

# Low Flow of Streams in the Susquehanna River Basin of New York



Scientific Investigations Report 2010–5063

**Cover.** Cuthrie Run, 3 miles north of Big Flats, Chemung County, New York, within an upland valley that drains 5.3 square miles. Photograph by Matthew B. Jennings, U.S. Geological Survey, June 29, 2007, when streamflow in this region was moderately low (exceeded about 80 percent of the time).

# **Low Flow of Streams in the Susquehanna River Basin of New York**

By Allan D. Randall

Scientific Investigations Report 2010–5063

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

**U.S. Department of the Interior**  
KEN SALAZAR, Secretary

**U.S. Geological Survey**  
Marcia K. McNutt, Director

U.S. Geological Survey, Reston, Virginia: 2010

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

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

To order this and other USGS information products, visit <http://store.usgs.gov>

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

Although this report is in the public domain, permission must be secured from the individual copyright owners to reproduce any copyrighted materials contained within this report.

Suggested citation:

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., at <http://pubs.usgs.gov/sir/2010/5063>.

# Contents

Abstract.....	1
Introduction.....	1
Watershed Properties and Processes that Control Spatial Variation in Low Flow of Streams.....	3
Amount of Water Available for Runoff .....	3
Annual Precipitation and Runoff .....	3
Seasonal Precipitation and Runoff .....	5
Extent and Properties of Surficial Sand and Gravel .....	8
Modeling Properties of Surficial Sand and Gravel that Affect Groundwater	
Discharge to Streams .....	10
Analytical Model .....	10
Numerical Model .....	11
Model Configuration and Operation .....	14
Model Evaluation .....	15
Evapotranspiration from Wetlands .....	18
Underflow beneath Sites of Streamflow Measurement .....	19
Less Important Factors that Affect Low Flows .....	22
Watershed Area.....	22
Extent and Properties of Till .....	22
Extent and Properties of Bedrock Units.....	22
Watershed Relief and Slope .....	23
Degree of Stream Regulation and Watershed Urbanization .....	24
Regression Analysis.....	24
Data Set .....	24
Low-Flow Statistics .....	24
Watershed Properties.....	28
Format of Regression Equations .....	32
Standard-Value Equations to Estimate 7-Day, 10-Year Low Flows .....	33
Effect of Sand-and-Gravel Area and Wetland Area .....	33
Effect of Amount of Water Available .....	33
Other Variables and Data Manipulations Tested.....	35
Per-Square-Mile Equations to Estimate 7-Day, 10-Year Low Flows.....	38
Comparison of Per-Square-Mile Equations with Standard-Value Equations .....	38
Evaluation of the Data Set and Equations .....	38
Effect of Underflow .....	42
Logarithmic Equations to Estimate 7-Day, 10-Year Low Flows .....	42
A Critique of the Rationale for Logarithmic Transformation .....	42
Comparison of Logarithmic Equations with Standard-Value Equations.....	43
Equations to Estimate 7-Day, 2-Year Low Flow and 90-Percent Flow Duration .....	44
Suggested Methods for Estimating Low-Flow Characteristics at Ungaged Sites.....	46
Watersheds Larger Than 30 Square Miles.....	47
Watersheds Smaller Than 30 Square Miles.....	47
Summary.....	48
Acknowledgments .....	49
References Cited .....	49

## Figures

1. Map showing location of the Susquehanna River basin and physiographic regions in New York and adjacent States.....	2
2. Diagram showing principal environmental factors that control the low flow of streams in the glaciated northeastern United States.....	4
3. Maps showing spatial variation in A, mean annual precipitation and B, mean annual runoff in the Susquehanna River basin of New York, 1951–80.....	6
4. Map showing ratios of mean August–October runoff to mean annual runoff for the period April 1937 to March 1963 in the Susquehanna River basin and adjacent areas ...	9
5. Diagram showing idealized aquifer represented by analytical equations of Rorabaugh (1963).....	11
6. Graphs showing effect of A, valley width and B, aquifer transmissivity on groundwater discharge to a stream during a period of low flow, as calculated from equations of Rorabaugh (1963).....	12
7. Diagram showing dimensions and properties of a two-dimensional groundwater-flow model that represents an idealized vertical section across an upland hillside and the adjacent valley.....	13
8. Graph showing simulated groundwater discharge, to a stream and as evapotranspiration, from an idealized upland hillside and valley fill after 4 months without recharge, as a function of hydraulic conductivity of the valley fill.....	16
9. Graph showing simulated groundwater discharge to a stream from an idealized valley fill after 4 months without recharge, as a function of valley half-width .....	18
10. Map showing locations of streamflow-measurement stations in the Susquehanna River basin of New York.....	20
11. Graph showing correlation of 7Q10 calculated from streamflow measurements with 7Q10 estimated using equation 3.....	34
12. Graph showing distribution of residuals from equation 5.....	36
13. Graphs showing correlation of low-flow indices calculated from streamflow measurements A, 7Q10 and, B, 90% FD, with selected watershed properties .....	37
14. Graph showing correlation of 7Q10 per square mile calculated from streamflow measurements with 7Q10 per square mile estimated using equation 9.....	39
15. Graph showing correlation of 7Q10 per square mile calculated from streamflow measurements with 7Q10 per square mile estimated using equation 11.....	40
16. Graph showing distribution of residuals from equation 11.....	41
17. Graph showing correlation of 7Q10 calculated from streamflow measurements with 7Q10 converted to cubic feet per second from values estimated using equation 11.....	41
18. Graph showing distribution of data values for independent variables in equation 12: A, sand and gravel exclusive of wetlands, and B, mean annual runoff .....	45
19. Graph showing correlation of 7Q10 calculated from streamflow measurements with 7Q10 converted to units of cubic feet per second from logarithmic values estimated using equation 12.....	46

## Tables

1. Ratios of August–October runoff to annual runoff in the Susquehanna River basin and adjacent areas.....	8
2. Model parameters that varied with time during simulations of groundwater discharge to the stream in a hypothetical valley typical of the Susquehanna River basin of New York.....	14
3. Streamflow-measurement stations in the Susquehanna River basin of New York, and selected indices of low flow. ....	25
4. Watershed properties compiled for regression analysis of low flow in the Susquehanna River basin of New York.....	29
5. Regression equations tested for estimating low flow of streams in the Susquehanna River basin of New York.....	53

## Conversion Factors and Datums

Multiply	By	To obtain
Length		
inch (in.)	2.54	centimeter (cm)
inch (in.)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
Area		
square foot (ft <sup>2</sup> )	929.0	square centimeter (cm <sup>2</sup> )
square foot (ft)	0.09290	square meter (m <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		
cubic foot (ft <sup>3</sup> )	28.32	cubic decimeter (dm <sup>3</sup> )
cubic foot (ft <sup>3</sup> )	0.02832	cubic meter (m <sup>3</sup> )
Flow rate		
foot per second (ft/s)	0.3048	meter per second (m/s)
foot per day (ft/d)	0.3048	meter per day (m/d)
cubic foot per second (ft <sup>3</sup> /s)	0.02832	cubic meter per second (m <sup>3</sup> /s)
cubic foot per second per square mile [(ft <sup>3</sup> /s)/mi <sup>2</sup> ]	0.01093	cubic meter per second per square kilometer [(m <sup>3</sup> /s)/km <sup>2</sup> ]
Hydraulic conductivity		
foot per day (ft/d)	0.3048	meter per day (m/d)
Transmissivity*		
foot squared per day (ft <sup>2</sup> /d)	0.09290	meter squared per day (m <sup>2</sup> /d)

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.

\*Transmissivity: The standard unit for transmissivity is cubic foot per day per square foot times foot of aquifer thickness [(ft<sup>3</sup>/d)/ft<sup>2</sup>]ft. In this report, the mathematically reduced form, foot squared per day (ft<sup>2</sup>/d), is used for convenience.



# Low Flow of Streams in the Susquehanna River Basin of New York

by Allan D. Randall

## Abstract

The principal source of streamflow during periods of low flow in the Susquehanna River basin of New York is the discharge of groundwater from sand-and-gravel deposits. Spatial variation in low flow is mostly a function of differences in three watershed properties: the amount of water that is introduced to the watershed and available for runoff, the extent of surficial sand and gravel relative to till-mantled bedrock, and the extent of wetlands. These three properties were consistently significant in regression equations that were developed to estimate several indices of low flow expressed in cubic feet per second or in cubic feet per second per square mile. The equations explain 90 to 99 percent of the spatial variation in low flow. A few equations indicate that underflow that bypasses streamflow-measurement sites through permeable sand and gravel can significantly decrease low flows. Analytical and numerical groundwater-flow models indicate that spatial extent, hydraulic conductivity and thickness, storage capacity, and topography of stratified sand-and-gravel deposits affect low-flow yields from those deposits. Model-simulated discharge of groundwater to streams at low flow reaches a maximum where hydraulic-conductivity values are about 15 feet per day (in valleys 0.5 mile wide) to 60 feet per day (in valleys 1 mile wide). These hydraulic-conductivity values are much larger than those that are considered typical of till and bedrock, but smaller than values reported for productive sand-and-gravel aquifers in some valley reaches in New York. Differences in the properties of till and bedrock and in land-surface slope or relief within the Susquehanna River basin of New York apparently have little effect on low flow.

Three regression equations were selected for practical application in estimating 7-day mean low flows in cubic feet per second with 10-year and 2-year recurrence intervals, and 90-percent flow duration, at ungaged sites draining more than 30 square miles; standard errors were 0.88, 1.40, and 1.95 cubic feet per second, respectively. Equations that express low flows in cubic feet per second per square mile were selected for estimating these three indices at ungaged sites draining less than 30 square miles; standard errors were 0.012, 0.018, and 0.022 cubic feet per second per square mile, respectively.

## Introduction

The Susquehanna River basin encompasses south-central New York, central Pennsylvania, and a small part of northern Maryland (fig. 1). The part of the basin in New York is mostly an upland area of till-covered bedrock hills. The hills are separated by many narrow valleys that are floored with thin deposits of sand and gravel laid down by modern streams, and by a few broad valleys that contain thick deposits of sand and gravel, silt, and clay laid down by glacial meltwater. The exchange of water between the sand-and-gravel deposits and the streams that flow across them is a vital aspect of the hydrology of this region, as it is throughout the glaciated northeastern United States (Randall, 2001). Most streamflow during periods of low flow consists of water discharged from sand-and-gravel deposits, and infiltration from streams is the largest potential source of recharge to the thick sand-and-gravel deposits that constitute the most productive aquifers in the region.

Accurate estimates of minimum streamflow for periods of several days, weeks, or months are required to evaluate the suitability of streams for fish habitat, recreation, water supply, wastewater dilution, and recharge to aquifers or well fields. Commonly used indices of the magnitude and frequency of low flow have been computed for more than 150 sites along streams in the Susquehanna River basin of New York, by analyzing multiple streamflow measurements at those sites (Hunt, 1967; Ku and others, 1975; Eissler, 1979), but information is often needed at sites where a suitable array of measurements has not been made and cannot be made within cost or time constraints. The primary objective of this investigation was to identify and explain concepts that describe the hydrologic processes and properties that control groundwater discharge to streams during periods of low flow, and to test those concepts by means of multiple-regression analysis. A secondary objective was to develop several regression equations that could be used to estimate indices of low flow at ungaged sites in this region. Accordingly, this report first presents a conceptual appraisal of watershed properties and processes that are considered likely to affect streamflow during periods of low flow, as interpreted from the literature and from quantitative groundwater-flow modeling done as part of this investigation. The report then describes

2 Low Flow of Streams in the Susquehanna River Basin of New York



Base from U.S. Geological Survey National Atlas, digital line graph data files, 2005  
 Source scale 1:2,000,000 decimal degrees projected to Universal Transverse Mercator 1983 (UTM 83)  
 Susquehanna River basin boundary, river mainstem, and tributaries coverages from Susquehanna  
 River Basin Commission (140d) 12-01-2003

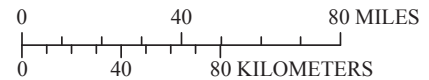


Figure 1. Location of the Susquehanna River basin and physiographic regions in New York and adjacent States.

the testing of these concepts through regression analysis, and presents a set of 18 equations that demonstrate the significance of particular watershed properties and computational procedures in estimating low-flow magnitude and frequency. Finally, six equations deemed conceptually and statistically sound are selected for practical application in estimating low flows in the Susquehanna River basin of New York.

This study refines a low-flow regionalization analysis by Ku and others (1975). The study was undertaken as part of the Northeast Glacial Aquifers Regional Aquifer-System Analysis project (NEGA-RASA), one of 28 projects designed by the U.S. Geological Survey to evaluate extensive aquifer systems that collectively cover much of the United States (Sun, 1986). Although the study was initially completed in 1989, revisions were postponed for several years to accommodate other priorities. Other studies of the interchange of water between aquifers and streams were undertaken as part of NEGA-RASA, including evaluations of watershed properties as controls on low streamflow in central New England (Wandle and Randall, 1994) and Connecticut (R.L. Melvin, U.S. Geological Survey, written commun., 1983) and studies of hydraulic properties and infiltration rates of streambeds (Lyford, 1986; Lapham, 1989; Dysart and Rheume, 1999; Williams, 1991; Williams and Morrissey, 1996; Yager, 1986).

## Watershed Properties and Processes that Control Spatial Variation in Low Flow of Streams

Previous studies of the interaction of groundwater and surface water (cited farther on) have, collectively, led to a provisional conclusion that the magnitude of unregulated streamflow in the glaciated northeastern United States during summer periods of low flow is controlled chiefly by four factors: (1) the amount of water that is available to become streamflow—that is, precipitation minus evapotranspiration; (2) the spatial extent and hydraulic properties of surficial sand and gravel, much of which (unlike other earth materials) can store a large proportion of the water available and gradually release that water to streams; (3) the extent of wetlands (especially wetlands bordered by sand and gravel), where evapotranspiration is greater than elsewhere during summer periods of low flow; and (4) the magnitude of underflow through riparian sand and gravel, whereby water is transmitted downvalley without being measured as streamflow. These four principal controls on low flow of streams are illustrated in figure 2 and discussed below, followed by a brief discussion of other factors that have less effect on low flow: watershed area, the extent and properties of till and of bedrock, watershed relief and slope, and streamflow regulation and watershed urbanization. The intent of the following sections is to identify and explain the watershed properties and processes that control groundwater discharge to streams during periods of

low flow, rather than to simply select a few easily measured watershed properties that collectively are closely correlated with low flow.

### Amount of Water Available for Runoff

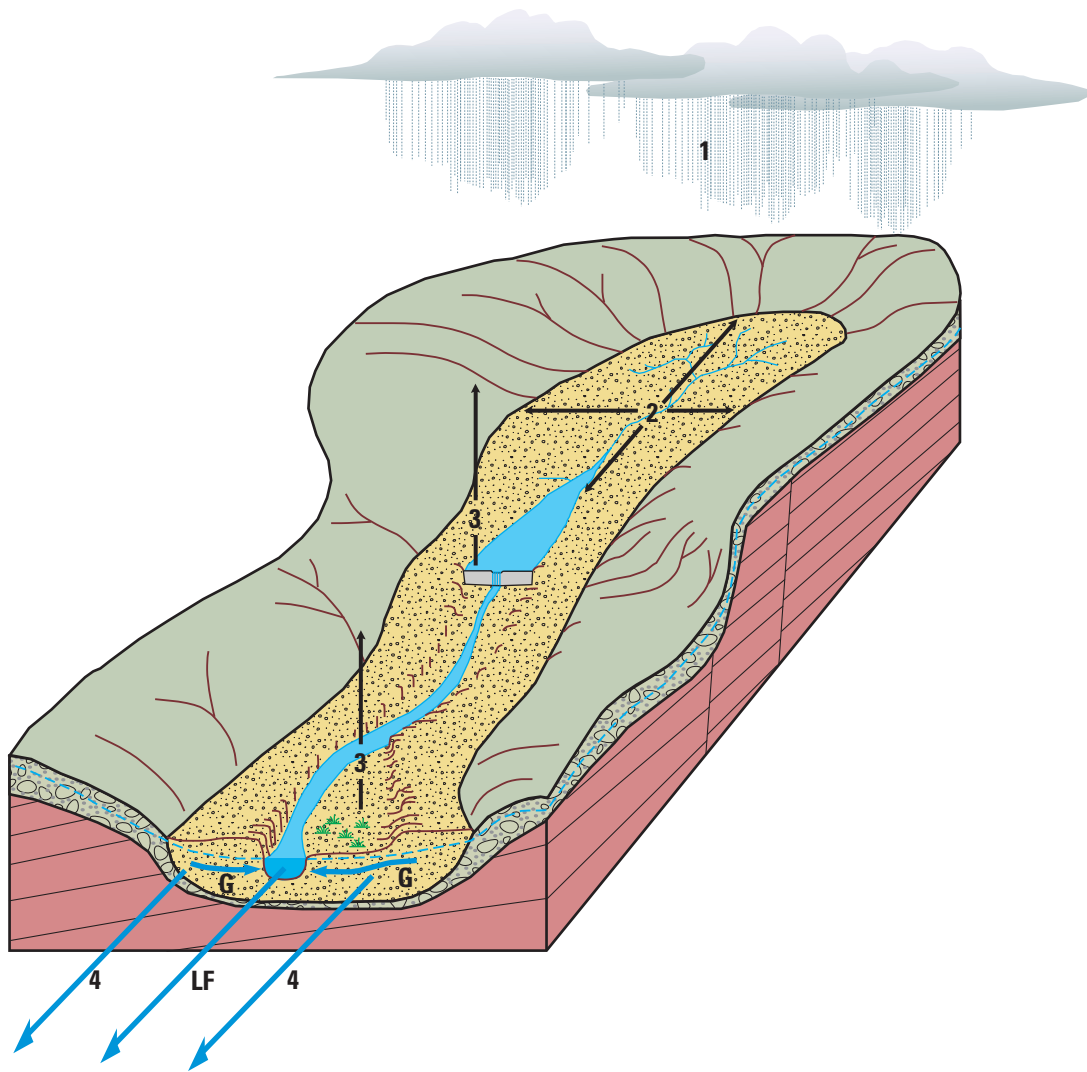
Water input to any watershed is entirely from precipitation, which varies in space (from one watershed to another) and in time (from one season or year to another). A substantial fraction of that input is lost to evapotranspiration, which also varies in space and time; the remainder becomes available for immediate storm runoff to streams or for groundwater recharge, which is eventually discharged to streams.

### Annual Precipitation and Runoff

Low streamflow consists predominantly of groundwater discharge, which is necessarily a function of prior groundwater recharge. Because most recharge in the glaciated Northeast occurs from late fall through early spring each year, as evidenced by an unsteady net rise in water levels during this period (Kontis and others, 2004, and references therein), annual low flows could be expected to correlate with the magnitude of recharge over the preceding year. Mean annual runoff from any watershed equals the average annual input to the watershed of water available for recharge or storm runoff (Lyford and Cohen, 1988), and closely approximates actual recharge in areas of surficial sand and gravel because precipitation can readily infiltrate nearly everywhere in these permeable soils.

Annual runoff and annual precipitation are strongly correlated with each other throughout the glaciated Northeast, and both generally increase with altitude (Knox and Nordensen, 1955; Randall, 1996). Runoff is expected to increase more rapidly than precipitation with altitude, because evapotranspiration should decrease with altitude as temperature decreases (Kontis and others, 2004, p. 11–13). An increase in evapotranspiration with altitude was calculated (as precipitation minus surface runoff) by DeAngelis and others (1984) and by Dingman (1981), but this increase could be attributed to an increase in forest cover with altitude, to greater deep groundwater circulation, or to underestimation of precipitation. Several studies have reported that the low flow of streams in the glaciated Northeast is in part a function of one or another of these three indices of water availability. For example, altitude was incorporated in low-flow-estimation equations for New Hampshire and Vermont by Dingman (1978) and by Wandle and Randall (1994), for Connecticut by Thomas and Cervione (1970), and for southeastern New York by Barnes (1986). Graphs first presented by Thomas (1966) for estimating the duration of low and high flows in Connecticut incorporated the ratio of mean annual runoff at any site of interest to the statewide average. Other studies have presented evidence for correlation between low flow

#### 4 Low Flow of Streams in the Susquehanna River Basin of New York



#### EXPLANATION

 Sand and gravel

 Till

 Bedrock

 Water table

- 1 Water input from precipitation
- 2 Area of sand or gravel (length x width)
- 3 Evapotranspiration from wetlands (lakes, swamps)
- 4 Groundwater flow downvalley (underflow)
- G Groundwater flow to stream through sand and gravel
- LF Low flow of stream

**Figure 2.** Principal environmental factors that control the low flow of streams in the glaciated northeastern United States.

and mean annual runoff (Ku and others, 1975; Johnson, 1970) or low flow and precipitation (Darmer, 1970; Male and Ogawa, 1982).

Maps of mean annual precipitation and mean annual runoff in the Susquehanna River basin in New York were presented by Ku and others (1975) for 1931–60 and were included within larger maps by Randall (1996) for 1951–80. In each of these reports, the two maps were drawn using a procedure (Randall, 1996) that forced the precipitation contours to parallel the runoff contours and assured that each set of contours was constrained both by precipitation data from measurement stations and by runoff data from stream-gaging stations. The 1951–80 maps (excerpts reproduced here as figure 3) were used to estimate mean annual precipitation and runoff for watersheds examined as part of this study. According to these maps, mean annual runoff is about 19 inches less than mean annual precipitation throughout the Susquehanna River basin of New York. Long-term mean precipitation is strongly correlated with altitude in western and central New York, as indicated by regression analysis (Ku and others, 1975, p. 10), but the correlation may have been controlled by large differences in altitude and precipitation between major physiographic regions in and near the Susquehanna River basin—the Appalachian Plateau, the much higher Catskill Mountains to the east, and the much lower Ontario lowland to the north (fig. 1). The narrow floors of major valleys in the Appalachian Plateau are commonly 800 to 1,000 feet lower in altitude than the extensive surrounding upland, but whether precipitation varies as a function of this localized relief could not be ascertained from the data available.

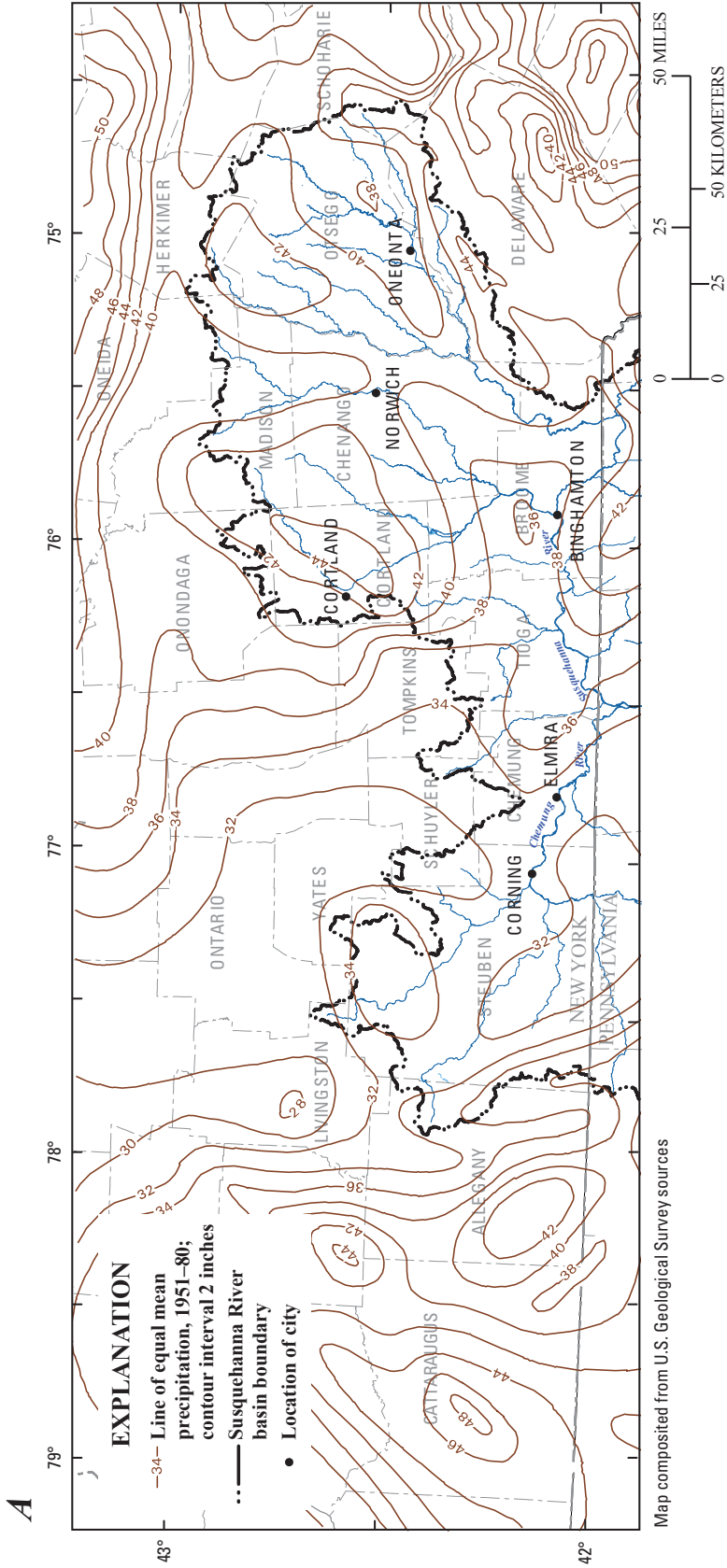
## Seasonal Precipitation and Runoff

The possibility that the effect of precipitation in late summer on low flow might exceed the effect of annual precipitation or runoff—because summer storms interrupt base-flow recessions, augment runoff, and cause localized recharge—is not supported by Ku and others (1975) nor by the present study. Ku and others (1975, p. 35) computed a variable that they called frost-free precipitation (total precipitation during the growing season) for 73 watersheds. They reported that low flow correlated more closely with mean annual runoff than with either mean annual precipitation or frost-free precipitation. In the present study, records of daily flow from several large watersheds were examined to ascertain which month contained the lowest average flow for 7 consecutive days in each year from 1938 through 1963; 95 percent of these values fell in August, September, or October. Therefore, precipitation during these 3 months at 41 precipitation stations used by Ku and others (1975) was averaged for 1931–60 and for 1941–70. (For 27 of these stations, 30-year monthly normals were available from National Weather Service publications or could be computed by estimating a few missing monthly values through correlation with other nearby stations. Monthly means for 12 to 27 years of record were

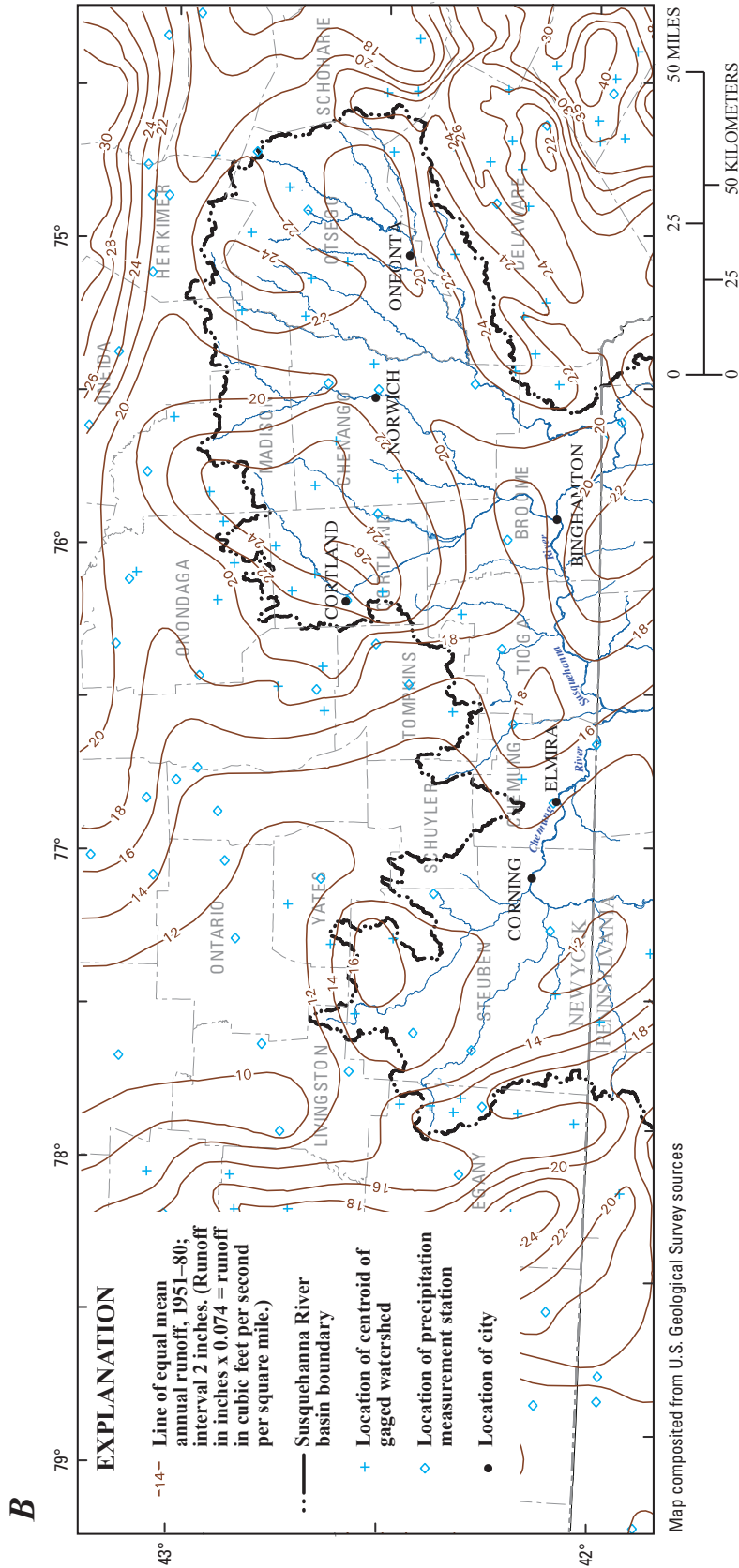
used for the remaining 14 stations.) The ratios of 3-month precipitation for 1931–60 and 1941–70 to mean annual precipitation for 1931–60 were then computed for each of the 41 precipitation stations. All ratios fell within a narrow range (0.24–0.28 for 1931–60, 0.23–0.27 for 1941–70) and did not show any apparent spatial trend across the region. Accordingly, mean annual precipitation was judged to be as suitable as mean August–October precipitation as an index of the amount of water available for runoff to be used in regression analyses. Maps of annual and seasonal precipitation (Dethier, 1966) support this judgment, in that distribution patterns for annual precipitation across the Susquehanna River basin in New York are similar to those for summer and those for the entire growing season.

