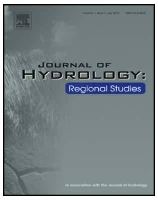
Journal of Hydrology: Regional Studies 4 (2015) 583–594

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh



Watershed area ratio accurately predicts daily streamflow in nested catchments in the Catskills, New York

Chris C. Gianfagna[a], Chris E. Johnson[a,*], David G. Chandler[a], Charlie Hofmann[b]

[a] Department of Civil and Environmental Engineering, 151 Link Hall, Syracuse University, Syracuse, NY 13244, USA
[b] Stone Environmental, Inc., 535 Stone Cutters Way, Montpelier, VT 05602, USA

ARTICLE INFO

Article history:
Received 21 January 2015
Received in revised form 28 August 2015
Accepted 6 September 2015
Available online 19 September 2015

Keywords:
Catskills
Flow scaling
Nested watersheds
New York
Streamflow
Watershed

ABSTRACT

Study region: The Catskills region of New York State is largely forested and dominated hydrologically by stream watersheds with few natural lakes. The area experiences intensive water resources management and ecosystem monitoring due to its strategic role as the principal water supply for New York City.
Study focus: We analyzed average daily flows in nested and non-nested pairs of gaged watersheds in the Catskills to assess whether daily flow in ungaged watersheds can be calculated based on watershed area ratios.
New hydrological insights for the region: Watershed area ratio was the most important basin parameter for estimating flow at upstream sites based on downstream flow. The area ratio alone explained 93% of the variance in the slopes of relationships between upstream and downstream flows. Regression analysis indicated that flow at any upstream point can be estimated by multiplying the flow at a downstream reference gage by the watershed area ratio. This method accurately predicted upstream flows at area ratios as low as 0.005. We also observed a very strong relationship ($R^2 = 0.79$) between area ratio and flow–flow slopes in non-nested catchments. Our results indicate that a simple flow estimation method based on watershed area ratios is justifiable, and indeed preferred, for the estimation of daily streamflow in ungaged watersheds in the Catskills region.

1. Introduction

The estimation and modeling of water availability and quality for water supply and ecological assessment requires reliable estimation of flow (Vogel et al., 1997). The U.S. Geological Survey (USGS) maintains an extensive network of stream gages for this purpose. However, recent budget cuts have resulted in reductions in the total number of gages in the network, especially in headwater catchments. Consequently, future water resources development projects, and studies of chemical fate and transport in surface waters are likely to require streamflow data at ungaged sites. The ability to estimate flow in ungaged catchments is therefore important for water resources planning and environmental management.

Historically, flow rates in ungaged catchments have been estimated using a variety of techniques. Perhaps the earliest and most common technique for estimating daily flow in an ungaged catchment is the watershed area ratio method. The area ratio method is used to estimate flow in an ungaged catchment when a nearby gaged watershed is present for use as

---

[*] Corresponding author.
E-mail address: cejohns@syr.edu (C.E. Johnson).

http://dx.doi.org/10.1016/j.ejrh.2015.09.002

a reference. The method estimates flow at an ungaged location by multiplying the measured flow at the nearby reference gage by the area ratio of the ungaged to gaged watersheds (Archfield and Vogel, 2010):

$$Q_{\text{ungaged}} = Q_{\text{gaged}} \times \frac{A_{\text{ungaged}}}{A_{\text{gaged}}} \tag{1}$$

in which Q represents streamflow and A represents watershed area. A major assumption of the area ratio method is that flow scales directly with watershed area. That is, as watershed area increases, flow rate increases at some fixed rate per unit area. This means that the flow per unit area is the same at both the ungaged location and gaged reference location. Other techniques include empirical regional regression models (Riggs, 1990), use of flow duration curves (FDCs) (Castellarin et al., 2004), and models developed from rainfall-runoff relationships (Post and Jakeman, 1999).

The choice of reference gage in the area ratio method has generally been determined by geographic proximity to the ungaged watershed of interest, or by locating a watershed that should share a similar hydrologic response as the ungaged watershed of interest (Archfield and Vogel, 2010). Mohamoud (2008) suggests choosing the closest stream gage, while Smakhtin (1999) suggests that several reference stream gages should be used in order to smooth out any timing-related issues between the ungaged and reference locations. Recently, Archfield and Vogel (2010) suggested a “Map Correlation Method”, a new technique for identifying the most correlated stream gage based on watershed characteristics and hydrologic response.

The watersheds in the Catskills Mountain region of New York State feed into the principal water supply reservoirs for New York City. Consequently, New York has a keen interest in monitoring streamflow in the watersheds. To this end, the city provides financial support to augment the network of gages maintained by the USGS. This dense network provides an opportunity to examine the scaling of flow in watersheds with nested gages. We hypothesized that the watershed area ratio (Eq. (1)) could accurately predict mean daily streamflow at the upstream locations in nested pairs of stream gages (Hypothesis 1). If true, then daily flow at any ungaged site can be easily estimated since all of the major streams in the region are gaged near the water-supply reservoirs. Additionally, we hypothesized that the prediction of flow using the area ratio method would be better in nested stream gage pairs than in non-nested stream gage pairs (Hypothesis 2), and that prediction in non-nested pairs would be best when the gages are closest to each other (Hypothesis 3).

2. Setting

This study is set in the Catskills Park of New York State (Fig. 1). The Catskills region is a mountainous area that contains many small streams, and a very high concentration of currently and historically active USGS stream gaging stations. The bedrock is comprised of relatively flat-lying sedimentary rocks (primarily sandstones and mudrocks) of Devonian age, which have been uplifted and tilted slightly to the west (Ver Straeten, 2013). Subsequent erosion produced a network of narrow river valleys. The geologically recent glacial activity in the Catskills is largely responsible for the region’s surficial bedrock, soil, and hydrologic characteristics. Glacial scour and erosion caused by meltwater deepened and re-routed existing drainages, creating a dense network of streams with few natural lakes (Fig. 1) (Rich, 1934).

Most of the region’s soils are underlain by glacial till, which has had significant influence as a parent material on the development of the soils, as well as their corresponding hydrologic response (Kudish, 2000). Although plot-scale heterogeneity in soil texture is common, the overwhelming majority of soils in the Catskills are classified as inceptisols, characterized by a sandy loam texture and poor horizon development (Kudish, 2000). Fragipans, dense cement-like layers that impede root growth and water infiltration, are also fairly common and widespread throughout the region (Kudish, 1979). Average soil depth to C horizon or bedrock in 25 upland catchments was estimated to be 57 ± 2.5 cm (Johnson, 2013), though soils in valley bottoms can be much deeper. Shallow upland soils produce a relatively uniform and flashy hydrologic response to rainfall and snowmelt events.

