JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION

Vol. 49, No. 5

AMERICAN WATER RESOURCES ASSOCIATION

October 2013

ESTIMATION OF DAILY STREAMFLOW OF SOUTHEASTERN COASTAL PLAIN WATERSHEDS BY COMBINING ESTIMATED MAGNITUDE AND SEQUENCE[1]

Herbert Ssegane, Devendra M. Amatya, E.W. Tollner, Zhaohua Dai, and Jami E. Nettles[2]

ABSTRACT: Commonly used methods to predict streamflow at ungauged watersheds implicitly predict streamflow magnitude and temporal sequence concurrently. An alternative approach that has not been fully explored is the conceptualization of streamflow as a composite of two separable components of magnitude and sequence, where each component is estimated separately and then combined. Magnitude is modeled using the flow duration curve (FDC), whereas sequence is modeled by transferring streamflow sequence of gauged watershed(s). This study tests the applicability of the approach on watersheds ranging in size from about 25-7,226 km^2 in Southeastern Coastal Plain (U.S.) with substantial surface storage of wetlands. A 19-point regionalized FDC is developed to estimate streamflow magnitude using the three most selected variables (drainage area, hydrologic soil index, and maximum 24-h precipitation with a recurrence interval of 100 years) by a greedy-heuristic search process. The results of validation on four watersheds (Trent River, North Carolina: 02092500; Satilla River, Georgia: 02226500; Black River, South Carolina: 02136000; and Coosawhatchie River, South Carolina: 02176500) yielded Nash-Sutcliffe efficiency values of 0.86-0.98 for the predicted magnitude and 0.09-0.84 for the predicted daily streamflow over a simulation period of 1960-2010. The prediction accuracy of the method on two headwater watersheds at Santee Experimental Forest in coastal South Carolina was weak, but comparable to simulations by MIKE-SHE.

(KEY TERMS: flow magnitude; flow sequence; ungauged watersheds; flow duration curve; regional equations; watershed variables.)

Ssegane, Herbert, Devendra M. Amatya, E.W. Tollner, Zhaohua Dai, and Jami E. Nettles, 2013. Estimation of Daily Streamflow of Southeastern Coastal Plain Watersheds by Combining Estimated Magnitude and Sequence. *Journal of the American Water Resources Association* (JAWRA) 49(5): 1150-1166. DOI: 10.1111/jawr.12077

INTRODUCTION

Hydrologic modeling at a watershed scale is a scientific practice that aids decision making in water resource planning, flood forecasts, management of surface runoff, sediment, nutrient leaching, and pollutant transport processes. However, the modeling process is complicated by limited understanding of how physical processes scale from point and hillslope observations to integrated complex watershed interactions (Sivapalan *et al.*, 2003; Wagener and Montanari, 2011). Even with limited understanding of physical processes that drive hydrological processes

[1]Paper No. JAWRA-12-0149-P of the *Journal of the American Water Resources Association* (JAWRA). Received June 18, 2012; accepted February 6, 2013. © 2013 American Water Resources Association. **Discussions are open until six months from print publication**.

[2]Respectively, Postdoctoral Research Associate (Ssegane), College of Engineering, University of Georgia, Driftmier Engineering Center, 597 DW Brooks Drive, Athens, Georgia 30602; Research Hydrologist (Amatya), Center for Forested Wetlands Research, USDA-FS, Cordesville, South Carolina 29434; Professor and Graduate Coordinator (Tollner), College of Engineering, University of Georgia, Athens, Georgia 30602; Postdoctoral Research Associate (Dai), University of New Hampshire, Durham, New Hampshire 03824; and Research Hydrologist (Nettles), Weyerhauser Company, Columbus, Mississippi (E-Mail/Ssegane: hssegane@fs.fed.us).

at a watershed scale, there are several watershed hydrological models with different model structures, process conceptualization, and required data input and resolution. Singh and Frevert (2002a, b; 2006), and USEPA (2008) discuss over 40 watershed models commonly used by water resource managers, engineers, and hydrologists. Regardless of the complexity of physically based and conceptual models, they may fail to model physical processes at watershed scale because of heterogeneity of watershed descriptors and complex flow dynamics (He *et al.*, 2011). This challenge is minimized at gauged watersheds by calibration of the models. Calibration determines model parameters using historical observations at a specified watershed location such that future predictions of system response can be inferred. The concept of model calibration is based on the assumption that past observations and watershed responses are strong predictors of future system response (Gupta *et al.*, 2003). These modeling challenges are compounded in ungauged and poorly gauged watersheds, in watersheds where monitoring has been discontinued, and in watersheds with few years of observations.

Some of the approaches for simulating streamflow time series at ungauged and poorly gauged watersheds include: (1) use of parameterized rainfall-runoff models (e.g., Bárdossy, 2007; Götzinger and Bárdossy, 2007; Bastola *et al.*, 2008; Hughes *et al.*, 2010), where hydrologic properties of ungauged watersheds are inferred from calibrated rainfall-runoff model parameters of nearby gauged watersheds or use of regionalized model parameters; and (2) statistical regionalization (e.g., Engeland and Hisdal, 2009; Zhu and Day, 2009), where regression analysis is used to correlate hydrological responses of watersheds to physical and climatic attributes at a regional scale. The regionalization may be based on regional frameworks such as hydrologic landscape regions (Wolock *et al.*, 2004) and ecoregions (Omernik and Bailey, 1997), or on watersheds classified as hydrologically similar using cluster analysis. According to Rao and Srinivas (2008), other methods of regionalization include: (1) method of residuals; (2) canonical correlation analysis; and (3) region of influence. He *et al.* (2011) review and generalize approaches for continuous flow predictions in ungauged watersheds using rainfall-runoff models in two groups: one generalization is the distance-based regional analysis that uses geographic proximity or some functional distance and spatial interpolation (e.g., McIntyre *et al.*, 2005), and the second is the regression-based regional analysis that uses multiregression analysis (e.g., Fernandez *et al.*, 2000; Pechlivanidis *et al.*, 2010). The downside of rainfall-runoff models is their relatively higher learning curve with respect to large number of parameters, their estimation, and higher computation time in contrast to regionalized empirical models (e.g., Guimares and Bohan, 1992; U.S. Geological Survey, U.S. Dept. of Interior, 2000; Grover *et al.*, 2002; Kroll *et al.*, 2004; Schilling and Wolter, 2005; Castellarin *et al.*, 2007; Verdin and Worstell, 2008; Feaster *et al.*, 2009; Zhu and Day, 2009).

One modeling approach that has not yet been fully explored is the conceptualization of daily streamflow as a composite of magnitude and sequence, such that each component (magnitude and sequence) can be modeled independently. This approach is referred to as the streamflow separation (SFS) technique (Mohamoud, 2008; Ssegane, 2011). The magnitude is modeled using the flow duration curve (FDC) whereas the streamflow sequence is modeled by transferring hydrological sequence of one or more neighboring gauged (donor) watersheds to the ungauged (target) watershed. Vogel and Fennessey (1994) describe the FDC as a complement of the cumulative distribution of streamflow that relates streamflow magnitude and frequency on a daily or weekly or monthly or annual basis. Therefore, this study interchangeably uses FDC to refer to streamflow magnitude and magnitude to refer to FDC. The streamflow sequence is defined as the timing or the temporal occurrence of specific streamflow magnitudes and therefore determines the date or the Julian day number when a specific magnitude occurred during the period of interest. The magnitude and sequence are then combined to form daily streamflow series by sorting the estimated magnitude by date associated with the same frequency of exceedence of the donor watershed. The assumption behind transfer of streamflow sequence between neighboring watersheds is that geographic proximity may infer similarity of climate, hydrology, watershed form, and geology such that timing of different flow regimes of neighboring watersheds is similar. Earlier work by Fennessey (1994) referred to the approach as QPPQ transform method whereas Hughes and Smakhtin (1996) and Smakhtin and Masse (2000) referred to the approach as spatial interpolation. However, Smakhtin and Masse (2000) used an index derived from daily precipitation to predict sequence. The use of the term SFS in this study is to emphasize the compartmentalization of daily streamflow into two independent components, such that each component is independently optimized for streamflow predictions at ungauged sites. The above studies use a parameter-based regression to estimate streamflow percentiles and a single neighboring gauged watershed to translate FDC to time series. Work by Archfield *et al.* (2010) and Mohamoud (2008) directly regressed watershed variables against streamflow percentiles instead of parameter-based regression. However, they still use a single neighboring gauged watershed. Shu and

Ouarda (2012) and Ssegane (2011) used more than one neighboring gauged watershed to estimate streamflow sequence. Shu and Ouarda (2012) used regression-based logarithmic interpolation to predict regional flow duration curve (RFDC) at ungauged sites and spatial interpolation to transfer streamflow sequence at 109 catchments in Quebec, Canada. Their approach is similar to work by Hughes and Smakhtin (1996), Smakhtin *et al.* (1997), and Smakhtin (1999). However, they used sequential stepwise regression to generate RFDC and multiple nearby gauges to transfer sequence from donor to target watersheds. Their results showed better performance of RFDC than use of the ratio of drainage areas. The results also showed an improvement in 79% of watersheds when using multiple sites compared with using a single site for spatial interpolation. Patil and Stieglitz (2012) assessed the regional influence on the ability to transfer hydrologic information between neighboring watersheds using 756 watersheds across the continental United States (U.S.). They showed high transferability of hydrologic information for the Appalachian Mountains (Eastern U.S.), the Rocky Mountains, and Cascade Mountains (Pacific Northwest). However, they found lower transferability of hydrologic information between watersheds dominated by evapotranspiration (ET) than those dominated by runoff (e.g., below the Mississipi River). Ssegane (2011) showed that the improvement in transferability of hydrologic information between watersheds in three Mid-Atlantic ecoregions using multiple neighboring gauged watersheds compared with the use of a single closest watershed varied between physiographic provinces. The improvement was consistent with the level of hydrological homogeneity of watersheds in each physiographic province. Their results showed improvement in 57% of the watersheds in Appalachian Plateaus, 81% of watersheds in Piedmont, and 55% of watersheds in Ridge and Valley. The differences in the improvements were also attributed to gauge density of the study areas because more nearby donor watersheds provided better sequence prediction.