Seasonal runoff was also evaluated during the present study as an index of water availability. The ratio of mean August–October runoff to mean annual runoff for 1938–63 was computed for each of eight watersheds or combinations of watersheds that range in size from 700 to 1,700 square miles (table 1) and were reasonably equidimensional (fig. 4). The runoff ratios varied more widely than the ratios of seasonal to annual precipitation. Furthermore, these runoff ratios tend to increase to the south and east (fig. 4)—that is, the percentage of annual runoff that occurs in August through October is slightly greater in Broome, Otsego, and Delaware Counties than in Cortland, Madison, and Steuben Counties. The reason for this pattern is not known, but spatial differences in the following three factors could be responsible:

1. *Precipitation minus evapotranspiration.* Ratios of August–October precipitation to annual precipitation for 1931–60 were also computed, at 41 measurement stations; they ranged only from 0.24 to 0.28 and exhibited no tendency to increase in any direction. This result indicates that orographic effects or storm-track patterns are probably not a cause of the southeastward increase in runoff. Although annual and seasonal precipitation are less west of Tioga County than that farther east (fig. 3A), annual evapotranspiration is uniformly about 19 inches across the region (Randall, 1996) and occurs chiefly during the growing season, from April through October (Olmsted and Hely, 1962). Therefore, the amount of water available for runoff is necessarily less west of Tioga County than to the east, especially during the growing season. This contrast could account for the low ratios of August–October runoff to annual runoff west of Tioga County (fig. 4).
2. *Surficial sand and gravel.* Eleven watersheds larger than 145 square miles east of Tioga County were reported by Ku and others (1975, appendix B) to average 17 percent sand and gravel, substantially greater than the 11-percent average for seven watersheds of similar size west of Tioga County. Therefore, the greater seasonal runoff east of Tioga County might be attributed to greater discharge from relatively extensive surficial sand and gravel. The reported sand-and-gravel percentages for



**Figure 3.** Spatial variation in *A*, mean annual precipitation and *B*, mean annual runoff in the Susquehanna River basin of New York, 1951-80. (From Randall, 1996.) (These two maps were drawn interactively, with both precipitation and runoff data used to control both maps. Distribution of data values is shown in figure 3B. Precipitation at each measurement site was partitioned into runoff and evapotranspiration. Watershed outlines, rather than centroids, were used to define locations of runoff data values.)



**Figure 3.** Spatial variation in A, mean annual precipitation and B, mean annual runoff in the Susquehanna River basin of New York, 1951-80. (From Randall, 1996). (These two maps were drawn interactively, with both precipitation and runoff data used to control both maps. Distribution of data values is shown in figure 3B. Precipitation at each measurement site was partitioned into runoff and evapotranspiration. Watershed outlines, rather than centroids, were used to define locations of runoff data values.)—Continued

## 8 Low Flow of Streams in the Susquehanna River Basin of New York

**Table 1.** Ratios of August–October runoff to annual runoff in the Susquehanna River basin and adjacent areas.

[ft<sup>3</sup>/s, cubic feet per second; watershed locations are shown in figure 4]

U.S. Geological Survey station numbers	U.S. Geological Survey station names and combinations of stations	Drainage area, in square miles	Mean runoff for April 1937–March 1963			Watershed number in figure 4
			August–October (ft <sup>3</sup> /s)	Annual (ft <sup>3</sup> /s)	Ratio	
014265 014135	West Branch Delaware River plus East Branch Delaware River at Margaretville	756	520	1,328	0.392	8
015005 015025	Susquehanna River at Unadilla plus Unadilla River at Rockdale	1,502	806	2,425	.332	7
015125	Chenango River at Chenango Forks	1,492	717	2,408	.298	6
015150	Susquehanna River near Waverly, less three upstream stations (5005, 5025, 5125)	1,786	899	2,667	.337	5
015205	Tioga River at Lindley	770	224	815	.275	2
015265 015295	Tioga River near Erwins plus Cohocton River near Campbell less upstream station (5205)	1,072	283	1,041	.272	3
015310	Chemung River at Chemung less two upstream stations (5265, 5295)	688	229	727	.315	4
042275	Genesee River at Jones Bridge, near Mt. Morris	1,417	443	1,621	.273	1

watersheds larger than 200 square miles are in part estimates, however, and the easternmost and highest ratio of seasonal runoff to annual runoff is in the adjacent Delaware River basin (table 1, fig. 4), where sand-and-gravel deposits on valley floors are relatively narrow and constitute a smaller percentage of watershed area than is typical of the Susquehanna River basin (Coates, 1971, table 1). Therefore, further study would be required to define the extent to which differences among large watersheds in percent area sand and gravel affect ratios of seasonal to annual runoff.

3. *Bedrock geology.* Grain size and fracture development in bedrock, and hence grain size of till, tend to increase to the south and east within the Appalachian Plateau of New York. These changes probably result in a modest southeastward increase in hydraulic conductivity in the uplands, which might be enough to augment groundwater discharge to streams in summer.

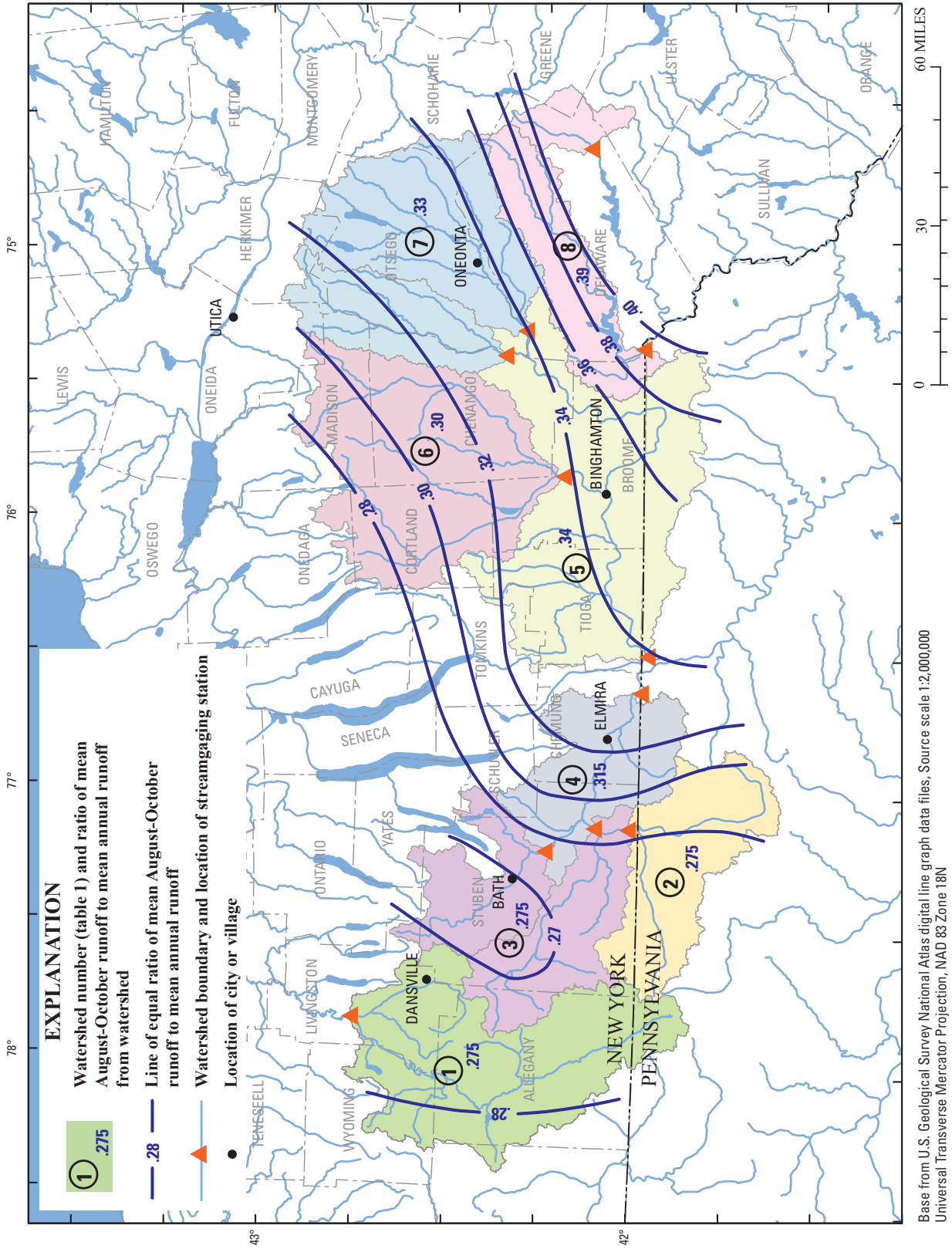
The foregoing discussion has shown that mean annual runoff is superior to mean annual or seasonal precipitation as an index of the amount of water available for purposes of estimating low flow, because low flow is derived from the part of precipitation that becomes runoff (rather than evapotranspiration). Mean annual runoff also is superior to seasonal runoff, because seasonal runoff is in part a function of surficial geology, whereas regional flow-duration graphs (Thomas, 1966; Kontis and others, 2004, fig. 10) indicate that

mean flows are negligibly affected by percent area sand and gravel. The superiority of mean annual runoff as an index of water availability is consistent with the concept that low streamflow is derived from groundwater stored during the winter and spring, chiefly in sand and gravel, and released slowly during the following summer. The ratio of August–October runoff to annual runoff varies systematically across the region, however, and may warrant consideration as a secondary control on low flow, but only if it can be computed as a function of precipitation minus evapotranspiration rather than from recorded streamflow, which is affected by surficial geology.

### Extent and Properties of Surficial Sand and Gravel

Many studies of streamflow in glaciated regions of the northeastern and north-central United States have concluded that low flow is strongly correlated with the spatial extent of surficial sand and gravel (expressed as the area of glacial and alluvial sediments that are primarily sand and gravel and that immediately underlie land surface, or as the fraction of watershed area occupied by such sediments). The present investigation augments a previous study (Ku and others, 1975) by adding new data and developing a numerical model to simulate several properties of surficial sand and gravel that affect low flow.





**Figure 4.** Ratios of mean August–October runoff to mean annual runoff for the period April 1937 to March 1963 in the Susquehanna River basin and adjacent areas. Watersheds 2–7 lie within, or drain into, the New York part of the Susquehanna River basin.

*Insights from previous studies.* Morse (1946) and Schneider (1957) ascribed the relatively constant flow from certain watersheds in Ohio, and particularly the large yield per square mile at 90-percent flow duration, to the predominance of coarse-grained stratified drift (sand and gravel deposited by glacial meltwater) in those watersheds. Thomas (1966) presented a family of flow-duration curves that express flow per square mile of unregulated streams in Connecticut as a function of the percentage of watershed area underlain by coarse-grained stratified drift. These curves were applied in a series of water-resources reports covering that State (for example, Mazzaferro and others, 1979) to estimate flows at ungaged sites. Thomas and Cervione (1970) incorporated the extent of coarse-grained stratified drift in regression equations, where it proved to be an exceptionally powerful predictor of low flow of unregulated streams in Connecticut (Riggs, 1973, p. 12). Coates (1971), Flint (1968), Ku and others (1975), and Pagano (1987) presented regression analyses that showed the percentage of watershed area occupied by coarse-grained stratified-drift valley fill to be a powerful predictor of low flow per square mile in the Susquehanna River basin of New York. Barnes (1986) reached similar conclusions for southeastern New York. Wandle and Randall (1994) found that the area of sand and gravel and the area of till plus fine-grained sediments are each significant in explaining variations in low flow in New England.

Tasker (1972) and Male and Ogawa (1982) found low flow of streams in Massachusetts to be positively correlated with a “ground-water factor” computed from the spatial extent of stratified glacial drift subdivided and weighted according to potential well yield; this factor was considered roughly proportional to average transmissivity of stratified drift in each watershed. Neither study addressed whether correlation of low flow with the ground-water factor was stronger or weaker than correlation with area of sand and gravel or with other indices of coarse-grained stratified drift. Allen and Cowan (1985) found low flow in northern Illinois to be a function of watershed area and a streamflow-recession index that is in turn a function of percentage of watershed area underlain by sand and gravel. Gebert (1979) reported that low flow in northern Wisconsin can be estimated from watershed area, transmissivity (a weighted average of standard values for five surficial geologic units), and an estimate of 90-percent flow duration; Gebert’s maps indicate that not only transmissivity but also the 90-percent flow-duration term are functions of the extent of surficial sand-and-gravel outwash relative to the extent of end moraine and lake clay.

The studies cited above and others have demonstrated that variation among watersheds in the extent of sand and gravel correlates strongly with, and presumably causes, variation in the low flow of streams. Several questions remain, however. What attributes of surficial sand and gravel are responsible for the large effect of these deposits on low streamflow? Could equations for estimating low flow be improved by subdividing stratified drift into categories that reflect those attributes—or can imperfect prediction of low

flow by regression equations be ascribed mostly to errors in calculating low-flow indices or in mapping the extent of sand and gravel, or to failure to consider other relevant watershed properties?

*Emphases of the present study.* Several approaches were used in this investigation to address the foregoing questions: (1) verify the interpretations of surficial geology by Ku and others (1975) in selected watersheds, as described in the next paragraph, (2) use analytical and numerical models to discern how certain properties of surficial sand and gravel affect groundwater discharge to streams, as described in the following section on modeling, (3) refine the selection of watersheds and the computation of low-flow indices, as described in the section “Data set,” and (4) test several watershed properties along with the extent of surficial sand and gravel in regression analysis; if the additional watershed properties improve estimates of low flow, the significance of the extent of sand and gravel is also affirmed.

*Verification of surficial geology.* Four watersheds whose low flows had been poorly estimated by the regression equations developed by Ku and others (1975, p. 38) were revisited during the present study to search for errors in the interpretation of surficial geology or other anomalies that might account for the outlier status of these watersheds in the regression analysis. An investigation of the watershed of Fivemile Creek at Kanona (station 01528000), whose low flow had been overestimated by a factor of 8, was described by Randall (1986b). Remapping of the surficial geology revealed that till and fine-grained lake sediments were much more abundant on the valley floor, and surficial sand and gravel were correspondingly less abundant, than had been realized previously. After localities where till or lake sediments cap the valley fill were removed from the area mapped as sand and gravel (even though gravel aquifers may be present at depth), the overestimation of low flow decreased to a factor of 3 (Randall, 1986b, p. 253–254). Remapping in the watersheds of Cohocton River at Cohocton (01527000), Castle Creek at Wallace (01527450), and a small part of the watershed of Charlotte Creek (01498500) resulted in little net change in area of sand and gravel, and hence little improvement in estimation of low flow.

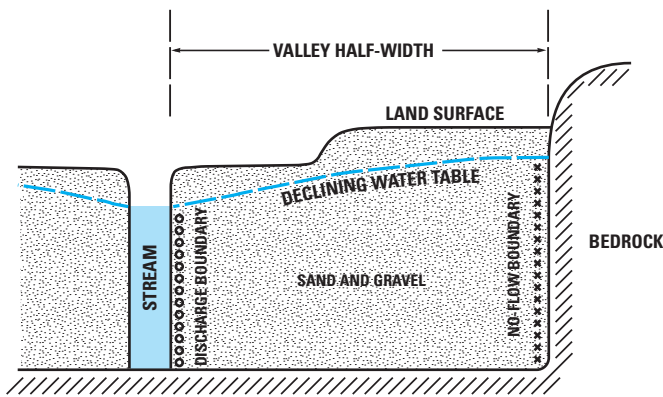
## Modeling Properties of Surficial Sand and Gravel that Affect Groundwater Discharge to Streams

Modeling groundwater discharge to a stream from a surficial sand-and-gravel aquifer in an idealized valley showed how variations in aquifer properties affect low streamflow. A simple analytical model is described first; discussion of a more realistic numerical model follows.

### Analytical Model

Rorabaugh (1960, 1963) developed equations to calculate groundwater discharge to streams as a function of transmissivity, storage coefficient, time, aquifer width from

stream to valley wall, and prior recharge. These equations apply to an idealized aquifer bounded on one side by a discharge boundary and on the opposite side by a no-flow boundary (fig. 5) and are used to calculate aquifer response to two conditions: (1) cessation of recharge that had taken place at a constant rate (Rorabaugh, 1963, equation 7), and (2) an instantaneous water-table rise, or an instantaneous stream-stage decline (Rorabaugh, 1963, equation 3 or 4).



**Figure 5.** Idealized aquifer represented by analytical equations of Rorabaugh (1963).

Rorabaugh's simple analytical model corresponds reasonably well to surficial sand-and-gravel aquifers in the glaciated Northeast, many of which are bordered by a stream in mid-valley (a discharge or recharge boundary, according to classical analytical conceptualizations) and by much less permeable till and bedrock along the valley side (a no-flow boundary). A typical year can be approximately represented as a long period of fairly constant recharge (October or November through April or May) followed by a period of no recharge (May or June through August, September, or October). Stream stage normally declines during the summer. For several large streams in the Susquehanna River basin, the median decline was 1.5 feet from May through August 1964, a dry summer with only occasional small pulses of storm runoff. Although the summer decline in stage is a result of declining groundwater discharge upstream, it can be treated as a cause of decline in water level in an aquifer bordering the stream at any given location downstream. Also, although the decline in stage is a gradual process, it can be simulated as several instantaneous steps.

For this report, Rorabaugh's equations were applied to an idealized set of aquifers and a plausible year, described as follows:

- constant recharge rate = 14.3 inches in 265 days = 0.0045 feet per day,
- length of subsequent period of no recharge = 100 days,

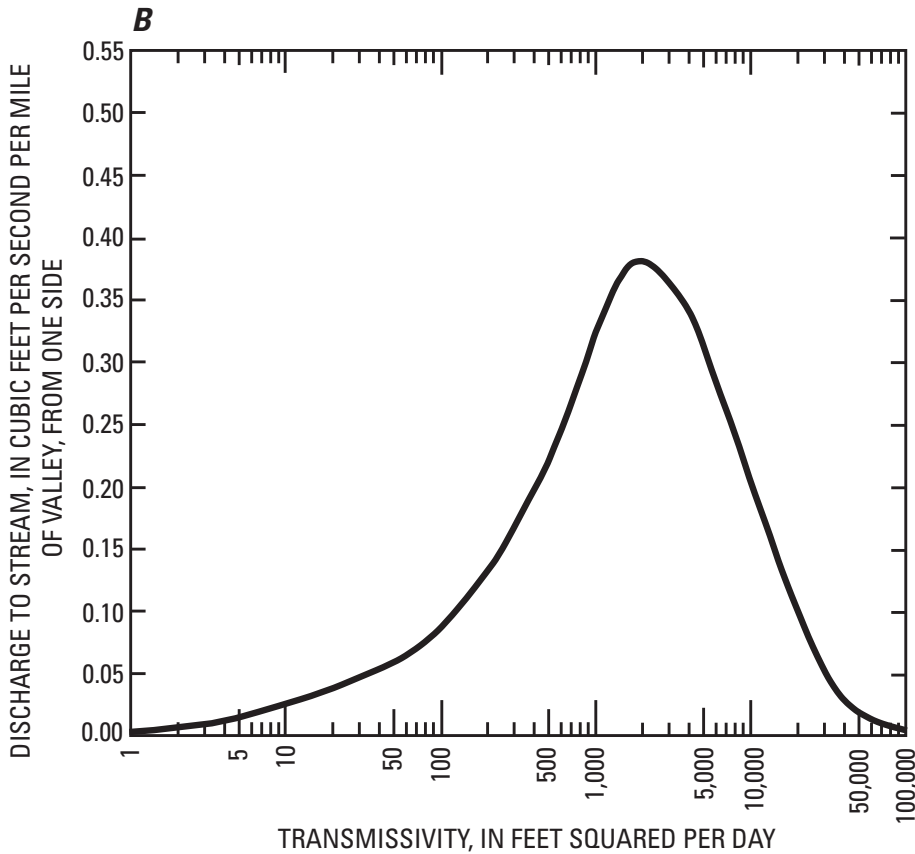
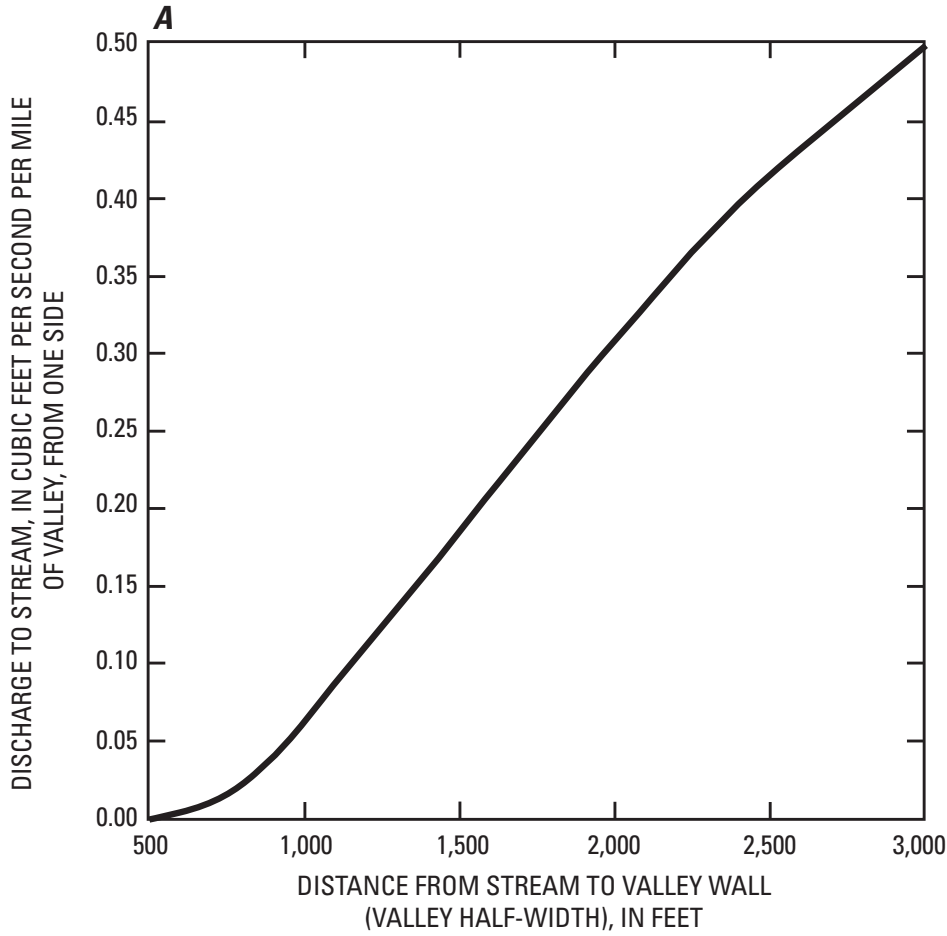
- stream-stage decline = 1 foot on the 40th day of no recharge,
- stream-stage decline = 0.5 foot on the 70th day of no recharge,
- specific yield = 0.2, and
- transmissivity was held constant at 5,000 feet squared per day while valley half-width was varied from 500 to 3,000 feet; then
- valley half-width was held constant at 2,000 feet while transmissivity was varied from 5 to 100,000 feet squared per day.

The equations were applied to calculate discharge per mile of valley length to a stream located midway between the valley sides, at the end of 100 days with negligible precipitation and no recharge. Results are summarized in figure 6. Low-flow discharge is approximately a linear function of valley half-width, for half-widths of 700 to 2,500 feet (fig. 6A). It is a nonlinear function of the logarithm of transmissivity, and peaks at a transmissivity of about 2,000 feet squared per day (fig. 6B). The general form of the relation in figure 6B is intuitively correct: as transmissivity approaches zero, little water can flow through the aquifer to the stream; as transmissivity approaches infinity, the aquifer resembles a lake in that water input from each storm drains immediately down to stream grade after a storm and therefore is not available to sustain low flow.

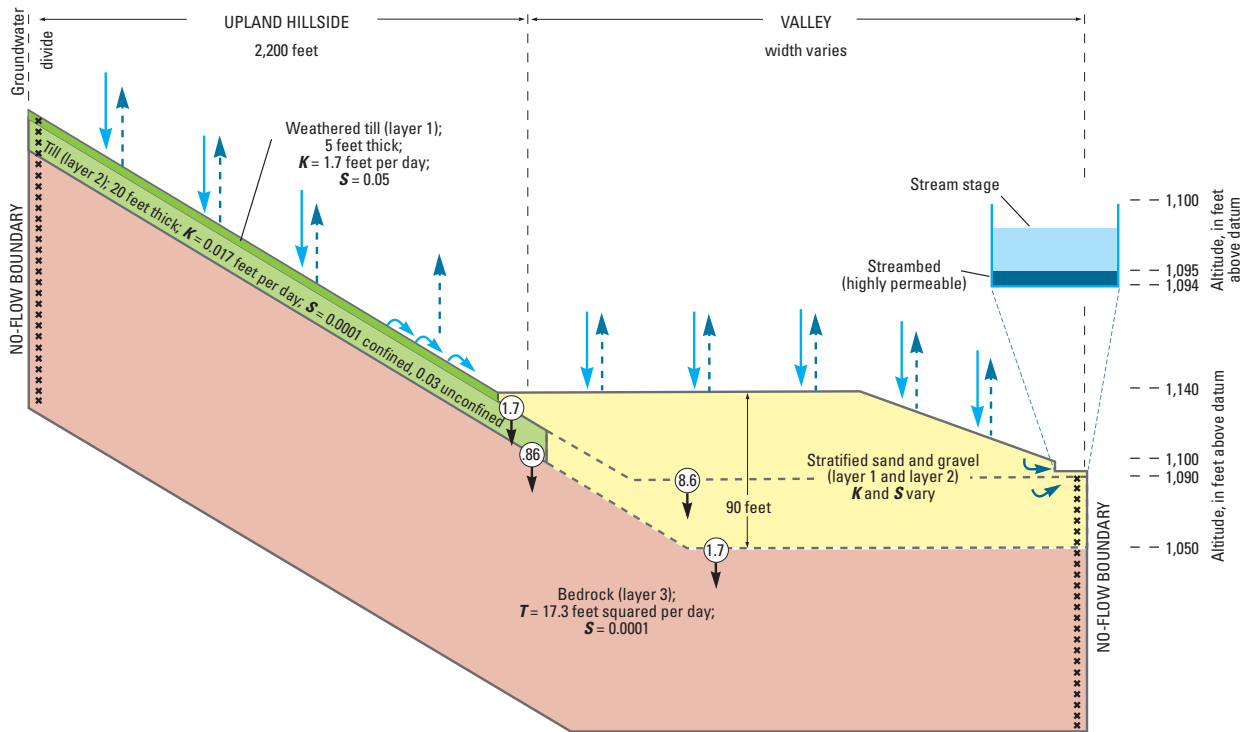
## Numerical Model

The analytical model by Rorabaugh does not fully describe the hydrology of valley-fill sand-and-gravel aquifers in that it does not allow for evapotranspiration, lateral flow from the adjacent upland, nor vertical flow near streams that only partially penetrate the aquifers. Therefore, a numerical groundwater-flow model was developed to incorporate these factors. The design, properties, and application of this model are explained on the following pages, followed by an analysis of the influence of several valley-fill properties on groundwater discharge to streams as revealed by results of model simulations. A summary of model design and principal results, including two figures, was published in Kontis and others (2004, p. C23–25).

The modular three-dimensional finite-difference digital code described by McDonald and Harbaugh (1988) was adapted to represent a two-dimensional vertical section across half of a hypothetical valley in which the valley floor is underlain by sand and gravel and is bordered by an upland hillside of till-mantled bedrock (fig. 7). Initially, the valley floor was assumed to be only 100 feet wide (50 feet from valley wall to stream). The part of the model that represented an upland hillside was calibrated by assigning hydraulic properties that were deemed reasonable for upland till and bedrock and that generated a plausible head distribution within



**Figure 6.** Effect of *A*, valley width and *B*, aquifer transmissivity on groundwater discharge to a stream during a period of low flow, as calculated from equations of Rorabaugh (1963).



**EXPLANATION**

- |   |   |                                       |   |
|---|---|---------------------------------------|---|
| <p>①.7<br/>↓</p> <p>↘ ↙</p> <p>↑</p> <p>↓</p> | <p>Vertical leakage between model layers, in feet per day per foot (x 10<sup>-3</sup> omitted)</p> <p>Discharge to stream</p> <p>Evapotranspiration</p> <p>Recharge</p> | <p>↘ ↙</p> <p>κ</p> <p>τ</p> <p>s</p> | <p>Seepage</p> <p>Hydraulic conductivity</p> <p>Transmissivity</p> <p>Storage coefficient</p> |
|---|---|---------------------------------------|---|

**Figure 7.** Dimensions and properties of a two-dimensional groundwater-flow model that represents an idealized vertical section across an upland hillside and the adjacent valley. (From Kontis and others, 2004.)

the upland and a discharge to the stream that was typical of the low flow of upland streams bordered by only narrow bands of alluvial gravel. Then, the width, hydraulic conductivity, and specific yield of the valley fill were varied, and the effects on simulated discharge to the stream were recorded.

