTH*: the distance, in miles, measured along stream channels from a streamflow measurement site upstream to the most distant point from which runoff would flow to that measurement site, as depicted on a 30-meter digital elevation model (U.S. Geological Survey, 2004)
- *CSL1085UP*: slope of the upper half of the longest stream channel in the watershed above a measurement site of interest, calculated as the difference in elevation, in feet, between points 10 percent and 85 percent of the distance from the midpoint of that channel to the most distant point from which runoff would reach that channel, divided by the distance in miles between the two points
- *CSL1085UP\_06*: *CSL1085UP* taken to the 0.6 power (which resulted in slightly better predictive accuracy than other mathematical manipulations tested)

## Evaluation of Watershed Properties

The formulation and testing of various conceptualizations for each of the several categories of watershed properties represented in the equations in table 3 are described in the following sections.

### Surficial Geology

Preliminary regression equations confirmed that area of surficial sand and gravel (all soils derived from glacial outwash or postglacial alluvium) correlated positively and significantly with low flow ( $p < |t| = < 0.0001$ ), as expected. Several alternative combinations of soil series were tested in an effort to refine and best represent this concept.

*Alluvium relative to sand and gravel deposited by glacial meltwater*.—Area of alluvium was tested as an independent variable, separate from area of sand and gravel deposited by

**Table 3.** Ordinary least squares regression equations for estimating low flow of streams in the Lower Hudson River Basin, New York.

[An explanation of the columns follows, on p. 16. <, less than; --, value for the equation could not be computed]

Equation number	Independent variable	Ordinary least squares regression equation										
		Tests of independent variables		Tests of accuracy of equation in estimating low flow					Tests of influence of data values		Tests related to normality	
				Standard error				Median percent error of estimate				
		$pr> t $	Variance inflation factor (VIF)	Adjusted $R^2$	Cubic feet per second	Cubic feet per second per square mile	Percent mean of dependent variable					
1												
	$SV\_Sp$	<0.0001	1.07	0.64	0.91	--	77.6	94	48.1	1.44	0.967	-0.165
	$MAQ/DRNAREA$	0.0002	1.07									
2												
	$SV\_Sp$	<0.0001	1.77	0.76	0.74	--	63.2	63	33.8	1.62	0.973	-0.227
	$MAQ/DRNAREA$	<0.0001	1.40									
	$SQFYpSWsqfy$	<0.0001	1.82									
	$CSL1085UP\_06$	0.0127	1.43									
3												
	$SV\_Sp$	<0.0001	1.81	0.77	0.73		61.8	66	35.7	2.03	0.988	-0.209
	$MAQ/DRNAREA$	<0.0001	1.40									
	$SQFYpSWsqfy$	<0.0001	1.88									
	$CSL1085UP\_06$	0.0153	1.44									
	$TTH$	0.0829	1.19									
4												
	$SV\_Sp$	<0.0001	1.07	0.75	1.68	--	62.4	51	159.7	1.02	0.988	-0.192
	$MAQ/DRNAREA$	<0.0001	1.07									
5												
	$SVQ$	<0.0001	2.50	0.85	1.29	--	47.9	41	100.3	1.35	0.990	-0.206
	$MAQ/DRNAREA$	<0.0001	1.40									
	$SQFYpSWsqfy$	<0.0001	2.30									
	$CSL1085UP\_06$	0.0115	1.45									
6												
	$SV\_Sp/DRNAREA$	<0.0001	1.04	0.78	--	0.032	63.9	46	0.054	0.79	0.981	-0.105
	$MAQ/DRNAREA$	0.0013	1.04									



**Table 3.** Ordinary least squares regression equations for estimating low flow of streams in the Lower Hudson River Basin, New York.—Continued

## Explanation of Columns in Table 3

### Independent variable

Acronyms for variables are defined in the text section “Regression Analysis Using Ordinary Least Squares,” p. 12.

### Tests of independent variables

$pr>|t|$

This statistic was computed for each independent variable in each regression equation. It expresses the probability that the true coefficient for that variable might actually be zero rather than the value computed by the regression program (in which case the variable would not be significant and should not be included in the equation). The  $pr>|t|$  statistic is typically assessed as follows (SAS Institute, 2008, online help):

<0.01: Strong evidence that the variable is significant

0.01–0.05: Appreciable evidence that the variable is significant

0.05–0.1: Weak evidence that the variable is significant

(Technically, the regression program computes the Student’s  $t$  statistic, which is the variable coefficient divided by its standard error; then the program calculates the probability of obtaining by chance alone a  $t$  statistic greater in absolute value than what was computed, if the true coefficient were zero. This calculation assumes the data to have an approximately normal distribution.) The  $pr>|t|$  statistic was also computed for the intercept term in each equation, which generally ranged from <0.0001 to 0.02.

### Variance inflation factor

Variance inflation factor was computed for each independent variable. It measures the degree to which that variable is correlated with the other independent variables in the equation. It is computed as  $1/(1 - R^2)$ , where  $R^2$  is the coefficient of determination from a regression of that variable on all other independent variables (Marquardt, 1970). Variables that are totally uncorrelated with (orthogonal to) one another, which is ideal, will have a variance inflation factor of 1.0. A variance inflation factor near or exceeding 10 (which implies  $R^2 \geq 0.9$ ) indicates that variable to be closely correlated (collinear) with one or more other independent variables. As collinearity increases, the magnitudes of regression coefficients become less stable and less representative of the true influence of those watershed variables on low flow.

### Tests of accuracy of equation in estimating low flow

#### Adjusted $R^2$ (coefficient of determination)

Adjusted  $R^2$  can range from zero to 1 and shows what fraction of the variation in the dependent variable (low flow) is explained by the regression equation. It is adjusted for the number of variables in the equation and the number of observations, according to the following formula (terms are defined in the subsequent discussion “Standard error”):

$$\text{Adjusted } R^2 = 1 - (SSE/TSS)(n - 1/n - p)$$



**Table 3.** Ordinary least squares regression equations for estimating low flow of streams in the Lower Hudson River Basin, New York.—Continued

### Standard error

Standard error ( $SE$ ) is expressed in units of the dependent variable (cubic feet per second, cubic feet per second per square mile), and in percent of the mean of the dependent variable. Standard error in units of the dependent variable is computed as follows for ordinary least squares equations:

$$SE = [ \sum (q_o - q_e)^2 / (n - p - 1) ]^{0.5};$$

standard error in units of the dependent variable is computed as follows for weighted least squares equations:

$$SE = [ \sum \{ (q_o - q_e) w \}^2 / (n - p - 1) ]^{0.5};$$

and standard error in percent of the mean of the dependent variable ( $\%SE$ ) is computed as follows:

$$\%SE = SE / (\sum q_o / n);$$

where  $n$  is the number of observations (measurement sites),  
 $p$  is the number of independent variables in the equation,  
 $q_o$  is observed low flow,  
 $q_e$  is low flow estimated by regression equation,  
 $TSS$  is total sum of squares,  
 $SSE$  is sum of squares of error  $[ \sum (q_o - q_e)^2 ]$ , and  
 $w$  is the weight assigned to each observation.

### Median percent error of estimate

$$\text{Percent error of estimate} = 100 [(q_o - q_e)/q_o]$$

The median value is selected from the array of absolute values of percent error of estimate for stations in the dataset.

### Tests of influence of data values

#### Prediction sum of squares (PRESS)

A “prediction residual” can be calculated for each streamflow station by removing that station from the dataset, then estimating low flow for that station from regression analysis based on the remaining stations; the difference (error) between that estimate and the observed low flow constitutes the “prediction residual,”  $e_{(i)}$ , for that station. The PRESS statistic can then be calculated as the sum of the squared prediction residuals:

$$\text{PRESS} = \sum_{i=1}^n e_{(i)}^2$$

Low values of this statistic imply that the regression relation includes few outliers that fall far from the average regression line (and therefore generate large squared prediction residuals). The PRESS statistic is scaled in units of the dependent variable; therefore, comparisons between equations scaled in different units (cubic feet per second, cubic feet per second per square mile, or logarithmic units) are not meaningful.

**Table 3.** Ordinary least squares regression equations for estimating low flow of streams in the Lower Hudson River Basin, New York.—Continued

**Maximum absolute value of DFFITS**

The DFFITS statistic is computed for each station in each equation. It is a scaled measure of the change in the estimated value of the dependent variable for that station that results from regenerating the equation after deleting data for that station from the dataset. The largest absolute value for any station is listed in table 3 for each equation; the larger the value, the more influence a single station has within its neighborhood in the data array.

**Tests related to normality**

**Correlation of residuals with normal scores**

One of the assumptions on which least-squares regression analysis is based is that residuals or errors (observed minus estimated low flows) have a normal distribution. This column reports the coefficients of Pearson product-moment correlations between the residuals and normal scores computed from the ranks of the residuals, using PROC RANK, NORMAL=BLOM option, and PROC CORR (SAS Institute, 1990). For regressions with 50 stations, the distribution of residuals is accepted as normal at the 0.05 alpha level (5-percent significance) if the correlation coefficient equals or exceeds 0.977, or at the 0.025 level (2.5-percent significance) if the correlation coefficient equals or exceeds 0.972 (Looney and Gullledge, 1985).

**Nonparametric test for slope**

This column reports Spearman correlation coefficients ( $\rho$ ) between ranks of residuals and ranks of estimated values. If the absolute value of the coefficient is less than 0.279 (for 50 stations), the slope or fit of the regression equation is accepted as correct at the alpha level of 0.025. This test is valid whether or not the residuals are normally distributed (Iman and Conover, 1983, p. 342, 381).

**Table 4.** Values of watershed properties that represent water available for runoff, stream geometry, and urban land cover for watersheds in the Lower Hudson River Basin, New York.

Watershed number (fig. 1)	U.S. Geological Survey streamflow measurement station		Mean annual runoff 1951–80, in cubic feet per second per square mile (MAQ/DRNAREA)	Length of main stream channel, in miles (LENGTH)	Slope of upstream half of main channel, in (feet per mile) <sup>0.6</sup> (CSL1085UP_06)	NLCD <sub>92</sub> urban land cover, in percent area (NLCDEURBN) <sup>a</sup>
	Number	Name				
1	01358500	Poesten Kill	1.30744	21.87985085	68.02605787	1.65
2	01359100	Wynants Kill	1.22858	14.78143744	149.5028626	5.65
3	01359155	Mill Creek	1.15562	10.85580798	28.34735126	4.01
4	01359200	Normans Kill	1.1652	16.51934713	73.86091779	1.64
5	01359340	Bozen Kill	1.33618	13.6088725	121.9204603	1.92
6	01359513	Hunger Kill (plus Blockhouse Creek)	1.43568	5.56	49.83	18.66
7	01359600	Vloman Kill	1.43125	16.89162397	45.82940377	5.33
8	01359750	Moordener Kill	1.18141	15.01495959	39.65821037	3.49
9	01359902	Coeymans Creek	1.54991	18.39732471	140.3537125	4.69
10	01359990	East Brook	1.74079	6.530158874	279.9320636	0.26
11	01360530	Trout Brook	1.11656	5.407654418	169.3895416	1.37
12	01361200	Claverack Creek	1.233	17.64469644	57.98906832	1.38
13	01361250	Taghkanic Creek	1.48358	7.031072604	109.4949201	0.25
14	01361500	Catskill Creek	1.36492	18.64098832	100.7384539	0.18
15	01361550	Ten Mile Creek 1	1.44526	12.13063118	111.9040201	0.17
16	01361570	Ten Mile Creek 2	1.53738	14.85015819	114.2883355	0.21
17	01361760	Wolf Fly Creek	1.64572	8.067307773	50.92157279	0.19
18	01362005	Bell Brook	1.6015	3.29479993	47.67107867	3.59
19	01362040	Marys Glen	2.18078	1.445261353	1374.145637	0
20	01362100	Roeliff Jansen Kill	1.65235	10.87125079	80.06442887	0.69
21	01362155	Preechey Hollow	1.77322	4.032987578	511.8456461	0.37
22	01362168	Fall Kill	1.47916	5.476375775	98.02108539	0.44
23	01362198	Esopus Creek	2.48516	13.61636209	196.8121831	0.16
24	01362400	Stony Clove Creek	2.21763	11.48775683	264.3277306	0.45
25	01364400	Platte Kill	1.92357	14.72738882	372.6210846	1.32
26	01364700	Saw Kill	1.46	5.497116887	76.57345642	0.24
27	01365500	Chestnut Creek	1.95231	6.528283138	111.8415439	0.29
28	01366650	Sandburg Creek	1.81965	18.74193581	59.0192225	0.81
29	01366750	Beer Kill	1.90809	13.3846107	49.24061685	0.47
30	01366800	Vernooy Kill	2.03191	14.16039833	144.9864431	0
31	01368810	Wawayanda Creek	1.77175	21.59372435	33.13310187	1.8
32	01369650	Stony Creek	1.4342	3.692907432	135.5029792	0.17
33	01370600	Crystal Brook	1.29712	6.850495246	25.06873385	0.44
34	01370800	Tin Brook	1.28975	11.04519359	14.48594213	2.85
35	01370836	Dwaar Kill	1.33471	10.88637441	23.07470592	0.39
36	01372030	Fallsburg Creek	1.30891	4.742160487	44.59289542	0.23
37	01372050	Fall Kill	1.30154	12.55620052	31.04973544	3.51
38	01372065	Casper Creek	1.34797	9.648371102	27.07192209	14.48
39	01372100	East Branch Wappinger Creek	1.46	13.67549481	52.82445156	0.86