Above-mentioned works show that transferability of hydrologic information between neighboring watersheds varies across regions and physiographic provinces. Therefore, the main objective of this study was to assess the capability of the SFS method to predict daily streamflow of watersheds in the Southeastern Coastal Plain, U.S., which are dominated by low gradient topography, higher ET, and high percent of surface storage (percent areal coverage under wetlands and open water surfaces). The method showed satisfactory performance when applied to watersheds in Mid-Atlantic physiographic provinces of Appalachian Plateaus, Piedmont, and Ridge and Valley (Mohamoud, 2008; Ssegane, 2011). This study is different from the above studies in that it directly regresses three watershed variables and 19 streamflow percentiles, and uses linear interpolation as opposed to the stepwise regression, weighted least squares regression, sequential stepwise regression (sequential forward or backward variable selection), and log interpolation used by the above authors to develop RFDC. The three variables were selected using a greedy-heuristic search process (Atallah, 1998) that searches for a local optimal at each streamflow percentile with the assumption that this will generate a global optimal across the entire FDC. Prediction performance is validated using four randomly selected watersheds not used in the development of the RFDC. Also, the developed RFDC is then used to estimate long-term FDC of two first-order streams in the coastal region.

METHODS

Study Area

The study area comprises the U.S. EPA Level III ecoregions of the Southeastern Plains, the Southern Coastal Plain, the Middle Atlantic Coastal Plain, and the Southern Florida Coastal Plains covering the states of North Carolina, South Carolina, Georgia, and Florida (Figure 1; cyan region on the map). Although the area excludes the states of Alabama, Kentucky, Louisiana, Mississippi, Tennessee, Texas, and Virginia, for this analysis, the study area is referred to as the Southeastern Coastal Plains (U.S.). According to Henderson and Grissino-Mayer (2009), the Southeastern Coastal Plain spreads from the Gulf and Atlantic coasts to eastern Texas characterized by low relief ranging from the sea level to about 90 m in elevation. The dominant land use and land cover is the Southeastern pine forest, whereas the climate is classified as a humid subtropical that supports an annual rainfall of 1,170-1,650 mm with mean annual temperatures of 16-23°C (Henderson and Grissino-Mayer, 2009). The mean annual precipitation is generally higher than the mean annual potential evapotranspiration, giving rise to excess soil moisture. The region has many swamps, marshes, and poorly drained soils (Feaster *et al.*, 2009).

Selection of Gauged Sites

The gauged station data was selected from a dataset of 943 stream gauges compiled by Feaster *et al.*

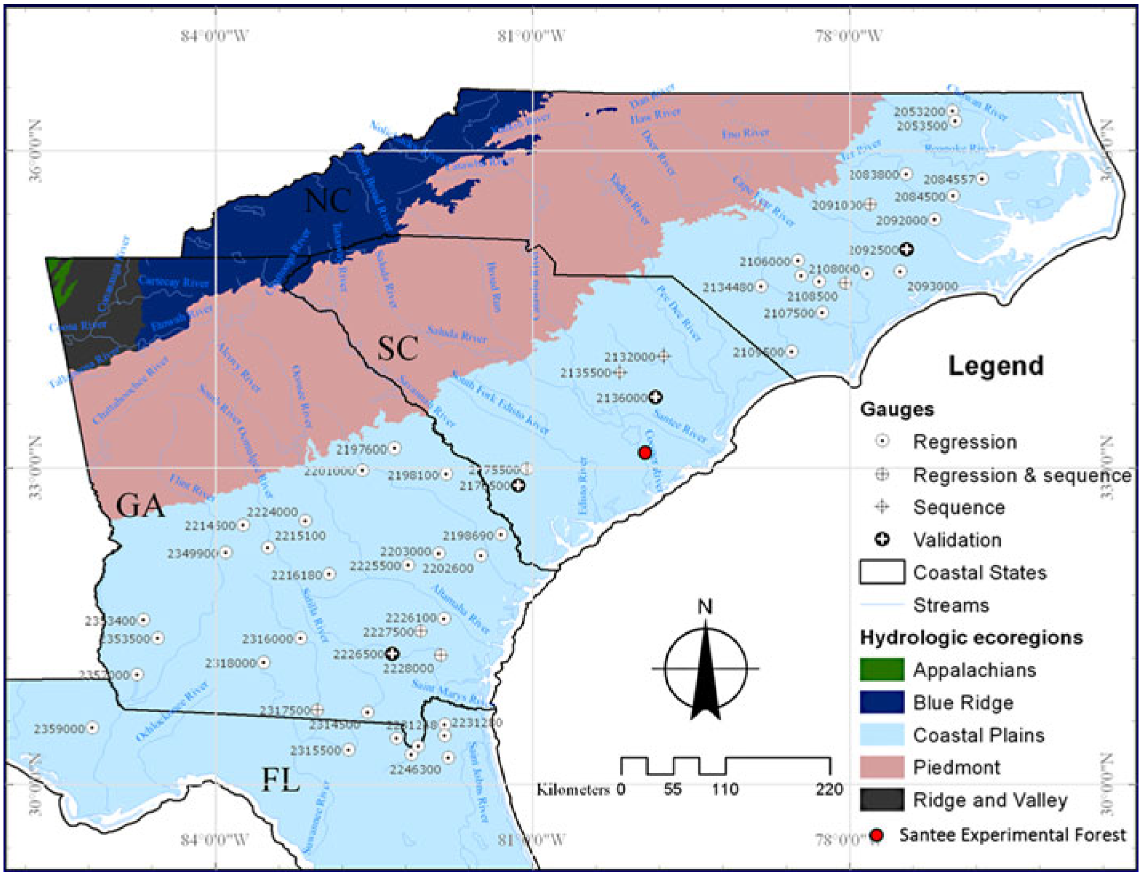


FIGURE 1. States of the Southeastern Coastal Plains and Gauge Locations. The "Regression," "Regression & sequence," "Sequence," and "Validation" gauge numbers correspond to the U.S. Geological Survey (USGS) gauges in Tables 1 and 3 without the leading zero digit.

(2009) for the Southeastern U.S. Two criteria were used to select the gauged sites for regionalization. The first criterion ensured that at least 90% of the selected watershed drained the Southeastern Coastal Plain. Thus, greater than 90% of the drainage area of each selected watershed drained the Coastal Plain compared to draining the Piedmont or Blue Ridge or Ridge and Valley (Figure 1). This reduced the dataset to 214 gauges, of which 206 gauges totally (100%) drained the Southeastern Coastal Plain. The second criterion selected watersheds with long-term data of at least 20 years (20-84 years). This reduced the total number of gauged watersheds to 51, of which 51% had a record period of at least 40 years. The average record period of selected watersheds was 46 years and the range of the respective drainage areas varied from 24.8 to 7,226.1 km^2 (Tables 1 and 2). The preference of such long-term data is because at least 28 years of daily data are required to calculate exceedence probabilities of 0.01 and 99.9% using the Weibull plotting position (Cunnane, 1978). Shu and Ouarda (2012) used a minimum of 10 years whereas Archfield *et al.* (2010) used a minimum of 20 years. Table 1 describes the watersheds used to develop the daily RFDC for the Southeastern Coastal Plain and Figure 1 depicts the geographical location of centroids of each watershed. The majority of selected watersheds are located in North Carolina and Georgia with only three in South Carolina. Of the three watersheds in South Carolina, two were among those randomly selected to validate the accuracy of RFDC and the SFS method. The average of the mean annual precipitation (MAP) (1,263 mm) for the watersheds falls within the long-term range for the region (1,170-1,650 mm).

Watershed Descriptors

Table 2 defines the climatic and physical watershed characteristics used in this study. The selected variables were based on readily available data compiled by Feaster *et al.* (2009) for the Southeastern U.S. The 20 variables are: 10 topographic variables (drainage area, main channel length, watershed perimeter, main channel slope, mean watershed slope, watershed shape factor, mean watershed elevation, maximum watershed elevation,

TABLE 1. Description of Watersheds Used in the Development and Validation of Regional Flow Duration Curve (RFDC).