#### Model Configuration and Operation

Properties and dimensions of the model are indicated in figure 7. Three layers were modeled in both upland and valley. The upland part of the model includes a relatively permeable top layer 5 feet thick that represents the surficial weathered zone that is typical of till in the Susquehanna River basin. Soils reports, such as Giddings and others (1971), describe many upland soils in the Appalachian Plateau of New York as 18 to 40 inches of friable, variably well-drained, relatively permeable channery silt loam above dense fragipan or bedrock. Ku and others (1975, p. 56) report that openings at shallow depth in upland soils are a principal avenue of runoff for several days after storms. The unweathered till underlying this top layer in the upland was assigned a thickness of 20 feet and a hydraulic conductivity equal to 1 percent of that in the top layer. The upper two model layers beneath the valley floor represent stratified drift and (or) alluvium, and were assigned identical hydraulic properties. The third or lowest layer represents bedrock and is uniform across the model. The streambed in the center of the valley, which is superimposed on the top layer at the far right of the model (fig. 7), was assigned a vertical hydraulic conductance large enough that the water level in the top layer directly below the stream would always equal the specified stream stage, because the intent of this exercise was to simulate effects of variations in properties of the sand-and-gravel aquifer as a whole, rather than variations in hydraulic conductivity or thickness of the streambed.

All simulations represented 1 year. For the first 6 months (November through April) evapotranspiration was specified as zero, river stage was constant, and water was made available for recharge at a constant rate through a computer code developed by A. L. Kontis (U.S. Geological Survey, written commun., 1984) and subsequently incorporated in a computer program described in detail by Kontis (2001). This code functioned much like the code that has been used to represent streams, in that each cell in the uppermost model layer was covered by an imaginary streambed that was assigned a conductance and thickness such that a vertical flux of 0.0104 feet of water per day (about 22 inches in 6 months) would be transmitted to the uppermost model layer as long as head in that layer remained well below land surface. The flux decreased as head rose to within 1 foot of land surface and ceased wherever head reached land surface and groundwater discharge occurred (simulating the natural development of seepage faces). In these simulations, seepage faces always developed on the lower part of the upland hillside, and occasionally developed in the adjacent part of the valley fill.

Simulation of the next 6 months (May through October) was designed to represent a dry summer that would result in low streamflow. Recharge was eliminated, although seepage discharge could still occur where heads were above land surface. River stage was lowered a total of 2.0 feet<sup>1</sup> in two steps (table 2), to approximate typical seasonal stream behavior. Evapotranspiration was specified at maximum rates that varied seasonally (table 2). The model subtracted evapotranspiration from cells in the top layer at rates that decreased linearly from the specified maximum where the water table was at land surface to zero where the water table was deeper than 5 feet. The specified maximum rate for each time period simulated was calculated by multiplying

<sup>1</sup> Reported incorrectly as 1.7 feet in Kontis and others (2004).

**Table 2.** Model parameters that varied with time during simulations of groundwater discharge to the stream in a hypothetical valley typical of the Susquehanna River basin of New York.

Simulation period	Duration of period (days)	Cumulative time simulated (days)	Maximum areal recharge rate, where water table is less than 1 foot below land surface (feet per day)	Stream stage (feet)	Maximum evapotranspiration, where water table is at land surface (feet per second)
November–April	180	180	$1.04 \times 10^{-2}$	1,097.0	0
May	30	210	0.0	1,097.0	$.10 \times 10^{-6}$
June	30	240	0.0	1,096.0	$.14 \times 10^{-6}$
July	30	270	0.0	1,096.0	$.16 \times 10^{-6}$
August	30	300	0.0	1,095.0	$.14 \times 10^{-6}$
September–October	65	365	0.0	1,095.0	$.065 \times 10^{-6}$

an annual lake evaporation or potential evapotranspiration rate for south-central New York of 24 inches (averaged from Thornthwaite and Mather, 1955; Hely and Nordenson, 1961; and Todd, 1970) by an estimate of the percentage of annual evapotranspiration occurring in that time period near Binghamton, New York (Randall, 1986a). That estimate treated evapotranspiration as proportional to the product of mean air temperature and hours of daylight (Olmsted and Hely, 1962).

A simulation was considered stable if water levels in the valley fill at the end of the simulated year were within 1 foot of those assumed at the start of the simulation, or if the water table ceased rising at some time during the 180-day period of recharge (which happened at high hydraulic conductivities). The rate of discharge to the stream at the end of August (that is, after 180 days with recharge followed by 120 days without recharge) was then recorded as the result of the simulation. If a simulation was not considered stable, a second simulation was run in which the starting heads in the valley fill were the same as those at the end of the initial simulation. Starting heads in cells representing the upland hillside were the same as starting heads used in the initial simulation, however, because (1) the model code did not allow for rewetting cells that had gone dry during the initial simulation, and (2) till in the uplands of the Appalachian Plateau becomes saturated nearly to land surface every spring.

Model Evaluation

Results of 38 simulations were used to plot the curves in figure 8, which depict groundwater discharge under low-flow conditions, at the end of 120 days without recharge. Two curves are plotted for each valley half-width simulated. The lower curve represents groundwater discharge to the stream: the upper curve represents total groundwater discharge (discharge to the stream plus evapotranspiration of groundwater). At the higher hydraulic conductivities, the two curves for each valley half-width are parallel and separated by about  $0.04 \times 10^{-4}$  cubic foot per second; the difference represents evapotranspiration of groundwater from the upland part of the model. This set of simulations indicated that groundwater discharge to the stream is controlled by four properties of the valley fill—hydraulic conductivity, specific yield, topography (which affects groundwater evapotranspiration), and width (or area) of valley-fill sand and gravel. Each of these properties is discussed in turn in the following paragraphs.

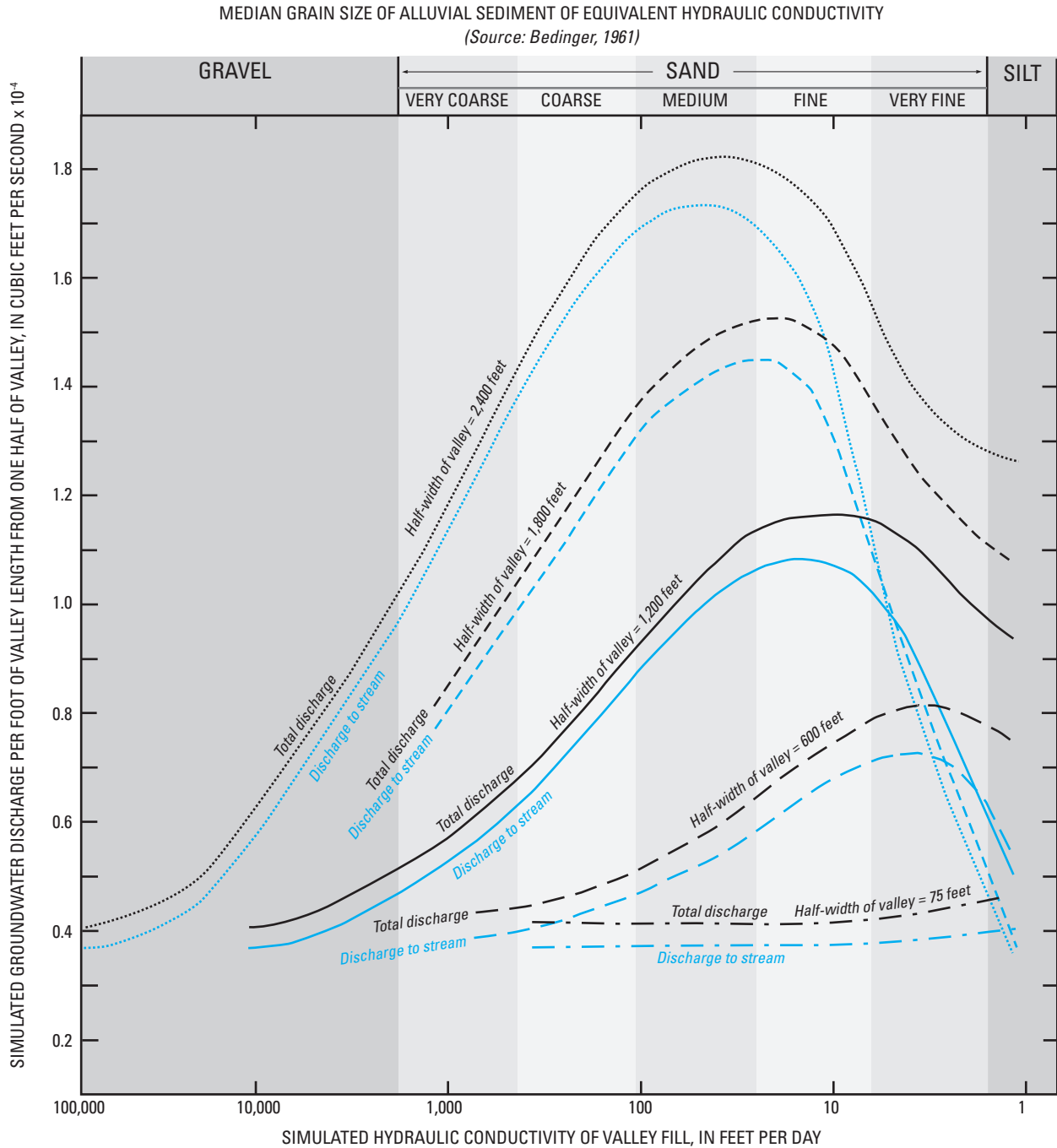
1. *Hydraulic conductivity.* The lowest hydraulic-conductivity value assigned to the valley fill (1.1 feet per day, fig. 8) is comparable to the value assigned to the uppermost till layer in the upland (1.7 feet per day, fig. 7). As the hydraulic conductivity of the valley fill was increased incrementally, discharge to the stream at the end of the recession increased at first, then decreased. At large hydraulic-conductivity values (typical of clean gravel) or at narrow valley widths, discharge to the

stream approached  $0.37 \times 10^{-4}$  cubic foot per second, which was the simulated rate of discharge from the upland. Thus, under either condition the valley fill served only as a conduit from upland to stream at times of low flow. Groundwater discharge to the stream at the end of the simulated period of low flow was greatest where hydraulic conductivity of the valley fill ranged from about 15 feet per day, equivalent to that of fine sand (in valleys with a half-width of about 0.25 mile) to about 60 feet per day, equivalent to that of medium sand (in valleys with a half-width of about 0.5 mile) (fig. 8). Saturated thickness could vary from 45 feet, with the water table at stream stage, to 90 feet, with the water table at land surface (fig. 7, table 2). Much larger average values of hydraulic conductivity have been reported for some valley-fill aquifers in New York (Randall, 1977; Reynolds, 1987; Bergeron, 1987); figure 8 suggests that such productive aquifers may contribute less water to streams during periods of low flow than fine-sand aquifers that drain more slowly. Such an inverse correlation between hydraulic conductivity of surficial sand and magnitude of low flows has been observed in the Coastal Plain of Delaware, where surficial fine to coarse sand (comparable to the valley component of figure 7) underlies entire watersheds (Johnston, 1971).

2. *Specific yield.* Specific yield of the valley fill was 0.2 in all simulations depicted in figure 8. In another simulation (not shown in figure 8), specific yield was decreased to 0.035 (a value similar to that assigned to weathered upland till); valley half-width was 2,400 feet, and hydraulic conductivity of the valley fill was 112 feet per day. The decrease in specific yield caused water-table altitude and discharge to the stream to increase early in the summer, but, as shown in the following table, also caused discharge to the stream to decrease during the period of low flow in late summer.

Mode of discharge	Groundwater discharge, in cubic feet per second, after 120 days without recharge	
	Specific yield = 0.2	Specific yield = 0.035
To stream	$1.641 \times 10^{-4}$	$0.557 \times 10^{-4}$
Evapotranspiration	$0.060 \times 10^{-4}$	$0.050 \times 10^{-4}$
Total	$1.701 \times 10^{-4}$	$0.607 \times 10^{-4}$

The model code used for this analysis treated specific yield as constant with time, and drainage as occurring immediately in response to a decline in head, although some delay in complete drainage of granular sediments is commonly observed (Rasmussen and Andreasen, 1959, p. 83; Kruseman and deRidder, 1990, p. 99). Delayed drainage could be expected to decrease groundwater runoff early in any period without recharge, and increase it later, relative to results of these simulations.



**Figure 8.** Simulated groundwater discharge, to a stream and as evapotranspiration, from an idealized upland hillside and valley fill after 4 months without recharge, as a function of hydraulic conductivity of the valley fill. Successive curves illustrate the effect of half-width of valley, which is the distance from the base of the upland hillside to a stream flowing along the valley axis. Total discharge for each half-width is groundwater discharge to the stream plus groundwater evapotranspiration. (From Kontis and others, 2004.)



3. *Topography and groundwater evapotranspiration.* Stream stage was 43 to 45 feet below most of the valley floor in the model (fig. 7, table 2), and in many simulations the water table remained far below the valley floor throughout the year. Under these conditions, simulated groundwater evapotranspiration was small. As hydraulic conductivity was decreased, however, the water-table gradient toward the stream increased and the water table rose closer to land surface seasonally, resulting in progressively greater groundwater evapotranspiration for longer periods and a corresponding decrease in discharge to the stream. This trend is expressed in figure 8 by an increase in separation of the two curves for each simulated valley half-width as hydraulic conductivity decreases. If land surface on the valley floor were higher than specified in figure 7, the water table would seldom approach land surface even in fine sand; if land surface were lower than specified, the loss of groundwater through evapotranspiration would decrease low flows even where the valley fill consisted of coarse, permeable sand. Trial simulations were run with a valley half-width of 2,400 feet and a hydraulic conductivity of 112 feet per day, first with the topography in figure 7, then with land surface lowered to 1 foot above the simulated water-table profile at the start of the 120-day period of no recharge. Results were as follows:

Mode of groundwater discharge	Groundwater discharge, in cubic feet per second, after 120 days without recharge	
	Land surface as depicted in figure 7	Land surface 1 foot above maximum water table
To stream	$1.641 \times 10^{-4}$	$1.231 \times 10^{-4}$
Evapotranspiration	$0.061 \times 10^{-4}$	$0.217 \times 10^{-4}$
Total	$1.702 \times 10^{-4}$	$1.448 \times 10^{-4}$

The topography of many valley reaches in the Susquehanna River basin of New York is similar to that depicted in figure 7 in that terraces or irregular collapsed landforms of permeable sand or gravel are many feet higher than the stream and cover a substantial part of the valley floor. Some valley reaches, however, contain low-lying areas where the water table is close enough to land surface that evapotranspiration limits discharge to streams during dry periods. In contrast, the water table throughout the uplands rises nearly to land surface every spring, allowing substantial evapotranspiration, localized groundwater discharge, and much runoff from rejected recharge—processes that together leave little water available for discharge to streams in late summer.

4. *Width, or area, of valley-fill sand and gravel.* Figure 8 indicates that in coarse-sandy valley fills, the maximum rate of groundwater discharge to the stream increases in proportion to the increase in valley width (and therefore in proportion to the increase in area of valley fill, inasmuch as the model represents a unit length of valley). At low hydraulic conductivity, however, this relation of discharge to width is masked by the effects of topography and groundwater evapotranspiration, as explained earlier. Figure 9 depicts the relation of valley half-width to groundwater discharge derived from the valley fill (that is, simulated discharge to the stream at the end of the 120-day recession minus the  $0.37 \times 10^{-4}$  cubic foot per second derived from the upland part of the model). That relation is nearly linear, except where the water table is close enough to land surface to allow appreciable groundwater evapotranspiration. Evapotranspiration of groundwater becomes appreciable where valley half-width exceeds about 1,800 feet in medium to fine sand with a hydraulic conductivity of 30 feet per day, or about 700 feet in fine to very fine sand with a hydraulic conductivity of 10 feet per day.

Although the numerical-model simulations indicate that groundwater discharge from valley-fill sand and gravel to streams is generally proportional to the area of sand and gravel, total low-flow yield per square mile from a long, moderately narrow valley-fill aquifer may exceed that from a wide aquifer of equal area because the long, narrow valley fill is likely to receive, store, and subsequently discharge more runoff from adjacent uplands (Morrissey and others, 1988). Accordingly, a regression equation of the form:

$$\text{Low flow} = a(\text{area of sand and gravel}) + b(\text{valley-wall perimeter of sand and gravel})$$

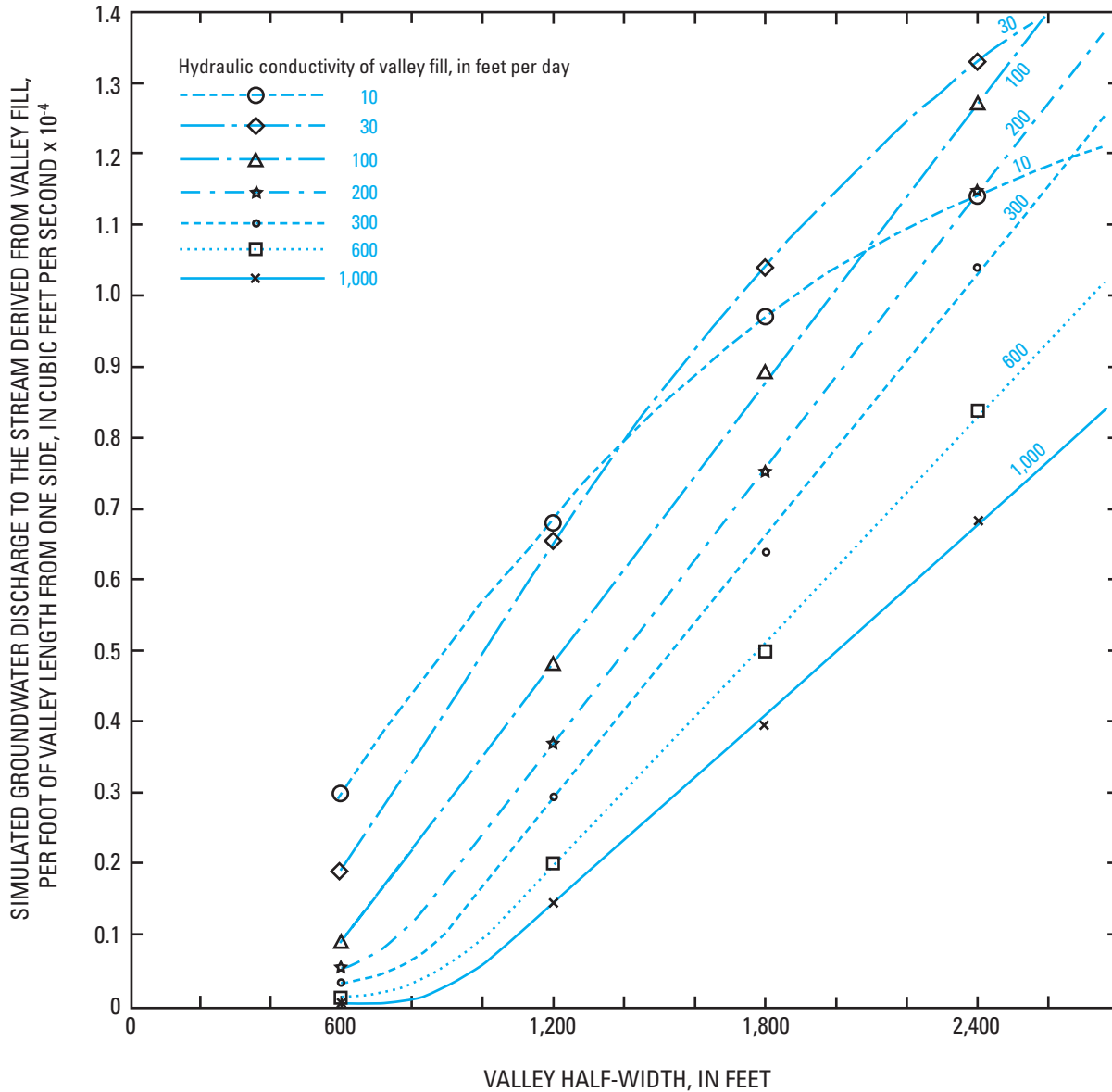
where

*a* and *b* are regression coefficients,

might estimate low flow more accurately than if area of sand and gravel were the only independent variable. However, nearly every watershed in the glaciated Northeast contains many stream reaches that are bordered only by till and receive small amounts of groundwater discharge from the till during periods of low flow. Therefore, an equation of the form

$$\text{Low flow} = a(\text{area of sand and gravel}) + b(\text{area of till})$$

could be expected to account approximately for the small low-flow yields from till areas whether discharged directly to streams or transmitted through sand and gravel. Equations of this form, in which regression coefficient *a* is much larger than *b*, were developed for streams in New England by Cervione and others (1982) and Wandle and Randall (1994).



**Figure 9.** Simulated groundwater discharge to a stream from an idealized valley fill after 4 months without recharge, as a function of valley half-width. Upward flattening of some curves reflects evapotranspiration from a shallow water table.

### Evapotranspiration from Wetlands

Although low flow in the Susquehanna River basin of New York is strongly correlated with the spatial extent of sand and gravel, that correlation consistently overestimates low flow from watersheds that contain extensive lakes or swamps (Ku and others, 1975, fig. 16). The area occupied by lakes and swamps was found to have a significant negative correlation with low flow in Connecticut (R.L. Melvin, U.S. Geological Survey, written commun., 1983), in Massachusetts (Male and Ogawa, 1982), and in central New England (Wandle and Randall, 1994). Barnes (1986) deleted areas of sand and gravel overlain by lakes or swamps in southeastern New York

before correlating the remaining areas of sand and gravel with low flow. Johnson (1970) demonstrated that the percentage of upland area underlain by poorly drained soils in northern Vermont is negatively correlated with low flow.

An inverse correlation between the extent of wetlands (such as lakes, ponds, or swamps) and low flow in the Susquehanna River basin could be expected for two reasons: (1) Precipitation cannot infiltrate where the water table is at land surface, so most precipitation on wetlands becomes surface runoff rather than groundwater recharge and discharge. (2) Most swamps and lakes border or are part of the stream network. During the growing season, these wetlands are sites of intense evapotranspiration that captures and returns to the

atmosphere groundwater that was recharged nearby and would have been discharged to streams in the absence of wetlands.

The area of riparian wetlands also commonly correlates inversely with high flow (for example, Zembrzuski and Dunn, 1979), presumably because wetlands provide temporary storage for floodwaters, thereby damping flood peaks and delaying runoff. This process may be operative to some extent during ordinary storms as well as major floods. Therefore, wetland area may correlate positively with increased streamflow over a wide range of moderate flow conditions, when the stored water is gradually released.

For this study, the extent of lakes and swamps shown on U.S. Geological Survey 7.5-minute topographic quadrangle maps was taken to be representative of wetland area. At least three alternative conceptualizations of wetland area, discussed below, may ultimately prove more useful as indices of evapotranspiration:

1. *Include all poorly and very poorly drained soils as part of wetland area.* Modern soils maps, prepared by the Natural Resources Conservation Service and cooperating agencies for large areas of the glaciated Northeast including the Susquehanna River basin, classify each soil-mapping unit with respect to drainage. Poorly drained soils are saturated to land surface periodically during the growing season; very poorly drained soils have free water on land surface during most of the growing season. Pagano (1987) measured areas of these two drainage classes in several watersheds in the Susquehanna River basin, and reported that in each watershed the area of poorly and very poorly drained soils was much larger than the area of swamps and lakes shown on topographic maps. Pagano (1987, p. 95) found that correlation of stratified valley-fill deposits with low flow was improved by deleting any areas of poorly or very poorly drained soils from the area of valley-fill deposits. He also reported a few equations in which area of upland soils with poor drainage correlated positively with low flow (probably because these soils were quite extensive and served as surrogates for drainage area).
2. *Include alluvium as part of wetland area.* Wandle and Randall (1994) distinguished alluvium (flood plains and alluvial fans) from sand and gravel deposited by glacial meltwater in 49 watersheds in central New England, and found that the inverse correlation of wetlands with low flow was improved by including alluvium with wetlands. An inverse correlation of alluvium with low flow might reflect the high hydraulic conductivity of coarse alluvium as well as abundant evapotranspiration from flood plains.
3. *Use National Wetlands Inventory delineation of wetlands.* National Wetlands Inventory maps, prepared by the U.S. Fish and Wildlife Service at a scale of 1:24,000, cover large areas of the Northeast. These maps delineate permanent (saturated) and temporary (seasonal) wetlands of several types (Cowardin and others, 1979) on the

basis of detailed photointerpretation, field studies, and interagency review. The maps have achieved accuracies above 95 percent (Swartwout, 1982) and, in central New England, depict a different and generally larger extent of wetlands than those shown on U.S. Geological Survey topographic maps (S.W. Wandle, U.S. Geological Survey, written commun., 1988).

## Underflow beneath Sites of Streamflow Measurement

Total runoff from a watershed consists of the flow of the stream that drains the watershed, plus any groundwater that leaves the watershed below land surface. In the uplands of the Susquehanna River basin, groundwater divides coincide with topographic divides, so no groundwater leaves a watershed beneath its upland perimeter. In valleys, however, the water-table gradient has two near-horizontal components, one oriented toward the main stream, the other oriented downvalley parallel to that stream and approximately equal to the stream gradient. Groundwater flow in response to this downvalley gradient is termed underflow.

Underflow can range widely from site to site along streams in the Susquehanna River basin. Ku and others (1975) measured streamflow and reported low-flow statistics at 40 upland sites where tributary streams are incised in till or bedrock, both of which are so poorly permeable that underflow is virtually nil. Most reaches of upland streams are either incised in till or bordered by flood plains that are only a few hundred feet wide and are underlain by alluvial gravel, sand, and silt probably less than 10 feet thick, through which underflow is likely to be small. The larger valleys, however, commonly contain several tens of feet of permeable sand and gravel that could transmit appreciable underflow. Rates of underflow in reaches of three broad valleys that are underlain by coarse, permeable stratified drift near the modern Susquehanna River basin divide were estimated by Randall and others (1988, p. 25, p. 68, p. 79). Underflow in one valley was estimated by two methods to average about 4 cubic feet per second over 11 months, and underflow in two other valleys was estimated less precisely from streamflow measurements to be nearly 3 cubic feet per second and several cubic feet per second, respectively. Ku and others (1975, p. 14, p. 33) noted that wherever a tributary leaves its own valley to cross sand and gravel within a larger valley, water seeps out of the tributary channel into the sand and gravel, and is transmitted as underflow within the larger valley. The seepage losses typically exceed 1 cubic foot per second per 1,000 feet of channel (Randall, 1978). Dry channel reaches have been observed during periods of extreme low flow along a few major streams, including Tuscarora Creek (near station 01526000, fig. 10), where drainage area is 114 square miles (U.S. Geological Survey, 1970) and East Branch Owego Creek (near Newark Valley, fig. 10), where drainage area is about 90 square miles. Watershed runoff under extreme

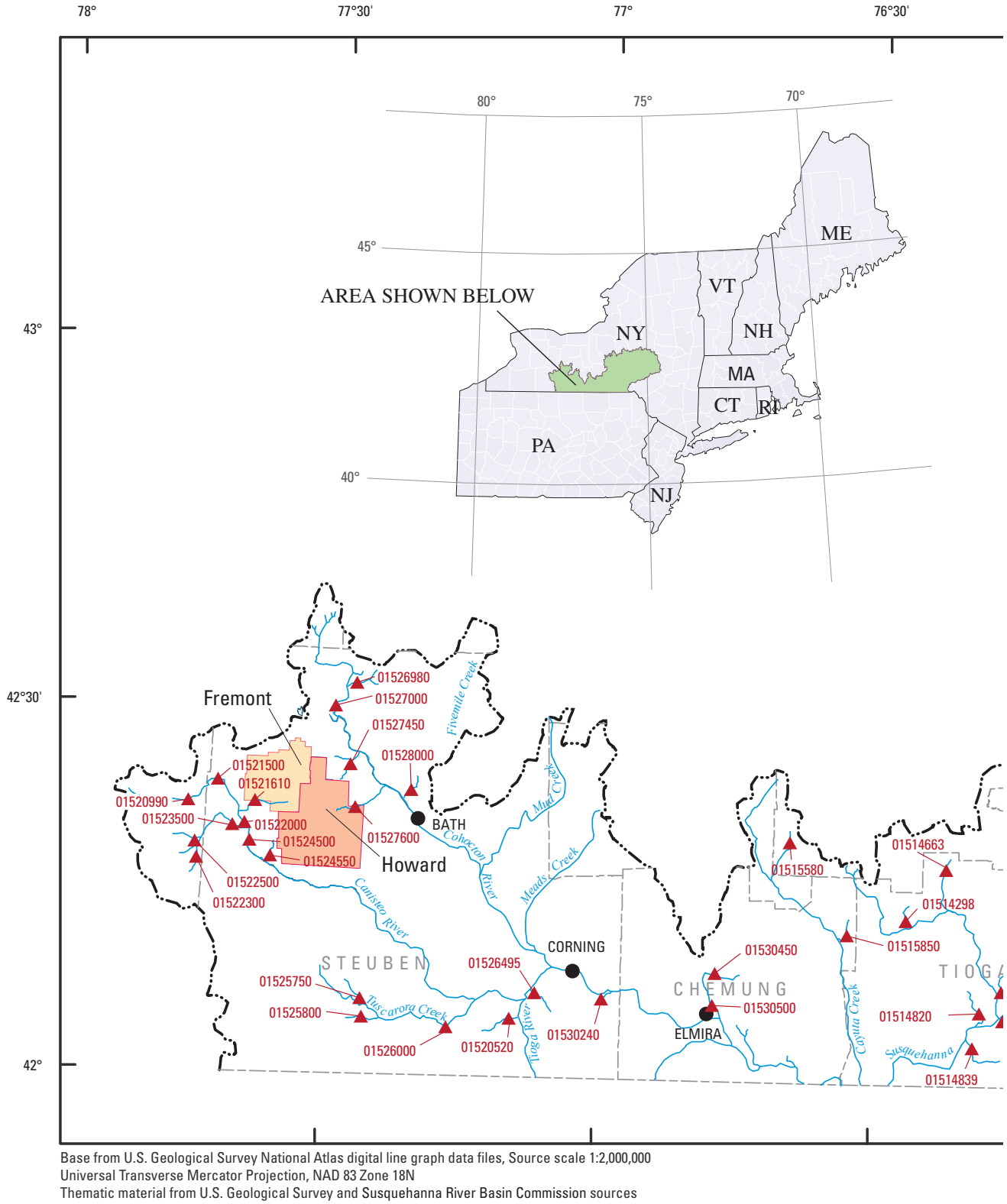


Figure 10. Locations of streamflow-measurement stations in the Susquehanna River basin of New York.