**Table 4.** Values of watershed properties that represent water available for runoff, stream geometry, and urban land cover for watersheds in the Lower Hudson River Basin, New York.—Continued

Watershed number (fig. 1)	U.S. Geological Survey streamflow measurement station		Mean annual runoff 1951–80, in cubic feet per second per square mile (MAQ/DRNAREA)	Length of main stream channel, in miles (LENGTH)	Slope of upstream half of main channel, in (feet per mile) <sup>0.6</sup> (CSL1085UP_06)	NLCD <sub>92</sub> urban land cover, in percent area (NLCDEURBN) <sup>a</sup>
	Number	Name				
40	01372200	Wappinger Creek	1.47032	17.44128559	3.623584081	0.56
41	01372300	Little Wappinger Creek	1.44231	17.30074558	32.43019476	0.36
42	01372400	Great Spring Creek	1.32292	7.867547055	40.87680017	1.03
43	01372800	Fishkill Creek	1.56907	21.00179526	27.52778633	1.61
44	01372850	Whortlekill Creek	1.49095	8.13703466	116.6027088	4.67
45	01372900	Sprout Creek	1.44821	20.58528115	53.17715404	0.72
46	01372950	Clove Creek	1.85945	9.980182172	143.6179267	0.87
47	01373690	Woodbury Creek	1.96926	6.050533624	283.8316293	4.88
48	01374440	Cedar Pond Brook	2.07023	8.273149613	161.9053113	2.09
49	01374460	South Branch Minisceongo Creek	2.06507	5.695557125	76.69134799	11.83
50	01376100	Pocantico River	1.81376	9.497365561	26.8004333	5.79
51	01376270	Sparkill Creek	1.69731	4.820701391	181.6056161	12.62

<sup>a</sup>Enhanced National Land Cover Data 1992 (Nakagaki and others, 2010). Although this watershed property slightly improved predictive accuracy of some regression equations, its significance was weak, so it is not included in equations presented in table 3.

glacial meltwater, in numerous trial regression equations. In equations that estimated 7Q2 in cubic feet per second or in cubic feet per second per square mile, alluvium was generally highly significant ( $pr>|t| = <0.01$ ) and was given a regression coefficient comparable in magnitude to that given to glacial sand and gravel. Combining alluvium with area of glacial sand and gravel (as *SVQ*) reduced the number of independent variables in the equation with little or no loss of accuracy. In test equations that estimated 7Q10, however, alluvium was generally not significant as an independent variable ( $pr>|t| = >0.1$ ), degraded predictive accuracy, and was given a regression coefficient much smaller than that given to glacial sand and gravel. Wandle and Randall (1994) combined area of alluvium with area of sand and gravel deposited by meltwater in several regression equations to estimate 7Q2 or 7Q10, but in the same equations they included alluvium with areas of high evapotranspiration (swamps, lakes) in a term that correlated negatively with low flow; in a few equations, they simply excluded alluvium from the sand-and-gravel term. Most areas of alluvium border modern streams and either consist largely of permeable gravels (in regions of high relief) or consist of silt overlying permeable channel-bar gravels that are in hydraulic contact with gravel in the bed of the stream. On the basis of the foregoing information, we infer that some groundwater generally remains stored above stream grade in alluvium when streamflow is near 7Q2 magnitude and continues to drain into the stream, but that groundwater storage in alluvium is essentially depleted before streamflow declines to 7Q10 magnitude.

*Valois soil.*—The Valois soil series consists of very deep, well-drained soil formed in till dominated by sandstone, siltstone, or shale, distributed on lateral moraines along lower valley sides. Valois soils are mapped in the western part of the Lower Hudson River Basin, within the Catskill Mountains and Appalachian Plateau. In this region, and in the Delaware River and Susquehanna River Basins farther west, upper valley sides are steep, whereas many lower valley sides have gentler slopes or rounded benches that lack the planar tops and hummocky topography typical of deltas and kame terraces deposited by meltwater. Borehole data in the Schoharie Creek Basin, immediately to the west of the Lower Hudson River Basin, reveal that the thick glacial drift on the lower valley sides consists of till interbedded with lacustrine fines and scattered lenses of sand and gravel (Randall, 2001, p. 146). The till that caps the rounded benches may have been deposited atop kame terraces by ice readvance, or by downslope creep of colluvium (Denny and Lyford, 1963). Regression equations were improved slightly by combining Valois soil with sand and gravel soils rather than with other till soils. Equations 1–9 in table 3 incorporate this formulation.

*Upland till.*—During periods of low flow, groundwater discharge to streams per unit area of till-mantled uplands is much smaller than the discharge per unit area of surficial sand and gravel. Studies of low flow of streams in Connecticut (Cervione and others, 1982) and central New England (Wandle and Randall, 1994) developed regression equations that reflect this concept, predicting 7Q10 and 7Q2 as the sum

of a large regression coefficient times the area of surficial sand and gravel plus a much smaller regression coefficient times the area of till. Some of the many soil series derived from till in the Lower Hudson River Basin typically overlies bedrock within 6 feet of land surface, whereas in other upland soil series the till extends beyond the 5- to 6-foot depth typically evaluated by soils scientists. The area of upland till more than 6 feet thick in each watershed was readily aggregated from SSURGO (see appendix 1, table 1–1) and was tested as an independent variable added to equations in table 3; in each test the regression coefficient for the till variable was positive and very much smaller than the coefficient for glacial sand and gravel, as expected, but in each test the significance of the till variable was either weak or negligible. The only four-term equation that was improved by adding thick till as an independent variable was equation 2, as illustrated by equation 3 (table 3).

**Hydrologic soil groups.**—The NRCS classifies each soil series into 1 of 4 Hydrologic Soil Groups or combinations thereof, according to the expected rate of water infiltration when the soils are unprotected by vegetation, thoroughly wet, and receiving precipitation from long-term storms. This classification is reported in each county soil survey and is incorporated in SSURGO.

Hydrologic Soil Group A consists mainly of deep, well-drained to excessively drained sands or gravels that have a high infiltration capacity. In the 51 Lower Hudson River Basin watersheds studied, the area of soils assigned to Hydrologic Soil Group A (*HSG\_A*) is equal to 87 percent of the area classified as outwash sand and gravel for this study, but to a much smaller percentage of the area classified as outwash plus alluvial sand and gravel. Despite this variation, the correlation of *HSG\_A* with each of the two variables that appear in several regression equations in table 3 as alternative representations of surficial sand and gravel [*SV\_Sp* and *SVQ\_Sp*] is close to 0.96, as shown in table 5.

Four regression equations that incorporate *HSG\_A*, along with mean annual runoff, are listed in table 3. 7Q2 in cubic feet per second is estimated slightly more accurately by using *HSG\_A* to represent surficial sand and gravel (equation 11) than by using *SV\_Sp* (equation 4). However, 7Q10 and 7Q2 are estimated less accurately by using *HSG\_A* (in equations 10, 12, and 13) rather than *SV\_Sp* (in equations 1, 6, and 8). Furthermore, the additional watershed properties related to evapotranspiration and watershed geometry that were strongly or appreciably significant and improved predictive accuracy when included in equations 2, 3, 5, 7, and 9 were weakly significant at best and had little effect on predictive accuracy when added to equations 10 to 13, and no better alternatives were found. *HSG\_A* apparently does not capture the effect of surficial geology on low flow quite as well as the terms *SVQ*, *SVQ\_Sp*, and *SV\_Sp* devised in this study. Trial equations that also incorporated combinations of Hydrologic Soil Groups B, C, and D (with negative regression coefficients) were comparable in accuracy to equations 10–13, so these trial equations were not included in table 3.

## Amount of Water Available for Runoff

Three measures of long-term average water input to a watershed were included in the dataset: 1951–80 mean annual runoff, 1951–80 mean annual precipitation, and mean watershed altitude. Trial equations demonstrated that mean annual runoff, expressed in cubic feet per second per square mile, resulted in the greatest predictive accuracy and was highly significant. Combining surficial geology with water availability in a single term (sand and gravel area, in square miles, multiplied by mean annual runoff per square mile) seemed logical inasmuch as doubling the aquifer area within the watershed or doubling the annual water input to that aquifer might be expected to double annual discharge to streams. However,

**Table 5.** Comparison of area of Hydrologic Soil Group A, as delineated by the Natural Resources Conservation Service, with areas of two surficial geology variables compiled for this report in the Lower Hudson River Basin, New York.

[*HSG\_A*, *SV\_Sp*, and *SVQ\_Sp* are defined in the “Regression Analysis Using Ordinary Least Squares” section of the report, p. 12]

Area of Hydrologic Soil Group A ( <i>HSG_A</i> )	Surficial geology variable	
	<i>SV_Sp</i> (outwash sand and gravel, not poorly drained)	<i>SVQ_Sp</i> (outwash sand and gravel, not poorly drained, plus alluvium)
Median of the areas of <i>HSG_A</i> in 51 watersheds, each expressed as percent of the area of the indicated surficial geology variable	87	52
Range in <i>HSG_A</i> areas, expressed as percentages of the area of each surficial geology variable, excluding the two largest and two smallest areas	100 to 20	77 to 11
Correlation coefficient ( <i>R</i> ) of <i>HSG_A</i> area with area of each surficial geology variable	0.967	0.948



such combinations degraded predictive accuracy. Mean annual runoff in cubic feet per second was significant in predicting 7Q10 and 7Q2 in cubic feet per second, but mean annual runoff in cubic feet per second per square mile was even more significant and effective, probably because area of sand and gravel multiplied by mean annual runoff per square mile is nearly proportional to recharge to and eventual discharge from sand and gravel, whereas area of sand and gravel multiplied by mean annual runoff from entire watersheds (much of which is storm runoff from uplands that largely fails to recharge groundwater) is less proportional to recharge.

## Riparian Evapotranspiration Loss

Previous studies have suggested that high rates of evapotranspiration from riparian wetlands and poorly drained soils, particularly in areas of sand and gravel, can divert to the atmosphere a significant fraction of groundwater discharge that otherwise would have become streamflow during periods of low flow. Several combinations of pertinent soil types and properties were extracted from SSURGO and tested as independent variables in regression equations. The third variable in equations 2, 3 and 5 (table 3), which estimate 7Q10 and 7Q2 in cubic feet per second, represents the spatial extent of somewhat poorly to very poorly drained soils in outwash sand and gravel, alluvium, and lacustrine silt and clay, plus the areal extent of swamps bordered at least in part by sand and gravel and (or) silt and clay. This variable proved to be highly significant ( $pr > |t| = < 0.0001$ ) in all three equations and was inferred to reflect the likelihood of intense groundwater discharge from these soils by evapotranspiration. In equations 7 and 9, which estimate 7Q10 or 7Q2 in cubic feet per second per square mile, the third variable [ $SWsqFY/DRNAREA$ ] represents the area of swamps bordered at least in part by sand and gravel plus the area of surficial clay, silt, and fine sand. This variable is nearly as significant ( $pr > |t| = < 0.01$ ) in equations 7 and 9 as the comparable wetland variable [ $SQFYpSWsqfy$ ] in equations 2, 3, and 5. Area of lakes bordered at least in part by outwash sand and gravel or alluvium was expected to have a similar effect on 7Q10 and 7Q2, as reported in other studies (Wandle and Randall, 1994; Randall, 2010; L. Weiss and R. Melvin, USGS, written commun., 1983). However, these lake areas turned out to have a weak *positive* correlation with 7Q2 and 7Q10. Several large manmade lakes in the Lower Hudson River Basin watersheds studied for this report are bordered entirely by till except for tiny areas of alluvium along a tributary stream; perhaps the illogical positive correlation could have been eliminated by a revised selection algorithm that rejected lakes if less than some small percentage of the lake perimeter were bordered by sand and gravel.

Deletion of somewhat poorly to very poorly drained soils from the area of glacial sand and gravel (indicated by the suffix “\_Sp” after *SV* or *SVQ*) also generally improved predictive accuracy slightly, and is incorporated in most equations in table 3. This improvement was inferred to reflect the

likelihood that the shallow water table in these soils often rejects recharge.

## Watershed Geometry

Several variables that describe elements of watershed geometry were tested in regression equations during this study. One such variable, the slope, in feet per mile, of the upstream half of the principal stream channel in each watershed, correlated negatively, as expected, with 7Q10 and 7Q2 in cubic feet per second. The maximum predictive accuracy was achieved when channel slope was expressed to the 0.6 power. The  $pr > |t|$  statistic indicates that this variable is strongly significant in estimating 7Q2 or 7Q10 in cubic feet per second (equations 2, 3, 5 in table 3). A different geometric variable, the length of the principal stream channel divided by watershed area, proved comparably significant in estimating 7Q10 and 7Q2 in cubic feet per second per square mile. The negative sign and the improved predictive accuracy resulting from this variable imply that a long, narrow deposit of sand and gravel parallel to a stream channel yields less water per square mile during periods of low flow than an equidimensional deposit of equal area, presumably because the wider deposit takes longer to drain. This interpretation is consistent with results of groundwater flow modeling (Randall, 2010).