#	USGSID	Station Name	DA (km^2)	Watershed Centroid LAT (DD)	Watershed Centroid LONG (DD)	Record Years	MAP[1] (mm)
1	2053200	Potecasi Creek near Union, North Carolina	582.7	36.36051	−77.24474	52	1,197
2	2053500	Ahoskie Creek at Ahoskie, North Carolina	163.9	36.26039	−77.13303	60	1,205
3	2083800	Conetoe Creek near Bethel, North Carolina	202.3	35.87000	−77.39311	45	1,205
4	2084500	Herring Run near Washington, North Carolina	24.8	35.58708	−76.99246	29	1,270
5	2084557	Van Swamp near Hoke, North Carolina	59.6	35.71655	−76.80142	33	1,270
6	2091000	Nahunta Swamp near Shine, North Carolina	208.2	35.49399	−77.93544	56	1,240
7	2092000	Swift Creek near Vanceboro, North Carolina	471.4	35.45134	−77.30809	38	1,269
8	2092500[2]	Trent River near Trenton, North Carolina	435.1	35.04093	−77.59321	60	1,323
9	2093000	New River near Gum Branch, North Carolina	243.5	34.90944	−77.55496	23	1,361
10	2106000	Little Coharie Creek, North Carolina	240.4	35.09660	−78.50505	41	1,234
11	2106500	Black River near Tomahawk, North Carolina	1,750.8	35.01365	−78.33317	59	1,257
12	2107000	South River near Parkersburg, North Carolina	981.6	35.12884	−78.60815	34	1,225
13	2107500	Colly Creek near Kelly, North Carolina	279.7	34.61893	−78.41884	20	1,273
14	2108000	N.E. Cape Fear River, North Carolina	1,551.4	35.03035	−77.93131	70	1,306
15	2108500	Rockfish Creek near Wallace, North Carolina	179.5	34.81220	−78.10263	25	1,324
16	2109500	Waccamaw River at Freeland, North Carolina	1,761.2	34.30207	−78.54694	71	1,314
17	2134480	Big Swamp near Tarheel, North Carolina	593.1	34.81934	−78.99461	25	1,226
18	2136000[2]	Black River at Kingstree, South Carolina	3,242.7	33.90164	−80.20200	80	1,226
19	2175500	Salkehatchie River near Miley, South Carolina	883.2	33.20536	−81.27669	59	1,219
20	2176500[2]	Coosawhatchie River, South Carolina	525.8	32.87244	−81.24770	59	1,227
21	2197600	Brushy Creek near Wrens, Georgia	72.5	33.20322	−82.37816	46	1,198
22	2198100	Beaverdam Creek near Sardis, Georgia	79.8	32.96424	−81.87424	24	1,080
23	2198690	Ebenezer Creek at Springfield, Georgia	419.6	32.47279	−81.36640	20	1,080
24	2201000	Williamson Swamp Creek, Georgia	282.3	33.00976	−82.72879	30	1,204
25	2202600	Black Creek near Blitchton, Georgia	600.9	32.28188	−81.64072	30	1,225
26	2203000	Canoochee River near Claxton, Georgia	1,437.4	32.44832	−82.11052	73	1,207
27	2214500	Big Indian Creek at Perry, Georgia	264.2	32.49651	−83.85260	27	1,189
28	2215100	Tucsawhatchee Creek, Georgia	422.2	32.28676	−83.68362	24	1,080
29	2216180	Turnpike Creek near Mcrae, Georgia	127.4	32.03298	−83.02423	27	1,080
30	2224000	Rocky Creek near Dudley, Georgia	162.9	32.54132	−83.21904	24	1,198
31	2225500	Ohoopee River near Reidsville, Georgia	2,874.9	32.51580	−82.47134	73	1,197
32	2226100	Penholoway Creek near Jesup, Georgia	466.2	31.50159	−81.88718	42	1,266
33	2226500[2]	Satilla River near Waycross, Georgia	3,108.0	31.43172	−82.78063	73	1,277
34	2227500	Little Satilla River, Georgia	1,673.1	31.65065	−82.26078	59	1,228
35	2228000	Satilla River at Atkinson, Georgia	7,226.1	31.47670	−82.49819	80	1,264
36	2229000	Middle Prong St. Marys River, Florida	323.7	30.33038	−82.39226	24	1,397
37	2230500	S. Prong St. Marys River, Florida	404.0	30.20386	−82.24248	20	1,397
38	2231000	St. Marys River, Florida	1,813.0	30.36981	−82.27204	84	1,397
39	2231268	Alligator Creek at Callahan, Florida	36.3	30.57661	−81.90391	22	1,397
40	2231280	Thomas Creek near Crawford, Florida	77.4	30.45556	−81.89385	38	1,397
41	2246300	Ortega River at Jacksonville, Florida	80.0	30.30587	−81.84995	37	1,397
42	2314500	Suwannee River at U.S. 441, Florida	2,926.7	30.95954	−82.47591	73	1,300
43	2315500	Suwannee River, Florida	6,293.7	30.80950	−82.57314	83	1,380
44	2316000	Alapaha River near Alapaha, Georgia	1,717.2	31.73860	−83.50924	38	1,228
45	2317500	Alapaha River at Statenville, Georgia	3,626.0	31.48872	−83.30934	79	1,251
46	2318000	Little River near Adel, Georgia	1,494.4	31.44243	−83.66545	30	1,268
47	2349900	Turkey Creek at Byromville, Georgia	116.5	32.24096	−83.83901	52	1,204
48	2353400	Pachitla Creek near Edison, Georgia	486.9	31.68627	−84.74378	22	1,341
49	2353500	Ichawaynochaway Creek, Georgia	1,605.8	31.68175	−84.64807	71	1,329
50	2357000	Spring Creek near Iron City, Georgia	1,364.9	31.29253	−84.79395	28	1,385
51	2359000	Chipola River near Altha, Florida	2,022.8	30.88588	−85.27911	67	1,397

Notes: Four watersheds were randomly selected to validate the RFDC and predicted daily streamflow (Table 3). The remaining 47 watersheds were used for development of RFDC. The actual U.S. Geological Survey gauge number (USGSID) is preceded with a zero.
[1]MAP refers to mean annual precipitation.
[2]Watersheds used for validation of RFDCs and validation of estimated daily and monthly streamflow.

minimum watershed elevation, and drainage density); two land cover variables (% impervious and % forest); six climatic variables (mean annual precipitation and maximum 24-h precipitation at five probabilities of exceedence [50, 10, 4, 2, and 1%]); and two soil variables (soil drainage index and hydrologic soil index).

TABLE 2. Watershed Descriptors (variables).

Variable[1]	Units	Description	Minimum	Maximum	Santee Experimental Forest WS77[2]	Santee Experimental Forest WS80[2]
DA	km^2	Drainage area	24.8	7,226.1	1.55	2.06
MCL	km	Main channel length	12.6	262.7	1.34	1.55
MCS	m/km	Main channel slope	0.095	1.97	1.137	2.468
BP	km	Basin perimeter	41.9	707.4	5.394	6.239
BSF	-	Basin shape factor	2.68	19.03	1.158	1.166
EMEAN	m	Mean elevation	10.4	133.4	8.58	8.44
EMAX	m	Maximum elevation	16.4	176.8	11.04	10.41
EMIN	m	Minimum elevation	1.4	94.2	4.97	3.55
SMEAN	%	Mean slope	0.211	3.665	2.39	2.15
IMPERV	%	Impervious surfaces	0.19	4.54	0.1	0.0
FOREST	%	Forest cover	18.9	56.9	89	69.8
SDI[3]	-	Soil drainage index	3.24	6.21	4.90	5.12
HSI[4]	-	Hydrologic soil index	2.12	3.93	3.50	3.47
MAP	mm	Mean annual precipitation	1,080	1,397	1,370	1,370
RF50[5]	mm	Rain exceeded 50%	88	124	98.7	98.7
RF10	mm	Rain exceeded 10%	136	194	145.6	145.6
RF4	mm	Rain exceeded 4%	168	227	170.0	170.0
RF2	mm	Rain exceeded 2%	188	256	185.8	185.8
RF1	mm	Rain exceeded 1%	205	283	203.2	203.2
DD	km/km^2	Drainage density	0.0	1.6	3.4	2.2

Notes: The minimum and maximum are based on watershed descriptors of 47 watersheds used to develop the regional equations.
[1]Watershed variables are based on work by Feaster *et al.* (2009).
[2]First-order stream watersheds at Santee Experimental Forest, Cordesville, South Carolina.
[3]Mean soil drainage index for the basin (range is 1-7, with 1 denoting excessively drained soils).
[4]Mean hydrologic soil index for the basin (range is 1-4, based on hydrologic soil group, $A = 1$ and $D = 4$).
[5]RFp, Maximum 24-h precipitation with a recurrence interval of ($1/p \times 100$) years.

For a detailed description and calculation methods for each variable, refer to Feaster *et al.* (2009). Table 2 also depicts the 20 watershed descriptors of two first-order coastal streams (WS77 and WS80) at Santee Experimental Forest in Cordesville, South Carolina. These watersheds were included to assess the applicability of RFDC for the estimation of long-term streamflow magnitudes of first-order streams.

Streamflow Magnitude

The SFS method conceptualizes the FDC as streamflow magnitude because the FDC is a graphical representation of magnitude and corresponding frequency. No sequence information (timing of specific magnitudes) is contained in the FDC. Nineteen streamflow percentiles representing high, medium, and low streamflow regimes are used to regionalize the flow duration curve. The 19 streamflow percentiles used in this study ranged from high flows ($Q_{0.01}$-Q_{10}), medium flows (Q_{20}-Q_{70}), to low flows (Q_{80} to $Q_{99.9}$). Where, Q_p represents the flow magnitude equaled or exceeded p percent of the daily flow record. Long-term daily streamflow data for 22 gauged watersheds with over 42 years of record were used to generate streamflow percentiles at the extremes of the FDC ($Q_{0.01}$ and $Q_{99.9}$) whereas daily streamflow for all 51 watersheds (Table 1) was used to generate the other streamflow percentiles for regionalization. Although, only 28 years of daily streamflow data are required to estimate probabilities of exceedence at the extremes ($Q_{0.01}$ and $Q_{99.9}$) using the Weibull plotting position, 42 years were used to minimize the effect of the record length on estimates of the streamflow percentiles at the extremes.