As a region, the Catskills are largely forested, though quite varied in composition (Kudish, 2000). At the lowest elevations, southern hardwoods are found, dominated by white and red oak, American chestnut and hickory. As elevation increases, southern hardwoods give way to northern hardwoods, dominated by yellow birch, American beech, and sugar maple, and at higher elevation, boreal forests with red spruce, balsam fir and paper birch. On the highest peaks, pockets of alpine meadow vegetation can still be found.

The climate in the Catskills is characterized by cold winters and moderately warm summers. Average annual temperature at the Winnisook site on Slide Mountain is approximately 5 °C (Stoddard and Murdoch, 1991). Precipitation is distributed evenly through the year, with an annual precipitation gradient from the northern Catskills (90–100 cm yr^{-1}) to the southern part of the region (150–160 cm yr^{-1} in the upper East Branch of the Neversink River watershed) (Stoddard and Murdoch, 1991). Precipitation comes from both coastal storms from the south and frontal systems from the west. At Biscuit Brook, in the southern Catskills, approximately 15% of the annual precipitation falls as snow (Stoddard and Murdoch, 1991).

The soils and forests of the Catskills region produce surface waters of exceptional quality. Beginning in the early 20th century, New York City built six reservoirs in the region, which now provides more than 90% of the city’s drinking water. The water provided by these reservoirs is sufficiently pure that it is delivered to residents without filtration (National Research Council, 2000).

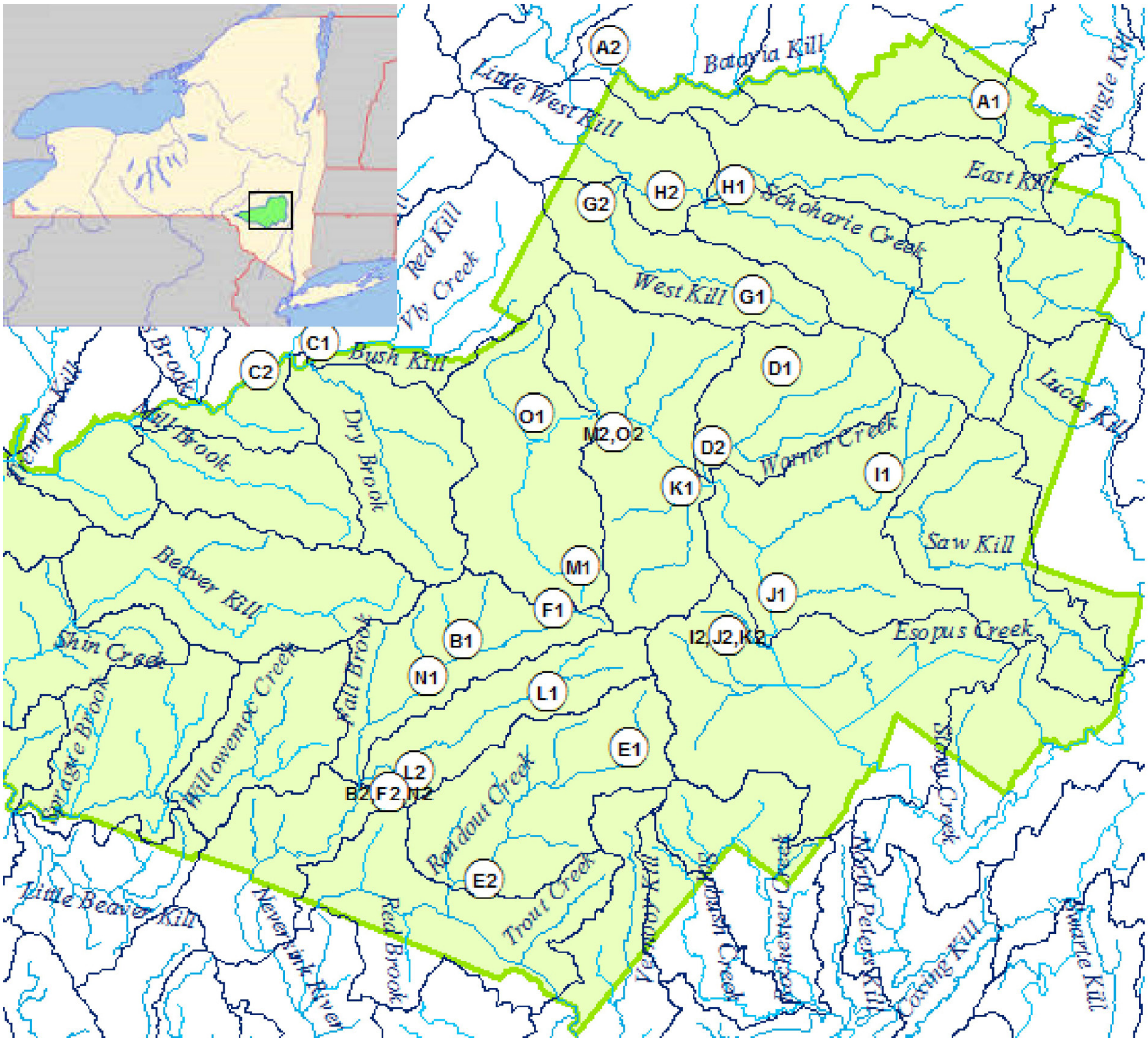


Fig. 1. Map of the Catskill Park (inset: New York State) showing the locations of the model development sites. Letters indicate the watershed, upstream gages are numbered ‘1′, and downstream reference gages are numbered ‘2′ (see Table 1). Note that some of the pairs share the same downstream reference gage.

3. Methods

To test Hypothesis 1 we examined the scaling of mean daily flow by locating nested pairs of stream gages in the study region and regressing the measured daily flows between paired gages. These pairs were located by inspection of a map containing active USGS gaging stations and with the help of the stream gage tool in the USGS StreamStats program (streamstats.usgs.gov). Criteria for selecting pairs included the requirements that: (1) both gaging stations be located within the political boundary of the Catskills Park; (2) the gaging stations have concurrent periods of record of at least 4 years; and (3) the watershed of neither gaging station be affected by man-made impoundments or other flow-altering devices.

Fifteen such pairs were identified and used for testing our hypothesis (Table 1). These pairs are henceforth referred to as model development pairs or model development sites. An attempt was made to use a representative set of nested pairs that covered the geographic region of the study site, and to select pairs that provided a variety of watershed area ratios. The model development pairs had concurrent periods of record that ranged from 4 years (pair 7) to 18 years (pair 12). The gage locations and names for each of the model development pairs are listed in Table 1 and Figure 1 shows their locations within the Park.