76°

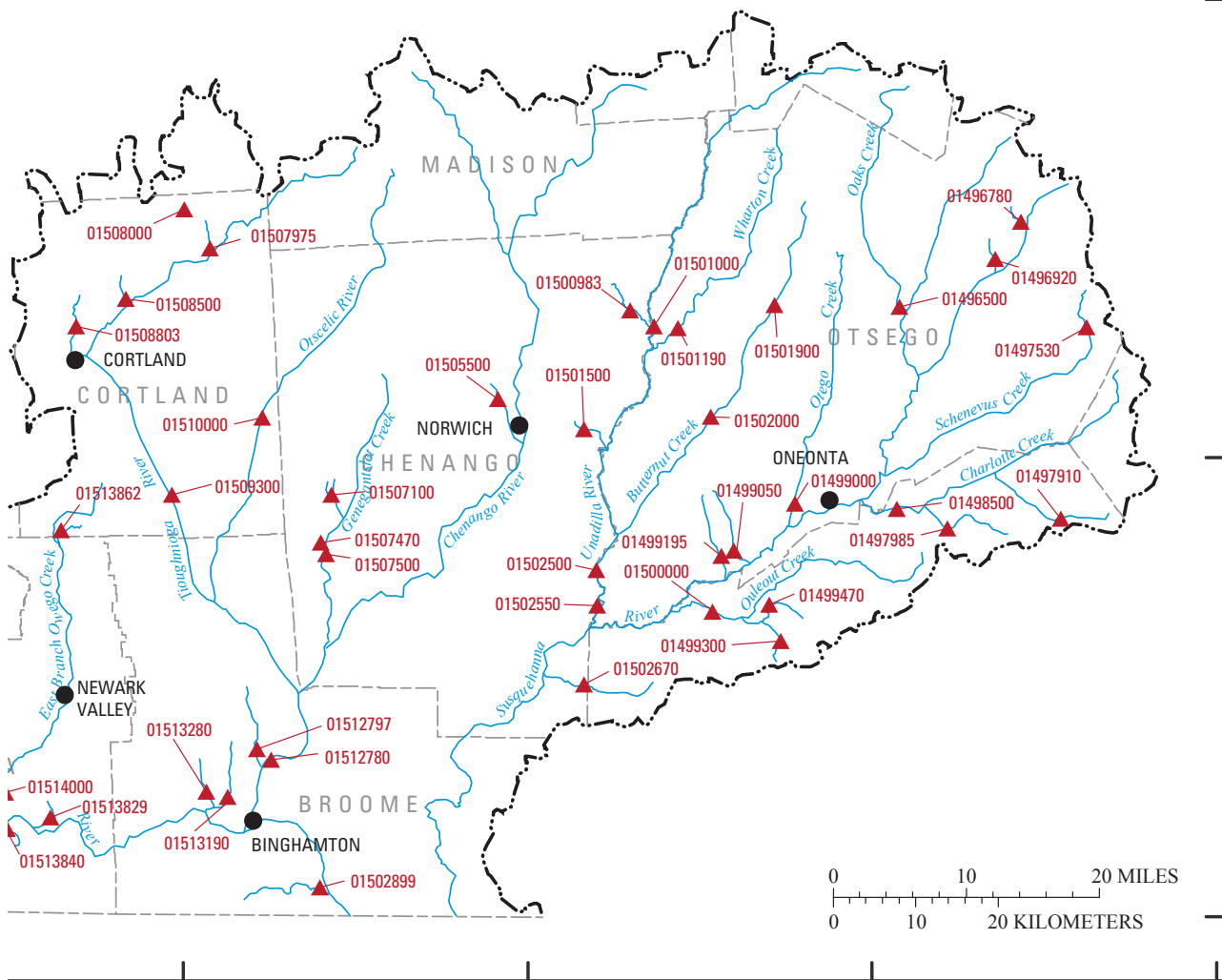
75°30'

75°

74°30'

**EXPLANATION**

- County boundary
- - - - - Susquehanna River basin boundary
- ▲ 01526980 U.S. Geological Survey streamflow measurement station and number
- Location of city or village



low-flow conditions at these two locations was probably in the range of 1 to 4 cubic feet per second, and occurred entirely as underflow.

If underflow is as great as a few cubic feet per second in some valley reaches, but virtually zero in others, this variation needs to be accounted for in estimating low flow, especially along small streams.

## Less Important Factors that Affect Low Flows

Five additional watershed properties or conditions that may affect low streamflow are watershed area, extent and properties of till, extent and properties of bedrock, watershed relief and slope, and degree of stream regulation and urbanization. Evidence briefly summarized on the following pages suggests that these properties were of less importance than water availability, surficial sand and gravel, evapotranspiration, and underflow in the Susquehanna River basin of New York during the period of streamflow record studied (1941–71).

## Watershed Area

Streamflow could logically be expected to increase in proportion to watershed area, all other factors being equal. All other factors are seldom equal, however, and a principal reason for developing equations for estimating low flow is to improve on the imprecise method of extrapolating low flows on a per-square-mile basis. According to studies in Connecticut and central New England, regression equations that incorporated area of sand and gravel and area of till as explanatory variables, or only area of sand and gravel, estimated low flow much more accurately than equations that relied simply on watershed area (Cervione and others, 1982; Wandle and Randall, 1994, p. 19). Trial equations tested as part of the present study confirmed this result. Therefore, drainage area alone is not used as an explanatory variable in the regression analyses presented in this report. Some of these equations account for the effect of area by dividing some or all watershed properties by drainage area. Other equations incorporate sand-and-gravel area and till area, which together equal drainage area but are more powerful individually as controls on groundwater discharge to streams.

## Extent and Properties of Till

Variations in till thickness, grain size, and permeability of the weathered soil zone had little effect on low flow from 14 watersheds in the Susquehanna River basin of New York and northernmost Pennsylvania (Pagano, 1987). Nine of these watersheds are included in the data set for this report. The spatial extent of surficial stratified drift and alluvium in these watersheds was strongly correlated with low flow. Pagano (1987) classified most of the till in these watersheds into

three categories according to properties reported in modern soils surveys:

1. till, less than 3 feet to bedrock;
- 2a. till, less than 20 percent silt and clay, permeability greater than 0.2 inches per hour; and
3. till, more than 20 percent silt and clay, permeability less than 0.2 inches per hour.

None of these three till categories proved significant in regression equations designed to estimate low flow. Pagano (1987) also identified two other categories of till, present in only a few watersheds, that had a significant inverse correlation with low flow:

- 2b. till, similar to 2a, but with thicker soil profiles and located only on valley sides or floors; and
4. till overlying or interbedded with outwash sand and gravel, located chiefly in valleys.

Abundant evapotranspiration from poorly drained till soils might result in an inverse correlation of those soils with low flow, but till categories 2b and 4 are not poorly drained. Because these two till categories are present largely in watersheds with valley floors 600 to 2,000 feet wide that are underlain by substantial thicknesses of stratified drift, they might have functioned in Pagano's regression analyses as surrogates for a decrease in low streamflow caused by underflow through stratified drift.

## Extent and Properties of Bedrock Units

The sedimentary bedrock of Devonian age that underlies the Susquehanna River basin of New York was classified into several facies by Rickard (1975). Arranged in order of decreasing grain size and percentage of detrital fragments, they are:

*Catskill*: shales, mudstones, siltstones, medium to coarse sandstones, pebble conglomerates;

*Cattaraugus*: shales, siltstones, sandstones;

*Chemung and Hamilton*: shales, mudstones, siltstones, fine to medium sandstones;

*Portage*: shales, mudstones, siltstones, rare fine sandstones;

*Portage and Marcellus*: shales, thin argillaceous limestones; and

*New Scotland*: argillaceous calcisiltites and calcilitites, calcareous shales.

Although these facies interfinger to a considerable extent within the Appalachian Plateau of New York, the coarsest facies are generally found to the southeast with successively finer facies to the north and west, as shown on charts by Rickard (1975) and on the Geological Map of New York (Fisher and others, 1970).

Two indices of bedrock lithology were computed by Coates (1971) for each of 13 watersheds in the Delaware River basin and 12 watersheds in the Susquehanna River basin, on the basis of measurements at outcrops. A "sandstone index" was defined as the percentage of total stratigraphic thickness made up of sandstones; a "massiveness index" was defined as the percentage of stratigraphic thickness made up of beds

in which the thickness between bedding planes was greater than 1 foot. Both indices were much higher in Delaware River basin watersheds than in most Susquehanna River basin watersheds. Within the Susquehanna River basin, these indices were highest along the Pennsylvania border and in the extreme eastern part of the basin, in agreement with the facies distribution presented by Rickard (1975).

These trends in bedrock lithology might be expected to affect groundwater discharge to streams, causing low flows to increase to the south and east across the Susquehanna River basin, for two reasons: (1) the open fractures through which groundwater flows tend to be larger and more persistent in massive sandstone layers than in shales, and (2) the grain size and lithologic composition of glacial till reflects those properties in the underlying bedrock. (For example, a notable decrease in carbonate content of till southward from the carbonate outcrops along the northern edge of the Appalachian Plateau was reported by Holmes (1952), Denny and Lyford (1963), and Moss and Ritter (1962).) Accordingly, one might expect till to become somewhat sandier and more permeable to the southeast, reflecting the increased percentage of sandstone in the bedrock. However, available evidence does not support the hypothesis that these trends cause low flow to increase southeastward. Coates (1971, table 6) showed that low flow correlated well with his sandstone index and massiveness index in 13 Delaware River basin watersheds, but not in 12 Susquehanna River basin watersheds. Furthermore, although spatial extent of valley-fill sand and gravel was strongly correlated with low flow in his Susquehanna River basin watersheds, no improvement in correlation resulted from adding the two bedrock indices as additional variables. In the present study, the spatial distribution of positive and negative residuals from several regression equations was compared with facies distribution in bedrock. If the inferred northwestward decrease in grain size of bedrock and till were a significant limitation on groundwater discharge to streams, one would expect equations that ignored these factors to underestimate low flow from southern and eastern watersheds and overestimate low flow from northern and western watersheds. No consistent southeastward increase in underestimation was obvious, however, nor were low flows in the few watersheds underlain by the Catskill (nonmarine) facies consistently underestimated. The distribution of positive and negative residuals generally seemed random, and some equations underestimated low flows in the western watersheds. The ratio of seasonal (August through October) runoff to annual runoff increases from northwest to southeast (fig. 4), but this ratio is not strongly correlated with either the sandstone index or the massiveness index computed by Coates (1971) for 12 Susquehanna River basin watersheds (correlation coefficient ( $R$ ) less than 0.58; Spearman rank correlation less than 0.56). Furthermore, as explained in the section "Seasonal Precipitation and Runoff," the trend in figure 4 might result from seasonal differences across the region in precipitation minus evapotranspiration, or

variation in the percentage of surficial sand and gravel among large watersheds.

## Watershed Relief and Slope

Topographic relief, average 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 favor greater storm runoff and (or) more rapid depletion of groundwater storage than gentle slopes (Paulsen and others, 1940, p. 440; Schumm, 1956; Zecharias and Brutsaert, 1988a, b). Although the low flows of several small upland tributaries in the Susquehanna River basin were found to decrease with increasing average land-surface slope (Pagano, 1987, p. 46), the low flows of larger streams consist chiefly of groundwater discharged from valley-fill sand and gravel, and are likely to increase with increasing watershed relief or slope because increased relief is associated with increased infiltration of upland runoff into the valley fill (Morrissey and others, 1988). Watershed slope and (or) relief indices have not proved consistently significant in regression equations designed to estimate low flow. For example, Sopper and Lull (1970) found main-channel gradient to be significant in estimating the duration of the longest period that includes 5 percent of annual streamflow in only two of four regions in the glaciated Northeast; moreover, its apparent influence was positive in one of these regions but negative in the other. This result implies that an increase in main-channel gradient (and hence in streamflow velocity, underflow, and alluvial grain size) does not inherently lengthen or shorten episodes of low streamflow; where main-channel gradient proved significant in their analyses, it may have functioned as a surrogate for untested watershed properties. Darmer (1970) included main-channel gradient in equations to estimate 7-day 10-year low flows over much of New York, but found relief to be more significant than main-channel gradient in the eastern Susquehanna and Delaware River basins; neither variable proved significant in estimating 7-day 2-year low flows anywhere in New York. On the other hand, Ku and others (1975) found valley gradient (similar to main-channel gradient) in the Susquehanna River basin to be significant in estimating 7-day 2-year low flows, but not 7-day 10-year low flows. Main-channel slope proved not to be significant in estimating 7-day 10-year and 7-day 2-year low flows in eastern New York (Barnes, 1986), Massachusetts (Male and Ogawa, 1982), and Wisconsin (Gebert, 1979). Comer and Zimmerman (1969) showed that a small till-covered basin with steep slopes in northern Vermont had more low flow per unit area than a larger adjacent basin with distinctly gentler slopes, but attributed the contrast to the greater extent of shallow poorly drained soils in the larger basin. Seasonal average runoff from eight till-covered upland watersheds

in northern Vermont in the summer and in the fall of an unusually dry year was not strongly correlated with the percentage of watershed area having a slope of 25 percent or more ( $R=0.67$  or less) (DeAngelis and others, 1984). Coates (1971), in a detailed geomorphic analysis of watersheds in the Susquehanna River basin of New York, computed four slope or relief properties: maximum watershed relief, stream relief (the average vertical distance between drainage divides and streams), average topographic slope in upland areas, and tributary stream gradients. Individually and as a group, these properties were not strongly correlated with 90-percent flow duration nor with base-flow recession rates ( $R=0.60$  or less). Given the inconsistent and inconclusive results of numerous studies cited above, measurements of watershed slope, relief, and gradient were not included in the data set for regression analysis in this study.

## Degree of Stream Regulation and Watershed Urbanization

Several of the larger streams in the Susquehanna River basin are affected by regulation, including interbasin transfers of water for purposes of power generation or canal operation, intermittent release of water stored in hydropower reservoirs, and seasonal release of water stored in lakes and reservoirs for recreational purposes. Most of these activities are mentioned in the notes accompanying published records of daily streamflow at gaging stations (U.S. Geological Survey, 1951–2005); information on regulation is also available from the Susquehanna River Basin Commission. Most of the streams analyzed for this report were unregulated. The effects of regulation on low flows of four streams were eliminated by adjustments based on reservoir-stage records or by deleting part of the watershed area, and for two other streams regulation was judged to be insignificant, as explained in footnotes to table 3.

Urbanization can have varied and complex effects on low flow. Buildings, pavement, roadside ditches on upland hillsides, and storm sewers can increase the volume and rate of storm runoff and thereby decrease the infiltration that sustains low flows. Groundwater development can lower the water table enough to locally eliminate groundwater discharge to streams. Conversely, some urban processes can augment low flows: buildings and pavement decrease evapotranspiration, leaky sewers and water mains can raise the water table locally, and discharge to streams from wastewater-treatment facilities continues during periods of low flow. Accordingly, and inasmuch as urbanization did not affect more than a small percentage of the land surface in any of the watersheds analyzed for this report during the period of data collection, the findings of this study are considered representative of nonurbanized conditions.

## Regression Analysis

Multiple regression analysis provides a means of testing and quantifying the significance of environmental factors that are hypothesized to affect the low flow of streams. Statistical tests measure the accuracy with which the equations estimate low flows and the relations among environmental variables and data values in those equations.

## Data Set

The regression analyses described in this report are based on data from 70 watersheds, including statistical indices of low streamflow and several watershed properties hypothesized to affect low streamflow.

## Low-Flow Statistics

Three commonly used statistical indices of low flow were computed for each watershed:

*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 a 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 a given year.

*90-percent flow duration (90%FD)*: The flow that was equaled or exceeded 90 percent of the time during the period of analysis.

The array of watersheds that make up the data set and the computation of low-flow statistics for each watershed were refined during four successive studies, including this one. The contribution of each study is summarized below.

Hunt (1967) adjusted records of streamflow through 1964 at all gaging stations in the Susquehanna River basin of New York to represent flow during 1931–60, and established many partial-record stations on small streams at which streamflow was measured occasionally during periods of low flow. He presented flow statistics for all stations, but made no attempt to regionalize low flow.

Ku and others (1975) noted that many older partial-record stations were located where tributaries begin to flow across sand and gravel in major valleys. Underflow in these localities is substantial and increases rapidly downstream (Ku and others, 1975, p. 14; Randall, 1978); hence, flow statistics for these stations represent only part of the yield from the watershed. Ku and others flagged these stations as unsuitable for extrapolating flow statistics to other sites on a per-square-mile basis, and excluded them from their subsequent regional analysis. They established many new stations on small streams at sites where underflow was expected to be minimal, and estimated flow statistics for each partial-record station through graphical correlation (Riggs, 1972) with one or two long-term



**Table 3.** Streamflow-measurement stations in the Susquehanna River basin of New York, and selected indices of low flow.

[Br, branch; Rd, road; SH, State Highway; FD, Flow Duration; flow values are in cubic feet per second; station locations are shown in figure 10; column-heading terms in brackets are variables used in regression equations shown in table 5 (at end of report)]

U.S. Geological Survey station number	Station name	Type of record <sup>1</sup>	Watershed area (square miles) [DAREA]	Lowest annual 7-day mean flow, April 1941–March 1971, with recurrence interval of		90-percent flow duration, October 1941– September 1971
				10 years [7Q10]	2 years [7Q2]	[90 % FD]
01496500	Oaks Creek at Index	G1 <sup>2</sup>	102	3.3	10.1	13.9
01496780	Cherry Valley Creek tributary at Roseboom	P	1.60	0.015	0.04	0.08
01496920	Shellrock Creek near Middlefield	P	5.45	0.03	0.13	0.21
01497500	Susquehanna River at Colliersville	G1 <sup>3</sup>				
01497530	Oak Creek near East Worcester	P	5.55	0.04	0.14	0.24
01497910	Center Brook at West Harpersfield	P	12.9	0.40	0.70	1.1
01497985	Kortright Creek at East Meredith	P	25.6	0.70	1.25	1.9
01498500	Charlotte Creek at West Davenport	G1	164.5	8.9	14.7	22.1
01499000	Otego Creek near Oneonta	G2	108	7.0	12.4	17.4
01499024	West Branch Otsdawa Creek near Otego	P <sup>4</sup>				
01499050	Flax Island Creek near Otego	G4	4.22	0.01	0.05	0.14
01499195	Brier Creek near Otego	P	6.96	0.03	0.11	0.26
01499300	West Branch Handsome Brook near Franklin	P	8.27	0.10	0.24	0.38
01499470	East Branch Handsome Brook near Franklin	G4	9.12	0.26	0.55	0.90
01500000	Ouleout Creek at East Sidney	G1 <sup>5</sup>	103	3.5	8.5	12.6
01500983	Center Brook near New Berlin	P	10.9	0.085	0.21	0.42
01501000	Unadilla River near New Berlin	G1	199	15.3	24.5	32.6
01501190	Wharton Creek at Pittsfield	P	84.4	7.0	11.0	14.0
01501500	Sage Brook near South New Berlin	G2	0.70	0.001	0.015	0.033
01501900	Butternut Creek near Garratsville	P <sup>6</sup>	16.0	1.1	1.8	2.5
01502000	Butternut Creek at Morris	G1	59.7	4.2	7.2	9.6
01502500	Unadilla River at Rockdale	G1	520	41.1	70.8	92.7
01502550	Guilford Creek at East Guilford	P	17.8	0.15	0.40	0.65
01502670	Big Brook above Bennetsville	P	25.4	0.32	0.68	1.2
01502899	Little Snake Creek above SH7 at Conklin	P	30.6	0.25	0.45	0.70
01505500	Canasawacta Creek near South Plymouth	G2	57.9	0.80	2.4	5.3
01506050	Bear Brook at Walker Corners	P <sup>4</sup>				
01507100	Five Streams near Smithville Flats	P	10.1	0.01	0.12	0.52
01507470	Red Brook at Smithville Flats	G4	7.06	0.017	0.06	0.13
01507500	Genegantslet Creek at Smithville Flats	G2	82.3	2.3	5.0	7.7
01507975	Muller Gulf Creek near Cuyler	G4 <sup>6</sup>	2.67	0.001	0.02	0.11
01508000	Shackham Brook near Truxton	G2	2.95	0.04	0.11	0.17
01508500	Albright Creek at East Homer	G2	6.80	0.07	0.30	0.57
01508803	West Branch Tioughnioga River at Homer	G3 <sup>7</sup>	71.5	9.3	15.5	20.5
01508803	West Br Tioughnioga River plus adjacent area	<sup>8</sup>	73.7	10.1	16.8	22.0

**Table 3.** Streamflow-measurement stations in the Susquehanna River basin of New York, and selected indices of low flow.

[Br, branch; Rd, road; SH, State Highway; FD, Flow Duration; flow values are in cubic feet per second; station locations are shown in figure 10; column-heading terms in brackets are variables used in regression equations shown in table 5 (at end of report)]

U.S. Geological Survey station number	Station name	Type of record <sup>1</sup>	Watershed area (square miles) [DAREA]	Lowest annual 7-day mean flow, April 1941–March 1971, with recurrence interval of		90-percent flow duration, October 1941– September 1971
				10 years [7Q10]	2 years [7Q2]	[90 % FD]
01509300	Hunts Creek at Marathon	P	10.8	0.15	0.40	0.6
01510000	Otselic River at Cincinnatus	G2	147	8.4	15.5	21.7
01511550	Halfway Brook near Triangle	P <sup>4</sup>				
01512780	Thomas Creek at Chenango Forks	P <sup>6</sup>	8.69	1.5	2.0	2.4
01512797	Castle Creek at Glen Castle	P	27.7	0.12	0.35	0.60
01513190	Little Choconut Creek at Stella	P	12.2	0.12	0.23	0.31
01513280	Finch Hollow Creek at Oakdale	P	3.96	0.04	0.07	0.095
01513829	Little Nanticoke Creek at Day Hollow Rd Owego	P	19.7	0.22	0.34	0.43
01513840	Pumpelly Creek at Owego	G4	8.59	0.03	0.06	0.08
01513862	East Br Owego Creek tributary at Harford Mills	P	5.77	0.05	0.10	0.16
01514000	Owego Creek near Owego	G1	185	9.9	14.6	17.2
01514298	Sulphur Springs Creek near Spencer	P	8.64	0.06	0.12	0.16
01514663	Willseyville Creek at Willseyville	P	8.49	0.75	1.15	1.3
01514820	Thorn Hollow Creek near Owego	P	4.13	0.025	0.055	0.065
01514839	Hunts Creek near Owego	P	6.78	0.027	0.055	0.075
01515580	Carter Creek near Cayuta	P	4.76	0.02	0.045	0.07
01515850	Langford Creek near Van Etten	P <sup>6</sup>	5.26	0.004	0.02	0.03
01520520	North Branch Glendenning Creek at Presho	P	9.27	0.05	0.14	0.18
01520990	Canisteo River at Bishopville	P	21.6	0.13	0.29	0.44
01521500	Canisteo River at Arkport	G1	30.6	0.60	1.1	1.5
01521610	Big Creek near North Hornell	P	16.8	0.25	0.60	0.90
01522000	Canisteo River at Hornell	G4 <sup>4</sup>				
01522300	Canacadea Creek near Almond	P	17.1	0.75	1.4	1.8
01522500	Karr Valley Creek at Almond	G2	27.4	0.11	0.55	0.95
01523500	Canacadea Creek near Hornell	G2 <sup>9</sup>	57.9	4.6	6.9	7.5
01524500	Canisteo River minus several tributaries	<sup>10</sup>	70.1	10.6	14.2	15.6
01524500	Canisteo River below Canacadea Creek Hornell	G2 <sup>9,11</sup>	145	11.6	16.5	19.0
01524550	Cunningham Creek near Canisteo	P	5.34	0.015	0.055	0.095
01525750	Tuscarora Creek tributary near Woodhull	G4	9.50	0.004	0.05	0.08
01525800	South Branch Tuscarora Creek near Woodhull	P	7.40	0.002	0.015	0.03
01526000	Tuscarora Creek near South Addison	G2	114	0.001	0.6	1
01526495	Mulholland Creek near Erwins	G4	5.06	0.01	0.04	0.055
01526980	Kirkwood Creek near Atlanta	G4	4.64	0.08	0.17	0.26
01527000	Cohocton River at Cohocton	G3	52.2	3.15	5.5	7.2
01527450	Castle Creek near Wallace	P	9.23	2.3	3	3.4

**Table 3.** Streamflow-measurement stations in the Susquehanna River basin of New York, and selected indices of low flow.

[Br, branch; Rd, road; SH, State Highway; FD, Flow Duration; flow values are in cubic feet per second; station locations are shown in figure 10; column-heading terms in brackets are variables used in regression equations shown in table 5 (at end of report)]

U.S. Geological Survey station number	Station name	Type of record <sup>1</sup>	Watershed area (square miles) [DAREA]	Lowest annual 7-day mean flow, April 1941–March 1971, with recurrence interval of		90-percent flow duration, October 1941–September 1971
				10 years [7Q10]	2 years [7Q2]	[90 % FD]
01527600	Goff Creek near Howard	P	17.9	1.4	2.3	2.8
01528000	Fivemile Creek near Kanona	G1	66.8	0.7	1.6	2.2
01530240	Gillette Creek near South Corning	P	3.77	0.01	0.03	0.04
01530450	Latta Brook at Horseheads	P	5.29	0.02	0.045	0.06
01530500	Newtown Creek at Elmira	G1	77.5	7.6	12.8	13.5

<sup>1</sup> Symbols defined as follows:

G1 Gaging station, continuous record 1941–71 or longer.

G2 Gaging station, record nearly complete 1941–71, 7-day low flows for missing years estimated by correlation prior to frequency analysis.

G3 Gaging station, less than 22 years of record 1941–71, low-flow indices estimated by correlation of 7-day low flows.

G4 Gaging station, about 2 years of record 1941–71, low-flow indices estimated by correlation of daily flows.

P Partial-record station, measurements chiefly 1966–68, low-flow indices estimated by correlation of individual measurements with daily flows at index stations.

<sup>2</sup> Low dam constructed at outlet of Canadatego Lake June 1964; thereafter, lake level was raised slightly each spring and lowered at end of summer. Records of lake stage 1968–71 and daily flow records indicated that this regulation had little or no effect on annual 7-day low flows of Oaks Creek.

<sup>3</sup> Deleted from data set. Regulation during periods of low flow, presumably by powerplant 0.25 mile upstream, typically resulted in several days of low, steady flow interrupted by single days of several times higher flow unrelated to rainfall. Daily flows between 30 and 50 ft<sup>3</sup>/s are rare in some years.

<sup>4</sup> This watershed was deleted from the data set compiled by Ku and others (1975, appendix B) because low-flow indices were poorly defined.

<sup>5</sup> Flow unregulated through 1949; thereafter, East Sidney Reservoir has regulated high flows but is reported not to significantly affect low flows.

<sup>6</sup> Correlation of this partial-record station with long-term index stations was rated as fair to poor because of large scatter of data. Accordingly, this station was deleted from the data set for several regression analyses, as described in text section “Per-square-mile equations to estimate 7-day 10-year low flow.”

<sup>7</sup> Continuous streamflow record 1967–68, 1973–85; 7-day low flows for these years as well as daily flows during base-flow periods were correlated with corresponding flows at three nearby stations.

<sup>8</sup> West Branch Tioughnioga River at Homer, station 01508803, plus a small area near Tully draining by way of numerous springs to Onondaga Creek (St. Lawrence River basin).

<sup>9</sup> Low flows unregulated through March 1965. Daily flows April 1965–March 1971 adjusted as necessary for changes in contents of Almond Reservoir, based on reservoir stage recorded daily at midnight (1965–86) or daily mean reservoir stage (1969–71) published by U.S. Geological Survey (1951–2005) and on capacity curve furnished by U.S. Army Corps of Engineers. Stages that were published only to the nearest 0.1 ft (1965–69) were adjusted as much as ± 0.04 ft during periods of changing stage to smooth the rates of change. In all but one of the 6 years of adjusted flows, the 7 consecutive days of lowest flow apparently occurred when adjustments were 0 to 40 percent of the recorded flow of Canacadea Creek, and on nearly the same days as the 7-day low flow of Canisteo River at Arkport (station 01521500).