Regionalization of Flow Duration Curves

Four watersheds were randomly selected from the pool of 51 (Figure 1) to validate the SFS method. The remaining 47 watersheds were used to regionalize the flow duration curve. Twenty watershed variables were regressed against each streamflow percentile using a greedy-heuristic search process. The greedy-heuristic search process selected a combination of three variables that gave the highest Nash-Sutcliffe coefficient of efficiency, NSE (Nash and Sutcliffe, 1970) from all possible combinations of three variables out of the given 20 variables. A total of 1,140 possible combinations were evaluated for each of the

19 streamflow percentiles. This was achieved by fitting Equation (1) for every possible three-variable combination at each streamflow percentile, where a, b, c, d, and e are regional coefficients, whereas X_1, X_2, and X_3 are three selected watershed variables. For each three-variable combination, Equation (1) was optimized by minimizing the largest singular value of the prediction error vector using an unconstrained nonlinear optimization algorithm based on Nelder-Mead Simplex method (Lagarias *et al.*, 1998). The approach is referred to as a greedy-heuristic search because it finds a local optimum at each streamflow percentile (optimal combination of three selected variables) in search of a global solution to the entire FDC (Atallah, 1998). This process selected a total of 13 unique variables across the 19 streamflow percentiles. To guarantee model parsimony and field application of the developed equations, the final RFDC (Table 4) equations were developed using only the three most frequently selected variables across the 19 streamflow percentiles. These included drainage area (DA; selected for all 19 percentiles), hydrologic soil index (HSI; selected for 10 percentiles), and maximum 24-h precipitation with a recurrence interval of 100 years (RF1; selected for 7 percentiles). Use of only drainage area did not yield satisfactory results especially for the low flows (NSE less than 0.30 for $Q_p < Q_{60}$).

$$Q_p = a10^b X_1^c X_2^d X_3^d \tag{1}$$

The above described procedure only generates 19 points on a flow duration curve with probabilities ranging from 0.01 to 99.9% and corresponding flow percentiles of $Q_{0.01}$-$Q_{99.9}$. However, the simulation duration of interest is 51 years (1960-2010) of daily streamflow, a total of 18,627 points (18,627 days) on the FDC with corresponding probabilities ranging from 0.005 to 99.995%. This study used linear interpolation and extrapolation to generate all points of the FDC for the period under consideration. Therefore, the total points of the FDC generated by linear interpolation and extrapolation depend on the period of interest.

Regionalized Flow Duration Curve of Two First-Order Streams

Twenty watershed variables (Table 2) of two first-order streams at U.S. Forest Service Santee Experimental Forest (Figure 1; SEF, which is part of Francis Marion National Forest near Charleston, South Carolina) were used to estimate their respective long-term FDC. The two watersheds are WS77 with a drainage area of 1.55 km^2 and WS80 with a drainage area of 2.06 km^2. The watersheds are adjacent to each other with a distance of 2.0 km between their centroids. The watersheds have low relief and are dominated by forest species of loblolly pine, longleaf pine, cypress, and sweet gum. Earlier work on the SEF showed that such first-order watersheds had short retention time of drainage outflow than other watersheds in Southeast U.S. (Young, 1967). Historical records show a higher percentage of flow (drainage/rainfall) from WS77 (27%) compared with WS80 (20%) during the pre-Hurricane Hugo period of 1965-1981 (Richter, 1980; Amatya *et al.*, 2006). As both watersheds were under similar land use, topography, climate, and soils, the observed difference is attributed to difference in hypsometry of the two watersheds. Therefore, this analysis assesses the applicability of regional equations (RFDC) on these two small first-order streams in the region. The results are compared to simulations of daily flow by process-based distributed hydrological model, MIKE-SHE (Dai *et al.*, 2010, 2011) and observed data between 1969 and 1980 (pre-Hurricane Hugo) (Amatya *et al.*, 2006).

Streamflow Sequence

The SFS method defines streamflow sequence as the timing or the temporal occurrence of streamflow magnitudes over a period of interest. Therefore, streamflow sequence determines the date or the Julian day number when a specific magnitude occurred for the period under consideration. This study adopts an approach outlined by Ssegane (2011) for the prediction of streamflow sequence. The approach uses the sequence of nearest gauged watershed (Figure 2), if the difference in distances of the nearest and second nearest donor watersheds from the target watershed is greater than 20 km. For neighboring watersheds where the difference in distances is less than or about 20 km, the approach generates the sequence from aggregated daily streamflows of the two or three closest watersheds. The aggregation is achieved using ensemble techniques such as Pythagorean means (arithmetic and geometric) and bootstrap resampling. Bootstrap resampling (Dixon, 2001) involves generating a new sample (bootstrap sample) by randomly drawing values from the initial sample with replacement and taking the arithmetic mean. This is performed a predefined number of times. For this study, ensemble techniques of geometric mean and bootstrap resampling were implemented for each Julian day.

An arbitrary period of interest for simulation of daily streamflow was chosen as 1960-2010; therefore only calibration watersheds whose period of record covered 1960-2010 were considered for sequence pre-

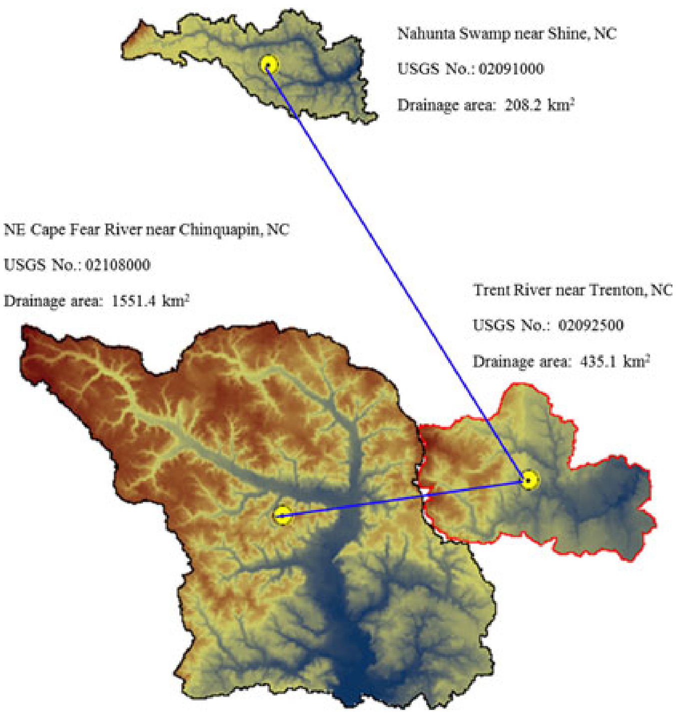


FIGURE 2. Geographic Proximity and Orientation of Two Closest Watersheds to trent River Near Trenton, North Carolina (USGS No. 02092500) and the Respective Euclidean Distances Between Watershed Centroids. Geographic proximity is based on only the sample data used in this study.

diction (Figure 1). For each validation watershed, the nearest watersheds that met the above criteria, and whose Euclidean distance between centroids of target and donor was less than 75 km, were selected (Table 3). The term "nearest" is restricted to only the dataset used in this study and to watersheds that met the above criteria. Daily streamflow was estimated by assigning a date at a donor watershed when a specific flow magnitude at a given probability of exceedence occurred, to the predicted magnitude (RFDC) at the target watershed with the same exact probability of exceedence. Thus, only sequence not the magnitude is transferred from donor to the target watershed.

Performance Criteria

Accuracy of estimated magnitude and estimated daily streamflow during the validation period was evaluated using NSE, NSE based on log-transformed data (log$E$), root mean square error (RMSE), and mean absolute error (MAE). These performance indices are defined by Equations (2)-(5) below (e.g., Perrin *et al.*, 2003; Krause *et al.*, 2005; Reusser *et al.*, 2009). For NSE and log$E$, a value of 1.0 is optimum whereas a value less than zero is indicative of poor model performance because the average of the observed data is a better predictor than the model. For RMSE and MAE, the smaller the value, the better the model estimates.

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}\left(Q_i^o - Q_i^m\right)^2}{\sum_{i=1}^{n}\left(Q_i^o - \overline{Q^o}\right)^2} \tag{2}$$

$$\log E = 1 - \frac{\sum_{i=1}^{n}\left[\log\left(Q_i^o + 1\right) - \log\left(Q_i^m + 1\right)\right]^2}{\sum_{i=1}^{n}\left[\log\left(Q_i^o + 1\right) - \log\left(\overline{Q^o}\right)\right]^2} \tag{3}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Q_i^o - Q_i^m\right)^2} \tag{4}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|Q_i^o - Q_i^m\right| \tag{5}$$

where ${Q_i}^o$ is the i^{th} observed flow (magnitude or time series), ${Q_i}^m$ is the i^{th} modeled or predicted flow; $\overline{Q^0}$ is

TABLE 3. U.S. Geological Survey (USGS) Stream Gauges to Validate the Streamflow Separation (SFS) Method and the Corresponding Nearest Donor Watersheds.

	Neighboring Donor Stations					
	Nearest		Second Nearest		Third Nearest	
Validation USGS No.	**USGS No.**	**Distance[1] (km)**	**USGS No.**	**Distance[1] (km)**	**USGS no.**	**Distance[1] (km)**
02226500	02228000	27.3	02317500	50.6	02227500	55.0
02092500	02108000	30.9	02091000	59.2	-	-
02136000	02135500	3.4	02132000	44.6	-	-
02176500	02175500	37.0	-	-	-	-

Note: Nearest watersheds were selected based on geographic proximity and availability of observed streamflow for the period under consideration (1960-2010).
[1]Euclidean distance (km) between centroids of donor and target watersheds. Euclidean distances greater than 75 km were not used.

the mean of the observed flow; and n is the total number of observations.

RESULTS AND DISCUSSION

Regionalized Flow Duration Curve

Table 4 depicts equations of the RFDC for the Southeastern Coastal Plain (U.S.) based on daily streamflow data of 47 USGS gauged watersheds with a record period spanning 20-84 years. The RFDC is based on only three most selected variables of drainage area (DA), hydrologic soil index (HSI), and maximum 24-h precipitation with a return period of 100 years (RF1), which are generally independent. The selection was achieved using a greedy-heuristic search process, which searches for a global optimal solution by finding locally optimal solutions. Therefore, the three most selected variables from a pool of variables selected at each percentile were used to generate models for the entire FDC (Table 4). The drainage area, which controls the scale of exchange and interaction between climatic and watershed processes, was selected for all 19 percentiles; the HSI, which is a measure of the control of water movement through the soil matrix (drainage potential), was selected for 10 of the 19 percentiles; and the RF1, which is a measure of extreme climatic events, was selected for 7 of the 19 percentiles. Variables of BP, MCL, BSF, SMEAN, FOREST, RF50, and RF10 (refer to Table 2) were not selected for any of the 19 percentiles. This was attributed to their intracorrelation with selected variables and the low information gain (Schroedl, 2010) between the variables and the respective streamflow percentiles. Other selected variables were not used to guarantee a parsimonious model of only three unique independent variables across all 19 percentiles. The selection of DA is consistent with results of other studies that demonstrated the relevance of DA as a major hydrologic scaling parameter, especially for peak and annual flows (e.g., Gupta *et al.*, 1994; Vogel and Sankarasubramanian, 2000; Ogden and Dawdy, 2003; Furey and Gupta, 2005; Segura and Pitlick, 2010). Selection of HSI is supported by the work of Brandes *et al.* (2005), who selected HSI as a significant predictor of recession rates of low flows. Also, the HSI index is used by hydrologic models such as soil and water assessment tool (SWAT) that implement the Curve Number method for flow generation (Borah and Bera, 2004).