Concurrent average daily flow data for each of the model development pairs were downloaded from the USGS National Water Information System website and analyzed using spreadsheet software. The flow at the upstream site was plotted against the flow at the downstream site and a linear functional relation (slope and intercept) was developed for each pair.

Table 1
Gage names and locations for model development pairs.

			Upstream gage/catchment				Downstream gage/catchment			
Pair	Name	Area ratio	Location (Lat/Long)	USGS station number	Gage name	Area (km^2)	Location (Lat/Long)	USGS station number	Gage name	Area (km^2)
1	Batavia Kill	0.0296	42°17′22”, 74°06′59”	01349840	Batavia Kill near Maplecrest, NY	5.26	42°18′30”, 74°23′25”	01329950	Batavia Kill at Red Falls, NY	177.67
2	Biscuit Brook	0.1101	41°59′46”, 74°30′01”	01434025	Biscuit Brook above Pigeon Brook at Frost Valley	9.63	41°55′13”, 74°34′30”	01434498	W. Br. Neversink River at Claryville, NY	87.54
3	Bush Kill	0.5681	42°09′03”, 74°36′06”	01413398	Bush Kill near Arkville NY	120.95	42°08′48”, 74°37′25”	01413408	Dry Brook at Arkville	212.90
4	Hollow Tree	0.0631	42°08′32”, 74°15′55”	01362342	Hollow Tree Brook at Lanesville, NY	5.05	42°06′07”, 74°18′39”	01362370	Stony Clove Creek below Ox Clove	80.03
5	Rondout Creek	0.1399	41°56′13”, 74°22′30”	01364959	Rondout Creek above Red Brook at Peekamoose, NY	13.88	41°51′59”, 74°29′15”	01365000	Rondout Creek near Lowes Corners, NY	99.20
6	Winnisook Creek	0.0228	42°00′40”, 74°24′53”	01434021	W. Br. Neversink at Winnisook Lake	1.99	41°55′13”, 74°34′30”	01434498	W. Br. Neversink River at Claryville, NY	87.54
7	West Kill	0.1841	42°11′06”, 74°16′38”	01349711	West Kill below Hunter Brook near Spruceton, NY	12.87	42°13′49”, 74°23′36”	01349810	West Kill near West Kill, NY	69.93
8	East Kill	0.3678	42°14′57”, 74°18′11”	01349700	East Kill near Jewett Center, NY	92.20	42°14′13”, 74°20′26”	01349705	Schoharie Creek near Lexington, NY	250.71
9	Beaver Kill Trib.	0.0051	42°04′59”, 74°10′59”	01362465	Beaver Kill Tributary above Lake Hill, NY	2.54	42°00′51”, 74°16′16”	01362500	Esopus Creek at Cold Brook, NY	497.28
10	Little Beaver Kill	0.0859	42°01′10”, 74°16′00”	01362497	Little Beaver Kill at Beechford near Mt. Tremper, NY	42.73	42°00′51”, 74°16′16”	01362500	Esopus Creek at Cold Brook, NY	497.28
11	Woodland Creek	0.1073	42°04′47”, 74°20′05”	0136230002	Woodland Creek above Mouth at Phoenicia, NY	53.35	42°00′51”, 74°16′16”	01362500	Esopus Creek at Cold Brook, NY	497.28
12	E. Br. Neversink	0.3900	41°58′01”, 74°26′54”	0143400680	E. Br. Neversink River Northeast of Denning, NY	23.13	41°55′31”, 74°32′26”	0122434017	E. Br. Neversink River near Claryville, NY	59.31
13	Esopus River	0.0242	42°02′01”, 74°25′15”	01362192	Panther Mtn. Trib. to Esopus near Oliverea, NY	3.99	42°07′01”, 74°22′50”	01362200	Esopus Creek at Allaben, NY	164.98
14	High Falls Brook	0.0811	41°58′38”, 74°31′21”	01434105	High Falls Brook at Frost Valley, NY	7.10	41°55′13”, 74°34′30”	01434498	W. Br. Neversink River at Claryville, NY	87.54
15	Birch Creek	0.1962	42°06′32″, 74°27′08″	013621955	Birch Creek at Big Indian	32.37	42°07′01”, 74°22′50”	01362200	Esopus Creek at Allaben, NY	164.98

The slopes of the functional relations were then regressed against the watershed area ratios for each pair to determine the importance of the area ratio in predicting flow at the upstream location based on the downstream reference gage. The utility of the area ratio method was also examined by normalizing the average daily flows by their respective watershed areas and determining the functional relation for each model development pair. If flow scales solely by area, the upstream and downstream flow per unit area should be the same and the slope of the functional relation of these area-normalized plots should be equal to 1. For the purposes of a general model, the functional relation slopes for all fifteen of the area-normalized model development pairs were averaged.

Since the nature of this study involves regressing published flow data, it was determined that a functional relation should be used rather than a least-squares regression relationship. Functional relations are used when the regression assumption that there is no error in the independent variable is unacceptable (Webster, 1997). When relating measured flow data at two gaged sites there is obviously error in both the dependent and independent variables. Webster (1997) offers several alternatives for the variance structure in this situation. In this study, it was assumed that the errors in the dependent and independent variables were proportional to their respective variances. The form of the linear functional relations derived in this study is the familiar equation:

$$y = \beta_0 + \beta_1 x \tag{2}$$

where x = flow at the downstream reference gage (ft^3 s^{-1}).

y = Flow at the upstream gage (ft^3 s^{-1}).

β_0 = Intercept term (ft^3 s^{-1}).

β_1 = Slope of the functional relation (dimensionless).

The slope of the functional relation in the proportional error case is calculated using the following equation (Webster, 1997):

$$\hat{\beta}_1 = \sqrt{\lambda} \tag{3}$$

where $\hat{\beta}_1$ = estimated slope of the functional relation.

$\lambda = {s_y}^2/{s_x}^2$.

${s_y}^2$ = Variance of the flow data for the upstream gage.

${s_x}^2$ = Variance of the flow data for the downstream gage.

The physical meaning of the intercept of the functional relation is the flow at the upstream location when there is no flow at the downstream location. Hydrologically, one would expect the intercept term to be near zero, or perhaps negative, if flow in the upstream catchment ceases before flow in the larger downstream watershed. Computationally, the intercept of the relation is found by the following equation (Webster, 1997):

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} \tag{4}$$

where $\hat{\beta}_0$ = estimated intercept of the functional relation (ft^3 s^{-1}).