Several other geometric variables were repeatedly tested in trial equations but failed to improve the predictive capability of the equation and (or) had only weak or negligible significance ( $pr > |t| = 0.05$  or more) so were discarded. They included watershed relief, average watershed slope (in feet per mile or in percent), ratio of main channel slope to watershed slope, average or maximum or minimum altitude, and lag factor, which is defined as channel length / [(upstream channel slope + 1)(downstream channel slope + 1)]<sup>0.5</sup>.

## Evaluation of Sources of Predictive Error in Regression Equations

OLS equations 7 and 9, the most accurate equations in table 3, explain 83 and 86 percent of the variation in 7Q10 and 7Q2, respectively. Standard errors are 0.028 and 0.044 cubic feet per second per square mile, respectively; the differences between low flows calculated from streamflow data and low flows estimated from these two equations were less than or nearly the same as the standard error at 73 and 77 percent of the 51 measurement sites. Comparable equations developed for the Susquehanna River Basin in New York (Randall, 2010) explained 90 percent of the variation in 7Q10 and 7Q2 in cubic feet per second per square mile. Comparable equations developed for central New England (Wandle and Randall, 1994) explained 96 percent of the variation in 7Q10 and 7Q2 in cubic feet per second. The lesser predictive accuracy of the Lower Hudson River Basin equations could be attributed to three factors, discussed in the following paragraphs.

1. *Calculated 7Q10 and 7Q2 values may be imprecise at some streamflow measurement stations.*—Of the 51 stations studied in the Lower Hudson River Basin dataset, 37 (72 percent) were partial-record stations or short-term gaging stations with less than 8 years of record. Of the 65 stations used for per-square-mile regression analysis in the Susquehanna River Basin dataset, 40 (61 percent) were partial-record stations or short-term gaging stations with only 2 years of record. Of the 49 stations in the central New England dataset, 3 (6 percent) were partial-record stations. Low-flow indices for partial-record stations or short-term gaging stations were calculated by correlation procedures that are less precise than procedures used at long-term gaging stations. To the extent that imprecise calculations of low-flow indices are responsible for error in estimation of those indices by regression equations, the equations (which predict average conditions) might be somewhat more (or less) accurate than the statistical tests in table 3 suggest. However, the correlation of partial-record stations with long-term stations commonly introduces a bias, in which case the statistical test overestimates accuracy.
2. *Underflow may result in streamflow at some measurement stations being less than total runoff from the watershed.*—Where streams occupy narrow valleys incised into lacustrine silt or clay, till, or bedrock, all runoff from the watershed necessarily appears in the stream channel. Where streams flow across flood plains or alluvial fans underlain and (or) bordered by sand and gravel deposited by the modern stream or by meltwater, some of the runoff can flow downvalley through the sand and gravel as “underflow,” bypassing the stream channel. Streamflow from 18 of the 51 watersheds evaluated in this investigation was measured at sites where underflow is probably negligible (table 1), as inferred from soils maps or observations of channel conditions in USGS files. At the other 33 sites, the width of alluvial and (or) glaciofluvial sand and gravel perpendicular to the stream is 300 to 4,800 feet; information on the saturated thickness and hydraulic conductivity of the sand and gravel was not readily available. A few pertinent streamflow measurements were available at five sites:
  - *Trout Brook (station 01360530).*—On August 14, 1981, flow was measured as 0.27 cubic feet per second at this site, which is located on an alluvial fan; that rate is twice the calculated 7Q2 of 0.13 cubic feet per second. On the same date, flow was measured as 0.32 cubic feet per second 800 feet upstream within a gorge incised into till. These data suggest that underflow could account for the smaller flow at the measurement site on that date, and the zero flow observed there on August 4, 1964.
  - *Wolf Fly Creek (station 01361760).*—On August 11, 2010, flow was measured as 0.10 cubic feet per second at this site, which is located where Wolf Fly Creek begins to cross alluvial and glaciofluvial gravel in the valley of a larger stream. Flow decreased downstream to zero 100 feet from the mouth. On the same date, flow was measured as 0.12 cubic feet per second about 1,500 feet upstream, where the county soils survey depicts Wolf Fly Creek as incised in till. Barnes (1986) calculated 7Q10 and 7Q2 at the measurement site to be zero. These data indicate that underflow was about 0.02 cubic feet per second at the measurement site and increased downstream on August 11, 2010, and was probably larger during periods of 7Q2 and 7Q10 low flow.
  - *Stony Clove Creek (station 01362400).*—On September 17, 1964, flow was 0.03 cubic feet per second (0.0009 cubic feet per second per square mile) at this measurement site, where Stony Clove Creek enters the valley of Esopus Creek. On the same date, mean daily flow of Esopus Creek was 4.2 cubic feet per second (0.07 cubic feet per second per square mile) nearby at Shandaken, N.Y. (station 01362198), a watershed of similar terrain with a similar percentage area of surficial sand and gravel. The much smaller runoff per square mile in Stony Clove Creek suggests that on this date, during a severe drought, much of the runoff from Stony Clove Creek watershed may have bypassed the measurement site as underflow. On August 21, 2010, by contrast, flow at the measurement site on Stony Clove Creek was 6.1 cubic feet per second, more than twice the calculated 7Q2 of 2.8 cubic feet per second and slightly larger than the concurrent flow of 5.8 cubic feet per second at a streamgaging station on Stony Clove Creek 1.8 miles upstream (station 013623800).
  - *Woodbury Creek (station 01373690).*—On September 23, 2010, flow was measured as 0.58 cubic feet per second at this site and 0.57 cubic feet per second 1,500 feet upstream near bedrock outcrops in the channel. On this date, several larger streams in the central part of the Lower Hudson River Basin were flowing at rates between 7Q2 and 7Q10 as calculated by Eissler (1979), and the measured flow of Woodbury Creek was also between calculated 7Q2 and 7Q10 (table 1). The county soils survey depicts surficial sand and gravel extending 1,200 feet west from Woodbury Creek at the measurement site, but quarry operations suggest that bedrock underlies the gravel at shallow depth, above stream grade. These observations indicate that underflow may be negligible at this measurement site.
  - *Blockhouse Creek (station 01359517).*—On October 28, 1966, streamflow was 0.76 cubic feet per second at the measurement site, where the creek traverses alluvium, but 1.66 cubic feet per second 1,000 feet upstream where the principal channels are incised in clay. Calculated 7Q2 at this site is 0.90 cubic feet per second.

Although the magnitude of underflow relative to the range of low flow has not been documented at most measurement sites in the Lower Hudson River Basin, significant depletion of 7Q10 by underflow at some measurement sites remains a possibility.

3. *Extent of surficial sand and gravel delineated by county soil surveys may be locally incorrect and (or) an imperfect index of groundwater discharge to streams.*—The regression equations developed in this study confirm previous studies in indicating that the spatial extent of surficial sand and gravel has much greater effect on low flows than any other watershed property tested. However, delineation of the areal extent of surficial sand and gravel in New England studies (Cervione and others, 1982; L. Weiss and R. Melvin, USGS, written commun., 1983; Wandle and Randall, 1994) was based largely on detailed maps prepared by geologists qualified to map glacial deposits. Delineation in the Susquehanna River Basin of New York (Randall, 2010) was based on less detailed reconnaissance maps by qualified geologists, facilitated by typically strong topographic contrasts between glacial or alluvial sand and gravel in valleys and till or bedrock on valley sides.

Interviews with several experienced geologists revealed that they usually compared their observations with published soils surveys as a means of quality control; their consensus was that the accuracy of county soils surveys varies widely from one county to another, depending on who did the mapping, and that map units delineated by soil scientists on aerial photographs are often less constrained by altitude and landforms than map units delineated by geologists on topographic maps.

The only detailed, published surficial geology map of a part of the Lower Hudson River Basin is that of LaFleur (1965), which covers all or parts of watersheds 1–3 and 8 (table 1), all in Rensselaer County. Two widespread, systematic differences in interpretation are evident between this map and the Rensselaer County soils survey:

- A region of moderately low relief lies between the much higher Rensselaer Plateau to the east and the low-lying lacustrine sediments to the west in the Hudson River valley. Part of this region was mapped by LaFleur as ice-contact (kame or outwash) gravels, interrupted by numerous large bedrock outcrops in some localities. The county soils survey depicts most of the localities with bedrock outcrops as thin till (less than 6 feet to bedrock).
- The soils survey depicts nearly all of the Rensselaer Plateau as till, except for organic soils in several large swamps and alluvium along lower reaches of the Poesten Kill. LaFleur interpreted several hummocky or terracelike topographic features low on hillsides in the plateau as kame or outwash gravel. He also interpreted as outwash gravel several broad, flat valley-floor reaches along tributary streams, which were interpreted

by the soils survey as till (on the plateau) or as alluvium (west of the plateau). In general, the soils survey probably underrepresents the surficial extent of glacial and alluvial sand and gravel in Poesten Kill, Wynants Kill, and Moordener Kill watersheds, but it is not known whether the areas of thinly saturated outwash or alluvial sand and gravel that were overlooked by the soils survey are important contributors to low flow.

Esopus Creek (watershed 23, fig. 1 and table 1) drains an area of high relief in the Catskill Mountains, mostly in Ulster County. The valley floor consists of a flood plain, generally 300 to 1,000 feet wide, bordered in many places by terraces 10 to 30 feet higher than the flood plain, and by alluvial fans near tributaries. The soils survey depicts the valley wall upslope from the terraces as mantled by lacustrine silt and clay in many places, up to an altitude of about 1,300 feet. The terraces and alluvial fans are depicted as Hoosic or Tunkhannock soils, which are described in soil surveys for all counties as derived from gravel deposited by glacial meltwater in deltas, kames, outwash terraces, and valley trains. In Ulster County only, these soils are also attributed to postglacial stream terraces and alluvial fans, as well as to glacial meltwater. Indeed, the terraces in Esopus Creek watershed cannot have been deposited by meltwater, because their upper surfaces grade downstream to the north and east (toward the retreating ice sheet). Clearly, these low terrace gravels were deposited by late-glacial or postglacial Esopus Creek and its tributaries, as they incised their valleys. Several exposures in terrace scarps and on the flood plain reveal a few feet of highly permeable cobble-boulder gravel, overlying lacustrine silt and clay or till that are sources of appreciable turbidity in Esopus Creek during episodes of high runoff (Davis and others, 2009; New York City Department of Environmental Protection, 2008). Recent geological mapping (Smith and others, 2001) supports this interpretation; the Esopus Creek valley bottom upstream from Allaben, N.Y., is depicted as alluvium, bordered in some places by alluvial terraces. Two test holes and logs of several water wells document that lacustrine silt and clay underlie the alluvial gravels all the way up the valley, although no surficial lacustrine silt or clay are mapped anywhere in the watershed. Thin, highly permeable terrace gravels could be expected to drain more quickly than the thick gravelly sands that typically underlie Hoosic and Tunkhannock soils elsewhere, and thus to yield little water during periods of low flow. The regression analyses in this report provide evidence that alluvium in the Lower Hudson River Basin yields less water per square mile during periods of 7Q10 low flow than sand and gravel deposited by glacial meltwater, as discussed in the subsection “Surficial Geology.” Unfortunately, SSURGO does not distinguish most of the postglacial stream terraces in Esopus Creek watershed from thicker meltwater deposits elsewhere that are also depicted as Hoosic and Tunkhannock soils.<sup>1</sup>

<sup>1</sup>Some Tunkhannock soils were distinguished by SSURGO as underlain by a “clayey substrate” at depths of 3.5 to 6 feet, so they were grouped for this study with other soil series described as a few feet of surficial sand over silt and clay.



Map units delineated by soil scientists on aerial photo-mosaic base maps and compiled by SSURGO were projected by GIS onto a USGS topographic base for each watershed in this study. Examination of the resulting maps revealed two widespread inconsistencies:

- The deposition of sand and gravel by meltwater along valleys in the glaciated northeastern United States generally conformed to a succession of depositional profiles (termed morphosequences) that were graded to lake surfaces or bedrock spillways. Some meltwater deposits delineated by SSURGO are separated by intervals within which small terracelike features close to the altitude of the depositional profile are present on the valley sides but interpreted as soils derived from till.
- Many tracts of flat valley floor several hundred feet wide that are present locally along upland tributaries are mapped by SSURGO as till, despite altitudes approximately at stream grade that would seem to allow deposition of alluvium.

None of these features were checked in the field for this study, and even if some of them consist of glacial or alluvial sand and gravel, their saturated thickness may be so slight that they contribute little water to streams at low flow.

## Logarithmic Transformations

Streamflow and watershed properties were transformed into logarithms prior to regression analysis in many studies directed to estimation of indices of low or high streamflow. Logarithmic transformations were deemed unnecessary and unsuitable in the present study, however, as explained in the following paragraphs.