The NSE at each estimated streamflow percentile during regionalization is greater than 0.80 (Table 4) therefore the functional form (Equation 1) and selected variables adequately represent flow dynamics. Both the NSE and $\log E$ for high and medium flows ($Q_{0.01}$-Q_{70}) are greater than 0.80. However, the

TABLE 4. RFDC for the Coastal Plain of the Southeastern Region of U.S.

Probability of Exceedence (%)	Regional Equation (L/s)	Prediction Performance	
		NSE	$\log E$
0.01	$Q_{0.01} = 1.68 \times 10^{0.88}\mathrm{DA}^{0.78}\mathrm{HSI}^{-0.28}\mathrm{RF1}^{0.94}$	0.81	0.81
0.05	$Q_{0.05} = 0.86 \times 10^{0.81}\mathrm{DA}^{0.81}\mathrm{HSI}^{-0.11}\mathrm{RF1}^{0.88}$	0.94	0.92
0.1	$Q_{0.1} = 1.95 \times 10^{0.87}\mathrm{DA}^{0.86}\mathrm{HSI}^{-0.03}\mathrm{RF1}^{0.63}$	0.95	0.93
0.5	$Q_{0.5} = 1.84 \times 10^{1.06}\mathrm{DA}^{0.89}\mathrm{HSI}^{-0.18}\mathrm{RF1}^{0.46}$	0.97	0.97
1	$Q_{1} = 2.25 \times 10^{0.82}\mathrm{DA}^{0.89}\mathrm{HSI}^{-0.17}\mathrm{RF1}^{0.47}$	0.98	0.97
5	$Q_{5} = 5.72 \times 10^{-1.23}\mathrm{DA}^{0.99}\mathrm{HSI}^{-0.27}\mathrm{RF1}^{0.93}$	0.99	0.99
10	$Q_{10} = 4.48e^{-9} \times 10^{6.58}\mathrm{DA}^{1.01}\mathrm{HSI}^{-0.40}\mathrm{RF1}^{1.41}$	0.99	0.99
20	$Q_{20} = 5.69e^{-8} \times 10^{3.63}\mathrm{DAHSI}^{-0.63}\mathrm{RF1}^{2.15}$	0.99	0.99
30	$Q_{30} = 1.04e^{-6} \times 10^{2.74}\mathrm{DA}^{1.02}\mathrm{HSI}^{-1.30}\mathrm{RF1}^{2.01}$	0.98	0.98
40	$Q_{40} = 8.99e^{-7} \times 10^{-1.27}\mathrm{DA}^{1.01}\mathrm{HSI}^{-1.24}\mathrm{RF1}^{3.67}$	0.97	0.96
50	$Q_{50} = 3.58e^{-13} \times 10^{2.52}\mathrm{DA}^{0.94}\mathrm{HSI}^{-1.19}\mathrm{RF1}^{4.78}$	0.94	0.91
60	$Q_{60} = 1.07 \times 10^{-10.47}\mathrm{DA}^{0.98}\mathrm{HSI}^{-2.08}\mathrm{RF1}^{5.03}$	0.92	0.87
70	$Q_{70} = 8.86e^{-18} \times 10^{6.33}\mathrm{DA}^{1.02}\mathrm{HSI}^{-3.11}\mathrm{RF1}^{5.21}$	0.91	0.78
80	$Q_{80} = 2.42e^{-16} \times 10^{4.19}\mathrm{DA}^{1.10}\mathrm{HSI}^{-4.30}\mathrm{RF1}^{5.53}$	0.92	0.75
90	$Q_{90} = 34.41 \times 10^{-10.50}\mathrm{DA}^{1.42}\mathrm{HSI}^{-7.23}\mathrm{RF1}^{4.48}$	0.95	0.62
95	$Q_{95} = 3.28e^{-16} \times 10^{6.12}\mathrm{DA}^{1.64}\mathrm{HSI}^{-8.79}\mathrm{RF1}^{4.52}$	0.97	0.55
99	$Q_{99} = 75.07 \times 10^{-18.17}\mathrm{DA}^{2.20}\mathrm{HSI}^{-9.87}\mathrm{RF1}^{6.70}$	0.98	0.46
99.5	$Q_{99.5} = 5.10e^{-13} \times 10^{-2.92}\mathrm{DA}^{3.17}\mathrm{HSI}^{-13.14}\mathrm{RF1}^{5.39}$	0.98	0.43
99.9	$Q_{99.9} = 34.61 \times 10^{-27.59}\mathrm{DA}^{2.73}\mathrm{HSI}^{-10.39}\mathrm{RF1}^{10}$	0.99	0.59

Note: The Coastal Plain comprises U.S. EPA Level III ecoregions: (1) Middle Atlantic Coastal Plain, (2) Southeastern Plain, (3) Southern Coastal Plain, and (4) Southern Florida Coastal Plain.

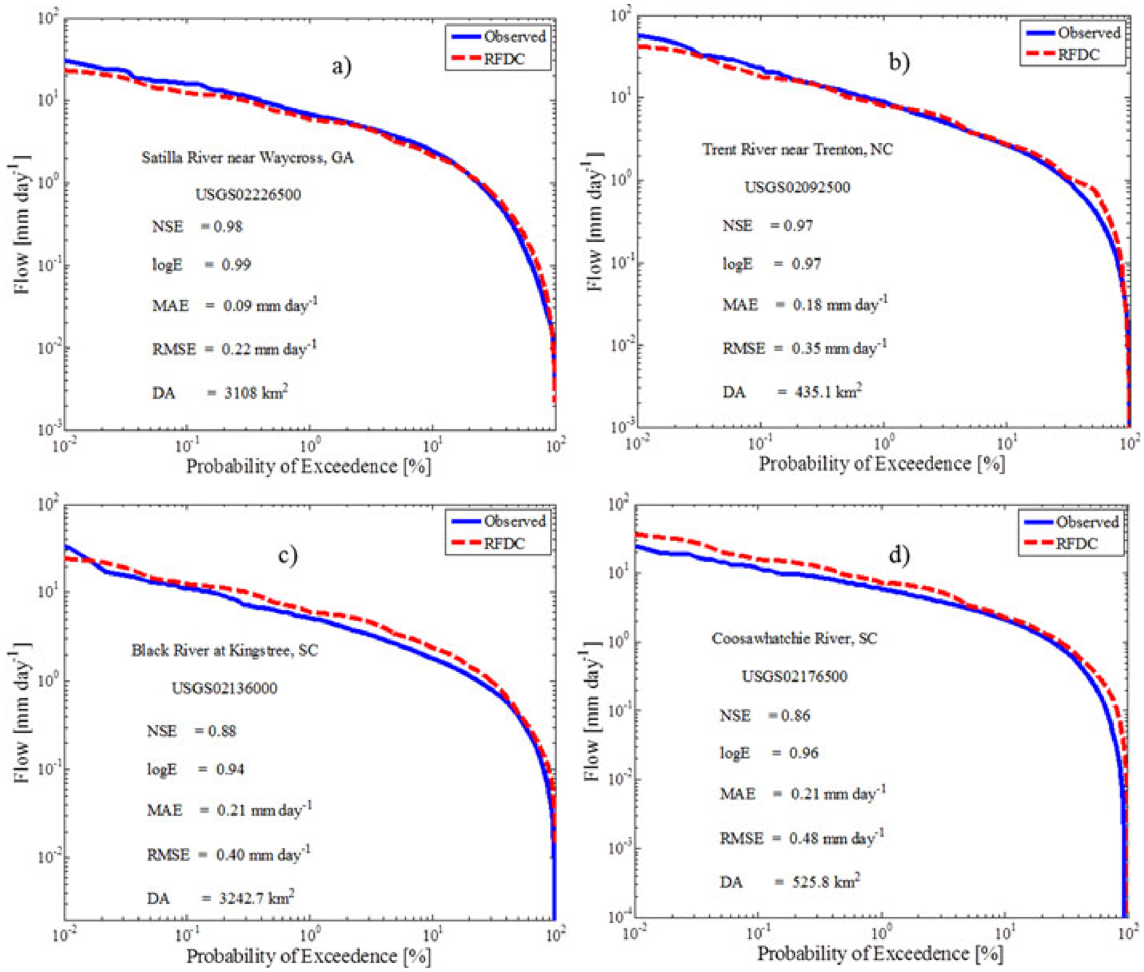


FIGURE 3. Observed and Estimated Daily Flow Duration Curves (FDC) for USGS Stream Gauges: (a) Satilla River Near Waycross, Georgia (02226500), (b) Trent River Near Trenton, North Carolina (02092500), (c) Black River at Kingstree, South Carolina (02136000), and (d) Coosawhatchie River Near Hampton, South Carolina (02176500). Observed and predicted values cover the period 1960-2010.

log$E$ for low flows (Q_{80}-$Q_{99.9}$) is less than 0.80 with log$E$ of Q_{99} and $Q_{99.5}$ less than 0.50. Thus, on average, the developed RFDC adequately estimates long-term high flows (NSE > 0.80), but gives poor to moderate estimates of long-term low flows (log$E$ < 0.80) on watersheds used to generate the RFDC. The poor to moderate performance of the RFDC on low flows is because some watersheds had zero-flow during low-flow regimes. For example, Turnpike Creek near Mcrae, Georgia (02216180) had zero flows for Q_{70}-$Q_{99.9}$ and Van Swamp near Hoke, North Carolina (02084557) had zero flows for Q_{90}-$Q_{99.9}$. Overall, 27.7% (13 of the 47 calibration watersheds) had zero flows during some low-flow regimes.