$\bar{Y}$ = Average of the upstream daily flow dataset (ft^3 s^{-1}).

$\hat{\beta}_1$ = Slope of the functional relation (dimensionless).

$\bar{X}$ = Average of the downstream daily flow dataset (ft^3 s^{-1}).

Once all of the functional relation slopes and intercepts were calculated, the slopes were plotted against their respective watershed area ratios. The watershed area of each model development site was taken from the USGS web page for each gage. In developing these relationships, it may be reasonably assumed that the dependent variable (the watershed area ratio) is known without error, so linear regression was used to determine the equation of the relationship. The statistical significance of the relationship was then tested and a coefficient of determination (R^2) calculated in order to quantify the importance of the area ratio in estimating daily flow at the upstream location of a pair of nested stream gages.

Hypothesis 2—that predictions of flow based on watershed area ratios would be better using nested gages than using non-nested gages — was tested by comparing the flows in 264 non-nested stream gage pairs within the Catskills Park. The same set of stream gages used to test Hypothesis 1 were used for this analysis. All possible pairs of the 25 individual stream gages were used for the non-nested analysis, after removing nested, inverse and self-same pairs. Flow relationships in the non-nested pairs were analyzed identically to the nested pairs. As with the nested pairs, the functional relation slopes of the flow–flow comparisons for the non-nested pairs were regressed against the watershed area ratio and the statistical significance of the relationship determined. The coefficient of determination (R^2) was calculated in order to quantify the importance of the area ratio in estimating daily flow in non-nested stream gages.

Additionally, the effect of the distance between non-nested gages on the quality of flow predictions (Hypothesis 3) was examined by comparing the coefficient of determination (R^2) for each non-nested pair to the distance between stream gages in the pair. Distance between gages was calculated using a variation of the Haversine Formula, which determines the great-circle distance between two points on a sphere:

$$d = \arccos(\cos(\text{Lat1})) \times \cos(\text{Lat2}) + \sin(\text{Lat1}) \times \sin(\text{Lat2}) \times \cos(\text{Long1} - \text{Long2}) \times r \tag{5}$$

where d = distance between gages (km).

Lat1, Lat2 = latitudes of gages 1,2 (radians).

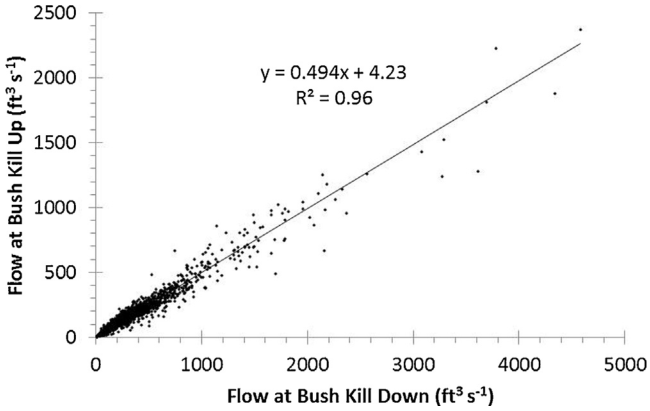


Fig. 2. Flow–flow relationship for the Bush Kill model development pair (number 3 in Table 1). The fitted line is a functional relation. See text and Webster (1997) for details.

Table 2
Functional relation slopes, intercepts, and coefficients of determination (R^2) for the model development pairs.

Pair	Name	Area Ratio	Functional Relation		
			Slope	Intercept	R^2
1	Batavia Kill	0.0296	0.044	−0.027	0.679
2	Biscuit Brook	0.1101	0.098	−0.058	0.877
3	Bush Kill	0.5681	0.494	4.234	0.962
4	Hollow Tree	0.0631	0.047	1.300	0.799
5	Rondout Creek	0.1399	0.171	−0.791	0.883
6	Winnisook	0.0228	0.030	−0.683	0.734
7	West Kill	0.1841	0.214	−0.066	0.904
8	East Kill	0.3678	0.319	1.628	0.966
9	Beaver Kill Tributary	0.0051	0.005	−0.360	0.815
10	Little Beaver Kill	0.0859	0.083	−0.916	0.824
11	Woodland Creek	0.1073	0.127	1.285	0.878
12	E. Branch Neversink	0.3900	0.499	−3.764	0.943
13	Esopus	0.0242	0.031	−1.344	0.530
14	High Falls	0.0811	0.050	2.069	0.873
15	Birch Creek	0.1962	0.142	4.402	0.906

Long1, Long2 = longitude of gages 1,2 (radians).
r = radius of the earth (km).

The relationship between the coefficient of determination for the flow–flow relationship and the distance between gages was determined by least-squares regression.

4. Results

The hypothesis that watershed area ratio is the dominant factor in estimating flows based on reference gages in upstream locations of nested catchments (Hypothesis 1) in the Catskills Park, was found to be reasonable. This is first demonstrated by the strength of the flow–flow relationships that were developed for each of the model development pairs. These plots generally produced strong linear relationships, all of which were statistically significant ($P < 0.05$). An example relationship for pair 3, Bush Kill, is shown in Fig. 2. Our results are consistent with very high correlations ($r = 0.83–1.0$) reported by Shaman et al. (2004) for daily flows in nested and non-nested catchments in the Neversink Basin in the Catskills.

The line displayed in Fig. 2 is the functional relation for the upstream-downstream flow comparison. Note the very high coefficient of determination for this particular example ($R^2 = 0.96$). The watershed area ratio for the Bush Kill model development pair is 0.568 (Table 1). Table 2 contains the functional relation equations and coefficients of determination for each model development pair.