*Data distribution.*—Riggs (1985, p. 84–85) gave two reasons for considering logarithmic transformation of variables prior to regression analysis: to linearize curved data distributions and to achieve equal variation about the regression line over its entire length. In some datasets, plots of dependent variables (such as low-flow indices calculated from streamflow measurements) against individual or combined independent variables show close grouping of data points near the origin but increasingly wide scatter as data values increase, a condition known as heteroscedasticity (Iman and Conover, 1983, p. 369). Logarithmic transformations tend to compress large values and expand small values, which can yield a more equal distribution of variances along the regression line. Natural values of the dependent variables used in this study (7Q10 and 7Q2, each in cubic feet per second and in cubic feet per second per square mile) are plotted in figure 2 against individual watershed properties that represent surficial sand and gravel area, wetland area, and mean annual runoff in equations 1–9 (table 3), and against the combined effects of several watershed properties as expressed by estimates of 7Q10 or 7Q2 generated by four of these equations. The four graphs of

observed low flow in relation to estimated low flow display no curvature and little or no increase in scatter of data points as low flow increases. The graph of  $SV\_Sp$ , which represents area of well-drained sand and gravel, does display some increase in scatter with increasing 7Q10. Little or no increase in scatter with increasing 7Q10 or 7Q2 is evident, however, when  $SV\_Sp/DRNAREA$  (well-drained sand and gravel per square mile) or  $SQFYpSWsqfy$  (wetland area) are similarly plotted. Thus, figure 2 (and other similar graphs, not shown) provides scant evidence that logarithmic transformations of natural values are needed to avoid curvature or heteroscedasticity.

*Conceptual Physical Relations.*—Each regression coefficient in the equations in table 3 converts the measured value of a watershed property into the positive or negative contribution of that property, in cubic feet per second (or cubic feet per second per square mile), to low flow. For example, equation 2 would be conceptually expressed as follows:

$$7Q10 = \text{intercept constant} + a(\text{area of well-drained surficial sand and gravel plus Valois soil}) + b(\text{net water input [precipitation minus evapotranspiration] per square mile}) - c(\text{area of swamps and poorly drained soils in valley fill}) - d(\text{slope of upper half of main channel})^{0.6}, \quad (A)$$

where

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

If all measured values were transformed to natural logarithms (base  $e$ ) before analysis, the resulting logarithmic regression equation would have a similar form:

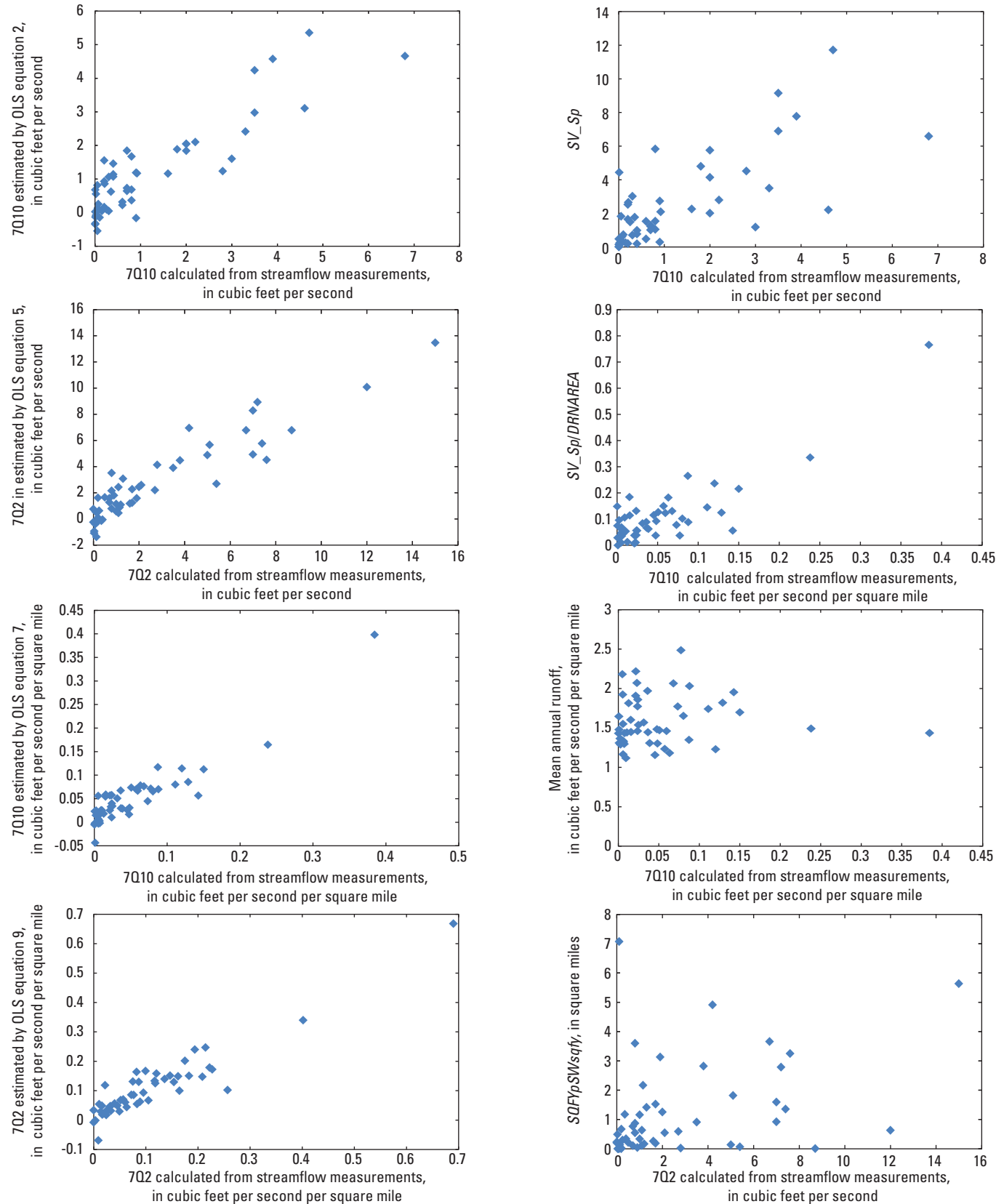
$$\ln(7Q10) = \ln(\text{intercept}) + a \ln(\text{area well-drained surficial sand and gravel plus Valois soil}) + b \ln(\text{net water input per square mile}) - c \ln(\text{area swamps and poorly drained soils in valley fill}) - d \ln(\text{slope of upper half of main channel})^{0.6} \quad (B)$$

which is equivalent to equation C:

$$7Q10 = (e^{\text{intercept}})(\text{area well-drained sand and gravel plus Valois soil})^a (\text{net water input per square mile})^b (\text{area swamp and poorly drained soils in valley fill})^{-c} ([\text{slope of upper half of main channel}]^{0.6})^{-d} \quad (C)$$

Equation C is altogether different conceptually from equation A. Equation C expresses 7Q10 as the *product* of four watershed properties, each raised to some power.

Let us assume a set of watersheds in which riparian wetland areas (specifically, the areas of swamps plus poorly drained soils within valley-fill deposits) range from 0.1 to 10 square miles. If the regression coefficient for the wetland term in a logarithmic regression equation is negative, so that the exponent for that term in a power equation such as equation C is also negative, then the value of that term in the power equation will be less than 1.0 for each watershed where total wetland area is greater than 1.0 square mile. Therefore, when multiplied by sand and gravel area and mean annual runoff



**Figure 2.** Low flows calculated from streamflow measurements compared with low flows estimated by four ordinary least squares (OLS) regression equations and with four watershed properties used in several equations. See table 3 for OLS equations. See the “Regression Analysis Using Ordinary Least Squares” section of the text for definitions of watershed properties. 7Q2, 7-day, 2-year low flow; 7Q10, 7-day, 2-year low flow.

in the regression equation, the wetland term will function to reduce the equation's estimate of 7Q10, as conceptualized. However, in watersheds where wetland areas total less than 1.0 square mile, the negative regression coefficient (or power) will cause the wetland term to be greater than 1.0, which will increase the equation's estimate of 7Q10. Conversely, if the exponent in equation C is positive, the value of the term will be less than 1.0 for watersheds with total wetland area less than 1 square mile, but greater than 1.0 for watersheds with total wetland area greater than 1 square mile. These illogical results could presumably be avoided by expressing wetland area in some smaller measurement unit (such as acres) that would be greater than 1.0 for any watershed, in which case a negative regression coefficient would result in a wetland term smaller than 1.0 for any watershed.

However, there is a further conceptual inconsistency in the representation of wetlands in logarithmic regression equations. If we assume a regression coefficient (or power) of -0.25, a wetland area of 1.5 square miles would result in a term with a value of 0.9 in the regression equation, regardless of the size of the sand-and-gravel area (the source of most low flow). That term would reduce the estimate of 7Q10 by 10 percent regardless of whether the sand and gravel area is 0.5 square miles or 10 square miles. This result does not make sense. (By contrast, in a regression equation scaled in natural values, the regression coefficient represents evapotranspiration loss, in cubic feet per second per square mile of wetland. Such a coefficient converts wetland area, in square miles, into riparian evapotranspiration loss, in cubic feet per second, from that wetland area, regardless of the area of sand and gravel or the magnitude of mean annual runoff, whose contribution of groundwater toward the stream prior to partial capture by riparian evapotranspiration is estimated by other terms in the natural-value regression equation.)

To test the foregoing conceptual reasoning, logarithmic equations were generated that used the same watershed properties as in equations 1 and 2 (table 3). Table 6 compares two

statistical indices for these logarithmic equations with the corresponding natural-value equations. In some equations, drainage area is included as an independent variable, with area of well-drained sand and gravel and area of poorly drained soils and swamps expressed as fractions of drainage area. The table indicates that the natural-value versions of both equation 1 and equation 2 have higher (better) values of adjusted  $R^2$  and correlation of residuals with normal scores, and that the wetland and channel-slope watershed properties, which caused equation 2 with natural values to outperform equation 1, proved not to be significant in the logarithmic equations.

Accordingly, logarithmic transformations were not investigated further. Similar conclusions regarding logarithmic transformations were reached in an earlier low-flow investigation (Randall, 2010). Several regional studies of low flows in Connecticut (including Cervione and others, 1982) developed regression equations for 7-day low flows using natural values. Male and Ogawa (1982, p. 29–30) concluded from analysis of streamflows in Massachusetts that “short-duration low flows may be better described by a linear model than by a log-linear model.”

## Regression Analysis Using Weighted Least Squares

The error in predictive ability of a regression equation consists of two components: “model error” (the extent to which the selected explanatory variables fail to account for all the watershed properties that actually influence low flow) and “time-sampling error” (the extent to which 7Q10 and 7Q2 calculated from observed streamflows at measurement sites fail to represent the true low-flow indices for the reference period at those sites). Time-sampling errors are likely to be smaller at streamgaging stations where a continuous record of streamflow is available for all or nearly all of the reference period

**Table 6.** Comparison of ordinary least squares equations 1 and 2 (table 3) with logarithmic transformations of those equations.

[Variables are fully defined in the section “Regression Analysis Using Ordinary Least Squares,” p.12. ln, natural logarithm; *DRNAREA*, drainage area]

Equation (table 3)	Format	Adjusted $R^2$	Correlation of residuals with normal scores
1	Natural values	0.64	0.967
1	ln (logarithmic) transformation	0.602	0.933
1	ln transformation, <i>DRNAREA</i> term included	0.596	0.925
2	Natural values	0.76	0.973
2	ln transformation <sup>a</sup>	0.57	0.923
2	ln transformation, <i>DRNAREA</i> term included <sup>a</sup>	0.58	0.93
2	ln transformation, <i>DRNAREA</i> term included <sup>b</sup>	0.61	0.961

<sup>a</sup>Wetlands [ln *SQFYpSWsqfy*] and channel slope [ln *CSLI085UP\_06*] also included but are not significant [ $pr>|t| = >0.7$ ].

<sup>b</sup>Wetlands incorporated in the term ln[*DRNAREA* – *SQFYpSWsqfy*], but that term is not significant ( $pr>|t| = 0.14$ ). Channel slope omitted.

than at stations where a continuous record of streamflow was obtained for only a few years and at partial-record stations where occasional streamflow measurements were made over a few years and correlated with continuous records elsewhere to estimate low-flow indices. Tasker (1980) and Stedinger and Tasker (1985) presented sophisticated methods to account for unequal lengths of continuous record at measurement sites, based in part on the skew coefficients and standard deviate generated by the Pearson Type III distribution commonly used to calculate low-flow indices. This approach was not feasible for the present study because records of the Pearson Type III analysis undertaken by Barnes (1986) were no longer available. However, a simpler alternative was undertaken, approximately as suggested in class lecture notes by G. Tasker and K. Eng (Flynn, 2007, section M): weights were equated to the number of years of continuous record at streamgaging stations and 0.75 times the number of years in which occasional streamflow measurements were made at partial-record stations. The use of weights should result in more realistic tests of accuracy of equations.

Several trial regression runs revealed that if the weight for every site was divided or multiplied by the same number (such that the ratio of weights from one site to another was not changed), the regression coefficients for independent variables and most statistical indices of accuracy, influence, and normality were unchanged—but variance and standard error varied in proportion to the sum of the weights. Therefore, each initial weight was divided by the sum of weights (376.25 years) and multiplied by 51 (the number of sites), thereby converting the sum of weights to 51, which is the implied sum of weights in the OLS regressions where each residual is assumed to have a weight of 1.0.