Figure 3 compares the observed FDC and estimated FDC based on regional equations (RFDC; Table 4) for four USGS gauged watersheds not used in development of the RFDC. NSE values for validation watersheds are greater than 0.85. Estimates of low-flow magnitudes for all validation watersheds are adequate because log$E$ values are greater or equal to 0.94 (Figure 3). Although the overall performance of the RFDC is strong (NSE > 0.85), it overpredicted high flows at 02136000 and 02176500 and underpredicted high flows at 02226500 and 02092500. The overprediction of high-flow magnitudes at 02176500 may be attributed to incidences of zero flow for intermittent type of streams because the RFDC is most suited for watersheds with minimal occurrences of zero flow such as perennial streams (Mohamoud, 2008). The overprediction and underprediction at other validation watersheds does not follow any discernible relationship with watershed size as the drainage areas are comparable for both cases (similar drainage areas for under- and overprediction). The above observations highlight the influence (although secondary) of other watershed descriptors not considered in the RFDC and in this study.

Performance of Regional Flow Duration Curve at Two First-Order Watersheds

Figure 5 compares estimates of flow duration curve by the RFDC to simulations by MIKE-SHE distributed model on two first-order watersheds (WS77 and WS80) at SEF with drainage areas less than 5.0 km^2. Examination of only the NSE and log$E$ model performance metrics for simulations by MIKE-SHE and RFDC for the period 1969-1980 gives a mendacious impression of strong prediction performance by the two models (NSE > 0.70 for MIKE-SHE and NSE > 0.95 for RFDC). The better performance by the RFDC compared with MIKE-SHE is due to better agreement between simulated and observed flows at very high flows (refer to Figure 4; $Q_{0.01}$-$Q_{0.1}$). However, examination of the MAE shows greater errors (MAE > 2.0 mm/day) for both MIKE-SHE and RFDC compared with similar errors for the validation watersheds (Figure 4; MAE < 1.0 mm/day). Thus, by visual inspection (Figure 5) and on basis of the MAE, both models poorly simulated daily streamflow for the period 1969-1980. The high NSE and log$E$ values for both models are due to high incidence of zero-flow periods for both watersheds (Q_{60}-$Q_{99.9}$ for WS77 and Q_{55}-$Q_{99.9}$ for WS80) that minimize influence of errors at high flows. The poor prediction performance by the RFDC on the two headwater watersheds was most likely because the DA and the RF1 variables were outside the range of values used to develop the RFDC (e.g., DA: 24.8 to 7,226.1 km^2 compared to 1.55 and 2.06 km^2; Table 2). The poor performance of the RFDC at the two watersheds is in accordance with observations by Niadas (2005), who highlighted failure of regionalized flow prediction methods to satisfactorily represent streamflow variability at small watersheds (<50 km^2).

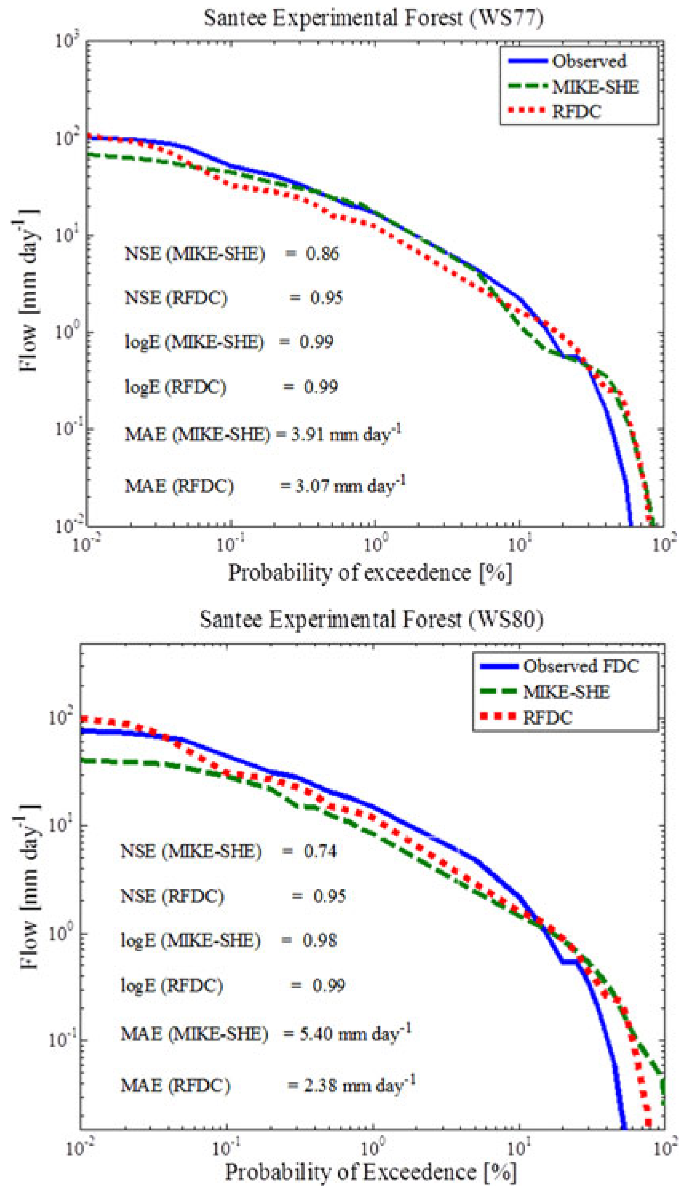
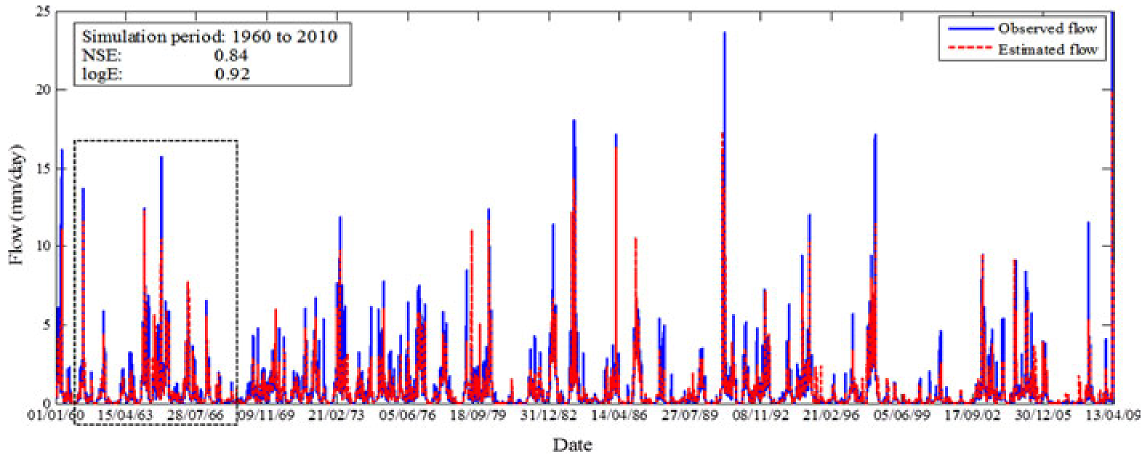


FIGURE 4. Comparison of Estimates by Regional FDC (RFDC) and MIKE-SHE Distributed Model at Two First-Order Streams at Santee Experimental Forest. Observed data and MIKE-SHE simulations cover the period January 1, 1969 and October 31, 1981: WS77 (top graph; DA = 1.55 km^2) and WS80 (bottom graph; DA = 2.06 km^2). The Nash-Sutcliffe efficiency value for MIKE-SHE estimates is greater than 0.90 compared to less than 0.30 for the RFDC.

Daily Streamflow

The accuracy of the SFS method significantly depends on accuracy of the estimated magnitudes (RFDC). Therefore, the accuracy of the daily streamflow is equal to the accuracy of estimated magnitudes, if the exact sequence (true streamflow sequence) is used (refer to Figure 3 and Table 5). Thereafter, the accuracy deteriorates based on the accuracy of the sequence. Table 5 shows difference in accuracy of estimated daily streamflows using different sequences to transform magnitude (RFDC) into daily streamflow. The sequences from the nearest and second nearest are based on streamflow per unit watershed area (mm/day). The results, on average, show a decrease in accuracy of estimated streamflow as the Euclidean distance between centroids of target and donor watersheds increases, (Figure 2, Table 3, and Table 5) indicating that this distance has more control on transferability of streamflow sequence than ratio of donor to target drainage area. This observation may be attributed to the fact that factors other than drainage area, such as drainage density, shape, and topography, also affect time of concentration. In addition, geographic proximity gives high probability for both target and donor watersheds to undergo similar flow regimes under similar hydro-