<sup>10</sup> Canisteo River below Canacadea Creek at Hornell, station 01524500, minus stations 01521500, 01521610, and 01522500, and minus 13 square miles above Hornell municipal reservoirs on Carrington Creek.

<sup>11</sup> City of Hornell diverts an average of 3.5 cubic feet per second for municipal supply from Carrington Creek, a tributary upstream from station; sewage enters Canisteo River downstream from station. Runoff from the 13 square miles above Hornell reservoirs is presumed to be zero during periods of low flow, so this part of the watershed is excluded from drainage area and other data tabulated in this report.

gaging stations that were nearby and (or) geologically similar to the partial-record station. They selected a data set of 73 stations for regression analysis, from which they developed equations to estimate 7Q2, 7Q10, and 90%FD for 1931–60.

Eissler (1979) estimated 7Q2 and 7Q10 for measurement sites throughout New York. He discerned that data for three of the partial-record stations used by Ku and others (1975) were insufficient for reliable analysis. Eissler's results for other stations differed somewhat from those reported by Ku and others, because Eissler (1) analyzed periods of record through 1975, and did not adjust records to a single common time period; (2) used a computer program to calculate and plot least-squares linear regressions relating measured flows at each partial-record station to concurrent flows at three or more long-term gaging stations, and averaged the resulting estimates of low-flow characteristics; and (3) plotted the data from a few partial-record stations incorrectly because pertinent information was not readily available.

Statistical indices of streamflow for differing periods of years commonly differ substantially because of trends or random variations in climate. Therefore, the usual practice when comparing streamflow records for different time periods is to adjust all records to a common reference period, typically 30 years in length (Searcy, 1959; Cervione and others, 1982). As part of the present study, low-flow indices for 70 watersheds in the Susquehanna River basin were recomputed to represent a reference period extending from October 1941 through September 1971 for flow duration, and from April 1941 through March 1971 for low-flow frequency. Results are presented in table 3. A 30-year period ending in 1971 was selected because (1) it corresponds more closely than any other 30-year period to the actual period of record of most gaging stations in the Susquehanna River basin of New York, and (2) it corresponds to the period used in similar regionalizations of low-flow indices in Connecticut (Cervione and others, 1982) and in central New England (Wandle and Randall, 1994). Unregulated streams in the northeastern United States have 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 unregulated 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).

The data set includes 10 gaging stations with 30 years of continuous record through 1971, 11 stations lacking only a few years of record from this period, and 2 stations with long records obtained mostly after 1971. Flow duration for these 23 stations was determined by using computer programs developed by the U.S. Geological Survey (1985); incomplete records were adjusted as described by Searcy (1959). Low-flow frequency for incomplete records was adjusted by first correlating 7-day annual minimum flows at each short-term station with 7-day annual minimum flows for the same years

at nearby gaging stations, then selecting the index station that provided the largest product-moment correlation coefficient exceeding 0.8 (Matalas and Jacobs, 1964). The data for each pair of stations selected were plotted on logarithmic scales and lines of relation fitted by eye. If only a few years were missing from the base period, these graphs were used to predict 7-day low flows for the missing years, and the reconstructed array of data for the 30-year period was then analyzed by computer programs that used the log-Pearson method. If many years were missing, 7Q2 and 7Q10 for the 30-year period were estimated directly from the graphs. Records for a few stations were adjusted individually to correct for local circumstances, as explained in footnotes to table 3.

The data set also includes 8 gaging stations operated for only a few years and 37 partial-record stations. For these 45 stations, graphs were available from earlier studies (Eissler, 1979; Ku and others, 1975) that related instantaneous, daily, and (or) monthly flows at each short-term station to corresponding flows at several long-term index stations, including some just outside the Susquehanna River basin. These graphs were re-examined, and data points were corrected or added for several stations. Trend lines were fitted by eye (Riggs, 1972) with emphasis on the data points that represent relatively low flow, and were used to transfer indices of low flow from the index stations. For a few of these short-term stations, one graph showed much better fit than others and was used exclusively, but for most stations the results from two to four graphical correlations were averaged.

Low-flow indices and drainage area for each watershed used in regression analyses described in this report are presented in table 3. Locations of the measurement stations are shown in figure 10. Table 3 also lists (without data) five watersheds that were analyzed by Ku and others (1975) but deleted from the data set during the present study.

## Watershed Properties

The data set includes 13 watershed dimensions or properties measured as part of the study by Ku and others (1975) or the present study. All measured values are listed in table 4 (except drainage area, which is listed in table 3). Each term is explained below. Each name is followed by an abbreviation or acronym (in parentheses) that is used in table 4 and also in the regression equations presented in table 5 (at end of report). All determinations of area were made on U.S. Geological Survey 7.5-minute topographic maps, measured at least twice by planimeter or electronic digitizer.

*Drainage area (DAREA).* The drainage area upstream from each streamflow-measurement site is measured routinely by the U.S. Geological Survey; these areas have been published in annual data reports (U.S. Geological Survey, 1951–2005) and interpretive studies (for example, Eissler, 1979). Watersheds with drainage areas greater than 200 square miles were generally excluded from the data set for this report because measurement of some watershed properties would have been quite time-consuming, and because the

**Table 4.** Watershed properties compiled for regression analysis of low flow in the Susquehanna River basin of New York.

[Column-heading terms in capital letters are acronyms used in regression equations]

Watershed or station number (from table 3)	Area of surficial sand and gravel (square miles)	Mean annual precipitation (inches)	Mean annual runoff (cubic feet per second)	Ratio of August–October runoff to mean annual runoff	Area of wetlands (square miles) bordered mostly by		Dimensions of sand and gravel aquifer (feet) at measurement station		Stream gradient at measurement station (dimensionless)	Underflow factor (width × thickness × slope) (feet squared)	Forested area (square miles)
					Till	Sand and gravel	Width	Thickness			
STATION	SGAREA	MAP	MAQ	RUNRATIO	LKSWTL	LKSWSD	WIDTH	THICK	SLOPE	UNDER1	FOREST
01496500	19.5	42.17	174.1	0.323	1.55	6.6	3,500	10	0.0032	112	129.21
01496780	0.02	40.2	2.5	0.330	0	0	20	1	0.0571	1.142	10.64
01496920	0.19	39.9	8.4	0.330	0.05	0	100	5	0.04	20	12.27
01497530	0.3	39.5	8.4	0.335	0.1	0.04	50	1	0.0143	0.715	12.93
01497910	0.95	42.4	22.2	0.360	0.04	0	1,000	5	0.0073	36.5	110.93
01497985	1.35	41.5	42.4	0.355	0.31	0	100	1	0.0143	1.43	17.64
01498500	17.7	40.21	257.0	0.345	2.06	0.09	3,500	80	0.0014	392	171.26
01499000	16.5	40.62	172.0	0.330	0.73	0.68	2,500	20	0.001	50	135.83
01499050	0.15	39.3	6.3	0.335	0	0	50	2	0.02	2	0.9
01499195	0.26	39.5	10.5	0.335	0	0	20	1	0.02	0.4	1.7
01499300	0.2	44.1	15.3	0.360	0.01	0	200	8	0.0118	18.88	2.7
01499470	0.32	43.6	16.5	0.355	0.05	0	250	8	0.009	18	2
01500000	6.9	42.21	176.1	0.350	0.19	0.06	200	12	0.0063	15.12	40
01500983	0.49	40	16.9	0.315	0.03	0	50	2	0.01	1	2.38
01501000	39.2	40.75	318.9	0.310	1.09	1.48	2,400	25	0.0009	54	165
01501190	14.2	42.3	144.9	0.315	0.5	0.35	3,800	100	0.002	760	21.1
01501500	0.001	40.75	1.1	0.323	0	0	10	1	0.067	0.67	0.7
01501900	1.25	43.2	28.5	0.323	0.2	0.06	1,200	12	0.008	115.2	4
01502000	8	42.44	103.1	0.325	0.48	0.16	3,000	70	0.0038	798	25
01502500	91.5	41.31	854.7	0.320	4.07	2.86	4,500	10	0.0007	31.5	173
01502550	0.56	40.2	27.8	0.330	0.21	0.06	50	2	0.05	5	3.1
01502670	0.98	43.1	45.1	0.350	0.13	0	500	8	0.0111	44.4	6.3
01502899	0.75	40.5	48.5	0.360	0.35	0	500	8	0.0077	30.8	4.5
01505500	3.2	42.85	101.7	0.305	0.33	0.13	500	15	0.005	37.5	29
01507100	0.4	41.5	16.7	0.310	0.18	0.06	1,100	14	0.02	308	4.58

**Table 4.** Watershed properties compiled for regression analysis of low flow in the Susquehanna River basin of New York.

[Column-heading terms in capital letters are acronyms for these properties used in regression equations]

Watershed or station number (from table 3)	Area of surficial sand and gravel (square miles)	Mean annual precipitation (inches)	Mean annual runoff (cubic feet per second)	Ratio of August–October runoff to mean annual runoff	Area of wetlands (square miles) bordered mostly by		Dimensions of sand and gravel aquifer (feet) at measurement station		Stream gradient at measurement station (dimensionless)	Underflow factor (width × thickness × slope) (feet squared)	Forested area (square miles)
					Till	Sand and gravel	WIDTH	THICK			
STATION	SGAREA	MAP	MAQ	RUNRATIO	LKSWTL	LKSWSD	WIDTH	THICK	SLOPE	UNDER1	FOREST
01507470	0.46	41	11.4	0.315	0.41	0.11	250	5	0.025	31.25	4.1
01507500	4.7	41.93	138.4	0.315	1.57	0.77	2,000	25	0.0049	245	48
01507975	0.03	44.1	4.9	0.280	0	0	100	5	0.0222	11.1	2.5
01508000	0.06	43.15	5.3	0.275	0	0	10	1	0.0154	0.154	2.7
01508500	0.2	43.03	12.0	0.280	0	0	500	10	0.0174	87	1.4
01508803	22.2	41.2	116.9	0.278	0.02	2.58	5,000	40	0.0011	220	126.15
01508803	23.9	41.1	120.0	0.278	0.02	2.59	5,000	40	0.0011	220	126.97
01509300	0.67	43	19.1	0.310	0	0	250	8	0.0133	26.6	2
01510000	21.3	43.17	261.8	0.295	0.45	0.49	3,600	70	0.0026	655.2	64
01512780	2	36	10.9	0.340	0.01	0.27	4,000	55	0.0008	176	13.32
01512797	1	37	36.7	0.330	0.21	0.01	10	1	0.0063	0.063	12.5
01513190	0.33	36	15.3	0.340	0.05	0	10	1	0.01	0.1	5.8
01513280	0.1	36.2	5.0	0.340	0.01	0	50	2	0.02	2	1.01
01513829	0.7	36.8	25.8	0.340	0.12	0	50	3	0.0086	1.29	7.7
01513840	0.4	38.5	12.3	0.340	0.02	0	100	2	0.0214	4.28	3
01513862	0.15	41.3	9.5	0.315	0.01	0	100	2	0.02	4	1.5
01514000	23.8	39.7	282.1	0.330	0.2	0.3	3,200	20	0.0027	172.8	95
01514298	0.33	35.4	10.4	0.327	0.01	0	350	8	0.01	28	2.2
01514663	2.8	35.4	10.3	0.323	0	0.32	1,900	20	0.0015	57	18.45
01514820	0.05	36.8	5.4	0.340	0.04	0	10	1	0.0235	0.235	11.58
01514839	0.11	37.9	9.4	0.345	0	0	10	1	0.0154	0.154	13.12
01515580	0.07	32.9	4.9	0.310	0	0	200	7	0.0166	23.24	14.48
01515850	0.12	35.9	6.5	0.328	0	0	200	3	0.0375	22.5	12.13
01520520	0.1	32	8.9	0.280	0.03	0	10	1	0.02	0.2	13.24
01520990	0.83	35.2	25.8	0.275	0.07	0	800	10	0.0083	66.4	113.61

**Table 4.** Watershed properties compiled for regression analysis of low flow in the Susquehanna River basin of New York.

[Column-heading terms in capital letters are acronyms for these properties used in regression equations]

Watershed or station number (from table 3)	Area of surficial sand and gravel (square miles)	Mean annual precipitation (inches)	Mean annual runoff (cubic feet per second)	Ratio of August–October runoff to mean annual runoff	Area of wetlands (square miles) bordered mostly by		Dimensions of sand and gravel aquifer (feet) at measurement station		Stream gradient at measurement station (dimensionless)	Underflow factor (width × thickness × slope) (feet squared)	Forested area (square miles)
					Till	Sand and gravel	Width	Thickness			
STATION	SGAREA	MAP	MAQ	RUNRATIO	LKSWTL	LKSWSD	WIDTH	THICK	SLOPE	UNDER1	FOREST
01521500	2.9	35.29	36.7	0.272	0.07	0.01	450	10	0.009	40.5	116.86
01521610	0.88	33.4	17.8	0.270	0.02	0	10	1	0.0118	0.118	14.84
01522300	3.2	36.3	21.8	0.273	0.06	0.01	1,700	45	0.0111	849.15	17.27
01522500	2.4	35.16	32.6	0.272	0.06	0	400	20	0.0077	61.6	116.49
01523500	9.5	34.3	65.3	0.272	0.12	0.21	700	12	0.0057	47.88	129.95
01524500	24.3	32.2	68.2	0.272	0.08	0.71	3,300	25	0.0011	90.75	127.3
01524500	30.5	32.5	144.1	0.272	0.23	0.72	3,300	25	0.0011	90.75	165.48
01524550	0.12	32.7	5.4	0.268	0	0	10	1	0.0125	0.125	1.65
01525750	0.47	31.3	8.6	0.272	0.07	0	100	2	0.0133	2.66	12.01
01525800	0.33	31.7	6.9	0.275	0	0	150	2	0.0125	3.75	11.94
01526000	7.1	31.17	102.2	0.272	0.21	0	1,500	10	0.0054	81	129.17
01526495	0.07	32.1	4.9	0.285	0.01	0	100	5	0.0333	16.65	13.98
01526980	0.14	35.1	5.5	0.269	0	0	300	8	0.0143	34.32	11.86
01527000	14.8	33.81	57.0	0.271	0	1.79	4,800	80	0.0026	998.4	116.43
01527450	2.92	34.2	10.3	0.268	0.02	0.17	50	2	0.025	2.5	12.92
01527600	4.84	33.6	19.3	0.268	0.1	0.18	100	2	0.0111	2.22	14.58
01528000	6.7	35	78.7	0.268	0.57	0.15	2,600	20	0.0044	228.8	118.75
01530240	0.01	32	3.6	0.300	0.01	0	50	1	0.05	2.5	12.38
01530450	0.08	34.3	6.0	0.322	0	0	150	2	0.0166	4.98	12.4
01530500	16.8	34.35	87.6	0.322	0.02	0.14	9,000	16	0.0009	129.6	122

<sup>1</sup>Forested area measured manually using an electronic digitizer; for other watersheds, area estimated by point-counter methods.

regression equations developed in this study are likely to be applied to small watersheds or parts of large watersheds, so it seemed prudent to rely chiefly on data from watersheds of comparable size.

*Area of surficial sand and gravel (SGAREA).* The extent of surficial sand and gravel was inferred largely from topographic maps and county soils maps; the maps were augmented in some localities by well records (Randall, 1972) and examination of earth materials exposed in excavations and roadcuts. SGAREA includes deposits of glacial meltwater and deposits of postglacial streams in alluvial fans and flood plains. Small areas along the valley sides in some watersheds could not be classified with confidence as sand and gravel or as till; therefore, maximum and minimum plausible areas of sand and gravel were delineated and an intermediate value was selected for use in this study. The areas of sand and gravel reported in table 4 include any wetlands within the boundaries of the sand and gravel.

*Area of till (TLAREA).* Till area was computed by subtracting the measured area of sand and gravel from the drainage area. Two watersheds, Fivemile Creek (01528000) and Newtown Creek (01530500, fig. 10), contain small areas of surficial silt and (or) clay that were deposited in late-deglacial lakes but not later capped by outwash and (or) alluvium; these areas were included with till in the regression analysis.

*Mean Annual Precipitation (MAP).* A value for each watershed was estimated by overlaying an outline of the watershed on contours of mean annual precipitation for 1951–80 (fig. 3A).

*Mean Annual Runoff (MAQ).* Values for long-term gaging stations were calculated from daily-flow records that had been adjusted to 1951–80, as explained by Randall (1996). Values for short-term gaging stations and partial-record stations were estimated by overlaying an outline of the watershed on contours of mean annual runoff for 1951–80 (fig. 3B).

*Ratio of mean August–October runoff to mean annual runoff (RUNRATIO).* A value for each watershed was estimated by interpolation of watershed location among the contours in figure 4, which denote the ratio of mean runoff from August through October to mean annual runoff, and are based on 1938–63 data from a few large watersheds.

*Area of wetlands underlain by sand and gravel (LKSWSL) and by till (LKSRTL).* Area of lakes and area of swamps shown on U.S. Geological Survey 7.5-minute topographic maps were measured separately, then aggregated as wetland areas for regression analysis. Wetlands mostly bordered by till were compiled separately from wetlands mostly bordered by sand and gravel. Where a large lake or swamp occupies the entire valley floor and is bordered upstream and downstream by stratified sand and gravel, but bordered on the sides mostly by till or bedrock, the entire area of that lake or swamp was included in the area of wetlands underlain by sand and gravel (LKSWSL), as well as in the area of surficial sand and gravel (SGAREA).

*Width of sand and gravel at measurement station (WIDTH).* The width of surficial sand and gravel was measured along a line perpendicular to the valley axis through the streamflow-measurement station.

*Thickness of sand and gravel (THICK).* The average saturated thickness of sand and gravel along the line used to measure width was estimated from well and test-boring records (where available) and interpretation of topographic maps. A minimum value of 1 foot was assigned where the stream is incised in bedrock or till; the maximum value allowed was 100 feet.

*Stream gradient (SLOPE).* The average stream gradient near the streamflow-measurement station was calculated from the topographic map by dividing the contour interval by the distance between successive contours where they intersect the stream, measured parallel to the valley axis. The stream gradient is presumed to be nearly identical to the downvalley component of the water-table gradient.

*Underflow factor (UNDERI).* This term was computed as the product of width of sand and gravel, thickness of sand and gravel, and stream gradient, and therefore encompasses all factors other than hydraulic conductivity that would control the rate of underflow downvalley through sand and gravel, according to Darcy's Law.

*Forested area (FOREST).* Areas with green overprint on U.S. Geological Survey 7.5-minute topographic maps were considered to be forested. The total area of green overprint in each of 38 watersheds was measured manually using an electronic digitizer as part of this study. The remaining 32 watersheds were assigned values previously compiled by Darmer (1970) or R.D. MacNish (U.S. Geological Survey, written commun., 1968), who counted whether forested or nonforested land lay beneath nodes of a grid placed over a map of the watershed. Results obtained from the two methods did not match closely, but agreed within 25 percent in 20 of 24 watersheds where both methods were used. How well the topographic maps represent forested area during the period of study is uncertain, inasmuch as many farm fields or pastures in the Susquehanna River basin were abandoned and allowed to revert gradually to brush and trees between 1941 and 1971. Because most of the topographic maps used were produced in the 1950s and 1960s, however, they may represent nearly average conditions for 1941–71.

## Format of Regression Equations

Three equation formats were used during this study:

1. Low-flow indices in cubic feet per second as a function of watershed properties expressed in common units of measurement: for example, area in square miles, mean flow in cubic feet per second. 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. These equations are referred to herein as “standard-value equations.”

2. Low-flow indices in cubic feet per second per square mile as a function of measured values of watershed properties divided by watershed area (referred to as “per-square-mile equations”). With this format, all watersheds in the data set have equal weight regardless of size.
3. Logarithms of low-flow indices as a function of the logarithms of measured watershed properties (referred to as “logarithmic equations”). Logarithmic transformations have commonly been applied to data sets to generate an array of values that are normally distributed and also consistent with other assumptions on which most statistical tests of estimation accuracy are based.

Many equations were tested, 18 of which are presented in table 5 (at end of report), along with statistical measures of their predictive accuracy and integrity. Each is described on the following pages. The rationale for presenting these several equations is not to offer alternative options for estimation of low streamflow, but rather to demonstrate how incorporating particular explanatory variables or data manipulations affect predictive accuracy and integrity. Eight standard-value equations for estimation of 7Q10 are described first, then three per-square-mile equations and a logarithmic equation for estimation of 7Q10, and finally six equations for estimation of 7Q2 or 90%FD.

## Standard-Value Equations to Estimate 7-Day, 10-Year Low Flows

This section first demonstrates how the predictive accuracy of standard-value regression equations is improved by successively incorporating sand-and-gravel area, wetland area, and water availability, and then considers the effects of several alternative variables and data manipulations.

The first five equations demonstrate how predictive accuracy and statistical integrity are progressively improved by incorporating sand and gravel area, wetland area, and mean annual runoff (a measure of the amount of water available) as explanatory variables. Equations 6 to 8 illustrate the effects of three data manipulations that offer some promise for further improvement, although only equation 8 is of practical use at present.

### Effect of Sand-and-Gravel Area and Wetland Area

Equation 1 (table 5, at end of report) indicates that area of sand and gravel is positively correlated with 7Q10 low flow, and that it explains most of the variation in low flow. As explained earlier, the area of sand and gravel in each watershed, as compiled in table 4, includes any wetlands that are bordered and underlain by sand and gravel. Equation 2 removes these wetland areas from the total area underlain by sand and gravel, resulting in significantly

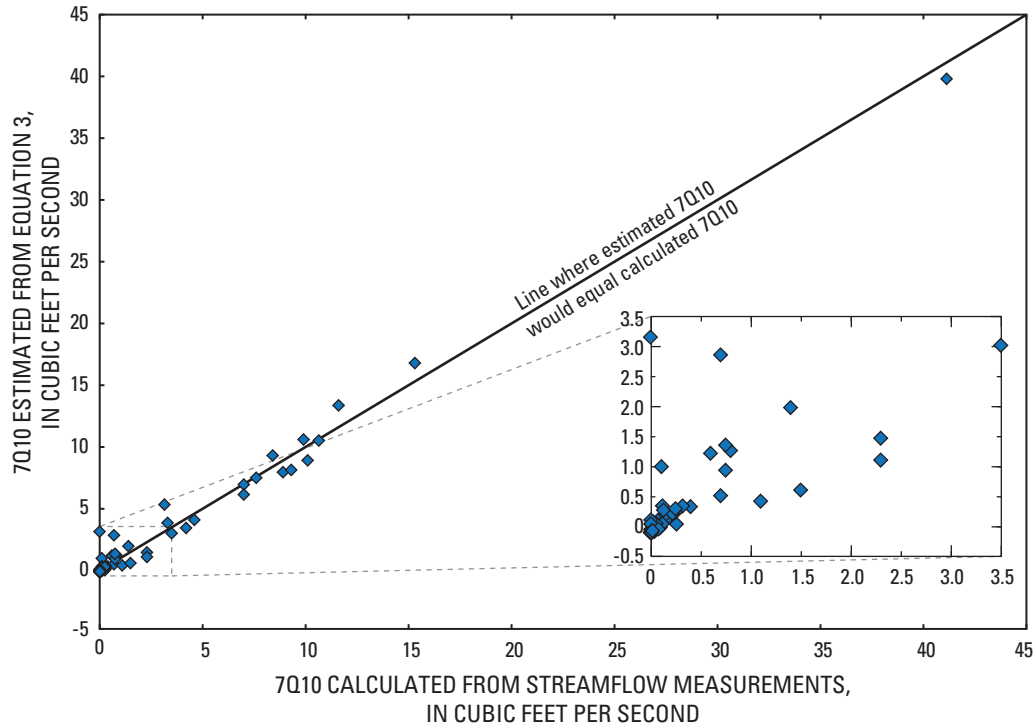
better estimation of 7Q10 low flow. Equation 3 separates the two mechanisms whereby wetlands diminish 7Q10 low flow. The first independent variable in equation 3, SGAREA minus LKSWSD, accounts for the failure of precipitation on wetlands to contribute to the amount of recharge that is stored in sand and gravel, because the water table is virtually at land surface in wetlands. The second independent variable in equation 3, LKSWSD, accounts for the intense evapotranspiration in wetlands that directs to the atmosphere some groundwater that was recharged nearby—water that would have become streamflow if the wetlands were absent. The regression coefficient for LKSWSD in equation 3 is smaller than the coefficient for SGAREA-LKSWSD, which implies that evapotranspiration has a smaller negative effect on 7Q10 low flow than the lack of recharge. Figure 11 depicts the accuracy of 7Q10 estimates from equation 3.

Equation 4 incorporates the same watershed variables as equation 3, but is applied to a slightly smaller data set created by deleting five partial-record stations whose low-flow indices may have been imprecisely calculated. (The purpose of this deletion is explained farther on, in the section on per-square-mile equations.) Most statistical test results for equations 3 and 4 are nearly identical.

One watershed has much larger low flow than the rest, as indicated in figure 11. That watershed, Unadilla River at Rockdale (01502500, fig. 10), has a drainage area of 520 square miles, whereas all other watersheds drain less than 200 square miles. The one large watershed was included in the data set because watershed properties for 343 of its 520 square miles had already been compiled to describe three smaller watersheds; therefore, the remaining 177 square miles were less than the 200-square-mile cutoff limit for data compilation. Unadilla River at Rockdale has the largest influence of any individual watershed on several of the standard-value equations, as indicated by the DFFITS statistic (table 5, column K), but it falls nearly on the trend of other watersheds in figure 11 (and similar graphs for other equations) and therefore is not unduly influential.

### Effect of Amount of Water Available

Equation 5 is the best of many standard-value equations tested that incorporates mean annual runoff (MAQ) as an index of annual water input to a watershed. Mean annual runoff had a positive sign, as expected, when incorporated without modification in preliminary trial equations. Its significance was weak, however (the  $p$ -t statistic in column H of table 5 (at end of report) was 0.05 or more) unless the equation also included till area, a term that was strongly collinear with mean annual runoff and that had a negative sign, which was illogical in that the contribution of groundwater to streams at 7Q10 low flow from upland till is small, but cannot be negative. (One such preliminary equation was presented by Randall and Johnson (1988, table 1).) Collinearity can be eliminated from some regression equations by subtracting a constant from a collinear independent variable, thereby



**Figure 11.** Correlation of 7Q10 calculated from streamflow measurements with 7Q10 estimated using equation 3. Inset graph expands scale at lower left corner of main graph. All data and analyses pertain to the Susquehanna River basin in New York. 7Q10, annual lowest streamflow averaged over 7 consecutive days that has a 1 in 10 chance of not being exceeded in a given year. See table 5 (at end of report) for equation 3; terms therein are defined in the “Data Set” section of the report.

resetting the datum from which that variable is measured (D. Helsel, U.S. Geological Survey, written commun., 1988). To test this approach, the mean annual runoff in cubic feet per second from each watershed was divided by drainage area in square miles, to obtain values that ranged from 0.90 to 1.85 cubic feet per second per square mile. Next, several arrays of 70 reset mean annual runoff values were generated by subtracting successively larger constants from the per-square-mile runoff value for each watershed, then multiplying the results by drainage area to convert back to units of cubic feet per second. Each data array thus generated was tested as an independent variable in a regression equation, along with area of sand and gravel, area of wetlands in sand and gravel, and till area. The collinear till-area term was eliminated (regression coefficient of zero) and the statistical indices reported in table 5 were optimized when the amount subtracted was 1.85 cubic feet per second per square mile, just large enough to convert the reset mean annual runoff term to values of zero or less for all watersheds. Therefore, this optimized term represents the spatial variation in mean annual water input per unit area, multiplied by drainage area to convert the term to units of cubic feet per second and with the datum shifted to place zero at one end of the array of values. Equation 5 (table 5) incorporates this term, and estimates 7Q10 substantially better than equations 1 through 4. The regression coefficient for sand-and-gravel area in equation 5

apparently represents groundwater discharge to streams per square mile of sand and gravel in localities where mean annual runoff is 1.85 cubic feet per second per square mile, and the mean annual runoff term in equation 5 serves to adjust the estimate of 7Q10 for each watershed downward to the extent that mean annual runoff from that watershed is less than 1.85 cubic feet per second per square mile.

Equation 6 (table 5) is similar to equation 5, but the mean annual runoff term incorporates the ratio of mean August-through-October runoff to mean annual runoff in large watersheds (fig. 4). In effect, this modification converts the term to an estimate of seasonal August-through-October runoff. The datum was reset by the same mechanism used in equation 5, although the constant subtracted from the estimated values of watershed runoff is only 0.6 cubic foot per second per square mile because the magnitude of August-through-October runoff is about one-third that of mean annual runoff. Nearly all statistical indices for equation 6 are slightly better than those for equation 5 (table 5). However, the same modification resulted in only insignificantly small statistical improvements in equations (described farther on) that estimate low flow per square mile—which probably means that only in some of the larger watersheds in the data set was low flow more closely correlated with estimated seasonal runoff than with annual runoff. Furthermore, as pointed out earlier, the reason for the spatial variation in the ratio of seasonal

runoff to annual runoff depicted in figure 4 is uncertain, and the contours in figure 4 are not tightly controlled by the sparse data values, all from large watersheds. Accordingly, the use of figure 4 and equation 6 to estimate low flow is probably not warranted. Nevertheless, the statistical indices for equations 5 and 6 suggest that some measure of average seasonal amount of water available (computed as precipitation minus evapotranspiration rather than by interpolation among measured runoff values) might ultimately prove useful as an index of water input in regression analysis to estimate 7Q10.