The four regression equations in table 3 that were deemed most promising for estimating low flows at unmeasured sites in the Lower Hudson River Basin were rerun by using weighted least squares; results are presented in table 7. WLS equations 2, 5, and 9 in table 7 include the same 4 watershed properties as their OLS equivalents in table 3; WLS equation 7 in table 7 includes only 3 watershed properties because the channel-length variable proved to not be significant in the WLS regression. WLS equations 7 and 9 explain nearly 84 and 87 percent, respectively, of the variation in 7-day, 10-year low flows; standard errors are 0.032 and 0.050 cubic feet per second per square mile (48.8 and 36.9 percent of the mean of the dependent variable). For all four WLS equations, the coefficients of determination (adjusted  $R^2$ ) are slightly larger than their OLS equivalents. Standard errors expressed in cubic feet per second are somewhat larger in the WLS equations than in the equivalent OLS equations, but are smaller in percentage of the mean

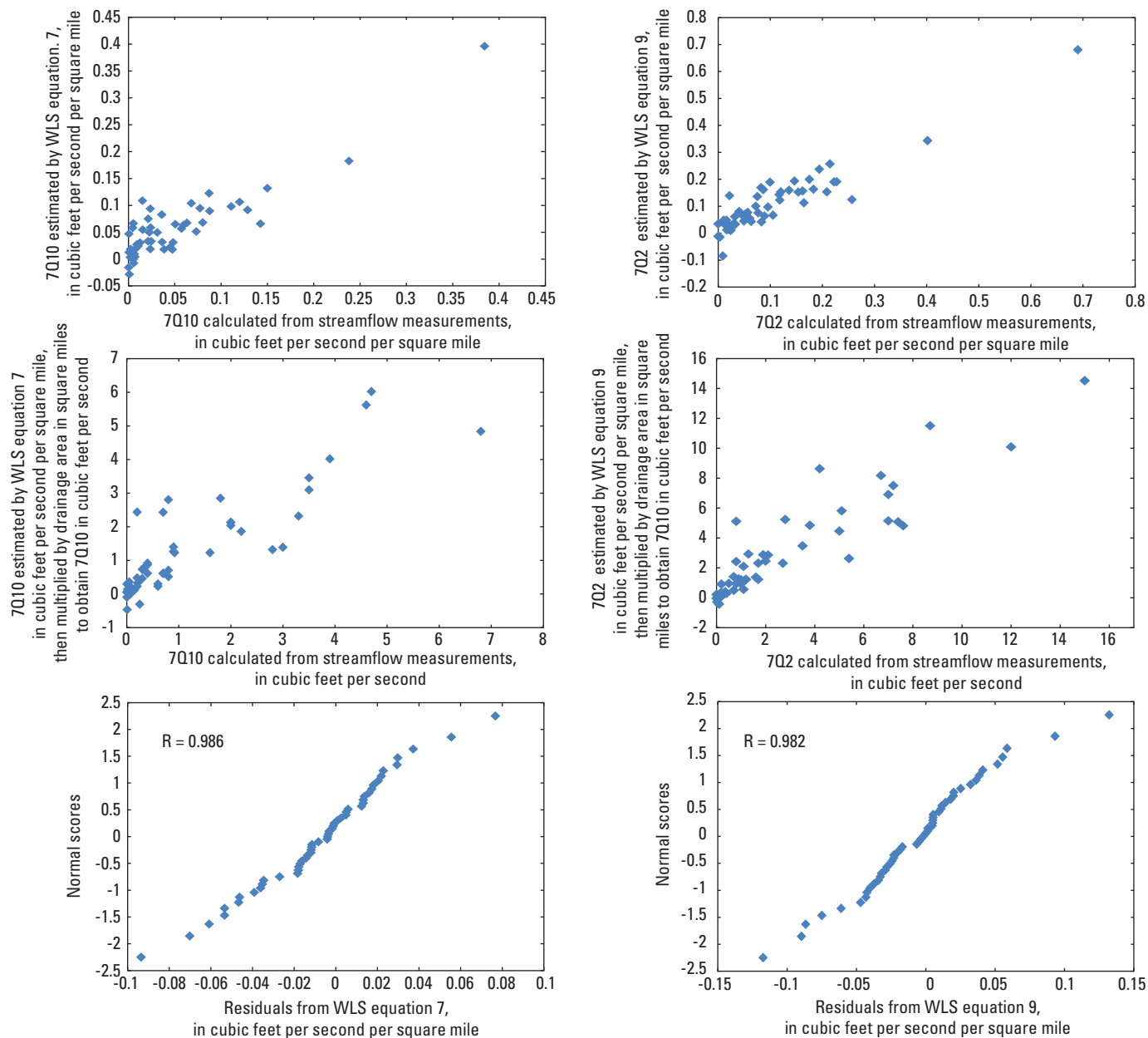
of the dependent variable, and median percent errors of estimate are larger. In all four WLS equations, the correlation of residuals with normal scores exceeds 0.977, indicating a normal distribution at the 5-percent significance level. Some individual measurement sites are more influential in the WLS equations than in equivalent OLS equations, as indicated by larger PRESS and maximum DFFITS statistics.

The statistical validity of a regression equation is degraded to the extent that the absolute value of residuals correlates with the magnitude of independent variables or of predicted low flows. As suggested by Gary Tasker (written commun., 1981), SAS PROC CORR was programmed to calculate the Spearman correlation coefficient for these relationships and its significance level. Only 1 of the 15 correlations of residuals with individual watershed properties in the four WLS equations proved to be significant at the 5-percent significance level. Absolute values of residuals were also not significantly correlated with predicted low flows in WLS equation 5. In WLS equations 2, 7, and 9, absolute values of residuals did correlate with predicted low flows at the 5-percent significance level, but the correlations were appreciably weaker than in the equivalent OLS equations in table 3. Graphs that compare observed low flows with low flows estimated from WLS equations, and that compare residuals with normal scores, are presented in figure 3.

Residuals from WLS equation 2 (7Q10 calculated from streamflow records minus 7Q10 estimated from the regression equation) were compared with residuals from OLS equation 2. The use of weights resulted in a net reduction in the sum of residuals at all 51 sites of -1.45 cubic feet per second. Twelve of the 13 measurement sites with records of 4 years or less had somewhat larger residuals in the WLS equation. These results would be expected from placing greater weight on sites with longer records. It is worth noting, however, that 7 of the 12 sites with relatively long continuous records (11 to 20 years) also had somewhat larger residuals—but 3 of those 12 sites had unusually large reductions that totaled -2.18 cubic feet per second and overwhelmed the much smaller increases at many other sites. These three sites—Esopus Creek, Chestnut Creek, and Sandburg Creek—all drain terrain of relatively high relief near the western margin of the Lower Hudson River Basin. This result might suggest an interpretation that the WLS equations predict 7Q10 more accurately than OLS equations in this high-relief terrain but less accurately in the remainder of the Lower Hudson River Basin. However, average, maximum, and minimum altitude, watershed relief, and mean watershed slope all proved not to be significant when tested in regression equations, which is evidence against such an interpretation.







**Figure 3.** Low flows calculated from streamflow measurements compared with low flows estimated from weighted least squares (WLS) equations 7 and 9, also comparing residuals from these equations with ranks of residuals (normal scores). See table 7 for WLS equations. 7Q2, 7-day, 2-year low flow; 7Q10, 7-day, 10-year low flow.

## Suggested Methods for Estimating Statistical Indices of Low Flow at Ungaged Sites in the Lower Hudson River Basin

The most accurate and reliable way to define the low flow of a stream at a particular site is to measure the full range of flow by operating a streamgaging station at that site. The accuracy of the calculated low-flow indices will increase with the length of the streamflow record. Less exact estimates of low-flow indices can be obtained by making 10 or more streamflow measurements at the site of interest that represent different periods of streamflow recession and that include a range of low flows, then correlating those measurements with concurrent daily mean flows at one or more continuous-record streamgaging stations. For best results, the streamgaging station(s) should be on a nearby unregulated stream that drains a watershed similar in size and surficial geology to the watershed of interest. Procedures for such correlations are explained by Cervione and others (1982), Hardison and Moss (1972), Matalas and Jacobs (1964), Reilly and Kroll (2002), Searcy (1959), Stedinger and Thomas (1985), and Zhang and Kroll (2007). The use of correlation procedures by Barnes (1986) to calculate many of the values of 7Q10 and 7Q2 that are listed in table 1 of this report was facilitated by the availability of records from the many streamgaging stations that were operated on unregulated streams in and near the Lower Hudson River Basin prior to 1978. Relatively few stations have been operated during subsequent years (through 2015), which may restrict correlations for those years to gaged watersheds that are relatively distant or dissimilar in size or geology.

Regional regression equations can be used to estimate low flows at sites where no measurements have been made nearby and there is no prospect of obtaining correlatable measurements over a suitable range of low flow. Such equations can be applied to entire unregulated watersheds. If a site of interest is on a stream where low-flow indices have been calculated for a streamgaging station some distance upstream or downstream, however, application of the regional equations only to the intervening area, then adding the resulting estimates to calculated low-flow indices at an upstream station (or subtracting from those indices at a downstream station), would allow any unique natural condition or regulation that might affect low flow from the remainder of the watershed to be preserved in the final estimate.

The goal of this study was to evaluate the extent to which a variety of watershed properties now (2016) available from SSURGO and other GIS databases control 7Q2 and 7Q10 at streamflow measurement sites in the Lower Hudson River Basin. This evaluation led to regional regression equations that estimate these low-flow indices with reasonable accuracy. The dataset for this study excluded watersheds larger than 100 square miles and most watersheds with urban land cover

of more than 10 percent, so the equations may not apply accurately to such watersheds.

The equations in this report supersede those by Barnes (1986). Replication of the equations presented by Barnes, using the data values in his table 1, resulted in slightly larger standard errors than those of equations 7 and 9 in table 3 of the present report, and resulted in some watershed properties being deemed not significant because  $pr > |t|$  values were greater than 0.10. Furthermore, Barnes stated (1986, p. 8) that area of stratified drift (a term he used to mean area of surficial sand and gravel deposited by glacial meltwater or postglacial streams) was compiled “from maps in a variety of publications ... County soils maps were the primary source of information.” For 25 of the 51 watersheds in the study area, the areas of sand and gravel reported by Barnes differed from those compiled from SSURGO for the present study by more than 3 percent of watershed area, and for 6 watersheds the areas reported by Barnes were 2 to 4 times the areas depicted on soils maps available prior to 1986 as well as on the recent soils maps incorporated in SSURGO. The ready availability of digital soils maps from SSURGO, such that a consistent interpretation of the area of surficial sand and gravel can be replicated by anyone using the following instructions, renders the equations in the present report much easier to apply reliably.

Equations 2, 5, 7, and 9 in table 3 are the most accurate of many ordinary least squares (OLS) equations tested. The equivalent weighted least squares (WLS) equations in table 7 are more sophisticated in that data from sites with longer records and more numerous measurements were assigned greater weights in the regression analysis; this weighting system is approximate and somewhat arbitrary, but the statistics for these WLS equations in table 7 are believed to be more realistic than the equivalent OLS equations in table 3 as indicators of the accuracy and statistical soundness of low flows estimated at unmeasured sites. WLS equations 2 and 5 should be applied only to drainage areas of about 26 square miles or more, because application to smaller areas would result in standard errors larger than those obtainable with WLS equations 7 and 9 after conversion to cubic feet per second.

The watershed properties required for application of WLS regression equations 2, 5, 7, and 9 (table 7) can be compiled as follows.

- *Drainage area.*—Import a topographic base map of the watershed upstream from the point of interest (or the watershed segment whose 7Q2 or 7Q10 is to be added to or subtracted from values available for another site). Alternatively, USGS StreamStats for New York ([https://water.usgs.gov/osw/streamstats/new\\_york.html](https://water.usgs.gov/osw/streamstats/new_york.html)) can be used to calculate drainage area.
- *Surficial geology and potential water loss to riparian evapotranspiration.*—Import the soils series data for the delineated watershed derived from SSURGO (Natural Resources Conservation Service, 2007), then modify the data as follows. Most of the following

steps can readily be automated in a GIS, and the sequence of these steps is illustrated in a flowchart in appendix 1 (fig. 1–1).