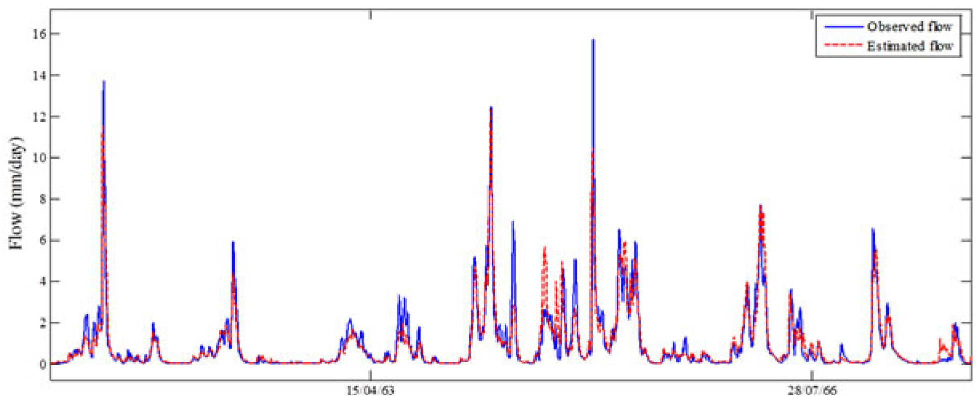


FIGURE 5. Observed and Estimated Daily Streamflow for USGS Stream Gauge at Satilla River Near Waycross, Georgia (02226500; DA = 3,108.0 km^2). The sequence used to generate daily streamflows is based on aggregation of daily streamflow of 02228000 (DA = 7,226.1 km^2), 02317500 (DA = 3,626 km^2), and 02227500 (DA = 1,675 km^2) at Euclidean distances of 27.3, 50.6, and 55 km, respectively.

climatic conditions, which are the main drivers for hydrologic response. The use of the nearest donor watershed is the simplest form of transferring hydrological information between watersheds (e.g., Mohamoud, 2008). However, as illustrated by Tables 3 and 5, the nearest donor watershed does not always provide the best streamflow sequence for the target watershed. For example, both 02226500 (target watershed) and 02228000 (nearest donor watershed) are in the same hydrologic unit code (HUC03070201; Satilla watershed-not shown in any figure) whereas 02227500 (third nearest donor watershed) is in a different hydrologic unit code (HUC03070202; Little Satilla watershed-not shown in any figure), yet 02227500 gave comparable estimates of streamflow sequence demonstrated by accuracy of predicted daily streamflow (Table 5; 0.68 for 02228000 *vs.* 0.71 for 02227500). Archfield and Vogel (2010) showed that use of the nearest donor watershed was the best choice on 79% of watersheds in their study. They also showed that by kriging (spatial interpolation) of correlation coefficients of log-transformed daily streamflows of predefined index watersheds, the accuracy of the best donor watershed improved from 79 to 90%. Use of donor watersheds for predicting sequence with prolonged periods of missing data was a major contributor of poor predictions at 02136000 and 0217500. Daily data for donor watershed 02135500 had missing data between 1992 and 2010 whereas donor watershed 02175500 had missing data between 1987 and 1990. Therefore, although the predicted streamflow magnitudes (RFDC) are strong (NSE > 0.80), the predicted daily streamflows are weak (NSE < 0.30) due to weak predictions of sequence.

Use of more than one donor watershed and ensemble methods significantly improved estimated daily streamflow for watershed 022226500 because of high density of gauged watersheds within its neighborhood (Table 5; rows 5-7) and minimal missing data for donor watersheds. However, use of two donor watersheds for sequence prediction did not improve the estimated daily streamflows for watershed 02092500 as the difference between the distances from the target to each of the nearest two donor watersheds was 28.3 km. Work by Ssegane (2011) for the Mid-Atlantic Appalachians, Piedmont, and Ridge and

TABLE 5. Effect of Sequence Prediction on Accuracy of Estimated Daily Streamflow.

		USGS No. 02226500				USGS No. 02092500				USGS No. 02136000				USGS No. 02176500			
Row #	Source of Sequence	NSE	Log*E*	MAE	RMSE	NSE	Log*E*	MAE	RMSE	NSE	Log*E*	MAE	RMSE	NSE	Log*E*	MAE	RMSE
1	True sequence	0.98	0.99	0.09	0.22	0.97	0.97	0.18	0.35	0.88	0.94	0.21	0.40	0.86	0.96	0.21	0.48
2	Nearest	0.68	0.84	0.35	0.87	0.74	0.82	0.44	1.01	0.25	0.67	0.40	0.98	0.09	0.69	0.48	1.23
3	Second nearest	0.61	0.79	0.42	0.97	0.23	0.42	0.80	1.76	0.29	0.74	0.35	0.97	-	-	-	-
4	Third nearest	0.71	0.82	0.37	0.83	-	-	-	-	-	-	-	-	-	-	-	-
5	Distance weighted mean of rows 2 to 4	0.81	0.90	0.28	0.67	0.60	0.73	0.53	1.27	0.33	0.73	0.36	0.94	-	-	-	-
6	Geometric mean of rows 2 to 4	0.84	0.92	0.26	0.62	0.63	0.73	0.55	1.21	0.35	0.73	0.36	0.94	-	-	-	-
7	Bootstrap aggregation based on values of rows 2 to 6	0.83	0.91	0.26	0.63	0.61	0.73	0.53	1.25	0.33	0.73	0.36	0.94	-	-	-	-

Notes: Performance is based on simulation period: 1960-2010. The nearest, second nearest, and third nearest donor watersheds are listed in Table 3 with the respective distances between donor and target watersheds.

Valley physiographic provinces on average, showed no improvement if the difference between Euclidean distances of the nearest and the second nearest was greater than 20 km. Similar observations are made at the sample validation watersheds for the Southeastern Coastal Plain. The difference in distance for 02092500 is 28.3 km compared to 23.3 km for 02226500. Other factors such as a significant difference in levels of surface and subsurface storage and urbanization between donor and target watersheds may contribute to accuracy of sequence prediction. On average, the use of geometric mean of multiple sequences from neighboring donor watersheds gave better estimates of streamflow sequence than other aggregation methods.

Graphical results of using sequence estimated by geometric mean aggregation of sequences from three nearest donor watersheds for 02226500 are given by Figure 5. The graph shows a strong agreement between estimated and observed low flows (Table 5; $\log E = 0.92$), however, the daily flow peaks were underpredicted (Table 5; NSE = 0.84). The underprediction of the peak daily flows was carried over from prediction accuracy of the RFDC (Figure 3). Table 6 depicts additional metrics of prediction accuracy at daily and monthly time scale, while Figure 6 depicts variability in prediction accuracy at peak flows and low flows on monthly time step. The accuracy of the predicted daily and monthly flows follow similar trends to accuracy of predicted magnitude (Table 6 and Figure 3), where strongest simulations were for 02226500 and the weakest for 02176500.

Comparison of our study results for Trent River to modeling results by Qi *et al.* (2008) shows a better simulation performance by our study approach at both the daily and monthly time steps. Qi *et al.* (2008) used U.S. Geological Survey's Precipitation Runoff Modeling System (PRMS) to model streamflow for Trent River over the period of 1980-2000. Their best PRMS model performance (NSE) was reported as 0.58 compared to 0.74; Table 6 on a daily time step and 0.79 compared to 0.85; Table 6 on a monthly basis, yet their model simulations covered 21 years compared to 51 years covered in this study. However, our method's current dependency on sequence of gauged neighboring watersheds does not provide for forecasting of hydrologic response to climatic, land use, or land cover changes compared with the process-based hydrologic models (e.g., PRMS) that are designed to model such scenarios. Therefore, although the approach developed herein provides a relatively easier method for simulating long-term daily streamflow data, it cannot assess impacts of climatic and land use change on future streamflows, a task that is not only appropriate but also increasingly being demanded in recent years for hydrologic

TABLE 6. Statistics of Model Performance in Estimating Daily and Monthly Streamflow at Four USGS Gauged Watersheds Used for Validation for the 1960-2010 Period.

Validation Watershed	Temporal Scale	NSE	Log*E*	MAE (mm)	Observed (mm)	Estimated (mm)
02226500	Daily	0.84	0.92	0.26	0.84	0.80
Satilla River, Georgia	Monthly	0.95	0.95	4.50	25.50	24.50
02092500	Daily	0.74	0.82	0.44	1.08	1.21
Trent River, North Carolina	Monthly	0.85	0.86	9.20	32.90	36.80
02136000	Daily	0.35	0.73	0.35	0.77	0.93
Black River, South Carolina	Monthly	0.67	0.83	8.15	23.40	28.20
02176500	Daily	0.09	0.69	0.48	0.78	1.08
Coosawhatchie River, South Carolina	Monthly	0.47	0.78	11.60	23.60	30.70

Note: The last two columns are the observed and estimated long-term averages on a daily and monthly time scale.