Fig. 3 shows the area-normalized flow relationship for the Bush Kill model development pair. The dashed lined in Fig. 3 is a 1:1 line, representing perfect scaling of streamflow with watershed area. Table 3 includes the functional relation slopes for the area-normalized flow comparisons for all of the model development pairs. The slopes for the 15 model development pairs vary from 0.61 to 1.50, with an average of 1.04. Together with the flow–flow relationships, this suggests that flow in Catskills streams generally scales according to watershed area. This is further demonstrated by Fig. 4, which compares the functional relation slope for each model development pair to its corresponding area ratio. The relationship in Fig. 4 was found to be statistically significant ($P < 0.05$) with the slope not significantly different from 1.0 ($P > 0.05$), and the intercept

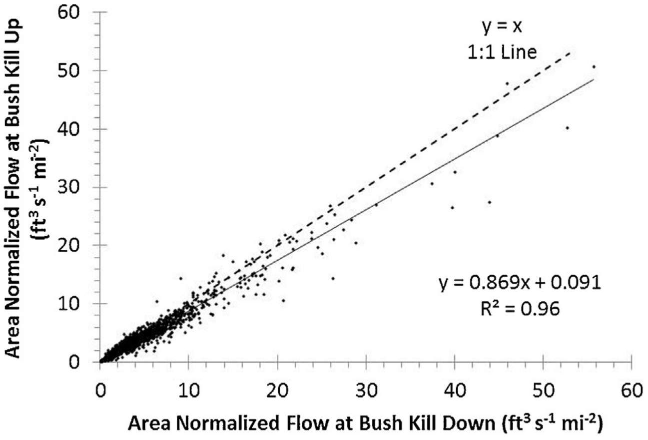


Fig. 3. Area-normalized flow–flow relationship for the Bush Kill model development pair (number 3 in Table 1). The fitted line is a functional relation. See text and Webster (1997) for details.

Table 3
Area-normalized functional relation slopes for the model development pairs.

Model development site	Area ratio	Area-normalized functional relation slope
Batavia Kill	0.0296	1.50
Biscuit Brook	0.1101	0.89
Bush Kill	0.5681	0.87
Hollow Tree	0.0631	0.74
Rondout Creek	0.1399	1.22
Winnisook Creek	0.0228	1.33
West Kill	0.1841	1.16
East Kill	0.3678	0.87
Beaver Kill Tributary	0.0051	0.96
Little Beaver Kill	0.0859	0.96
Woodland Creek	0.1073	1.18
E. Branch Neversink	0.3900	1.28
Esopus River	0.0242	1.26
High Falls Brook	0.0811	0.61
Birch Creek	0.1962	0.72
Average	–	**1.04**

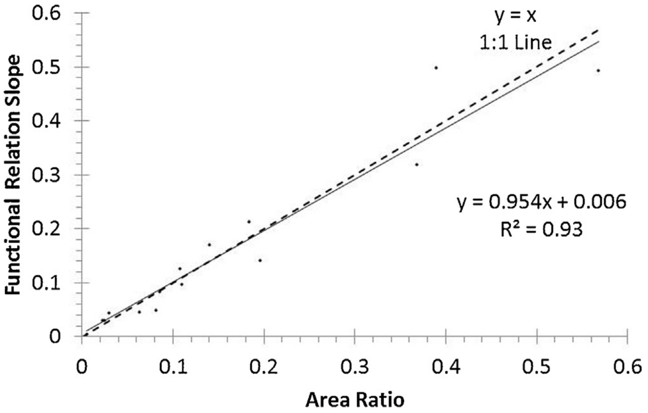


Fig. 4. Relationship between the slopes of the functional relations between upstream and downstream flows and watershed area ratio for the model development pairs. The fitted line is a least-squares regression.

not significantly different from 0 ($P > 0.05$). The coefficient of determination (R^2) for this relationship was 0.93, indicating that the area ratio alone accounts for 93% of the observed variation in the slopes of the functional relations relating upstream to downstream flow in the 15 model development pairs.

The results from the analysis of non-nested pairs (Hypothesis 2) demonstrate that watershed area ratio can also be used effectively to predict flow at an ungaged site based on data from a stream gage in another watershed. However, area ratio only accounted for 79% of the variation in the flow–flow relationships in non-nested pairs, compared to 93% in nested pairs

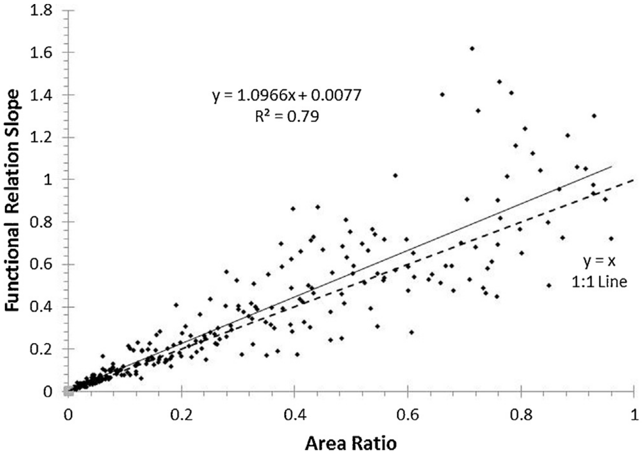


Fig. 5. Relationship between the slope of the functional relations between upstream and downstream flows and watershed area ratio for the non-nested pairs. The fitted line is a least-squares regression.

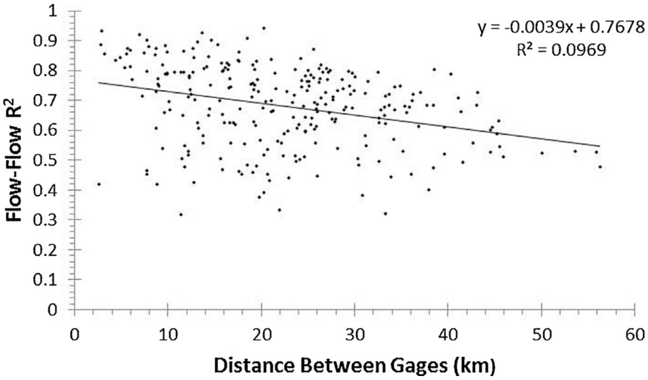


Fig. 6. Relationship between the flow–flow coefficients of determination and the relative distance between gages for the non-nested pairs. The fitted line is a least-squares regression.

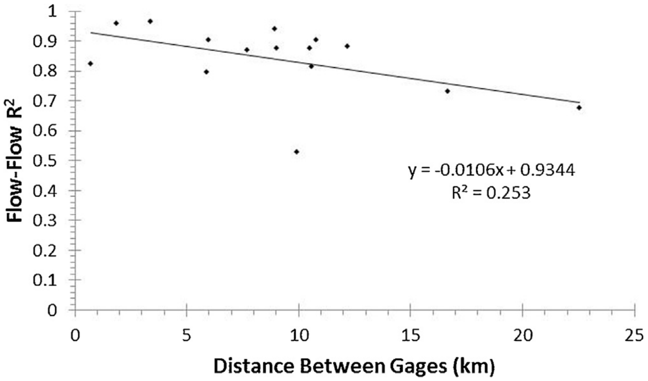


Fig. 7. Relationship between the flow–flow coefficients of determination and the relative distance between gages for the nested pairs. The fitted line is a least-squares regression.