1. Identify any polygons not classified as to soil series within the watershed of interest (such as urban areas, pits or quarries, cut and fill). Reclassify each as the soil series whose parent material probably exists here beneath the fill or excavation (as inferred from adjacent areas, topography, well records, or other data).
  2. Select an appropriate equation from table 7.
  3. Refer to appendix 1 (table 1–4) and note the soil series that make up the first term in that equation (which represents sources of groundwater discharge to streams from surficial geologic deposits) and the third term in the equation (which represents sources of riparian evapotranspiration). Sum the areas of all soil polygons within the watershed that are underlain by soils listed in appendix 1 (table 1–4) for each term that appears in the selected equation. Be sure to exclude from the third term (evapotranspiration) any swamp polygons that are not bordered at least in part by gravel, sand, or alluvium [for term *SWsq*] or by gravel, sand, alluvium, or fine sand to clay [for term *SWsqfy*].
- *Amount of water available for runoff*.—Calculate 1951–80 mean annual runoff for the watershed of interest from the mean annual runoff contours presented by Randall (1996). This can be done visually by superimposing the watershed outline on the contour map, or can be automated in a GIS. Values generated from either source will be in units of inches; multiply by 0.0737 to convert to cubic feet per second per square mile, which are the units required for the independent variable *MAQ/DRNAREA* in all equations in table 7.
  - *Watershed geometry*.—Import the watershed properties *CSL1085UP* and *LENGTH* from USGS StreamStats for New York ([https://water.usgs.gov/osw/streamstats/new\\_york.html](https://water.usgs.gov/osw/streamstats/new_york.html)).<sup>2</sup> Then, take the 0.6 power of *CSL1085UP* to generate the variable *CSL1085UP\_06* in WLS equations 2 and 5 (table 7).

<sup>2</sup>As noted in the section “Regression Analysis Using Ordinary Least Squares,” the watershed property *LENGTH* that appears in some regression equations was generated from a 30-meter digital elevation model. That model is no longer available in 2016. However, the U.S. Geological Survey StreamStats program cited in this paragraph will generate the same watershed property from a 10-meter digital elevation model. Results from the two models were compared and found to be similar for nearly all sites tested.

Note that in WLS equations 7 and 9, all watershed properties (not just *MAQ*) are divided by *DRNAREA*, and the equation estimates 7Q2 and 7Q10 in units of cubic feet per second per square mile.

## Summary

Two statistical indices of low streamflow, the 7-day, 2-year (7Q2) and 7-day, 10-year (7Q10) low flows, had been calculated in an earlier study from streamflow records at 53 sites in the Lower Hudson River Basin for the period 1958–78. For 3 of these watersheds the indices were revised, deleted, or combined, and the drainage areas of 5 watersheds were reduced slightly to reflect diversions that affect low flows (including natural diversions through permeable surficial gravel or cavernous limestone, and manmade diversions for public water supply). Watershed properties now (2016) readily available from geographic information systems were compiled and used to generate regression equations that estimate the low-flow indices.

Like studies of several other regions in the glaciated northeastern United States, this study demonstrates that 7Q10 and 7Q2 are largely a function of area of surficial sand and gravel, water input to the watershed (represented by mean annual runoff), and area of swamps and poorly drained soils adjacent to surficial sand and gravel (where groundwater recharge is unlikely and riparian water loss to evapotranspiration is substantial). Small but statistically significant improvements in estimation accuracy resulted from inclusion of an index of watershed geometry—either channel slope or channel length—in the regression equations. The most accurate and realistic estimation equations generated in this study employ weighted least squares and explain nearly 84 and 87 percent of the variation in 7Q10 and 7Q2, respectively, with standard errors of 0.032 and 0.050 cubic feet per second per square mile (48.8 and 36.9 percent of the mean of the dependent variable.) These equations are not as precise as those achieved in some similar studies elsewhere in the glaciated northeastern United States; the shortfall may reflect (1) a higher percentage of partial-record stations relative to long-term gaging stations in the Lower Hudson dataset than in the datasets for the similar studies, (2) the unavoidable reliance on soils maps rather than detailed hydrogeologic or surficial geologic maps to delineate area and type of surficial sand and gravel in the Lower Hudson River Basin, and (or) (3) part of the low flow bypassing some measurement sites as underflow through surficial sand and gravel. Nevertheless, these equations should be useful in estimating low flow at sites where there is no prospect of operating a streamgaging station or obtaining multiple low-flow measurements that could be correlated with a suitable nearby streamgaging station.



Three generalizations from this study may be useful in designing regression equations for low-flow estimation in the glaciated northeastern United States.

- In equations that estimated 7Q2, area of alluvium and area of sand and gravel deposited by glacial meltwater were equally and highly significant. In equations that estimated 7Q10, alluvium was generally not significant and was assigned small positive regression coefficients, whereas area of sand and gravel deposited by meltwater remained highly significant. This contrast is interpreted to indicate that groundwater storage in alluvium is essentially depleted before streamflow declines to the magnitude of 7Q10.
- Natural-value equations that included drainage area as a watershed property (along with area of sand and gravel and area of riparian wetlands, each expressed as a fraction of drainage area) had statistics of accuracy and normality much inferior to equations that omitted drainage area but included area of sand and gravel and area of riparian wetlands as watershed properties.
- A protocol for classifying soil series into 12 geologic units, which was developed for this study, represents the effect of surficial geology on low flow somewhat more accurately than Hydrologic Soil Group A (which was the best of several tested combinations of hydrologic soil groups A, B, C, and D, into which soil series are classified in the Soil Survey Geographic Database.)

## References Cited

- Baker, V.R., 1973, Geomorphology and hydrology of karst drainage basins and cave channel networks in east-central New York: Water Resources Research, v.13, p. 303–312.
- Baker, V.R., 1976, Hydrogeology of a cavernous limestone terrane and the hydrochemical mechanisms of its formation, Mohawk River basin, New York: Empire State Geogram, v. 12, no. 2, p. 2–65.
- 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.
- Burns, D.A., Klaus, Julian, and McHale, M.R., 2007, Recent climate trends and implications for water resources in the Catskill Mountain region, New York, USA: Journal of Hydrology, v. 336, p. 155–170.
- Cervione, M.A., Jr., Melvin, R.L., and Cyr, K.A., 1982, A method for estimating the 7-day 10-year low flow of streams in Connecticut: Connecticut Water Resources Bulletin 34, 17 p.
- Chazen Companies, 2006, Wappinger Creek watershed groundwater recharge and stream baseflow assessment: Poughkeepsie, N.Y., The Chazen Companies, 38 p.
- Coates, D.R., 1971, Hydrogeomorphology of Susquehanna and Delaware basins, in Morisawa, Marie, ed., Quantitative geomorphology, some aspects and applications: Binghamton, N.Y., State University of New York, Publications in Geomorphology, p. 273–306.
- Davis, Dan, Knuepfer, Peter, Vian, Mark, and Miller, Nicholas, 2009, Fluvial geomorphology of upper Esopus Creek and implications for water quality, in Vollmer, F.W., ed, Field trip guidebook, New York Geological Association 81<sup>st</sup> annual meeting, New Paltz, New York, p. 8.1–8.20.
- Denny, C.S., and Lyford, W.H., 1963, Surficial geology and soils of the Elmira-Williamsport region, New York-Pennsylvania: U.S. Geological Survey Professional Paper 379, 60 p.
- Eissler, B.B., 1979, Low-flow data and frequency analysis of streams in New York, excluding New York City and Long Island: New York Department of Environmental Conservation Bulletin 74, 176 p.
- Fisher, D.W., Isachsen, Y.W., and Rickard, L.V., 1970, Geologic map of New York: New York State Museum and Science Service, Map and Chart Series 15, scale 1:250,000, 5 sheets.
- Flynn, Kate (coordinator), 2007, Statistical approach to surface-water hydrologic analysis: U.S. Geological Survey lecture notes SW–2011–TC, unpaginated.
- Hardison, C.H., and Moss, M.E., 1972, Accuracy of low-flow characteristics estimated by correlation of base-flow measurements: U.S. Geological Survey Water-Supply Paper 1542B, 55 p. [Also available at <http://www.nysm.nysed.gov/gis/>.]
- Hayhoe, Katharine, and 11 others, 2007, Past and future changes in climate and hydrological indicators in the U.S. Northeast: Climate Dynamics, v. 28, p. 381–407.
- Heisig, P.M., 2010, Water resources of Rockland County, New York 2005–07, with emphasis on the Newark Basin bedrock aquifer: U.S. Geological Survey Scientific Investigations Report 2010–5245, 130 p. [Available at <http://pubs.usgs.gov/sir/2010/5245/>.]
- Hibbert, A.R., 1969, Water yield changes after converting a forested catchment to grass: Water Resources Research, v. 5 no. 3, p. 634–640.
- Hodgkins, G.A., Dudley, R.W., and Huntington, T.G., 2005, Summer low flows in Northeast over the 20th century: American Water Resources Association Journal, v. 41, p. 403–412.

- Homer, Collin, Dewitz, Jon, Fry, Joyce, Coan, Michael, Hosain, Nazmul, Larson, Charles, Herold, Nate, McKerrow, Alexa, vanDriel, J.N., and Wickham, James, 2007, Completion of the 2001 National Land Cover database for the conterminous United States: Photogrammetric Engineering and Remote Sensing, v. 73, no. 4, p. 337–341. [Also available at <http://www.mrlc.gov/nlcd2001.php>].
- Iman, R.I., and Conover, W.J., 1983, A modern approach to statistics: New York City, John Wiley and Sons, 497 p.
- Knox, C.E., and Nordenson, T.J., 1955, Average annual runoff and precipitation in the New England-New York area: U.S. Geological Survey Hydrologic Investigations Atlas HA-7, scale 1:1,000,000, 3 sheets.
- Kontis, A.L., Randall, A.D., and Mazzaferro, D.L., 2004, Regional hydrology and simulation of flow of stratified-drift aquifers in the glaciated northeastern United States: U.S. Geological Survey Professional Paper 1415-C, 156 p.
- Ku, H.F.H., Randall, A.D., and MacNish, R.D., 1975, Streamflow in the New York part of the Susquehanna River basin: New York Department of Environmental Conservation Bulletin 71, 130 p.
- LaFleur, R.G., 1965, Glacial geology of the Troy, New York quadrangle: New York State Museum and Science Service, Map and Chart Series 7, 22 p., 1 pl., scale 1:32,680.
- Law, Frank, 1957, The effect of afforestation on yield of water catchment areas: Institute Water Engineers Journal, v. 11, no. 3, p. 269–276.
- Lewis, D.C., 1968, Annual hydrologic response to watershed conversion from oak woodland to annual grassland: Water Resources Research, v. 4, no. 1, p. 59–72.
- Looney, S.W., and Gullledge, T.R., 1985, Use of correlation coefficient with normal probability plots: The American Statistician, v. 39, p. 75–79.
- Male, J.W., and Ogawa, H., 1982, Low flow of Massachusetts streams: Amherst, Massachusetts, University of Massachusetts Water Resources Center Publication 125, 152 p.
- Marquardt, D.W., 1970, Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation: Technometrics, v. 12, p. 591–612.
- Matalas, N.C., and Jacobs, B., 1964, A correlation procedure for augmenting hydrologic data: U.S. Geological Survey Professional Paper 434-E, 7 p.
- Morrissey, D.J., Randall, A.D., and Williams, J.H., 1988, Upland runoff as a major source of recharge to stratified drift in the glaciated Northeast, *in* Randall, A.D., and Johnson, A.I., eds., Regional aquifer systems of the United States, northeast glacial aquifers: American Water Resources Association Monograph Series no. 11, p. 17–36.
- Nakagaki, Naomi, Price, C.V., Falcone, J.A., Hitt, K.J., and Ruddy, B.C., 2010, Enhanced National Land Cover Data 1992 (NLCDe 92) (ver. 1.1): U.S. Geological Survey data release, accessed June 6, 2008, at <http://water.usgs.gov/lookup/getspatial?nlcde92>.
- Natural Resources Conservation Service, 2007, Soil Survey Geographic Database: Natural Resources Conservation Service database, accessed June 2007 at <http://sdmdataaccess.nrcs.usda.gov/>
- New York City Department of Environmental Protection, 2008, Evaluation of turbidity reduction potential through watershed management in the Ashokan basin: Kingston, N.Y., New York City Department of Environmental Protection, Bureau of Water Supply.
- Palmer, A.N., Palmer, M.V., Porter, C.O., Rubin, P.A., and Mylroie, J.E., 1991a, A geologic guide to the karst and caves of the Helderberg Mountains, Schoharie and Albany Counties, New York, *in* Nardacci, Michael, ed., Guide to the caves and karst of the Northeast—50th anniversary NSS convention, Cobleskill, New York, 1991: Huntsville, Ala., National Speleological Society, p. 105–167.
- Palmer, A.N., Rubin, P.A., and Palmer, M.V., 1991b, Interaction between karst and glaciation in the Helderberg Plateau, Schoharie and Albany Counties, New York, *in* Ebert, J.R., ed., Field trip guidebook: Oneonta, N.Y., New York State Geological Association, 63rd Annual Meeting, p. 161–190.
- Paulsen, C.G., Bigwood, B.L., Harrington, A.W., Hartwell, O.W., and Kinnison, H.B., 1940, Hurricane floods of September 1938: U.S. Geological Survey Water-Supply Paper 867, 562 p.
- Price, C.V., Nakagaki, Naomi, Hitt, K.J., and Clawges, R.M., 2006, Enhanced historical land-use and land-cover data sets of the U.S. Geological Survey: U.S. Geological Survey Digital Data Series 240, accessed July 14, 2008 at <http://pubs.usgs.gov/ds/2006/240>.
- Randall, A.D., 1977, The Clinton Street-Ballpark aquifer in Binghamton and Johnson City, New York: New York Department of Environmental Conservation Bulletin 73, 87 p.
- 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 pl. [The two plates were reprinted in Kontis and others (2004); digital versions are available through <https://pubs.er.usgs.gov/publication/ofr96395>.]
- Randall, A.D., 2001, Hydrogeologic framework of stratified-drift aquifers in the glaciated northeastern United States: U.S. Geological Survey Professional Paper 1415B, 179 p.
- 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.