### Other Variables and Data Manipulations Tested

Equation 7 (table 5, at end of report) is the best of several equations formulated to test the hypothesis that low-flow yield from sand and gravel, and evapotranspiration from wetlands, may each be proportional to the product of the area of that watershed property multiplied by mean annual runoff per unit area. Several fractional-power transformations of mean annual runoff were tested, because such transformations can improve the fit of regression equations by making relations more linear and residual variations constant (Montgomery and Peck, 1982, p. 25). The 0.4 power transformation in equation 7 implies that only a fraction of any increase in water input to a watershed translates to increased low-flow yield from sand and gravel and increased evapotranspiration loss from wetlands. The statistical tests of estimation accuracy of equation 7 are superior to those of equations 1 through 4 (table 5), again confirming that water input (as represented by mean annual runoff) significantly influences low flow, but these tests are inferior to those of equations 5 and 6.

Four additional watershed properties were tested during the development of standard-value equations 2 through 7, but were found to be insufficiently significant:

1. *Till area.* The low flow of streams in New England is best estimated by regression equations that include till area as well as sand-and-gravel area as explanatory variables (Cervione and others, 1982; Wandle and Randall, 1994), even though these two variables are not mutually independent (their sum is equal to drainage area). The equations express low flow as the sum of a large yield per unit area of sand and gravel plus a much smaller yield per unit area of till, which is entirely consistent with conventional concepts of groundwater discharge to streams. None of the equations presented in table 5 for estimating 7Q10 in cubic feet per second could be improved by including till area, however. This result seems to imply that groundwater discharge to streams from till-covered uplands in the Susquehanna River basin is negligible at 7Q10 low flow. Indeed, the low flows of upland tributaries in the Susquehanna River basin might be expected to be smaller than those of similar streams in New England, because the shale bedrock and silty clay tills derived from shale in the uplands of the Susquehanna River basin are likely to be less permeable than the metamorphic bedrock and coarser tills that predominate in
2. *Wetlands in till.* The area of wetlands in till proved significant in a few trial equations, but this variable, and the total area of wetlands in the watershed, generally were less significant than the area of wetlands in sand and gravel in regression equations. This result is consistent with the observation that many wetlands in till-mantled uplands are in headwater localities where streamflow often ceases in dry weather; hence, evapotranspiration from these wetlands cannot further decrease low flow.
3. *Mean annual precipitation.* This term proved to be less significant than mean annual runoff in regression analysis.
4. *Forested area.* Several studies (Weeks and Stangland, 1971; Lewis, 1958; Trousdell and Hoover, 1955) have shown that evapotranspiration is greater from forested area than from comparably located grassland or cropland. No significant differences were detected in other studies, however (Hibbert, 1969; Schneider and Ayer, 1961). The FOREST term, compiled from U.S. Geological Survey topographic maps as described in the section "Watershed Properties," did not significantly improve any equation in which it was tested.

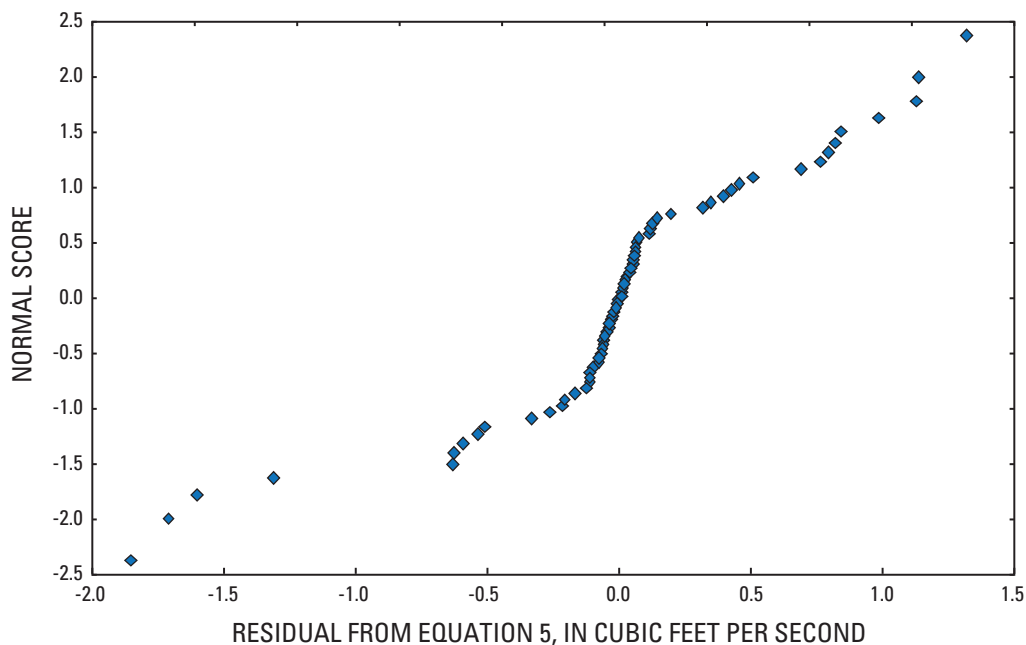
*Trimming the Data Set.* Evaluation of the statistical integrity of equations 1 through 7, explained below, led to a decision to apply the best of these equations to a smaller data set that provided a more nearly normal distribution of

residuals. (Residuals are the differences between the observed values of the dependent variable and the values estimated by the regression equation.) One of the assumptions on which least-squares linear regression analysis is based is that residuals are normally distributed (Iman and Conover, 1983, p. 368). Residuals from equations 1 through 7 are not normally distributed, as indicated by the coefficient of correlation of residuals with normal scores (table 5, column L, at end of report; normal scores are defined in table 5 explanation). A correlation coefficient of 0.983 or greater is consistent with a normal distribution for a data set of 70 stations, at a significance ( $\alpha$ ) level of 0.05 (Looney and Gullidge, 1985). The departure of residuals from a normal distribution is illustrated in figure 12 by the departure of data points from a straight line. If residuals are not normally distributed, the parametric statistical tests used to gauge the estimation accuracy of regression equations are not precisely valid (Ryan and others, 1985, p. 177). Therefore, the coefficient of determination and the standard error reported in table 5 for equations 1 through 7 should be considered only approximately correct.

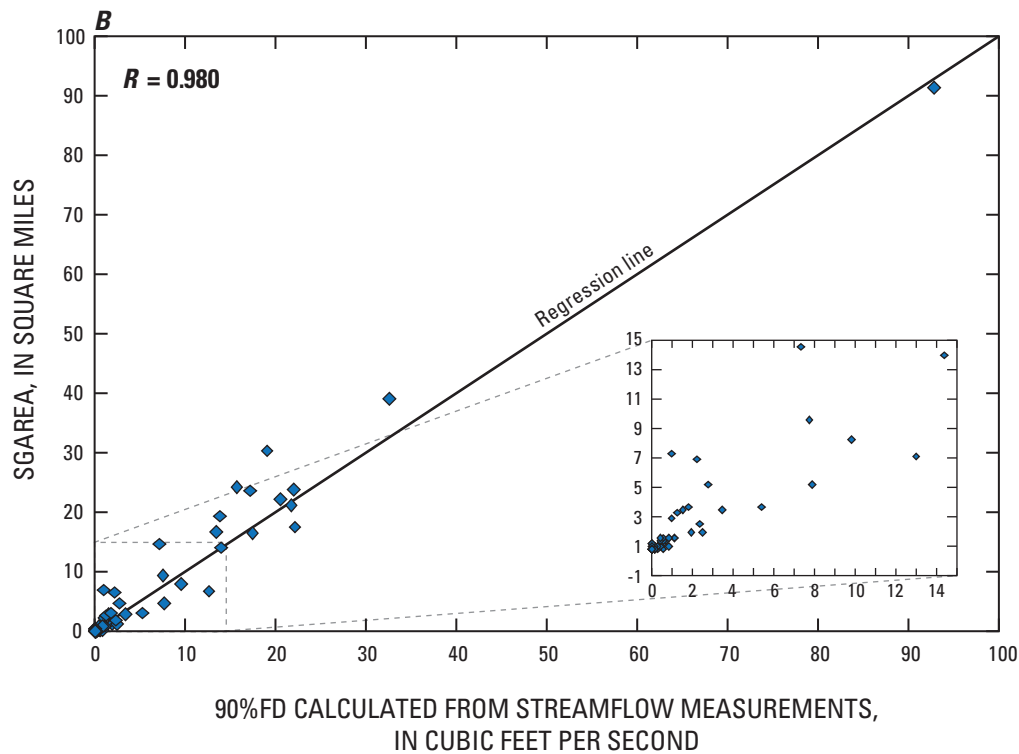
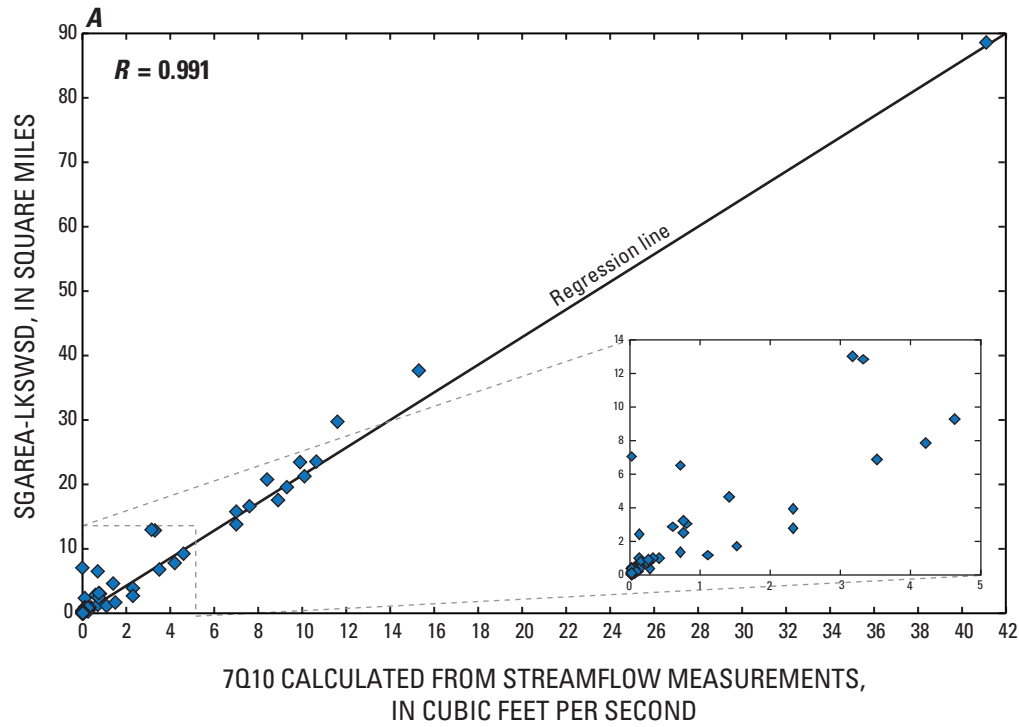
A method of testing the slope of a regression (such as the relation between observed low-flow indices and low-flow indices estimated by regression equations) is explained by Iman and Conover (1983, p. 381). This nonparametric method, which is valid whether or not residuals are normally distributed, is based on correlation between the ranks of estimated low-flow indices and the ranks of the residuals. If the absolute value of the Spearman rank correlation coefficient

is less than 0.235 (for 70 stations) or 0.244 (for 65 stations), the slope of the regression line is accepted as correct. Equations 5 through 7 easily meet this criterion, as indicated in the last column of table 5, as do several other equations described farther on.

Another assumption of least-squares linear regression is that the variance of the residuals is constant over the range of estimated values (Iman and Conover, 1983, p. 368). Casual examination of graphs in which low flows calculated from streamflow measurements are plotted against individual watershed properties (fig. 13) or against low flows estimated with a particular equation (fig. 11) suggests that this criterion may be met, in that departures of data points from the equality line or regression line seem to be of similar magnitude along most of the line. This is certainly true of the large watersheds, but the data set also includes many smaller watersheds, nearly all with flow measured at partial-record stations, that have small low flows and correspondingly small residuals. Because these many small watersheds with small residuals are clustered at one end of the array of low flows, the data set fails to meet a test for constant variance of residuals (Gary Tasker, U.S. Geological Survey, written commun., 1981), in which the values of the independent variables in a regression equation are correlated with the absolute values of residuals from the regression. For equations 1 through 7, the Spearman rank correlation coefficients for such correlations ranged from 0.45 to 0.75, and the null hypothesis of no correlation was rejected at all significance levels.



**Figure 12.** Distribution of residuals from equation 5. All points would fall on a straight line if their distribution were exactly normal. See table 5 (at end of report) for equation 5 and for explanation of normal scores.



**Figure 13.** Correlation of low-flow indices calculated from streamflow measurements *A*, 7Q10 and *B*, 90% FD, with selected watershed properties. Inset graph expands scale at lower left corner of main graph. All data and analyses pertain to the Susquehanna River basin in New York. 7Q10, annual lowest streamflow averaged over 7 consecutive days that has a 1 in 10 chance of not being exceeded in a given year; 90% FD, the flow that was equaled or exceeded 90 percent of the time during the period of analysis; SGAREA, area of surficial sand and gravel, including any overlying lakes or swamps; LKSWSD, area of lakes and swamps underlain by sand and gravel; R, Pearson product-moment correlation coefficient

The departure of standard-value equations from normal distribution and from constant variance of residuals was decreased greatly by deletion of 38 small watersheds from the data set. The remaining 32 watersheds include all long-term continuous-record gaging stations and all partial-record stations that represent watersheds larger than 30 square miles or watersheds in which more than 18 percent of the area is underlain by sand and gravel. Analysis of this smaller data set yielded equation 8 (table 5), which is similar to equation 5 in that all variables and regression coefficients are the same—probably because the large watersheds that predominate in the 32-station data set on which equation 8 is based include all 14 stations that are moderately influential in equation 5, as indicated by DFFITS values greater than 0.3. Mean watershed size is larger for equation 8 than for equation 5; therefore, standard error is also larger in cubic feet per second, but is smaller as a percentage of mean 7Q10. The distribution of residuals from equation 8 is normal, as evidenced by the correlation of residuals with normal scores (table 5, column L), which exceeds the minimum limit of 0.966 for normal distribution at the 5-percent significance level in a data set of 32 stations (Looney and Gullidge, 1985). The variance of residuals easily passes the test described in the previous paragraph at a 5-percent significance level. Therefore, equation 8 is conceptually and statistically acceptable, at least with respect to watersheds larger than about 30 square miles, which make up 72 percent of the data set.

### Per-Square-Mile Equations to Estimate 7-Day, 10-Year Low Flows

The large watersheds were more influential than the more numerous small watersheds in the development of standard-value equations 1 through 8 in table 5 (at end of report), but the chief practical application of regression equations is to estimate low flow from small ungaged watersheds, or small increments within larger watersheds between measurement sites and sites where an estimate of low flow is desired. Therefore, several equations were tested in which all variables were divided by watershed area to give equal weight to all watersheds, regardless of size. Three per-square-mile equations for estimating 7Q10 are presented in table 5 (equations 9, 10, and 11).

### Comparison of Per-Square-Mile Equations with Standard-Value Equations

Some of the statistical indices of estimation accuracy (table 5, at end of report) computed by the regression program for the per-square-mile equations are not directly comparable to indices computed for the standard-value equations because the variables are scaled in different units. To permit quantitative comparison, each per-square-mile equation was converted to standard values, as follows: (1) each predicted value of low flow in cubic feet per second per square mile was multiplied by watershed area; then (2) the resulting estimated

low flows in cubic feet per second were correlated with observed low flows in cubic feet per second. The statistical indices of estimation accuracy for this correlation are reported in parentheses in table 5.

The statistical indices of estimation accuracy reported for equation 9 after conversion to standard values are similar to those reported for equation 2, in which standard values of the same variables were used directly in the regression. Therefore, per-square-mile equations seem to be at least comparable to standard-value equations in estimation accuracy, despite the different values of some statistical indices that result from different numerical expression of watershed properties.

### Evaluation of the Data Set and Equations

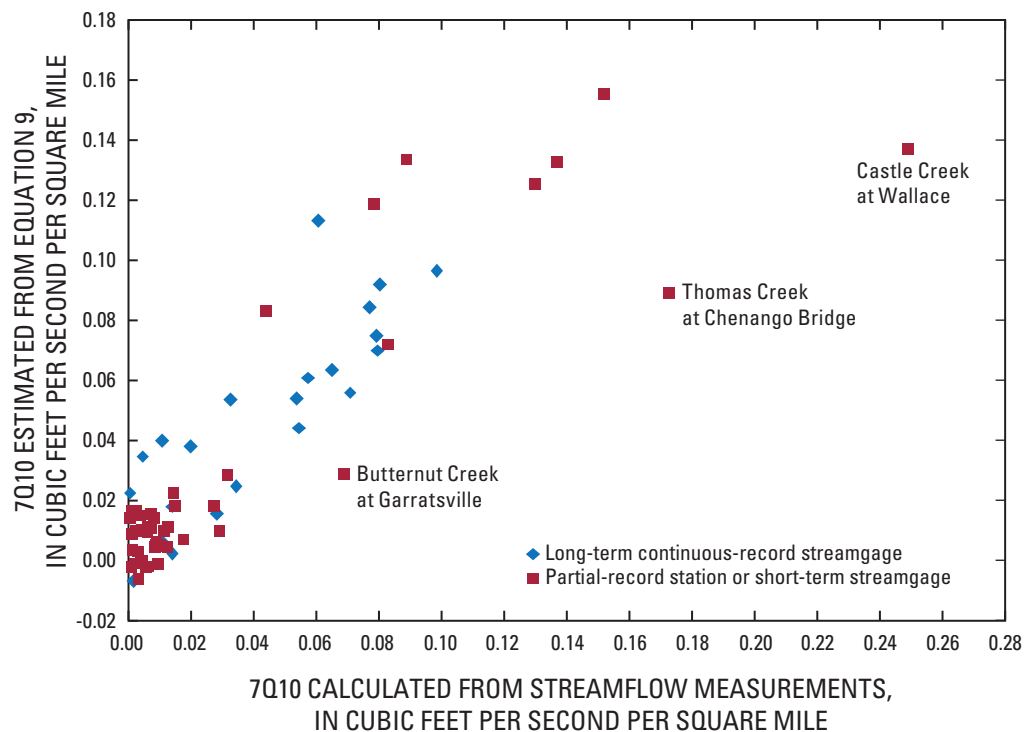
Graphs in which 7Q10 low flow per square mile calculated from streamflow records is plotted against 7Q10 estimated using equation 9 showed appreciable scatter (fig. 14). Two stations plot far from the trend of other data in figure 14 because the estimated low flows for these stations are much smaller than the calculated low flows. Both stations, Thomas Creek at Chenango Bridge and Castle Creek at Wallace, are partial-record stations on streams draining relatively small watersheds (about 9 square miles in area) that contain an unusually large proportion of sand and gravel. As might be expected, the DFFITS influence statistic shows these two watersheds to have greater influence on equation 9 than any other watersheds in the data set. Recalculation of 7Q10 low flow for these two watersheds and remapping the surficial geology of Castle Creek watershed did not result in significant changes in the input data. A hypothesis that low flow per square mile of sand and gravel might be greater in small watersheds than in large ones was tested by mathematical manipulations of the input data, and rejected. A re-evaluation of the data set, explained in the following paragraphs, led to a decision to delete Thomas Creek, Castle Creek, and three other small watersheds.

The graphical correlations used in previous studies to calculate low-flow indices at partial-record stations in the Susquehanna River basin were reviewed at the start of this study. Four stations, Butternut Creek near Garratsville (station 01501900), Muller Gulf Creek near Cuyler (station 01507975), Thomas Creek at Chenango Bridge (station 01512780), and Langford Creek at Van Etten (station 01515850), were rated as having large scatter of data and rather poorly defined correlations, but were retained in the data set to ensure inclusion of a wide variety of watersheds. Imprecise calculation of low-flow indices for these stations may, however, explain why regression equations could not closely reproduce the indices for Thomas Creek, nor for Butternut Creek, which also departs from the general trend as illustrated in figure 14. Thomas Creek drains a broad valley that is filled with stratified drift, much like many broad valleys in the Susquehanna River basin except that it is not traversed by a major river, a condition present in only a few scattered localities (Randall and others, 1988, fig. 2; Reynolds

and Brown, 1984; Randall, 1977). Consequently, the ratio of valley-floor width at the measurement site to watershed area is larger for Thomas Creek than for any other watershed in the data set, and one might expect underflow to constitute an unusually large fraction of the yield from Thomas Creek watershed. If this were the case, regression equations that ignore underflow would overestimate low streamflow for this watershed, and yet equation 9 does the opposite (fig. 14). Furthermore, this watershed had a large influence on regression equations that included a variable intended to represent underflow, causing other variables to have a sign opposite from what was expected and was achieved after deletion of this watershed from the data set. Either the 7Q10 low flow calculated from correlation of streamflow measurements at Thomas Creek is too large, or underflow here is unusually small for unknown reasons.

The graphical correlation of measured flows at Castle Creek at Wallace (partial-record station 01527450) with concurrent flows at two nearby gaging stations exhibited less scatter than the correlations cited in the previous paragraph. Nevertheless, six flow measurements in Castle Creek spaced over 22 days in 1967 ranged from 4.29 to 4.03 cubic feet per second, while daily flows on the same days at the nearby gaging stations ranged from 14 to 8.6 and from 13 to 7.4 cubic feet per second. The much smaller fluctuation in the flow of Castle Creek implies a significant difference in

rates or timing of recession that could lead to miscorrelation of the frequency of low flows. The watershed of Castle Creek (01527450, fig. 10) lies within a distinctive small upland terrane between the Cohocton and Canisteo Rivers in Steuben County, in which abundant stratified drift was deposited when a tongue of ice blocked the Cohocton valley at Bath (Randall, 1986b). The stratified drift was deposited as deltas in lakes graded to saddles on modern divides, but has been deeply incised by postglacial streams, which in some places have cut entirely through the stratified drift to till or bedrock. The incised stratified drift is mostly unsaturated, but springs and seeps near its base contribute to streamflow. Castle Creek and nearby Goff Creek are the only watersheds in the data set in which this type of terrane is prominent; in most upland watersheds, sand and gravel is limited to the alluvium on narrow valley floors. In New England, sand and gravel perched above stream grade on upland hillsides or capping fine-grained sediment in valleys apparently generates moderately greater low flows per square mile than coarse-grained stratified drift or alluvium on the floors of major valleys (Wandle and Randall, 1994, p. 24). Thus, the unusually large 7Q10 low flow per square mile calculated for Castle Creek might be an accurate representation of an uncommon upland terrane, or the result of an erroneous correlation of streams in which recession rates and timing differ.



**Figure 14.** Correlation of 7Q10 per square mile calculated from streamflow measurements with 7Q10 per square mile estimated using equation 9. All data and analyses pertain to the Susquehanna River basin in New York. 7Q10, annual lowest streamflow averaged over 7 consecutive days that has a 1 in 10 chance of not being exceeded in a given year. See table 5 (at end of report) for equation 9; terms therein are defined in the “Data set” section of the report.

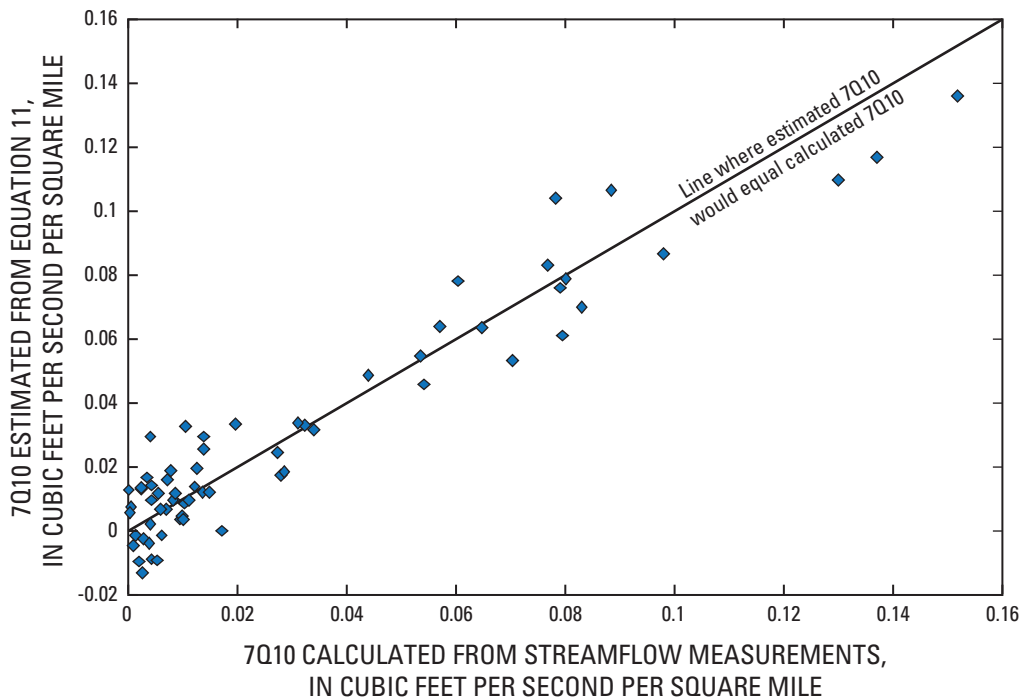
Accordingly, Castle Creek at Wallace and the four partial-record stations for which low-flow correlations were poorly defined were deleted from the data set used for per-square-mile regression analysis. Equations 10 and 11 in table 5 (at end of report) are based on this smaller data set of 65 stations.

Three per-square-mile equations are presented in table 5. Equations 9 and 10 show that much of the variation in 7Q10 per square mile can be explained by variation in the percentage of watershed area underlain by surficial sand or gravel exclusive of wetlands; the regression coefficients differ because equation 9 is based on the full data set of 70 stations whereas equation 10 is based on the data set of 65 stations. The deletion of five small watersheds significantly improved the standard error expressed in cubic feet per second per square mile or in percent of mean cubic feet per second per square mile, and improved the statistical indices of influence and normality presented in columns J through M of table 5. After equation 10 is converted to cubic feet per second, however, the standard error is appreciably larger than those for equations 1 through 9 (columns D through F, table 5).

Equation 11 is the most sophisticated of the per-square-mile regression equations. All terms in equation 11 are significant at the 0.01 level and have the expected sign; all statistical measures of estimation accuracy (table 5) are improved relative to equation 10. The relation appears to be

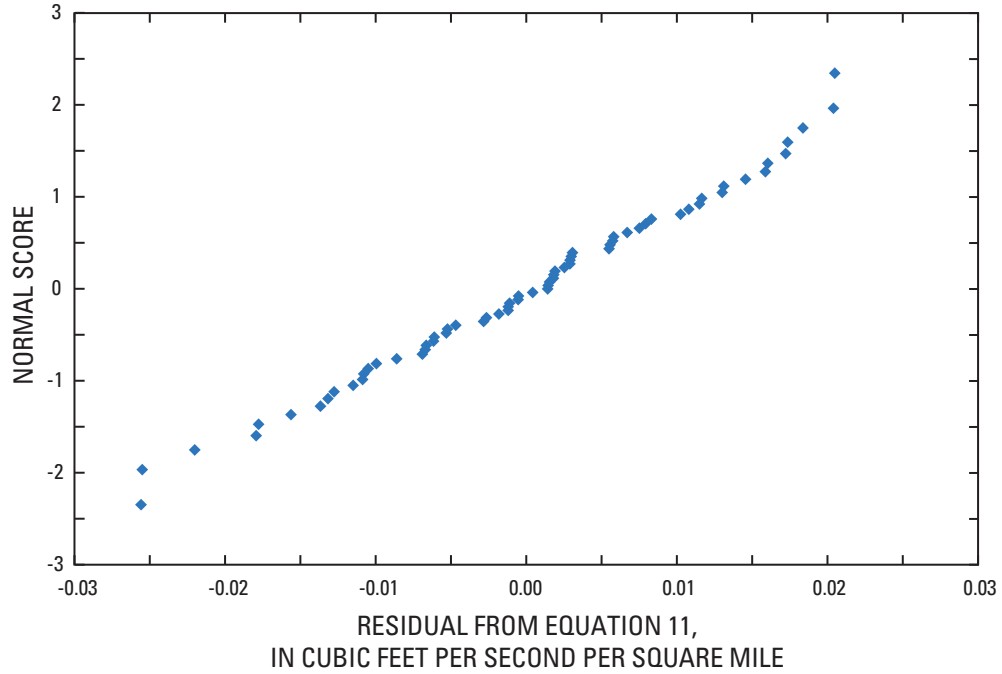
homoscedastic; that is, the spread of data appears relatively uniform throughout the range of values, as shown in figure 15. Residuals are normally distributed, as indicated by their correlation with normal scores (table 5, column L) and by the nearly straight array of data values in figure 16. The Variance Inflation Factor (table 5, column I) indicates the variables to be mutually independent, and the DFFITS statistic (table 5, column K) revealed influence to be widely distributed, with values for seven stations ranging from 0.78 to 0.91 but none higher, the lowest maximum of any 7Q10 equation in table 5. Low flows estimated with equation 11 and converted to cubic feet per second show no systematic departure from equality with observed values (fig. 17).

Equation 11 substantially underestimates the low flow of one of the watersheds deleted from the data set, Castle Creek at Wallace. Calculated 7Q10 for this station is 0.249 cubic foot per second per square mile (2.3 cubic feet per second), whereas equation 11 estimates a 7Q10 of 0.117 cubic foot per second per square mile (1.1 cubic feet per second). As explained earlier, this station lies within and may be typical of a small area in Steuben County where upland valleys are filled with stratified drift. Therefore, application of equation 11 to upland watersheds in the towns of Howard and Fremont (fig. 10) that contain substantial areas of stratified drift may result in underestimation of low flow.