(Fig. 5). The relationship in Fig. 5 was also found to be statistically significant ($P < 0.05$), with an intercept that was not significantly different than 0 ($P > 0.05$). The slope, however, was significantly different than 1.0 ($P < 0.05$), although it was within 10%.

We observed a weak, but statistically significant ($P < 0.05$) relationship between the coefficient of determination for the non-nested flow–flow relationships and the distance between gages. This is demonstrated in Fig. 6, which shows an inverse relationship with a coefficient of determination of 0.097. The strength of the flow–flow regression relationship also decreased with distance in nested catchments (Fig. 7). Nevertheless, gages separated by up to 50 km demonstrated flow–flow relationships with R^2 values of 0.50 or more.

5. Discussion

5.1. Estimation of flow in ungaged basins

Estimating flow in ungaged catchments has been an active area of research in hydrology over the past decade. With the predictions in ungaged basins (PUB) initiative, initiated by the International Association of Hydrological Sciences, many researchers have wrestled with the task of predicting flood statistics, real-time flow, and groundwater characterization in ungaged basins (Wagener and Wheater, 2006). Some of the models that have emerged recently have been quite complex, utilizing several basin characteristics and meteorological data that requires detailed information on both the basin and the climate.

The results of this study, however, suggest that the area ratio of the ungaged to gaged watersheds alone may be adequate for estimating average daily flow based on a reference gage in the Catskills region of New York. This is demonstrated by the statistically significant relationship between area ratio and functional relation slope for both the nested and non-nested pairs (Figs. 4 and 5), which had high coefficients of determination ($R^2 = 0.93$ and $R^2 = 0.79$, respectively). The fact that watershed area ratio alone accounted for 93% of the observed variance in the flow–flow slopes of nested gage pairs clearly indicates that the area ratio largely determines the flow in catchments upstream from a gage. This and the fact that the regression between the slope of the flow–flow relationship and the area ratio produced a slope that was not significantly different from 1.0 and an intercept that was not significantly different from zero indicates that Eq. (1) can be used to estimate daily streamflow at ungaged sites in the Catskills region based on a downstream reference gage.

Other recent studies provide conflicting evidence regarding the generality of this result. Mohamoud and Parmar (2006) found that non-linear regional regression equations based on drainage area alone could predict mean annual streamflow with coefficients of determination between 0.95 and 0.98. Their study considered 75 gaged watersheds in the Mid-Atlantic region, and while their results demonstrate how important drainage area is in regulating the annual flow regime, they also allude to its potential predictive power as an explanatory variable at shorter time scales. In a more recent study, Mohamoud (2008) attempted to predict daily streamflow in the Appalachian region by sequencing constructed FDCs with streamflow at a gaged reference site. In this study Mohamoud compared flow values predicted from his FDC method, and from various forms of the area ratio method, to the actual flow values in the study streams. His model utilized multiple regression to identify explanatory basin and climate characteristics from 26 catchments to develop region-specific FDC construction models. Although each point on the FDCs was generated using only two explanatory variables, the total number of variables used to construct all of the points on the curves exceeded 20 basin and climate characteristics. These characteristics included land use, geomorphology, soil, geology, and climate characteristics, which required the use of geographic information systems (GIS), digital elevation models (DEM), soil survey information, and detailed climate records.

After the development of such a complex model, requiring significant input data – over 20 explanatory variables – and a reference stream gage for streamflow sequencing, the model produced results comparable to those of the area ratio method for the prediction of daily streamflow in the three test watersheds (Mohamoud, 2008). Furthermore, both the predictions made by the FDC method and the area ratio method generally agreed well with the observed flows in the test streams. Although Mohamoud's FDC method does indeed produce good predictions of daily streamflow, the predictions were not significantly better than those made from the area ratio method.

Another example in which a complex model failed to consistently and accurately predict daily streamflow was a 1999 study by Post and Jakeman. This study involved a rainfall-runoff model with 6 explanatory variables that predicted daily streamflow with coefficients of determination ranging for 0.07–0.72 for their 16 test watersheds. Again, a complex model, requiring substantial basin and climate data, did not produce consistently good estimates of daily streamflow. For the Catskills region, where area ratio alone explained 79–93% of the variation in daily flow (Figs. 4 and 5), it is unlikely that more complex approaches, such as those suggested by Post and Jakeman (1999), would greatly improve predictions of flow.

Examples in which approaches based on flow duration curves significantly outperformed area ratio methods include a 2009 study by the Ohio EPA, which examined the White Oak Creek watershed (Ohio EPA, 2009). They concluded that the area ratio method was inadequate for predicting real-time flows. Predicted flows differed from actual flows by an average of 262% and 64% in two test watersheds using the area ratio method, while the FDC method they used showed an average error in predicted versus actual flows of 113% and 35% for the same test watersheds (Ohio EPA, 2009). The data used to assess the performance of the area ratio method were based on the difference in observed and predicted flow from 10 and 12 instantaneous flow measurements for the two test watersheds, while the FDC method utilized a model containing over 50 years of stream data from 10 watersheds in Illinois that has the ability to account for man-made flow-altering devices, such as water withdrawals, which were present in one of the test watersheds. The area ratio method is clearly inappropriate for such catchments, and may not have been appropriate for the estimation of instantaneous flow values in their test watersheds at all. Use of area ratio methods for estimation of instantaneous flow can be significantly affected by lag-time between the reference gage and the point of interest, caused by hydraulic gradient and in-stream storage. In our analysis, the effects of lag-time are diminished by using average daily flow values in streams without natural or artificial storage, and with relatively high gradients. Furthermore, the data for the area ratio analysis for the White Oak Creek study were collected

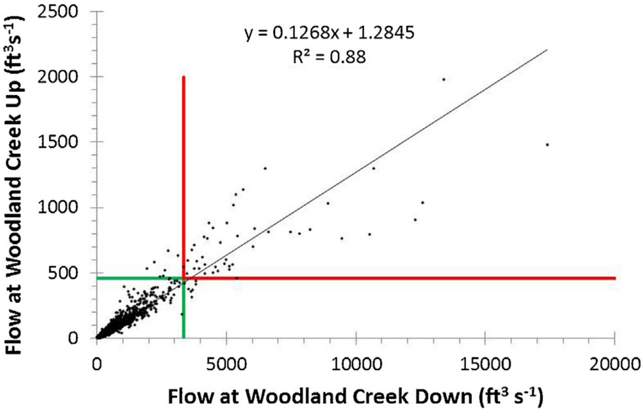


Fig. 8. Flow–flow relationship for the Woodland Creek model development pair (number 11 in Table 1). The vertical and horizontal lines indicate the 98th percentile for flow at both gages.

during summer low flow, which the authors acknowledged as being the worst season for predicting flows based on the area ratio method (Ohio EPA, 2009).