- Randall, A.D., and Johnson, A.I. 1988, The northeast glacial aquifers RASA project—An overview of results through 1987, *in* Randall, A.D., and Johnson, A.I. eds., Regional aquifer systems of the United States, northeast glacial aquifers: American Water Resources Association Monograph Series no. 11, p. 1–15.
- Reilly, C.F., and Kroll, C.N., 2002, Estimation of low flow frequency using cross correlation of baseflow measurements: Proceedings of the American Society of Civil Engineers Environmental and Water Resources Institute conference, Roanoke, Va., 2002.
- Rickard, L.V., 1975, Correlation of the Silurian and Devonian rocks in New York: New York State Museum and Science Service, Map and Chart Series 24, 4 sheets, 16 p.
- Riggs, H.C., 1972, Low-flow investigations: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. B1, 18 p.
- Riggs, H.C., 1985, Streamflow characteristics: Amsterdam, Elsevier Science Publishers, 249 p.
- Ryder, R.B., Thomas, M.P., and Weiss, L.A., 1981, Water resources inventory of Connecticut, part 7, upper Connecticut River basin: Connecticut Water Resources Bulletin 24, 76 p.
- Saft, Margarita, Western, A.W., Zhang, Lu, Peel, M.C., and Potter, N.J., 2015, The influence of multiyear drought on the annual rainfall-runoff relationship—An Australian perspective: Water Resources Research, v. 51, no. 4, p. 2444–2463.
- SAS Institute, 1990, SAS/STAT user's guide, version 6 (4th ed.): Cary, N.C., SAS Institute Inc., 1,675 p.
- SAS Institute, 2008, SAS/STAT version 9.2: Cary, N.C., SAS Institute, Inc.
- Schneider, W.J., and Ayer, G.R., 1961, Effect of reforestation on streamflow in central New York: U.S. Geological Survey Water-Supply Paper 1602, 61 p.
- Schumm, S.A., 1956, Evolution of drainage systems and slopes in badlands at Perth Amboy, New Jersey: Geological Society of America Bulletin, v. 67, p. 597–646.
- Searcy, J.K., 1959, Flow-duration curves: U.S. Geological Survey Water-Supply Paper 1542–A, 33 p.
- Smith, C.A., Kozlowski, A.L., and Stefanik, P.A., 2001, Progress on three-dimensional surficial geologic mapping in the Catskill mountains, New York, Shandaken Quadrangle: New York Geological Survey Open-File Report 2515, 17 p.
- 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., and Thomas, W.O., Jr., 1985, Low-flow frequency estimation using base-flow measurements: U.S. Geological Survey Open-file Report 85–95, 22 p.
- Stone, E.L., and Kalisz, P.J., 1991, On the maximum extent of tree roots: Forest Ecology and Management, v. 46, p. 59–102.
- Suro, T.P., and Gazoorian, Christopher, 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.
- Tasker, G.D., 1980, Hydrologic regression with weighted least squares: Water Resources Research, v. 16, no. 6, p. 1107–1113.
- Tolman, C.F., 1937, Ground water: New York and London, McGraw-Hill Book Company, 593 p.
- U.S. Geological Survey, 1978, Water resources data for New York, water year 1977, volume 1, New York, excluding Long Island: U.S. Geological Survey Water Data Report NY–77–1, 566 p.
- U.S. Geological Survey, 2004, The National Map, 3D Elevation Program: U.S. Geological Survey Web page, accessed 2004 at [http://nationalmap.gov/3dep\\_prodserv.htm](http://nationalmap.gov/3dep_prodserv.htm).
- U.S. Geological Survey, 2006a, USGS 01358500 Poesten Kill near Troy NY: U.S. Geological Survey National Water Information System, accessed 2006 at [http://waterdata.usgs.gov/nwis/inventory?agency\\_code=USGS&site\\_no=01358500](http://waterdata.usgs.gov/nwis/inventory?agency_code=USGS&site_no=01358500).
- U.S. Geological Survey, 2006b, USGS 01359902 Coeymans Creek near Selkirk NY: U.S. Geological Survey National Water Information System, accessed 2006 at [http://waterdata.usgs.gov/nwis/inventory/?site\\_no=01359902](http://waterdata.usgs.gov/nwis/inventory/?site_no=01359902).
- Wandle, S.W., Jr., and Randall, A.D., 1994, Effects of surficial geology, lakes and swamps, and annual water availability on low flows of streams in central New England, and their use in low flow estimation: U.S. Geological Survey Water-Resources Investigations Report 93–4092, 57 p. [Revised 2007.]
- Zecharias, Y.B., and Brutsaert, Wilfried, 1988a, Influence of basin morphology on groundwater outflow: Water Resources Research, v. 24, no. 10, p. 1645–1650.
- Zecharias, Y.B., and Brutsaert, Wilfried, 1988b, Recession characteristics of groundwater outflow and base flow from mountainous watersheds: Water Resources Research, v. 24, no. 10, p. 1651–1658.
- Zhang, Zhenxing, and Kroll, C.N., 2007, A closer look at the baseflow correlation method: Journal of Hydrologic Engineering, v. 12, no. 2, p. 190–196.



## Appendix 1. Aggregation of Soil Series Into Surficial Geology Units and Sources of Abundant Evapotranspiration for Use in Regression Analysis

This appendix specifies (in table 1–1) the soil series that were aggregated into each of several surficial geology units that were tested in regression analysis. It also specifies (in table 1–2) which of these soil series were included in three categories of somewhat poorly to very poorly drained soils with a high potential for evapotranspiration. It then specifies (in table 1–3) how the geologic units and high-evapotranspiration categories were incorporated in independent watershed variables tested in regression analyses. Finally, as a convenience to persons who may wish to apply the regression equations presented in this report to estimate 7Q2 or 7Q10 at ungaged sites in the Lower Hudson River Basin, the soil series that comprise each of the independent variables that represent surficial geology and evapotranspiration in these equations are

specified in table 1–4. A flow chart (figure 1–1) illustrates the steps in combining geologic units and poorly drained soils to formulate these independent watershed variables.

### References Cited

- Natural Resources Conservation Service, 2007, Soil Survey Geographic Database: Natural Resources Conservation Service database, accessed June 2007 at <http://sdmdataaccess.nrcs.usda.gov/>.
- Randall, A.D., 1978, Infiltration from tributary streams in the Susquehanna River basin, New York: U.S. Geological Survey Journal of Research, v. 6, no. 3, p. 285–297.



**Table 1–1.** Reclassification of soils in the Lower Hudson River Basin, New York, into 12 geologic units.

[Soil series are from the Soil Survey Geographic Database (Natural Resources Conservation Service, 2007). —, not applicable]

Geologic unit	Definition	Soil series included
<i>TTH</i>	Till, greater than 6 feet to bedrock	Alden, Allbrights, Amenia, Appleton, Aurelie, Bath, Bernardston, Brayton, Broadalbin, Buckland, Burdett, Busti, Cambridge, Canandaigua, Cattaraugus, Cayuga, Cazenovia, Charlton, Chautauqua, Cheshire, Chippewa, Churchville, Conesus, Culvers, Darien, Dummerston, Dutchess, Elka, Eric, Farmington, Fullam, Georgia, Hibernia, Honeoye, Ilion, Lackawanna, Lanesboro, Langford, Lansing, Leicester, Lewbeach, Lima, Lyons, Manheim, Mardin, Massena, Menlo, Mohawk, Monarda, Morris, Mosherville, Nellis, Neversink, Norwich, Nunda, Onteora, Ovid, Paxton, Pittsfield, Pittstown, Punxit, Ridgebury, Rockaway, Scriba, Stockbridge, Sun, Suny, Sutton, Swartswood, Venango, Volusia, Wellsboro, Wethersfield, Willowemoc, Woodbridge, Wurtsboro, <i>and combinations thereof or with “urban”</i>
<i>TSR</i>	Till, less than 6 feet to bedrock	Allis, Angola, Arnot, Brockport, Cardigan, Chatfield, Farmington, Galway, Glover, Greene, Halcott, Hawksnest, Hollis, Holyoke, Hornell, Juliet, Kearsarge, Lordstown, Macomber, Manlius, Mongaup, Nassau, Oquaga, Palatine, Pits (quarry), Rock outcrop, Stockbridge, Taconic, Tor, Torull, Tuller, Varick, Vly, Wassaic, Watchug, Yalesville, <i>and combinations of these with one another or with TTH soils, or with “rock outcrop,” or with “urban”</i>
<i>V</i>	Till, greater than 6 feet to bedrock, C-horizon very gravelly fine sandy loam; occurs on lower valley sides, some landforms are till-mantled valley terraces of sand and gravel	Maplecrest, Valois
<i>SG</i>	Sand and gravel, generally deposited by glacial meltwater (outwash or ice-contact deposits) <sup>1</sup>	Allard, Alton, Atherton, Barbour & Tioga (fan), Blasdel, Castile, Chenango, Copake, Fredon, Halsey, Haven, Herkimer, Hinckley, Hoosic, Howard, Otisville, Palmyra, Phelps, Pits (gravel), Pompton, Preakness, Red Hook, Riverhead, Tunkhannock <i>and combinations thereof</i>
<i>SD</i>	Sand, generally deposited by glacial meltwater (outwash or delta)	Colonic, Elnora, Granby, Junius, Knickerbocker, Lamson, Oakville, Plainfield, Pompton, Scarboro, Stafford, Udisamments, Walpole, Windsor, <i>and combinations of these with one another or with SG soils</i>
<i>FL</i>	Fine sand to silt	Birdsall, Canandaigua, Collamer, Raynham, Scio, Sudbury, Unadilla [compiled separately but similar], Wallington
<i>LY</i>	Silt and clay of glaciolacustrine origin	Birdsall, Covington, Fonda, Hudson, Kingsbury, Lakemont, Livingston, Madalin, Niagara, Odessa, Rhinebeck, Schoharie, Udorthents clayey, Vergennes, Williamson
<i>SY</i>	Thin sand over silt and clay	Cheektowaga, Claverack, Cossad, Elmridge, Phelps over clay, Shaker, Tunkhannock over clay
<i>QAL</i>	Alluvium, mostly gravel or sand, variable, commonly overlain by silt and fine sand, deposited by postglacial streams	Alluvial land, Barbour, Basher, Fluvaquents/Udifluvents, Holly, Limerick, Linlithgo, Middlebury, Occum, Ochrepts, Papakating, Pawling, Rippowam, Sloan, Suncook, Tioga, Wakeland, Wappinger, Wayland [some counties]
<i>QALL</i>	Alluvium, predominantly silt or very fine sand to depth of more than 5 feet, deposited by postglacial streams	Bash, Basher, Hamlin, Philo, Pope, Teel, Wayland [some counties]

**Table 1–1.** Reclassification of soils in the Lower Hudson River Basin, New York, into 12 geologic units.—Continued

[Soil series are from the Soil Survey Geographic Database (Natural Resources Conservation Service, 2007). —, not applicable]

Geologic unit	Definition	Soil series included
<i>SW</i>	Organic sediment (muck or peat) [ <i>SW</i> was subdivided into three subunits: <i>SWsq</i> — <i>SW</i> bordered in part by <i>SG</i> , <i>SD</i> , <i>QAL</i> , or <i>QALL</i> . <i>SWfy</i> — <i>SW</i> not bordered by <i>SG</i> , <i>SD</i> , <i>QAL</i> , or <i>QALL</i> but bordered in part by <i>FL</i> , <i>LY</i> , or <i>SY</i> . <i>SWtl</i> — <i>SW</i> bordered only by <i>TTH</i> , <i>TSR</i> , or <i>V</i> .]	Ackerman, Adrian, Beseman, Carlisle, Catden, Freshwater marsh, Greenwood, Histic Humaquents, Hydraquents, Ipswich, Loxley, Medisaprists, Muck, Muskego, Olentangy, Ossipee, Palms, Peat, Pinnebog, Saprist & Aquents, Walkill, Wawayanda
<i>LK</i>	Water (lakes, ponds, wide rivers) [ <i>LK</i> was subdivided into two subunits: <i>LKsg</i> — <i>LK</i> bordered in part by <i>SG</i> , <i>SD</i> , <i>QAL</i> , or <i>QALL</i> . <i>LKtl</i> — <i>LK</i> bordered only by <i>TTH</i> , <i>TSR</i> , or <i>V</i> .]	—

<sup>1</sup>Where a tributary stream draining an upland area enters a lowland or the valley of a larger stream, particularly in a region of high relief, it may deposit permeable gravel in a triangular or cone-shaped wedge that is termed an “alluvial fan.” Water in the tributary channel commonly seeps into the fan (Randall, 1978), recharges aquifers in the valley fill, and ultimately helps to sustain the low flow of the larger stream. Several alluvial fans in the Lower Hudson River Basin were mapped by county soil surveys as follows:

Albany County: Chenango channery silt loam, fan

Rensselaer County: Chenango gravelly loam, fan; Haven silt loam

Columbia County: Blasdell channery silt loam, fan

Dutchess County: Hoosic channery loam, fan

Greene County: Tunkhannock gravelly loam, fan

Ulster County: Hoosic cobbly loam; Tunkhannock and Chenango gravelly loam

Deposition on most alluvial fans has continued into modern times, so fans could reasonably be termed alluvium. However, each of the foregoing soil series names (without “fan” appended) is widely applied to deposits of glacial meltwater: outwash plains, terraces, deltas, kames, and eskers. Because most soil surveys in the Lower Hudson River Basin considered fans to be similar to glacial outwash, and because the capability of fans to store groundwater for gradual release to streams seems more akin to glacial outwash than to flood-plain alluvium, all these soil series were reclassified as *SG* for this report. In the Schoharie County soil survey only, fans are mapped as “Barbour and Tioga gravelly loam (fan),” which was also reclassified as *SG* for this report. Other Barbour and Tioga soils, however, are widely distributed in Schoharie and other counties on flood plains and alluvial terraces, so these soils were reclassified as *QAL*.