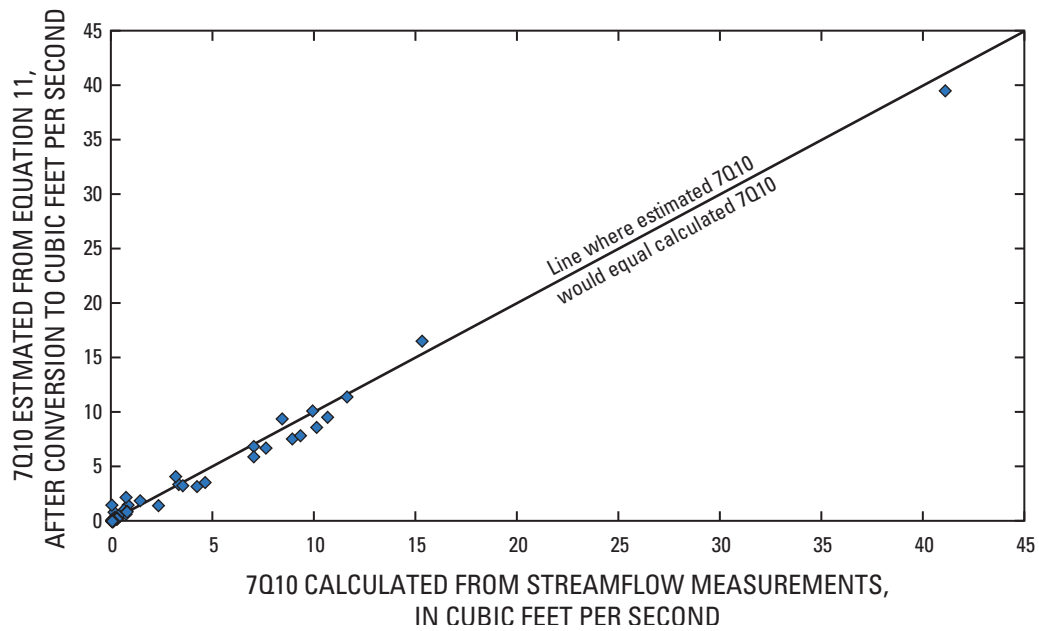


**Figure 15.** Correlation of 7Q10 per square mile calculated from streamflow measurements with 7Q10 per square mile estimated using equation 11. All data and analyses pertain to the Susquehanna River basin in New York. 7Q10, annual lowest streamflow averaged over 7 consecutive days that has a 1 in 10 chance of not being exceeded in a given year. See table 5 (at end of report) for equation 11; terms therein are defined in the “Data set” section of the report.





**Figure 16.** Distribution of residuals from equation 11. See table 5 (at end of report) for equation 11 and for explanation of normal scores.



**Figure 17.** Correlation of 7Q10 calculated from streamflow measurements with 7Q10 converted to cubic feet per second from values estimated using equation 11. All data and analyses pertain to the Susquehanna River basin in New York. (7Q10, annual lowest flow averaged over 7 consecutive days that has a 1 in 10 chance of not being exceeded in any one year. See table 5 (at end of report) for equation 11 and conversion to cubic feet per second.

### Effect of Underflow

Equation 11 is the only equation in table 5 (at end of report) that includes an explanatory variable intended to represent underflow, although this variable (UNDER1) also proved significant at the 0.05 or 0.1 level in a few other per-square-mile equations tested during this investigation. Comparison of equation 10 with equation 11 (table 5) shows that mean annual runoff and underflow together produce only a modest improvement in the statistical tests of estimation accuracy and normality. Nevertheless, the fact that the UNDER1 term proved significant in equation 11 supports the hypothesis that underflow can be one of the principal causes of low-flow variation. The regression coefficient obtained for this term is also consistent with this hypothesis. UNDER1 incorporates all the components of Darcy's Law that are required for computation of underflow except hydraulic conductivity, and is scaled in units of feet squared [length (feet)  $\times$  width (feet) of surficial sand and gravel  $\times$  downvalley gradient (feet per foot)]. Therefore, the regression coefficient of the UNDER1/DAREA variable in equation 11, 0.00059, must be scaled in units of hydraulic conductivity, feet per second, such that multiplication by UNDER1 (feet squared) and division by DAREA (miles squared) yields a number scaled in cubic feet per second per square mile, which is the unit in which the dependent variable in equation 11 is scaled. A hydraulic conductivity of 0.00059 foot per second (51 feet per day), which would be typical of medium-grained alluvial sand (fig. 8), is a plausible value for average hydraulic conductivity of surficial sand and gravel.

The failure of underflow indices to prove significant in other equations may result from any or all of the following reasons:

1. The saturated thickness of surficial sand and gravel and (or) the downvalley water-table gradient at some measurement sites may be inaccurately estimated.
2. Differences from place to place in hydraulic conductivity of surficial sand and gravel may cause large variations in underflow, but were ignored in this study because no reliable means of estimating hydraulic conductivity at each measurement station was devised.
3. Underflow may be so small relative to 7Q10 at nearly all sites that it rarely diminishes low streamflow significantly.

### Logarithmic Equations to Estimate 7-Day, 10-Year Low Flows

Streamflow and watershed properties have been transformed into logarithms prior to regression analysis in many studies directed to estimation of 7Q10 or other indices of low or high streamflow. Logarithmic transformations were judged to be unsuitable and unnecessary in the present study, as explained in the next section. To verify

this judgment, several logarithmic regression equations were developed; the best of these is included in table 5 (at end of report), and an evaluation of that equation confirms that it has serious limitations relative to standard-value and per-square-mile equations.

### A Critique of the Rationale for Logarithmic Transformation

Logarithmic transformations were deemed unsuitable and unnecessary in this study for three reasons: (1) logarithmic formulations are inconsistent with the conceptual physical relations of some watershed properties; (2) logarithmic transformations require artificial adjustments to data values that degrade the integrity of the data set; and (3) the distribution of values in this data set were statistically adequate without transformation. These reasons are explained in the following paragraphs.

*Conceptual physical relations.* As explained earlier, low flow of streams in the glaciated Northeast is conceived to be the sum of large groundwater discharge per square mile from sand and gravel and much smaller discharge per square mile from till, augmented or decreased by spatial variation in annual or seasonal water input, and decreased by groundwater evapotranspiration from wetlands and by underflow. These relations could be plausibly expressed in equations conceptually formulated in several ways. The simplest concept is expressed by the following equation:

$$7Q10 = a(\text{area sand and gravel}) + b(\text{area till}) + c(\text{water input}) - d(\text{area wetlands}) - e(\text{underflow}) \quad (\text{A})$$

where

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

Standard-value equations 1 through 6 (table 5, at end of report) and per-square-mile equations 9 through 11 are all variations on the general form of equation A. (The "area till" term in equation A did not prove significant in any of the equations in table 5 that estimate 7Q10, but appears in such equations for Connecticut (Cervione and others, 1982) and central New England (Wandle and Randall, 1994). An alternative formulation, also conceptually plausible and exemplified by equation 7 (table 5), is

$$7Q10 = a(\text{area sand and gravel})(\text{water input}) + b(\text{area till})(\text{water input}) - d(\text{area wetlands})(\text{water input}) - e(\text{underflow}) \quad (\text{B})$$

A conceptually plausible logarithmic formulation is

$$\log(7Q10) = a \log(\text{area sand and gravel}) + b \log(\text{water input}) \quad (\text{C})$$

which is equivalent to

$$7Q10 = (\text{area sand and gravel})^a (\text{water input})^b \quad (\text{D})$$

Equation D resembles equation B in that sand-and-gravel area multiplied by water input is used to estimate low flow. The other watershed properties considered in equations A and B (till, wetlands, and underflow) cannot be plausibly incorporated as independent variables in a logarithmic equation, however. That is,

$$\log(7Q10) = a \log(\text{area sand and gravel}) + b \log(\text{water input}) + c \log(\text{area till}) + d \log(\text{wetland area}) + e \log(\text{underflow}) \quad (E)$$

is equivalent to

$$7Q10 = (\text{area sand and gravel})^a (\text{water input})^b (\text{till area})^c (\text{wetland area})^d (\text{underflow})^e \quad (F)$$

Equation F is conceptually implausible because low flow should not be proportional to the product of sand-and-gravel area, till area, wetland area, and underflow.

*Data Integrity.* The data set for this investigation contains some zero values of watershed properties and a few values of low-flow indices that are virtually zero (0.001 cubic foot per second). Because zero cannot be expressed as a logarithm, and extremely small values are unduly influential, common practice is to substitute some small number for each zero value, or to add some small number to all values in the data set. During this investigation, 1.0 was added to each value of each independent and dependent variable tested (including variables derived as a sum or difference of individual watershed properties) before conversion to natural logarithms. Later, as a test, 0.1 was added instead of 1.0, and the magnitude of regression coefficients and the significance of watershed properties changed, as explained farther on. The sensitivity of logarithmic equations to such necessary but arbitrary adjustments was deemed undesirable.

*Data Distribution.* Riggs (1985, p. 84–85) gives two reasons for considering logarithmic transformations of variables prior to regression analysis: to achieve equal variation about the regression line over its entire length, and to linearize curved data distributions.

1. *Equal variation about the regression line over its entire length.* Plots of dependent against independent variables in some data sets show close grouping of 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, thereby producing a more nearly equal distribution of variances along the regression line. However, plots of low-flow indices in this data set against individual or combined watershed properties (for example, figures 11, 13, 15, and 17) depict a range of scatter that is similar all along the regression line. A quantitative test for heteroscedasticity was applied to the seven equations for which residuals are normally distributed (table 5, column L). For each equation, the variance of the residuals from the larger half of the low flows (in cubic

feet per second or in cubic feet per second per square mile) exceeded the variance of the residuals from the smaller half, thus confirming that some heteroscedasticity exists, but F-tests (Iman and Conover, 1983, p. 274–276) indicated that the variances were statistically equivalent at the 0.05 significance level for four of the seven equations.

2. *Curved data distributions.* Plots of the low-flow indices in this data set against watershed properties (fig. 13) show very little curvature, as do plots of these indices against estimated low flows (figs. 11, 15).

Accordingly, application of logarithmic transformations to this data set was judged to have little theoretical justification. Male and Ogawa (1982, p. 29–30) also concluded from analysis of streams in Massachusetts that “short-duration low flows may be better described by a linear model than by a log-linear model.”

### Comparison of Logarithmic Equations with Standard-Value Equations

The statistical measures of estimation accuracy for logarithmic equations are not directly comparable to those for standard-value equations because they are scaled in different units. To permit comparison, the following detransformation procedure was undertaken:

1. The logarithm of 7Q10 estimated for each station was converted to cubic feet per second.
2. Each value of estimated 7Q10 in cubic feet per second was multiplied by  $e^{0.5(\text{mean square error})}$  to correct for detransformation bias, as recommended by Choquette (1988) and Miller (1984).
3. The constant 1.0 that had been added to each value of 7Q10 in preparation for logarithmic regression was subtracted.
4. Standard error in cubic feet per second was computed from the following equation, then divided by the mean value of 7Q10 for the data set to obtain standard error in percent of the mean; results are given (in parentheses) in table 5 (at end of report) for logarithmic equation 12.

where

$$SE = \left[ \frac{\sum_1^n (a - p)^2}{(n - v - 1)} \right]^{0.5}$$

SE = Standard Error,  
n = number of stations in the data set,  
v = number of independent variables in the regression equation,  
a = observed value of 7Q10 at a station, and  
p = detransformed estimated value of 7Q10 at that station.

Equation 12, the only logarithmic equation in table 5, had the highest statistical measures of estimation accuracy of any logarithmic equation tested. Its two independent variables are normally distributed at an alpha level of 0.01 or higher

(fig. 18), highly significant ( $p < 0.0001$ , table 5), and independent ( $VIF = 1.1$ , table 5). No station had a DFFITS influence statistic greater than 1.0 (table 5), and residuals are normally distributed at the 0.01 alpha level (table 5, column L). Like the standard-value and per-square-mile equations described earlier, equation 12 is consistent with the concept that low flow is positively correlated with surficial sand-and-gravel area and with amount of water available. Nevertheless, equation 12 has several serious limitations:

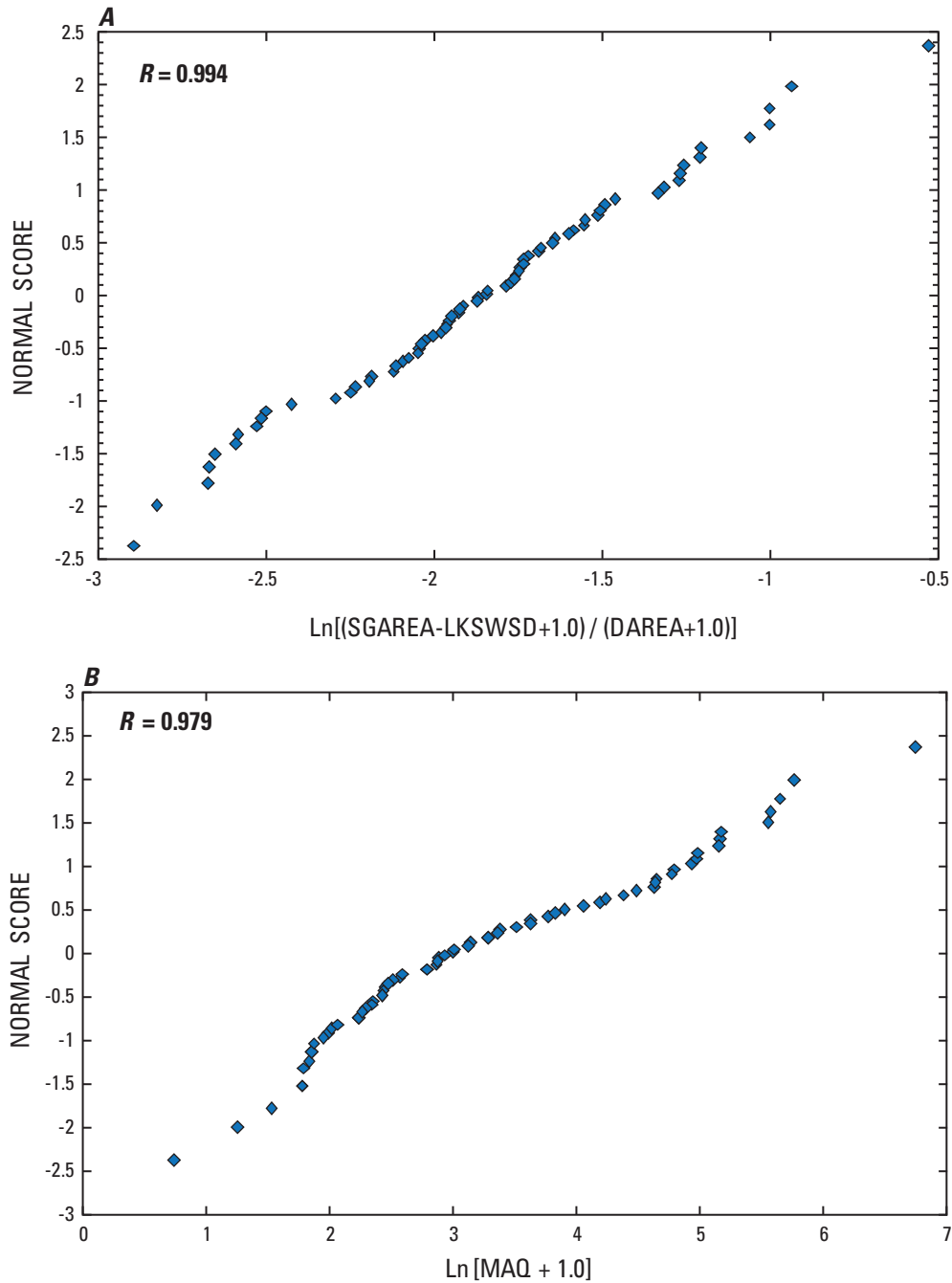
1. Wetland area was subtracted from the total area underlain by sand and gravel in formulating equation 12, as in equations 2 through 11. However, the estimation accuracy of equations 3 through 8 and 11 was enhanced by also including wetland area underlain by sand and gravel as a separate independent variable, whose negative regression coefficient was inferred to calibrate the loss of water through evapotranspiration. This refinement could not be logically incorporated in a logarithmic equation.
2. An index of underflow also could not be logically incorporated as an independent variable in a logarithmic equation. Its effect could be accounted for in logarithmic equations if underflow could be independently computed at each station by Darcy's Law, then added to 7Q10 low flow. This approach would not easily accommodate the iterative testing and weighting inherent in regression analysis, however, and was not considered feasible in this study because information on hydraulic conductivity at measurement stations was lacking.
3. The statistical measures of estimation accuracy for equation 12 after conversion to cubic feet per second are larger (less accurate) than those for other equations in table 5. Equation 12 systematically underestimates 7Q10 low flow in cubic feet per second at most stations, especially the larger values (fig. 19).
4. The addition of constant values to each variable, as required for logarithmic analysis, strongly influenced the results of the regression. An alternative test in which 0.1 rather than 1.0 was added to all data values yielded a logarithmic regression equation with much worse statistical indices of estimation accuracy and normality, although the conversion of that equation to cubic feet per second yielded low-flow estimates that were slightly closer to observed low flows. Most important, the low flows estimated with this alternative test differed substantially from those estimated with equation 12, solely because of the difference in magnitude of the arbitrary constant added to measured values to facilitate computation. Addition of constants to individual watershed properties such as SGAREA and LKSWSD rather than to combined terms such as (SGAREA-LKSWSD) or (SGAREA - LKSWSD)/DAREA also changed the results substantially. A computational technique that requires arbitrary data distortions that significantly alter the results was deemed unreliable.

## Equations to Estimate 7-Day, 2-Year Low Flow and 90-Percent Flow Duration

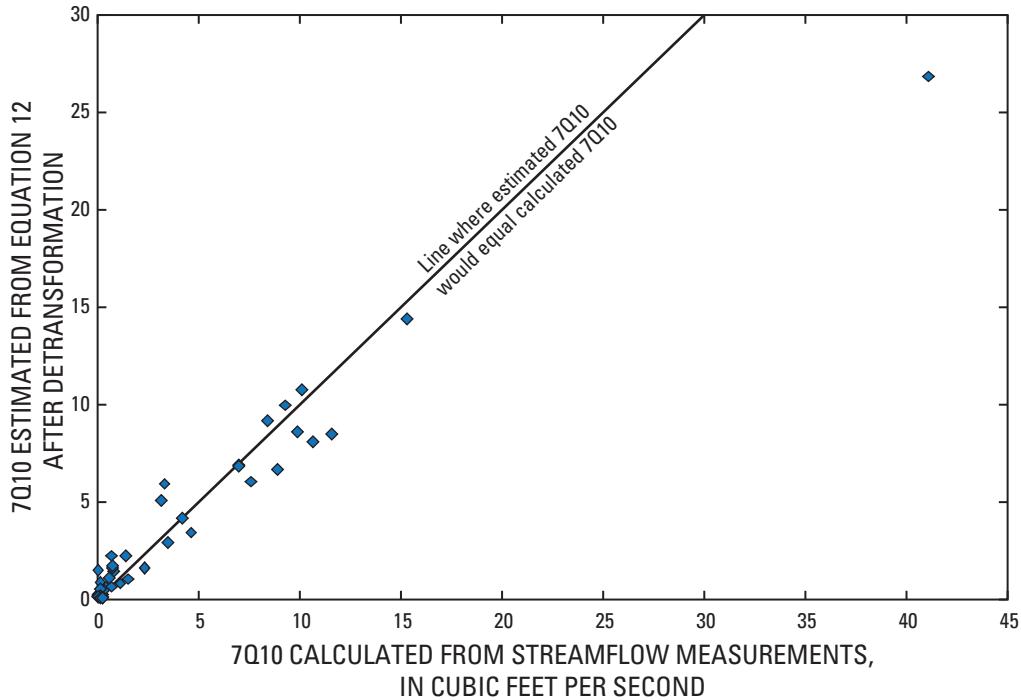
Watershed properties were also tested in regression analyses to estimate 7Q2 and 90%FD, which respectively average 1.7 and 2.2 times 7Q10 (table 3). In general, the watershed properties that proved most influential were the same properties that proved influential in the 7Q10 analysis. At these larger low flows, till area also had a significant positive effect, although the regression coefficients indicate that groundwater discharge to streams per square mile of till-mantled uplands was only a small percentage of the discharge per square mile of sand and gravel (equations 13–16, table 5, at end of report). Substitution of the term till area minus area of wetlands in till for till area had almost no effect on regression coefficients nor on any of the statistical indices for these equations in table 5. Area of wetlands in sand and gravel, which had proved significant as an independent variable in estimating 7Q10, was not significant in any 7Q2 or 90%FD equation, which implies that the loss of water through evapotranspiration from wetlands is too small to significantly diminish flows that are about twice as large as 7Q10.

Standard-value equations developed from the full data set of 70 stations to estimate 7Q2 and 90%FD resemble their 7Q10 counterparts in that residuals are not normally distributed, as indicated by the correlation with normal scores for equations 13 and 15 (table 5, column L). Furthermore, the one watershed in the data set that is larger than 200 square miles has a greater influence on regression equations for 7Q2 and 90%FD than on the 7Q10 equations. Values of the influence statistic DFFITS for this watershed from equations 13 and 15 ranged from 6.1 to 4.3. These statistical weaknesses were rectified by deleting 38 small watersheds (as in equation 8) and the one unusually large watershed; 31 stations remained. These deletions did not change the relative significance of watershed properties (compare equations 13 and 14, or 15 and 16, in table 5) and changed estimated values of low flow from individual watersheds only moderately (the median difference in estimated 7Q2 between equations 13 and 14 was only 0.34 cubic foot per second, or 6 percent). Equations 14 and 16 have normally distributed residuals and fairly low variance inflation. Therefore, these equations are suitable for estimating 7Q2 and 90%FD in moderately large unregulated watersheds.

Per-square-mile equations 17 and 18 are appropriate for estimating 7Q2 and 90%FD in watersheds or segments of watersheds smaller than about 30 square miles. These equations are based on data sets dominated by small watersheds; they also have smaller standard errors in cubic feet per second than equations 14 and 16, and better statistical indices of normality, influence, significance of variables, and collinearity (variance inflation). Like equations 10 and 11, however, equations 17 and 18 may underestimate low flows from a small area in Steuben County, between Bath and Hornell, where upland watersheds contain abundant stratified drift. Equations 17 and 18 incorporate the product



**Figure 18.** Distribution of data values for independent variables in equation 12: *A*, sand and gravel exclusive of wetlands, and *B*, mean annual runoff. See table 5 at end of report for equation 12, and for explanation of normal scores. (Ln, natural logarithm; SGAREA, area of surficial sand and gravel, including any overlying lakes or swamps; LKSWSD, area of lakes and swamps underlain by sand and gravel; DAREA, drainage area; MAQ, mean annual runoff; R = Pearson product-moment correlation coefficient)



**Figure 19.** Correlation of 7Q10 calculated from streamflow measurements with 7Q10 converted to units of cubic feet per second from logarithmic values estimated using equation 12. All data and analyses pertain to the Susquehanna River basin in New York. (7Q10, annual lowest streamflow averaged over 7 consecutive days that has a 1 in 10 chance of not being exceeded in a given year. See table 5 (at end of report) for equation 12 and detransformation to cubic feet per second.)

of sand-and-gravel area (exclusive of wetlands) multiplied by mean annual runoff per square mile, which proved better than sand-and-gravel area alone in estimating 7Q2 or 90%FD per square mile.

## Suggested Methods for Estimating Low-Flow Characteristics at Ungaged Sites

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 constructing and operating a streamflow-gaging station at that site. The accuracy of the calculated low-flow characteristics will increase with the length of the record. Less exact estimates of low-flow characteristics 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 flows at one or more continuous-record streamflow-gaging stations. For best results, the gaging station(s) should be on a nearby unregulated stream that drains a watershed similar in size and geology to the watershed of interest. Procedures for estimating low-flow characteristics at a site by correlating an array of individual streamflow measurements made at that site with daily

mean flows at nearby continuous-record gaging stations are explained by Cervione and others (1982), Hardison and Moss (1972), Matalas and Jacobs (1964), Riggs (1972), Searcy (1959), and Stedinger and Thomas (1985). Use of correlation procedures in this study was facilitated by the availability of records from the many gaging stations that were operated on unregulated streams in and near the Susquehanna River basin prior to 1972; fewer gaging stations have been operated during subsequent years (through 2010), which may limit the reliability of such correlation procedures in some localities.

Streamflow measurements had been made at more than 150 sites in the Susquehanna River basin of New York as of 1975. Site locations and low-flow characteristics computed for these sites are given by Eissler (1979) and by Ku and others (1975). Additional sites are established from time to time by the U.S. Geological Survey. Daily mean streamflows at gaging stations and individual flow measurements at other sites have been published in annual water-data reports (U.S. Geological Survey, 1951–2005); information can also be obtained from a U.S. Geological Survey Website (<http://ny.water.usgs.gov>). Low-flow characteristics computed for a site generally can be extrapolated to a site of interest nearby along the same stream in proportion to drainage area, but if either the site of interest or the previously established site happens to be where a tributary enters or crosses the valley of a larger stream, per-square-mile extrapolation is not advisable because low flows may vary greatly over short distances in this geologic setting (Ku and others, 1975; Randall, 1978).

The regional regression equations developed in this study can be used to estimate low flows at sites where no measurements have been made nearby and there is no prospect of obtaining measurements over a suitable range of low flow. These equations can be applied to entire unregulated watersheds. If a site of interest is on a stream where low-flow indices have been computed for a station some distance upstream or downstream, however, application of the regional equations to the intervening area 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 primary objective of the regression analysis in this investigation was to identify and quantify the watershed properties that control low streamflow, but several regression equations presented in table 5 (at end of report) are useful tools that can be applied throughout the Susquehanna River basin of New York to estimate three commonly used indices of low flow for the period 1941–70:

7Q10 (annual lowest streamflow averaged over 7 consecutive days that has a 1 in 10 chance of not being exceeded in a given year);

7Q2 (annual lowest streamflow averaged over 7 consecutive days that has a 1 in 2 chance of not being exceeded in a given year);

90%FD (flow equaled or exceeded 90 percent of the time during the 30-year period of analysis).

The six equations listed below are suggested for practical application. In these equations, areas are in square miles, length and width in feet, gradient in feet per foot, and mean annual runoff in cubic feet per second. Abbreviations for independent variables are defined in the section “Watershed properties.”

### Watersheds Larger Than 30 Square Miles

$$7Q10 = 0.047 + 0.51 (SGAREA - LKSWSD) - 0.42 (LKSWSD) + 0.029 [(MAQ/DAREA) - 1.85](DAREA)$$

[Equation 8 (table 5, at end of report); standard error 0.58 cubic foot per second, median percent error of estimate 22 percent]

$$7Q2 = 0.412 + 0.70 (SGAREA - LKSWSD) + 0.019 (TLAREA) + 0.059 [(MAQ/DAREA) - 1.85](DAREA)$$

[Equation 14 (table 5); standard error 1.40 cubic feet per second, median percent error of estimate 23 percent]

$$90\%FD = 0.58 + 0.83 (SGAREA - LKSWSD) + 0.052 (TLAREA) + 0.106 [(MAQ/DAREA) - 1.85] (DAREA)$$

[Equation 16 (table 5); standard error 1.95 cubic feet per second, median percent error of estimate 23 percent]

### Watersheds Smaller Than 30 Square Miles

$$7Q10 = DAREA \{ 0.0042 + 0.46 [(SGAREA - LKSWSD) / DAREA] - 0.40 (LKSWSD/DAREA) + 0.021 [(MAQ/DAREA) - 1.85] - 0.00059 (WIDTH)(THICK)(SLOPE) / DAREA \}$$

[Equation 11 (table 5, at end of report); standard error 0.0115 cubic foot per second per square mile, median percent error of estimate 36 percent]

$$7Q2 = DAREA \{ - 0.00069 + 0.475 [(SGAREA - LKSWSD) (MAQ/DAREA)/DAREA] \}$$

[Equation 17 (table 5); standard error 0.018 cubic foot per second per square mile, median percent error of estimate 29 percent]

$$90\%FD = DAREA \{ 0.026 + 0.56 [(SGAREA - LKSWSD)(MAQ / DAREA) / DAREA] + 0.034 [(MAQ / DAREA) - 1.85] \}$$

[Equation 18 (table 5), standard error 0.022 cubic foot per second per square mile, median percent error of estimate 31 percent]

The independent variables required by the foregoing equations can be compiled largely, but not entirely, from published sources. Drainage area, area of wetlands, and stream gradient at sites of interest can be measured from U.S. Geological Survey topographic maps. Mean annual runoff can be estimated from figure 3 of this report, or from maps in U.S. Geological Survey Open-File Report 96–395 (Randall, 1996) or Professional Paper 1415–C (Kontis and others, 2004), available at <http://ny.water.usgs.gov/publications>. Area of surficial sand and gravel, and width of sand and gravel at sites of interest, may be compiled from county soils maps (for example, Giddings and others, 1971); or from unpublished 1:24,000-scale reconnaissance maps of surficial geology in the files of the New York Geological Survey at Albany, New York; or from a series of U.S. Geological Survey aquifer map reports at 1:24,000 scale that cover several reaches of major valleys in the Susquehanna River basin and are cited (with locations) at <http://ny.water.usgs.gov/projects/bgag/aquifer.maps/maps.html>. A map at 1:38,400 scale of surficial geology in Fivemile Creek watershed was prepared during this investigation (Randall, 1986b). Geologic and hydrologic maps at 1:250,000 scale (for example, Cadwell and Dineen, 1987; Miller, 1988) are not sufficiently detailed for this purpose. Distinct boundaries are commonly evident on topographic maps where terraces and flood plains underlain by surficial sand and gravel on valley floors abut steep valley walls underlain by till and bedrock. Therefore, delineation of the extent of surficial sand and gravel from examination of 1:24,000-scale topographic maps and soils maps are likely to be reasonably accurate and reproducible. Thickness of

sand and gravel at sites of interest can be estimated from the same sources; pertinent information for sites in broad valleys may also be available from well or test-boring records (for example, Randall, 1972) or aquifer-thickness maps (MacNish and Randall, 1982, or more detailed studies of some localities). For a few small watersheds that contain little sand and gravel, some equations will predict negative low flows, which should be treated as zero flow. The equations for watersheds smaller than 30 square miles may underestimate low flows from upland watersheds in the towns of Howard and Fremont (between the Cohocton and Canisteo Rivers, fig. 10) that contain appreciable sand and gravel.