Although the FDC-based model used by Ohio EPA (2009) did outperform the area ratio method, it was based on a small sample size (2 catchments, with 10 or 12 instantaneous measurements in each catchment), with data collected during the worst predictive season (summer), and a comparison of the modeled results to area ratio results from a location inappropriate for its application (a catchment with flow-altering devices). These shortcomings call into question the conclusion that FDCs are significantly better at predicting daily flow in ungaged catchments than methods based on area ratios.

Based on our analysis, good predictions of flow in ungaged basins in the Catskills can be obtained by applying the area ratio method (Eq. (1)), regardless of whether the reference gage is in the same drainage basin or not. The quality of the predictions are better when the reference gage is in the same stream network (nested: Fig. 4) than when the reference gages is in a different basin (non-nested: Fig. 5). The strength of the flow–flow relationships decreases somewhat with distance between gages in both nested and non-nested cases (Figs 6 and 7), but the decrease is small. Therefore, it is better to use a distant reference gage in the same basin than a closer reference gage in a nearby basin.

The relatively minor importance of distance between stream gages was a surprising result of the non-nested analysis. We hypothesized that the closer the two non-nested gages were to each other, the better the predictions of flow would be. The results of this study indicate that distance between stream gages, in the range that we analyzed (0–60 km), is largely unimportant for prediction of flow in non-nested pairs ($R^2 = 0.097$), but somewhat more important when considering nested pairs ($R^2 = 0.25$). This suggests that basin characteristics and basin similarity are more important than geographic proximity in the selection of a reference gage in the Catskills. However, the distance between gages in this study was relatively small (less than 60 km); at larger spatial scales geographic proximity may be a more important factor in selecting a reference gage.

5.2. Prediction of extreme flows

Estimation of extreme flows is a challenge in hydrologic modeling and an important application of the principles for predicting flow in ungaged watersheds. We observed a decreasing quality of fit of the functional relation line to observed flow data at flows above the 98th percentile. As an example, the functional relation for the Woodland Creek model development pair is shown in Fig. 8. The region in the lower left of the graph represents days in which streamflow was below the 98th percentile in both the upstream and downstream gages. The region in the upper right of the graph includes daily flows that exceeded the 98th percentile in both the upstream and downstream gages. The scatter of the data points about the functional relation line increases substantially at flow values greater than the 98th flow percentile. This pattern was observed in most of the model development pairs, but tended to be most dramatic in pairs with low area ratios (<0.1). Although there are relatively few data points in this region of the graph, because they represent the highest flow days of the year, the decreasing quality of fit has important implications for other estimates based on flow data, such as chemical mass fluxes.

Improvements in high-flow prediction may be possible if meteorological data were incorporated into an area-ratio based model for the high-flow regime. Although this would increase the complexity of the model, it could potentially improve flow estimates in the highest flow percentiles. This suggestion is based on the assumption that the scaling of very high flow events may be best explained by meterological phenomena rather than basin characteristics. It is worth noting in this regard that the data used in this analysis included the largest flow event on record in the Catskills region — Hurricane Irene (2014). Local variations in rainfall intensity in such storms, detectable by radar data, are likely to cause variations in streamflow that are only weakly related to watershed area. Analysis of historical meteorological and/or radar data for extreme events could provide a basis for inclusion of a supplementary flow adjustment coefficient based on storm intensity and duration to improve estimation of high flows in ungaged basins.

5.3. Flow-scaling in Catskills watersheds

Hortness (2006) suggested that the watershed area ratio method only be used when the area ratio between the ungaged and reference watersheds is between 0.5 and 1.5. Others have both extended and restricted this range: Koltun and Shwartz (1987) suggested a very limited range of 0.85 to 1.15, while Ries and Friesz (2000) showed that the area ratio method can be used with area ratios as low as 0.3 for low flow estimates. Interestingly, however, only Ries and Friesz (2000) provide any scientific evidence for their suggested range. The other studies simply provided guidelines without any justification beyond common practice. The results of this study, however, indicate that reasonable predictions of average daily flow in the Catskills can be made using Eq. (1) at much lower area ratios. This is supported by the results, which show good correspondence between watershed area ratio and the slope of the relationship between upstream and downstream flow for ratios as low as 0.005 (Table 2, Fig. 4).

Additionally, our analysis suggests that good predictions of flow can be made not only at very low watershed area ratios, but also in relatively small watersheds. Of the seven nested pairs with upstream watershed areas less than 1000 ha (10 km^2) (Batavia Kill, Biscuit Brook, Hollow Tree, Winnisook Creek, Beaver Kill Trib., Esopus River, High Falls Brook), five had coefficients of determination greater than 0.7, despite having relatively small upstream catchment areas. Although we were not able to identify an absolute watershed area at which the relationship breaks down, based on our analysis we are able to speculate that it is likely less than 1000 ha, and possibly less than 200 ha.

The very strong scaling of daily flow according to watershed area, extending to very low watershed areas and watershed area ratios, is somewhat surprising. Much of the winter precipitation in the Catskills falls as snow, producing multiple snowmelt events in the winter and spring. One would expect a differential flow response during these events, with greater flow production at lower elevations than in upper reaches of the watersheds. Our results suggest that this phenomenon does not substantially affect the scaling of flow in Catskills watersheds.

The success of area-ratio-based methods in the Catskills may be the result of several factors. These factors are based primarily on the fact that the region is relatively small and is likely hydrologically homogenous. This is largely because the soil, climate, topography, and basin characteristics are broadly similar throughout the region, creating a fairly predictable hydrologic response, despite differences in distance between gages. The soils in the Catskills are almost exclusively inceptisols (sandy loams), with the presence of fragipans a common occurrence (Kudish, 1979). The soil texture and the presence of fragipans, along with the generally shallow soil depths found at higher elevations (Johnson, 2013), can act to decrease infiltration rates and soil water storage, and promote rapid runoff in upland catchments. Despite small-scale heterogeneities in soil texture and depth, they are likely fairly homogenous at the catchment scale, leading to generally flashy hydrologic response across the region. This is in line with what McDaniel et al. (2008) concluded in a study regarding flashy upland watersheds in Idaho, which also contained fragipans. They determined that shallow soil depths underlain by fragipans were responsible for the flashy hydrologic response observed in their study sites.