**Table 1–2.** Subdivision of some geologic units in the Lower Hudson River Basin, New York, into categories that have high potential for evapotranspiration.

[Each category includes only those soils within the listed geologic units that are appraised by the National Resources Conservation Service as somewhat poorly to very poorly drained. Geologic units are defined in table 1–1. Soil series are from the Soil Survey Geographic Database (Natural Resources Conservation Service, 2007)]

Evapotranspiration Category	Geologic units	Soil series included
<i>Sp</i>	<i>SG</i> and <i>SD</i>	Atherton, Fredon, Granby, Junius, Lamson, Preakness, Red hook, Scarboro, Stafford, Walpole and combinations thereof
<i>FYp</i>	<i>FL</i> , <i>LY</i> , and <i>SY</i>	Birdsall, Canandaigua, Cheektowaga, Covington, Cossad, Fonda, Kingsbury, Lakemont, Livingston Madalin, Niagara, Odessa, Raynham, Rhinebeck, Shaker, Wallington and combinations thereof
<i>Qp</i>	<i>QAL</i> and <i>QALL</i>	Bash, Fluvaquents, Holly-Papakating, Linlithgo, Limerick, Middlebury (some counties), Rippowam, Sloan, Wakeland, Wayland

**Table 1–3.** Independent variables in the Lower Hudson River Basin, New York, used in regression equations in this report, and the geologic units and evapotranspiration categories from which the variables were derived.

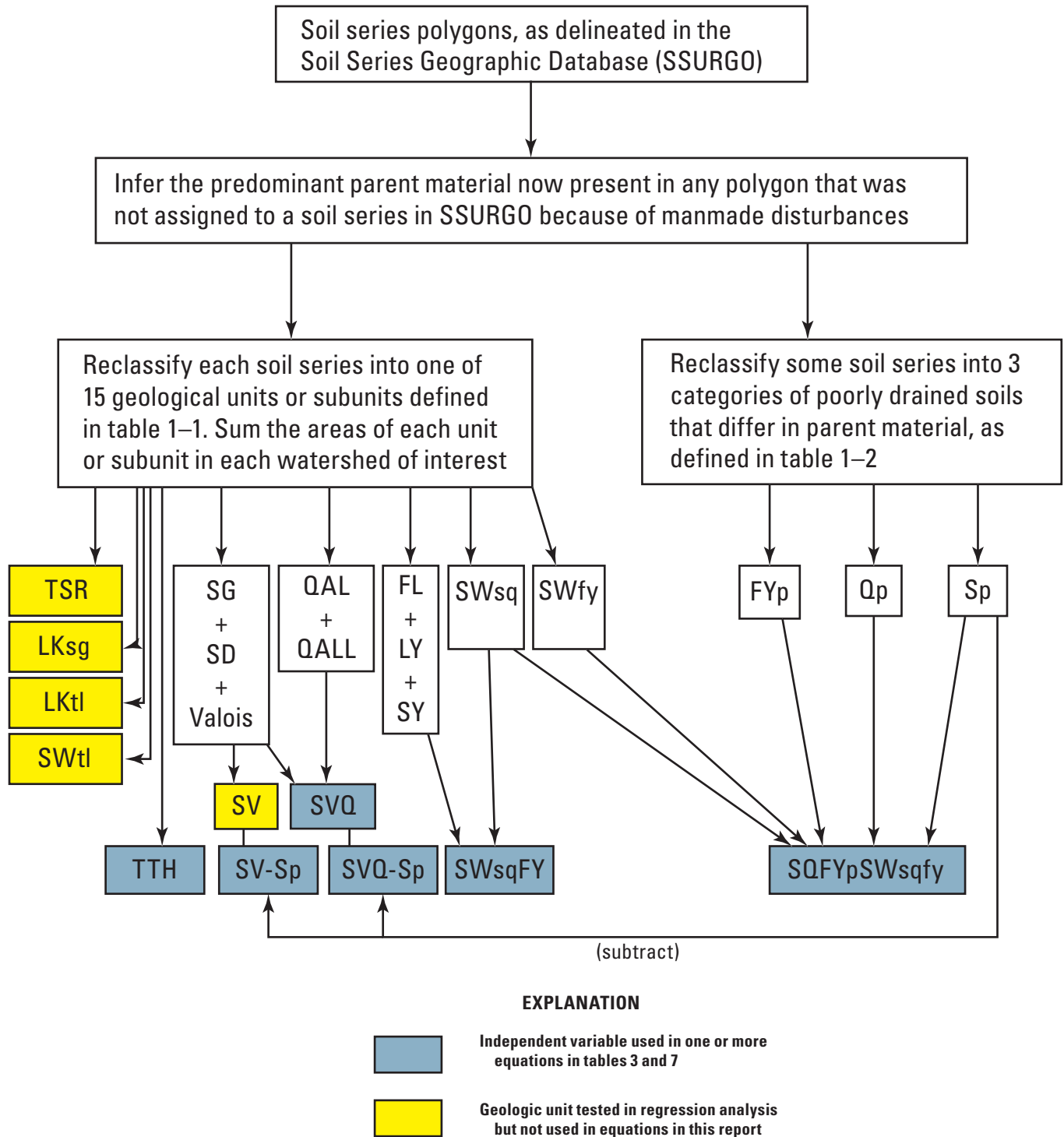
[Regression equations are in tables 3 and 7. Acronyms for variables are also defined in the text section “Regression Analysis Using Ordinary Least Squares,” p. 12. Geologic units are defined and equated to soil series in Soil Survey Geographic Database (SSURGO) (Natural Resources Conservation Service, 2007) in table 1–1. Categories with high potential for evapotranspiration are defined and equated to SSURGO soil series in table 1–2 (*Sp*, *FYp*, *Qp*) or in table 1–1 (*SWsq*, *SWfy*).]

Independent variable	Calculation of variable from geologic units and evapotranspiration categories
<i>SV</i>	<i>SG</i> plus <i>SD</i> plus <i>V</i>
<i>SVQ</i>	<i>SG</i> plus <i>SD</i> plus <i>V</i> plus <i>QAL</i> plus <i>QALL</i>
<i>SV_Sp</i>	<i>SV</i> minus <i>Sp</i>
<i>SVQ_Sp</i>	<i>SVQ</i> minus <i>Sp</i>
<i>SQFYpSWsqfy</i>	<i>Sp</i> plus <i>Qp</i> plus <i>FYp</i> plus <i>SWsq</i> plus <i>SWfy</i>
<i>SWsqFY</i>	<i>SWsq</i> plus <i>FL</i> plus <i>LY</i>

**Table 1–4.** Independent variables that represent surficial geology and high riparian evapotranspiration in the regression equations in this report, for use in estimating 7-day, 2-year and 7-day, 10-year low flows at ungaged sites in the Lower Hudson River Basin, New York.

[Regression equations are in tables 3 and 7. Soil series are from the Soil Survey Geographic Database (SSURGO) (Natural Resources Conservation Service, 2007). Independent variables are defined in the text section “Regression Analysis Using Ordinary Least Squares,” p. 12, and are equated in table 1–3 to the geologic units in table 1–1 and the categories of high potential evapotranspiration in table 1–2 or in table 1–1.]

Independent variable	Soil series included
<i>SV_Sp</i>	Allard, Alton, Barbour & Tioga (fan), Blasdell, Castile, Chenango, Colonie, Copake, Elnora, Halsey, Haven, Herkimer, Hinckley, Hoosic, Howard, Knickerbocker, Maplecrest, Oakville, Otisville, Palmyra, Phelps, Pits (gravel), Plainfield, Pompton, Riverhead, Tunkhannock, Udisamments, Valois, Windsor, <i>and combinations of these with one another but not with other soil series</i>
<i>SVQ</i>	Allard, Alluvial land, Alton, Atherton, Barbour, Barbour & Tioga (fan), Bash, Basher, Blasdell, Castile, Chenango, Colonie, Copake, Elnora, Fluvaquents/Udifluvents, Fredon, Granby, Halsey, Hamlin, Haven, Herkimer, Hinckley, Holly, Hoosic, Howard, Junius, Knickerbocker, Lamson, Limerick, Linlithgo, Maplecrest, Middlebury, Oakville, Occum, Ochrepts, Otisville, Palmyra, Papakating, Pawling, Phelps, Philo, Pits (gravel), Plainfield, Pompton, Pope, Preakness, Red Hook, Rippowam, Riverhead, Scarboro, Sloan, Stafford, Suncook, Teel, Tioga, Tunkhannock, Udisamments, Valois, Wakeland, Walpole, Wappinger, Wayland, Windsor, <i>and combinations of these with one another</i>
<i>SVQ_Sp</i>	Allard, Alluvial land, Alton, Barbour, Barbour & Tioga (fan), Bash, Basher, Blasdell, Castile, Chenango, Colonie, Copake, Elnora, Fluvaquents/Udifluvents, Halsey, Hamlin, Haven, Herkimer, Hinckley, Holly, Hoosic, Howard, Knickerbocker, Limerick, Linlithgo, Maplecrest, Middlebury, Oakville, Occum, Ochrepts, Otisville, Palmyra, Papakating, Pawling, Phelps, Philo, Pits (gravel), Plainfield, Pompton, Pope, Rippowam, Riverhead, Sloan, Suncook, Teel, Tioga, Tunkhannock, Udisamments, Valois, Wakeland, Wappinger, Wayland, Windsor, <i>and combinations of these with one another</i>
<i>SQFYpSWsqfy</i>	<i>This variable includes the following poorly drained soil series:</i> Atherton, Bash, Birdsall, Canandaigua, Cheektowaga, Covington, Cossad, Fonda, Fluvaquents, Fredon, Granby, Holly-Papakating, Junius, Kingsbury, Lakemont, Lamson, Linlithgo, Limerick, Livingston, Madalin, Middlebury, Niagara, Odessa, Preakness, Raynham, Red Hook, Rhinebeck, Rippowam, Scarboro, Shaker, Sloan, Stafford, Wakeland, Wallington, Walpole, Wayland, <i>and combinations of these with one another. This variable also includes swampy areas underlain by any of the following soil series, but only if these areas are bordered at least in part by one or more soil series listed under SG, SD, QAL, QALL, FL, LY, or S/Y, in table 1–1:</i> Ackerman, Adrian, Beseman, Carlisle, Catden, Freshwater marsh, Greenwood, Histic Humaquents, Hydraquents, Ipswich, Loxley, Medisapristis, Muck, Muskego, Olentangy, Ossipee, Palms, Peat, Pinnebog, Sapristis & Aquents, Wallkill, Wawayanda
<i>SWsqFY</i>	<i>This variable includes the following poorly drained soil series:</i> Birdsall, Canandaigua, Cheektowaga, Claverack, Collamer, Cossad, Covington, Elmridge, Fonda, Hudson, Kingsbury, Lakemont, Livingston, Madalin, Niagara, Odessa, Phelps over clay, Raynham, Rhinebeck, Schoharie, Scio, Shaker, Sudbury, Tunkhannock over clay, Udorthents clayey, Unadilla, Vergennes, Wallington, Williamson. <i>This variable also includes swampy areas underlain by any of the following soil series, but only if these areas are bordered at least in part by one or more soil series listed under SG, SD, QAL, or QALL in table 1–1:</i> Ackerman, Adrian, Beseman, Carlisle, Catden, Freshwater marsh, Greenwood, Histic Humaquents, Hydraquents, Ipswich, Loxley, Medisapristis, Muck, Muskego, Olentangy, Ossipee, Palms, Peat, Pinnebog, Sapristis & Aquents, Wallkill, Wawayanda



**Figure 1-1.** Flowchart of steps in calculating the watershed variables that represent surficial geology and riparian evapotranspiration in regression equations for estimating low flow of streams in the Lower Hudson River Basin, New York.



For more information concerning this report, contact:  
Director, New York Water Science Center  
U.S. Geological Survey  
425 Jordan Road  
Troy, NY 12180  
[dc\\_ny@usgs.gov](mailto:dc_ny@usgs.gov)  
or visit our Web site at:  
<https://ny.water.usgs.gov>

Publishing support provided by the  
Pembroke Publishing Service Center