## Summary

Low flow of streams in the Susquehanna River basin of New York increases with the amount of water available for runoff annually (precipitation minus evapotranspiration) and with the spatial extent of surficial sand and gravel within the watershed. Low flow is decreased by evapotranspiration from wetlands interspersed with the surficial sand and gravel, and by underflow beneath the valley floor that bypasses stream channels. Spatial variation in the properties of till and bedrock and in watershed relief or slope appears to have little effect on low flow within this region.

The goal of this study was to identify hydrologic processes or properties that control groundwater discharge to streams during periods of low streamflow, and to test and quantify their importance through multiple regression analysis. Such information is likely to improve our ability to understand and reliably model aquifers, runoff, and particularly the magnitude and quality of low streamflow. A data set was assembled that consists of low-flow characteristics at 23 long-term continuous-record streamflow-gaging stations and 47 short-term or partial-record stations or combinations thereof, along with measurements of several dimensions and properties of the corresponding watersheds. Regression analyses based on this data set led to an array of equations that estimate 7-day mean low flows with 2-year and 10-year recurrence intervals, and flow equaled or exceeded 90 percent of the time, from 1941 through 1971. These equations estimate low-flow characteristics as a function of area of sand and gravel, area of wetlands overlying sand and gravel, area of till-mantled uplands, mean annual runoff, and, in a few equations, aquifer properties associated with underflow. At the lowest low flow evaluated (7-day low flow with 10-year recurrence interval), groundwater discharge from till becomes insignificantly small, and evapotranspiration from wetlands consumes a larger fraction of groundwater discharge to streams than at larger low flows. Equations that estimate low flow in cubic feet per second and are based on a subset of 31 or 32 watersheds, mostly between 30 and 200 square miles in size, have standard errors between 16 and

21 percent of the mean values of low-flow characteristics. Equations that estimate low flow in cubic feet per second per square mile and are based on a subset of 65 watersheds have somewhat larger percent standard errors but estimate more accurately the low-flow characteristics in watersheds smaller than 30 square miles. These equations may underestimate low flows from the rare upland watersheds that contain extensive surficial sand and gravel, however. Logarithmic equations cannot express some inferred conceptual relations and are sensitive to the values of arbitrary constants added to avoid zero values of variables. Nevertheless, logarithmic equations indicate low flows to be a function of the same watershed properties that proved most significant in the absence of logarithmic transformation. Six equations deemed conceptually and statistically sound are suggested for practical application in estimating low flows at ungaged sites.

Variations in hydraulic conductivity of stratified valley-fill deposits are in principle a powerful but nonlinear influence on low flow, as confirmed by multiple simulations conducted with a cross-sectional model of a hypothetical aquifer. Groundwater discharge to streams is small where hydraulic conductivity of the valley fill is small (comparable to that of till) because the rate of groundwater flow is small and because the water table may be so close to land surface that much of the precipitation cannot infiltrate or is lost to evapotranspiration. Groundwater discharge is also small during periods of low flow where hydraulic conductivity is large, especially in narrow valleys, because the water table drains quickly down to stream grade, long before periods of low flow. Therefore, fine-sandy deltas, thin sand aquifers overlying fine-grained lacustrine sediment, and mostly unsaturated large stratified deposits on upland hillsides are likely to contribute more groundwater to streams per unit area during periods of low flow than thick, highly permeable aquifers in valleys or coarse alluvium beneath flood plains. This concept could not be tested by regression analysis, but some support is provided by one watershed in the data set (Castle Creek at Wallace) that is atypical in that it contains abundant stratified drift on upland hillsides and in upland valleys; low flow per unit area from this watershed was calculated to be considerably larger than that from many watersheds in which stratified sand and gravel are mostly below stream grade in major valleys. A parameter designed to represent underflow beneath streamflow-measurement sites incorporated the width, thickness, and downvalley gradient of surficial sand and gravel, but not its hydraulic conductivity. The reasonable accuracy of regression equations that incorporate this parameter suggests that, although the hydraulic conductivity of stratified drift commonly varies widely over distances of a few hundred feet, this variability may approach some uniform average value on a scale of entire valley systems or watersheds.



## Acknowledgments

The computation of low-flow indices for gaging stations in the Susquehanna River basin for 1941–71 was done chiefly by Charles Barnes of the U.S. Geological Survey.

## References Cited

- Allen, H.E., Jr., and Cowan, E.A., 1985, Low-flow characteristics of streams in the Kishwaukee River basin, Illinois: U.S. Geological Survey Water-Resources Investigations Report 84-4311, 35 p.
- Barksdale, H.C., O'Bryan, Deric, and Schneider, W.J., 1966, Effect of drought on water resources in the Northeast: U.S. Geological Survey Hydrologic Investigations Atlas HA-243, 1 sheet.
- Barnes, C.R., 1986, Methods 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.
- Bedinger, M.S., 1961, Relation between median grain size and permeability in the Arkansas River valley, Arkansas: U.S. Geological Survey Professional Paper 424-C, p. C31-C32.
- Bergeron, M.P., 1987, Effect of reduced industrial pumpage on the migration of dissolved nitrogen in an outwash aquifer in Olean, Cattaraugus County, New York: U.S. Geological Survey Water-Resources Investigations Report 85-4082, 39 p.
- Cadwell, D.H., and Dineen, R.J., 1987, Surficial geologic map of New York, Hudson-Mohawk sheet: New York State Museum Map and Chart series 40, scale 1:250,000.
- 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.
- Choquette, A.F., 1988, Regionalization of peak discharges for streams in Kentucky: U.S. Geological Survey Water-Resources Investigations Report 87-4209, 105 p.
- Coates, D.R., 1971, Hydrogeomorphology of Susquehanna and Delaware basins, in Morisawa, Marie, ed., Quantitative geomorphology, some aspects and applications: Binghamton, New York, State University of New York at Binghamton, Publications in Geomorphology, p. 273-306.
- Comer, G.H., and Zimmerman, R.C., 1969, Low flow and basin characteristics of two streams in Vermont: Journal of Hydrology, v. 7, p. 98-108.
- Cowardin, L.M., Carter, V., Golet, F.C., and LaRoe, E.T., 1979, Classification of wetlands and deepwater habitats of the United States: U.S. Fish and Wildlife Service, FWS/OBS-79-31, 103 p.
- Darmer, K.I., 1970, A proposed streamflow data program for New York: U.S. Geological Survey Open-File Report 70-94, 47 p.
- DeAngelis, R.J., Urban, J.B., Gburek, W.J., and Contino, M.A., 1984, Precipitation and runoff on eight New England basins during extreme wet and dry periods: Hydrological Sciences Journal, v. 29, no. 1, p. 13-28.
- 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.
- Dethier, B.E., 1966, Precipitation in New York State: Ithaca, N.Y., Cornell University Agricultural Experiment Station Bulletin 1009, 78 p.
- Dingman, S.L., 1978, Synthesis of flow-duration curves for unregulated streams in New Hampshire: Water Resources Bulletin, v. 14, no. 6, p. 1481-1502.
- Dingman, S.L., 1981, Elevation, a major influence on the hydrology of New Hampshire and Vermont, USA: Hydrological Sciences Bulletin, v. 26, no. 4, p. 399-413.
- Dysart, J.E., and Rheaume, S.J., Induced infiltration from the Rockaway River and water chemistry in a stratified-drift aquifer at Dover, New Jersey: U.S. Geological Survey Water-Resources Investigations Report 96-4068, 112 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.
- Flint, J.J., 1968, Hydrogeology and geomorphic properties of small basins between Endicott and Elmira, New York: Binghamton, New York, State University of New York at Binghamton, unpublished Masters thesis, 74 p.
- Gebert, W.A., 1979, Low-flow characteristics of streams in the Lake Superior basin, Wisconsin: U.S. Geological Survey Water Resources Investigations Open-File Report 79-38, 74 p.
- Giddings, E.B., Flora, D.F., and Olson, G.W., 1971, Soil survey of Broome County, New York: U.S. Soil Conservation Service, 95 p. plus maps.
- 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 1542-B, 55 p.
- 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.

- Hely, A.G., Nordenson, T.J., and others, 1961, Precipitation, water loss, and runoff in the Delaware River basin and New Jersey: U.S. Geological Survey Hydrologic Investigations Atlas HA-11, 1:100,000.
- 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 Assoc. Journal*, v. 41, p. 403-412.
- Holmes, C.D., 1952, Drift dispersion in west-central New York: *Geological Society of America Bulletin*, v. 63, no. 10, p. 993-1010.
- Hunt, O.P., 1967, Duration curves and low-frequency curves of streamflow in the Susquehanna River basin, New York: *New York Water Resources Commission Bulletin* 60, 52 p.
- Iman, R.L., and Conover, W.J., 1983, *A modern approach to statistics*: New York City, John Wiley & Sons, 497 p.
- Johnson, M.L., 1970, Runoff from eight watersheds in northern Vermont: Baltimore, MD, Johns Hopkins University, unpublished Ph.D. dissertation.
- Johnston, R.H., 1971, Base flow as an indicator of aquifer characteristics in the coastal plain of Delaware: U.S. Geological Survey Professional Paper 750-D, p. D212-D213.
- 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, 1:1,000,000, 3 sheets.
- Kontis, A.L., 2001, Computer program for simulation of variable recharge with the U.S. Geological Survey modular finite-difference ground-water flow model: U.S. Geological Survey Open-File Report 00-173, 75 p.
- 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 McNish, R.D., 1975, Streamflow in the New York part of the Susquehanna River basin: *New York Department of Environmental Conservation, Bulletin* 71, 130 p.
- Kruseman, G.P., and deRidder, N.A., 1990, Analysis and evaluation of pumping-test data: Wageningen, the Netherlands, International Institute for Land Reclamation and Development, 377 p.
- Lapham, W.W., 1989, Use of temperature profiles beneath streams to determine rates of vertical ground-water flow and vertical hydraulic conductivity: U.S. Geological Survey Water-Supply Paper 1578, 70 p.
- Lewis, D.C., 1958, 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 Gulledege, T.R., 1985, Use of correlation coefficient with normal probability plots: *The American Statistician*, v. 39, p. 75-79.
- Lyford, F.P., 1986, Northeast glacial regional aquifer system study, in Sun, R.J., ed., *Regional aquifer-system analysis program of the U.S. Geological Survey, Summary of Projects, 1978-84*: U.S. Geological Survey Circular 1002, p. 162-167.
- Lyford, F.P., and Cohen, A.J., 1988, Estimation of water available for recharge to sand and gravel aquifers in the glaciated Northeastern United States, 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. 37-62.
- McNish, R.D., and Randall, A.D., 1982, Stratified-drift aquifers in the Susquehanna River basin, New York: *New York Department of Environmental Conservation Bulletin* 75, 68 p.
- 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.
- Mazzaferro, D.L., Handman, E.H., and Thomas, M.P., 1979, Water-resources inventory of Connecticut, part 8, Quinnipiac River basin: *Connecticut Water-Resources Bulletin* 27, 88 p.
- McDonald, M.G., and Harbaugh, A.W., 1988, A modular three-dimensional finite-difference ground-water flow model: U.S. Geological Survey Techniques of Water-Resources Investigations, book 6, chap. A1, 586 p.
- Miller, D.M., 1984, Reducing transformation bias in curve fitting: *American Statistician*, v. 38, no. 2, p. 124-126.
- Miller, T.S., 1988, Unconsolidated aquifers in upstate New York—Finger Lakes sheet: U.S. Geological Survey Water-Resources Investigations Report 87-4122, 1:250,000.
- Montgomery, D.C., and Peck, E.A., 1982, *Introduction to linear regression analysis*: John Wiley & Sons, 504 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.

- Morse, H.H., 1946, Soil profile characteristics and stream flow behavior: Soil Science Society of America Proceedings, v. 11, p. 442.
- Moss, J.H., and Ritter, D.F., 1962, New evidence regarding the Binghamton substage in the region between the Finger Lakes and the Catskills: American Journal of Science, v. 260, p. 81–106.
- Olmsted, F.H., and Hely, A.G., 1962, Relation between ground water and surface water in Brandywine Creek basin, Pennsylvania: U.S. Geological Survey Professional Paper 417-A, 21 p.
- Pagano, T.S., 1987, Hydrologic and geomorphic properties of several tributary drainage basins of the Susquehanna River basin, New York: Binghamton, New York, State University of New York at Binghamton, unpublished Masters thesis, 153 p.
- 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.
- Randall, A.D., 1972, Records of wells and test borings in the Susquehanna River basin: New York Department of Environmental Conservation Bulletin 69, 92 p.
- 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., 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.
- Randall, A.D., 1981, Hydrology in relation to glacial geology along the Susquehanna River valley, Binghamton to Owego, New York, *in* Enos, Paul, ed., Guidebook for field trips in south-central New York: New York Geological Association, 53rd Annual Meeting, p. 147–170.
- Randall, A.D., 1986a, Aquifer model for the Susquehanna River valley in southwestern Broome County, New York: U.S. Geological Survey Water-Resources Investigations Report 85–4099, 38 p.
- Randall, A.D., 1986b, Ice readvance in the valley of Fivemile Creek, Steuben County, New York, and the effect of the resulting deposits on postglacial streamflow, *in* Cisne, J.L., ed., Field trip guidebook: New York Geological Association, 58th Annual Meeting, p. 239–260.
- 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.
- Randall, A.D., 2001, Hydrogeologic framework of stratified-drift aquifers in the glaciated northeastern United States: U.S. Geological Survey Professional Paper 1415-B, 179 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.
- Randall, A.D., Snively, D.S., Holecek, T.J., and Waller, R.M., 1988, Alternative sources of large seasonal ground-water supplies in the headwaters of the Susquehanna River basin, New York: U.S. Geological Survey Water-Resources Investigations Report 85–4127, 121 p.
- Rasmussen, W.C., and Andreasen, G.E., 1959, A hydrologic budget of the Beaverdam Creek basin, Maryland: U.S. Geological Survey Water-Supply Paper 1472, 106 p.
- Reynolds, R.J., 1987, Hydrogeology of the surficial outwash aquifer at Cortland, Cortland County, New York: U.S. Geological Survey Water-Resources Investigations Report 85–4090, 43 p.
- Reynolds, R.J., and Brown, G.A., 1984, Hydrogeologic appraisal of a stratified-drift aquifer near Smyrna, Chenango County, New York: U.S. Geological Survey Water-Resources Investigations Report 84–4029, 53 p.
- Rickard, L.V., 1975, Correlation of the Silurian and Devonian rocks in New York State: New York State Museum and Science Service Map and Chart Series No. 24, 16 p. plus 4 plates.
- Riggs, H.C., 1972, Low-flow investigations: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, Hydrologic analysis and interpretation, chap. B1, 18 p.
- Riggs, H.C., 1973, Regional analyses of streamflow characteristics: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, Hydrologic analysis and interpretation, chap. B3, 15 p.
- Riggs, H.C., 1985, Streamflow characteristics: Amsterdam, New York, Elsevier, 249 p.
- Rorabaugh, M.I., 1960, Use of water levels in estimating aquifer constants: International Association of Scientific Hydrology Publication no. 52, p. 314–323.
- Rorabaugh, M.I., 1963, Estimating changes in bank storage and ground-water contribution to streamflow: International Association of Scientific Hydrology Pub. 63, p. 432–441.
- Ryan, B.F., Joiner, B.L., and Ryan, T.A., 1985, Minitab handbook, 2nd ed.: Boston, Mass., PWS Publishers, 379 p.
- SAS Institute, Inc., 1990, SAS/STAT Users Guide, version 6, 4th ed.: Cary, North Carolina, SAS Institute Inc., 1675 p.
- Searcy, J.K., 1959, Flow-duration curves: U.S. Geological Survey Water-Supply Paper 1542-A, 33 p.
- Schneider, W.J., 1957, Relation of geology to streamflow in the upper Little Miami basin: Ohio Journal of Science, v. 57, p. 11–14.

- Schneider, W.J., and Ayer, G.A., 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.
- Sopper, W.E., and Lull, H.W., 1970, Streamflow characteristics of the northeastern United States: Pennsylvania State University College of Agriculture Bulletin 766, 129 p.
- 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.
- Sun, R.J., ed., 1986, Regional Aquifer-System Analysis Program of the U.S. Geological Survey, Summary of projects 1978–84: U.S. Geological Survey Circular 1002, 264 p.
- Swartwout, D.J., 1982, An evaluation of National Wetlands Inventory in Massachusetts: Unpublished Ms. thesis, Amherst, Mass., University of Massachusetts, 123 p.
- Tasker, G.D., 1972, Estimating low-flow characteristics of streams in southeastern Massachusetts from maps of ground-water availability, *in* Geological Survey Research 1972: U.S. Geological Survey Professional Paper 800–D, p. D217–D220.
- Thomas, M.P., 1966, Effect of glacial geology upon the time distribution of streamflow in eastern and southern Connecticut: U.S. Geological Survey Professional Paper 550–B, p. B209–B212.
- Thomas, M.P., and Cervione, M.A., Jr., 1970, A proposed streamflow data program for Connecticut: Connecticut Water Resources Commission Bulletin 23.
- Thornthwaite, C.W., and Mather, J.R., 1955, The water balance: Drexel Institute of Technology Publications in Climatology, v. 8, no. 1, 86 p.
- Todd, D.K., 1970, The water encyclopedia: Port Washington, N.Y., Water Information Center, 559 p.
- Trousdell, K.B., and Hoover, M.D., 1955, A change in groundwater level after clearcutting of loblolly pine in the coastal plain: Journal of Forestry, v. 53, no. 7, p. 493–498.
- U.S. Geological Survey, 1951–2005, Water resources data New York: Albany, NY, U.S. Geological Survey (issued annually).
- U.S. Geological Survey, 1985, Watstore Users Guide, v. 1, chap. 4, section G (revised), programs A969 and A193.
- Wandle, S.W., 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.
- Weeks, E.P., and Stangland, H.G., 1971, Effects of irrigation on streamflow in the central sand plain of Wisconsin: U.S. Geological Survey Open-File Report, 113 p.
- Williams, J.H., 1991, Tributary-stream infiltration in Marsh Creek valley, north-central Pennsylvania: U.S. Geological Survey Water-Resources Investigations Report 90–4052, 39 p.
- Williams, J.H., and Morrissey, D.J., 1996, Recharge of valley-fill aquifers in the glaciated northeast from upland runoff, *in* Ritchey, J.D., and Rumbaugh, J.O., eds., Subsurface fluid-flow (ground-water and vadose zone) modeling: American Society for Testing and Materials, Standard Publication 1288, p. 97–113.
- Yager, R.M., 1986, Simulation of ground-water flow and infiltration from the Susquehanna River to a shallow aquifer at Kirkwood and Conklin, Broome County, New York: U.S. Geological Survey Water-Resources Investigations Report 86–4123, 70 p.
- Zecharias, Y.B., and Brutsaert, W., 1988a, Influence of basin morphology on groundwater outflow: Water Resources Research, v. 24, no. 10, p. 1645–1650.
- Zecharias, Y.B., and Brutsaert, W., 1988b, Recession characteristics of groundwater outflow and base flow from mountainous watersheds: Water Resources Research, v. 24, no. 10, p. 1651–1658.
- Zembrzuski, T.J., Jr., and Dunn, Bernard, 1979, Techniques for estimating magnitude and frequency of floods on rural unregulated streams in New York excluding Long Island: U.S. Geological Survey Water-Resources Investigations Report 79–83, 66 p.

**Table 5.** Regression equations tested for estimating low flow of streams in the Susquehanna River basin of New York.

[Terms in column headings are explained following the table; acronyms for variables are defined in the “Data Set” section of the report; a dash (–) indicates value could not be computed for that equation; parentheses enclose values that were recomputed after equations were converted to cubic feet per second; ft<sup>3</sup>/s, cubic feet per second; <, less than; R<sup>2</sup>, coefficient of determination]

Equation number (A)	Number of stations (B)	Adjusted R <sup>2</sup> (C)	Tests of accuracy of equation in estimating low flow			Tests of independent variables			Tests of influence of data values			Tests related to normality	
			Cubic feet per second (D)	Cubic feet per second per square mile (E)	Percent of mean of dependent variable (F)	Median percent error of estimate (G)	Largest p>t (H)	Largest variance-inflation factor (VIF) (I)	Prediction sum of squares (PRESS) (J)	Maximum absolute value of DFITS (K)	Correlation of residuals with normal scores (L)	Non-parametric test for slope (M)	
1	70	0.973	0.97	--	37.4	95	<0.0001	1.0	88	4.2	0.826	-0.248	
2	70	0.982	0.80	--	30.7	80	<0.0001	1.0	57	3.8	0.882	-0.139	
3	70	0.983	0.765	--	29.4	68	0.0096	1.4	57	3.5	0.892	-0.193	
4	65	0.984	0.76	--	28.0	96	0.0098	1.4	52	3.4	0.880	-0.264	
5	70	0.990	0.58	--	22.3	55	<0.0001	2.9	27	1.4	0.926	-0.006	
6	70	0.992	0.54	--	20.9	51	<0.0001	2.8	25	2.7	0.940	-0.105	
7	70	0.986	0.70	--	27.0	65	0.0009	1.4	41	2.8	0.916	-0.073	
8	32	0.987	0.88	--	15.9	23	0.0041	2.2	27	1.1	0.978	0.089	
9	70	0.768	--	0.023	65.7	52	<0.0001	1.0	0.041	2.2	0.865	-0.254	
Equation 9 converted to ft <sup>3</sup> /s: 7Q10 = DAREA (equation above)													
(0.83) (30.7) (52)													
10	65	0.867	--	0.0135	44.7	52	<0.0001	1.0	0.013	0.9	0.983	-0.019	
Equation 10 converted to ft <sup>3</sup> /s: 7Q10 = DAREA (equation above)													
(1.22) (44.7) (52)													

**Table 5.** Regression equations tested for estimating low flow of streams in the Susquehanna River basin of New York.

[Terms in column headings are explained following the table; acronyms for variables are defined in the “Data Set” section of the report; a dash (–) indicates value could not be computed for that equation; parentheses enclose values that were recomputed after equations were converted to cubic feet per second; ft<sup>3</sup>/s, cubic feet per second; <, less than; R<sup>2</sup>, coefficient of determination]

Equation number (A)	Number of stations (B)	Tests of accuracy of equation in estimating low flow				Tests of independent variables			Tests of influence of data values			Tests related to normality	
		Adjusted R <sup>2</sup> (C)	Cubic feet per second (D)	Cubic feet per second per square mile (E)	Percent of mean of dependent variable (F)	Median percent error of estimate (G)	Largest pr>t (H)	Largest variance-inflation factor (VIF) (I)	Prediction sum of squares (PRESS) (J)	Maximum absolute value of DFITS (K)	Correlation of residuals with normal scores (L)	Non-parametric test for slope (M)	
11	7010/DAREA = 0.0042 + 0.46 [(SGAREA-LKSWSD) / DAREA] - 0.40 (LKSWSD / DAREA) + 0.021[(MAQ / DAREA) - 1.85] - 0.00059(UNDER1 / DAREA)	0.903	--	0.0115	38.1	36	0.0127	1.6	0.010	0.9	0.994	-0.193	Equation 11 converted to ft <sup>3</sup> /s: 7Q10 = DAREA (equation above)
12**	Ln(7010+1.0) = 0.173 + 0.996 Ln{[(SGAREA - LKSWSD)+1.0] / (DAREA+1.0)} + 0.722 Ln(MAQ + 1.0)	(0.63)	(23.4)	(36)	74	<0.0001	1.1	4.6	1.0	0.973	-0.204	Equation 12 converted to ft <sup>3</sup> /s: 7Q10 = 1.19 C (equation above) - 1.0	
13	702 = -0.028 + 0.74 (SGAREA-LKSWSD) + 0.025 (TLAREA) + 0.067 [(MAQ / DAREA) - 1.85] DAREA	0.990	1.01	--	22.7	42	<0.0001	8.2	109	5.1	0.903	-0.183	(1.94) (74.8) (104)
14	702 = 0.412 + 0.70 (SGAREA-LKSWSD) + 0.019 (TLAREA) + 0.059 [(MAQ / DAREA) - 1.85] DAREA	0.957	1.40	--	19.0	23	0.032	2.4	74	1.6	0.993	0.117	
15	90%FD = -0.013 + 0.876 (SGAREA-LKSWSD) + 0.059 (TLAREA) + 0.114 [(MAQ / DAREA) - 1.85] DAREA	0.989	1.35	--	23.3	30	<0.0001	8.2	184	3.9	0.897	-0.154	
16	90%FD = 0.58 + 0.83 (SGAREA-LKSWSD) + 0.052(TLAREA) + 0.106 [(MAQ / DAREA) - 1.85] DAREA	0.951	1.95	--	20.3	23	0.0001	2.4	149	1.6	0.982	0.154	
17	702/DAREA = -0.00069 + 0.475 [(SGAREA-LKSWSD)/(MAQ / DAREA) / DAREA]	0.898	--	0.018	32.7	29	<0.0001	1.0	0.022	0.8	0.987	-0.173	Equation 17 converted to ft <sup>3</sup> /s: 7Q2 = DAREA (equation above)
18	90%FD/DAREA = 0.026 + 0.56 [(SGAREA-LKSWSD)/(MAQ / DAREA) / DAREA] + 0.034[(MAQ / DAREA) - 1.85]	0.897	--	0.022	28.9	31	0.0007	1.0	0.033	0.8	0.993	-0.081	Equation 18 converted to ft <sup>3</sup> /s: 90%FD = DAREA (equation above)
		(1.54)	(25.2)	(31)									

\*For equation 12 only: standard error is in natural logarithm units, not in cubic feet per second per square mile.

\*\*Ln = natural logarithm (base e). For conversion to ft<sup>3</sup>/s: 1.19 = e<sup>0.173</sup>; C = detransformation bias correction (Miller, 1984) = e<sup>0.03105</sup> = 1.03154

## Explanation of Columns in Table 5

### Tests of Accuracy of Equation in Estimating Low Flow (columns C–G)

#### Adjusted R<sup>2</sup> (coefficient of determination)

Adjusted R<sup>2</sup> can range from 0 to 1 and shows the fraction of the variation in the dependent variable (low flow) that 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 below):

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

#### Standard error

Standard error is expressed in units of the dependent variable (cubic feet per second, cubic feet per second per square mile, or natural logarithm of cubic feet per second), and in percent of the mean of the dependent variable. It is computed as follows:

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

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

where:

- $n$  = number of observations (measurement sites)
- $p$  = number of independent variables in the equation
- $q_o$  = observed low flow
- $q_e$  = low flow estimated by regression equation
- $TSS$  = total sum of squares
- $SSE$  = sum of squares of error  $[\sum (q_o - q_e)^2]$

#### 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 data set.

For each equation that estimates low flow in cubic feet per second per square mile, or in the natural logarithm of low flow plus 1.0 cubic feet per second, the estimate for each station was converted to low flow in cubic feet per second. From the array of low-flow estimates thus generated and the corresponding array of observed low flows in cubic feet per second, standard error in cubic feet per second, standard error in percent of the mean of the observed values, and median percent error of estimate were computed as described above. The resulting values are given in parentheses in table 5.

## Tests of Independent Variables (columns H, I)

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, Inc., 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.) In table 5, the largest (least significant) pr>t statistic for any independent variable in each equation is listed. The pr>t statistics for all other independent variables in each equation indicated strong significance. The pr>t statistic was also computed for the intercept term in each equation, and commonly indicated a lesser significance for that term.

## 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  $R^2 > 0.9$ ) 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 Influence of Data Values (columns J, K)

### Prediction sum of squares (PRESS)

A "prediction residual" can be calculated for each streamflow station as follows: first remove that station from the data set, then estimate low flow for that station from a regression analysis that includes only 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:  $PRESS = \sum_{i=1}^n e_i^2$ .

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.

### Maximum DFFITS

The DFFITS statistic is computed for each station in the data set for 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 data set. The maximum absolute value thus computed is listed in table 5 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 (columns L, M)

### 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. Normal scores for a particular set of data values are an array of the same number of values that have exactly a normal distribution. If the distribution of values in the data set is nearly normal, a plot of the data values against normal scores (each ranked in order from large to small) would approach a straight line, and the coefficient of correlation of the data values with the normal scores would approach 1. 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=BLM option, and PROC CORR (SAS Institute, Inc., 1990). If the correlation coefficient is greater than 0.983 for regressions with 70 stations, 0.981 for 65 stations, or 0.965 for 31 stations, the distribution of residuals is accepted as normal at the 0.05 alpha (5 percent significance) level (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 0.235 (for 70 stations), 0.244 (for 65 stations), or 0.349 (for 32 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).

This page has been left blank intentionally.

Prepared by the Pembroke and West Trenton  
Publishing Service Centers

For additional information write to:  
New York Water Science Center  
U.S. Geological Survey  
425 Jordan Road  
Troy, NY 12180

Information requests:  
(518) 285-5602  
or visit our Web site at:  
<http://ny.water.usgs.gov>