Since the Catskills region is relatively small (2900 km^2), the climate, weather patterns, erosional settings, and soil development conditions are similar for the entire region. Basin characteristics, including topography, stream channel characteristics, watershed storage, and land-use conditions are also similar across the region and likely lead to the nearly uniform hydrologic response. Since the Catskills are not true mountains (in an orographic sense), the topography is best explained by alluvial and glacial erosion rather than mountain building processes. This has led to relatively similar channel slopes and stream channel characteristics across the region. Stream channels in the Catskills tend to be relatively straight, steep and well defined, therefore decreasing travel time in the channel and increasing the likelihood of correlation between upstream and downstream flows.

The nearly uniform lack of surface-water storage features in Catskills watersheds also influences hydrologic response. Catskills watersheds rarely contain lakes, large wetlands, or other natural water storage features that would act to slow the hydrologic response, therefore contributing to the consistently flashy response that characterizes the region. Land-use conditions also play an important role in regulating the hydrologic response and are one of the key reasons that area ratio methods are successful in the Catskills. The model development sites used in this study are largely forested and lack urban areas. This helps to increase the hydrologic homogeneity of the sites, and is representative of the region as a whole, which is generally forested and lacks urban centers.

These factors combine to control the relatively simple hydrologic response observed in the Catskills region, where watershed area ratio alone can be used to describe and predict flow at the upstream location of both nested and non-nested pairs of stream gages. This may not be true in hydrologically more complex systems with longer, more sinuous stream channels, regions with significant groundwater contributions, or regions with abundant lakes and wetlands. Nevertheless, the findings of this study strongly support the use of area ratio methods for estimation of daily flow in the Catskills.

6. Conclusions

Based on the results of this study, it is clear that watershed area ratio is the most important basin characteristic for estimating flow at the upstream location of both nested and non-nested pairs of stream gages based on a reference gage in the Catskills. The area ratio explains 93% of the variance in the slopes for the flow–flow relationships for nested stream gage pairs and 79% of the variance for non-nested pairs. Because of these high R^2 values, the area ratio is the only basin parameter required to make reasonable estimates of flow in ungaged catchments for the purpose of estimating daily flow in ungaged basins. The use of overly complex models is not likely to produce consistently better estimates of daily average flow than methods based on area ratios, making the added complexity unwarranted. An exception to this is the prediction of extremely high flow values (>98% flows). For these high flow values agreement between the functional relation line and the observed flow data decreases.

Conflict of interest

The authors confirm that there are no known conflicts of interest associated with this publication.

Acknowledgments

This work was supported by a grant from the New York State Energy Research and Development Authority (contract no. 16299). We thank Janet Marsden for her help in preparing Fig. 1.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ejrh.2015.09.002.

References

Archfield, S.A., Vogel, R.M., 2010. Map correlation method: selection of a reference streamgage to estimate streamflow at ungaged catchments. Water Resour. Res. 46, W10513.
Castellarin, A., Galeati, G., Brandimarte, L., Montanari, A., Brath, A., 2004. Regional flow-duration curves: reliability for ungauged basins. Adv. Water Resour. 27, 953–965.
Hortness, J.E., 2006. Estimating low flow frequency statistics for unregulated streams in Idaho. U.S. Geol. Surv. Sci. Invest. Rep., 06–5035.
Johnson, C.E., 2013. Chemical properties of upland forest soils in the Catskills region. Ann. N. Y. Acad. Sci. 1298, 30–42.
Koltun, G.F., Shwartz, R.R., 1987. Multiple-regression equations for low flows at ungaged streams in Ohio. U.S. Geol. Surv. Water Resour. Invest. Rep., 86–4354.
Kudish, M., 1979. Catskill Soils and Forest History. Catskill Center for Conservation and Development, Inc., Hobart, NY.
Kudish, M., 2000. The Catskill Forest: A History, First ed. Purple Mountain Press, Fleischmanns, NY.
McDaniel, P.A., Regan, M.P., Brooks, E., Boll, J., Barndt, S., Falen, A., Young, S.K., Hammel, J.E., 2008. Linking fragipans, perched water tables, and catchment-scale hydrological processes. Catena 73, 155–173.
Mohamoud, Y.M., 2008. Prediction of daily flow duration curves and streamflow for ungaged catchments using regional flow duration curves. Hydrol. Sci. J. 53, 706–724.
Mohamoud, Y.M., Parmar, R.S., 2006. Estimating streamflow and associated hydraulic geometry, the Mid-Atlantic region, USA. J. Am. Water Resour. Assoc. 42, 755–768.
National Research Council, 2000. Watershed Management for Potable Water Supply: Assessing the New York City Strategy. National Academy Press, Washington, DC.
Ohio EPA, 2009. Total Maximum Daily Loads for the White Oak Creek Watershed. Final Report.
Post, D.A., Jakeman, A.J., 1999. Predicting the daily streamflow of ungauged catchments in S.E. Australia by regionalising the parameters of a lumped conceptual rainfall-runoff model. Ecol. Modell. 123, 91–104.
Rich, J.L., 1934. Glacial Geology of the Catskills. N. Y. State Mus. Bull., 299.
Ries, K.G., Friesz, P.J., 2000. Methods for estimating low-flow statistics for Massachusetts Streams. U.S. Geol. Surv. Water Resour. Invest. Rep., 00–4135.
Riggs, H.C., 1990. Estimating flow characteristics at ungauged sites. In: Beran, M.A., Brilly, M., Becker, A., Bonacci, O. (Eds.), Regionalisation in Hydrology, Proceedings of a Symposium Held at Ljubljana. April 1990, IAHS Publication No. 191.
Shaman, J., Stieglitz, M., Burns, D., 2004. Are big basins just the sum of small catchments? Hydrol. Process. 18, 3195–3206.
Smakhtin, V.U., 1999. Generation of natural daily flow time series in regulated rivers using a nonlinear spatial interpolation technique. Regul. Rivers: Res.Manag. 15, 311–323.
Stoddard, J.L., Murdoch, P.S., 1991. Catskill mountains. In: Charles, D.F. (Ed.), Acidic Deposition and Aquatic Ecosystems: Regional Case Studies. Springe–Verlag, New York.
Ver Straeten, C.A., 2013. Beneath it all: bedrock geology of the Catskill mountains and implications of its weathering. Ann. N. Y. Acad. Sci. 1298, 1–29.
Vogel, R.M., Bell, C.J., Fenessey, N.M., 1997. Climate, streamflow and water supply in the northeastern United States. J. Hydrol. 198, 42–68.
Wagener, T., Wheater, H., 2006. Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty. J. Hydrol. 320, 132–154.
Webster, R., 1997. Regression and functional relations. Eur. J. Soil Sci. 48, 557–566.