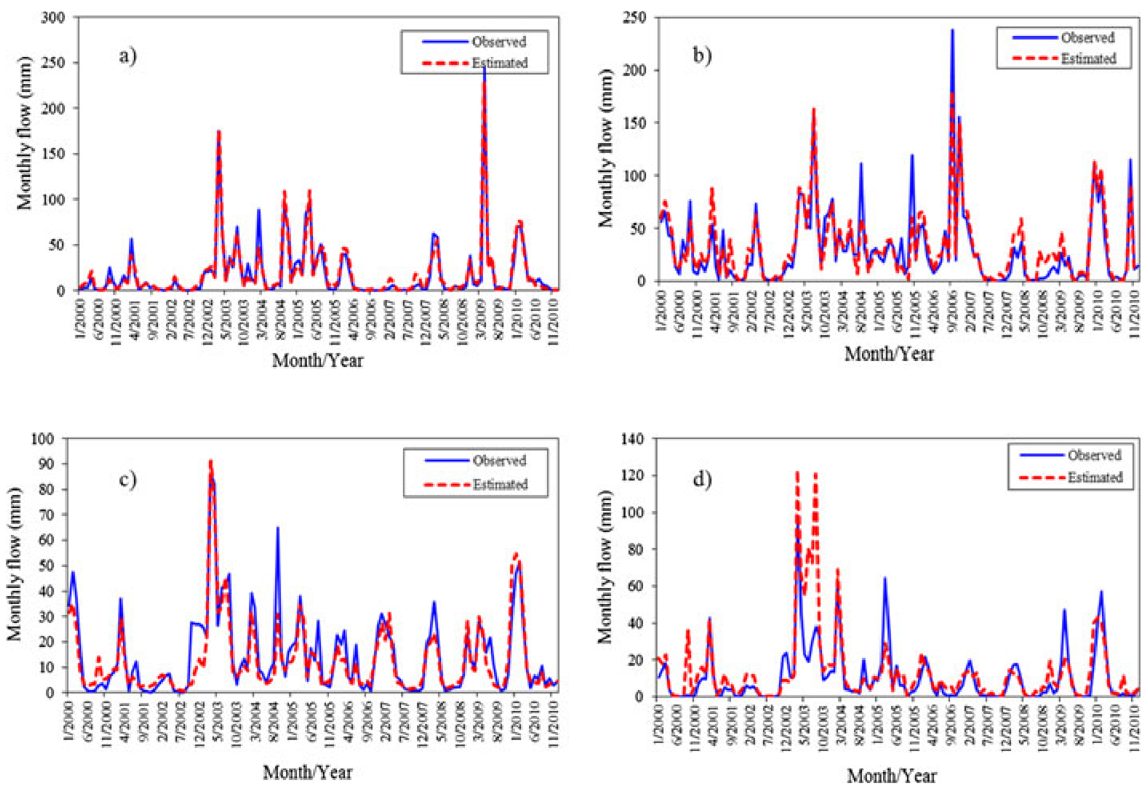


FIGURE 6. Observed and Estimated Monthly Streamflow of Four USGS Validation Watersheds. Only data for 2000-2010 is shown although the simulation covers the period 1960-2010. Refer to Table 6 for additional metrics of prediction accuracy.

models. Future work will concentrate on use of spatially distributed rainfall data, generic methods of choosing optimum donor watersheds, and aggregation techniques to utilize data from more than one donor watershed.

SUMMARY AND CONCLUSIONS

Independent modeling of daily streamflow as a composite of magnitude and sequence (SFS) for the Southeastern Coastal Plain watersheds (ranging in size from 24.8-7,226.1 km^2) provided very good to satisfactory results during model validation (NSE $\geq$ 0.86 for prediction of daily magnitudes-FDC and 0.09 $\leq$ NSE $\leq$ 0.84 for prediction of daily streamflow time series). This study directly regressed drainage area (DA), hydrologic soil index (HSI), and maximum 24-h precipitation with a return period of 100 years (RF1) to specific 19 streamflow percentiles ($Q_{0.01}$-$Q_{99.9}$) along a flow duration curve in combination with linear interpolation to predict long-term RFDC at ungauged sites. The three explanatory variables were the top three most selected watersheds across the 19 streamflow percentiles using a greedy-heuristic search process. The performance of the

RFDC for Southeastern Coastal Plain watersheds is limited to the watershed characteristics within the range used to develop them. This was demonstrated by poor performance of the RFDC on two small headwater forested watersheds. Comparison of MIKE-SHE model and RFDC simulations on the two headwater watersheds gave high NSE values (NSE > 0.70) for daily FDC. However, examination of the graphs and the respective MAE values (MAE > 2.0 mm/day) showed that the FDC predictions were weak for both models. Major conclusions drawn from above results include: (1) accuracy of estimated daily streamflow for an ungauged watershed can never be greater than accuracy of estimated magnitude (FDC); (2) high density of gauged (donor) watersheds in proximity of ungauged watershed provides better estimates of sequence and therefore use of more than one donor watershed whose distances from centroid of the ungauged are within 50 km, improves sequence prediction; (3) the Euclidean distance between the centroids of the donor and target watersheds is a better predictor of sequence than the ratio of the donor to target drainage areas, however, proximity does not always give the best prediction of sequence; (4) RFDC tends to overpredict high-flow magnitudes for watersheds with high incidences of zero flow, for example, intermittent streams, thus, SFS is most suited for predictions at continuously flowing watersheds with perennial streams; and (5) the prediction accuracy of both the magnitude and the sequence should be strong to give satisfactory simulations of daily streamflow at ungauged watersheds.

ACKNOWLEDGMENTS

The authors thank Weyerhaeuser Company for their financial support to the University of Georgia through the USDA-Forest Service Center for Forested Wetlands Research (Cordesville, South Carolina) and Toby D. Feaster of U.S. Geological Survey (South Calorina Water Science Center) for providing the 20 watershed characteristics for over 800 watersheds in the Southeastern United States. The authors also would like to acknowledge Andy Harrison, Hydrologic Technician at Center for Forested Wetlands Research for helping provide data for small watersheds (WS77 and WS80) at Santee Experimental Forest.

LITERATURE CITED

Amatya, D.M., M. Miwa, C.A. Harrison, C.C. Trettin, and G. Sun, 2006. Hydrology and Water Quality of Two First Order Forested Watersheds in Coastal South Carolina. Paper No. 06-2182, ASABE, St. Joseph, Michigan, 22 pp.

Archfield, S.A. and R.M. Vogel, 2010. Map Correlation Method: Selection of a Reference Streamgage to Estimate Daily Streamflow at Ungaged Catchments. Water Resources Research 46 (10):1-15.

Archfield, S.A., R.M. Vogel, P.A. Steeves, S.L. Brandt, P.W. Weiskel, and S.P. Garabedian, 2010. The Massachusetts Sustainable-Yield Estimator: A Decision-Support Tool to Assess Water Availability at Ungaged Stream Locations in Massachusetts. U.S. Geological Survey Scientific Investigation Report 2009-5227, 41 pp. plus [CD-ROM].

Atallah, M.J., 1998. Algorithms and Theory of Computation Handbook. CRC Press LLC, Boca Raton, Florida.

Bárdossy, A., 2007. Calibration of Hydrological Model Parameters for Ungauged Catchments. Hydrology and Earth System Sciences 11(2):703-710.

Bastola, S., H. Ishidaira, and K. Takeuchi, 2008. Regionalisation of Hydrological Model Parameters Under Parameter Uncertainty: A Case Study Involving TOPMODEL and Basins Across the Globe. Journal of Hydrology 357(3-4):188-206.

Borah, D.K. and M. Bera, 2004. Watershed-Scale Hydrologic and Nonpoint-Source Pollution Models: Review of Applications. Transactions of the ASAE 47(3):789-803.

Brandes, D., J.G. Hoffmann, and J.T. Mangarillo, 2005. Base Flow Recession Rates, Low Flows, and Hydrologic Features of Small Watersheds in Pennsylvania, USA. Journal of the American Water Resources Association 41(5):1177-1186.

Castellarin, A., G. Camorani, and B. Allen, 2007. Predicting Annual and Long-Term Flow-Duration Curves in Ungauged Basins. Advances in Water Resources 30(4):937-953.

Cunnane, C., 1978. Unbiased Plotting Positions - Review. Journal of Hydrology 37(3-4):205-222.

Dai, Z., D.M. Amatya, G. Sun, C.C. Trettin, C. Li, and H. Li, 2010. A Comparison of MIKE SHE and DRAINMOD for Modeling Forested Wetland Hydrology in Coastal South Carolina, USA. *In*: Proceedings of the XVII World Congress of the International Commission of Agricultural and Biosystems Engineering (CIGR), Québec City, Canada, June 13-17, 2010.

Dai, Z., D.M. Amatya, G. Sun, C.C. Trettin, C. Li, and H. Li, 2011. Climate Variability and Its Impact on Forest Hydrology on South Carolina Coastal Plain, USA. Atmosphere 2:330-357.

Dixon, P.M., 2001. Bootstrap Resampling. *In*: The Encyclopedia of Environmetrics, A.H. El-Shaarawi and W.W. Piegorsch (Editors). Wiley, New York, 9 pp.

Engeland, K. and H. Hisdal, 2009. A Comparison of Low Flow Estimates in Ungauged Catchments Using Regional Regression and the HBV-Model. Water Resources Management 23(12):2567-2586.

Feaster, T.D., A.J. Gotvald, and J.C. Weaver, 2009. Magnitude and Frequency of Rural Floods in the Southeastern United States, 2006, South Carolina. U.S. Geological Survey, Scientific Investigations Report, 2009-5156, 226 pp.

Fennessey, N., 1994. A Hydro-Climatological Model of Daily Streamflow for the Northeast United States. Ph.D. Dissertation, Department of Civil and Environmental Engineering, Tufts University, Medford, Massachusetts.

Fernandez, W., R.M. Vogel, and A. Sankarasubramanian, 2000. Regional Calibration of a Watershed Model. Hydrological Sciences Journal 45(5):689-707.

Furey, P. and V. Gupta, 2005. Effects of Excess Rainfall on the Temporal Variability of Observed Peak-Discharge Power Laws. Advances in Water Resources 28(11):1240-1253.

Götzinger, J. and A. Bárdossy, 2007. Comparison of Four Regionalisation Methods for a Distributed Hydrological Model. Journal of Hydrology 333(2-4):374-384.

Grover, P.L., D.H. Burn, and J.M. Cunderlik, 2002. A Comparison of Index Flood Estimation Procedures for Ungauged Catchments. Canadian Journal of Civil Engineering 29(5):734-741.

Guimares, W.B. and L.R. Bohan, 1992. Techniques for Estimating Magnitude and Frequency of Floods in South Carolina. USGS Water Resources Investigations Report 91-4157.

Gupta, H.V., S. Sorooshian, T.S. Hogue, and D.P. Boyle, 2003. Advances in Automatic Calibration of Watershed Models. *In*: Water Science and Application 6:9-28.

Gupta, V.K., O. Mesa, and D. Dawdy, 1994. Multiscaling Theory of Flood Peaks - Regional Quantile Analysis. Water Resources Research 30(12):3405-3421.

He, Y., A. Bárdossy, and E. Zehe, 2011. A Review of Regionalisation for Continuous Streamflow Simulation. Hydrology and Earth System Sciences 15(11):3539-3553.

Henderson, J.P. and H.D. Grissino-Mayer, 2009. Climate–Tree Growth Relationships of Longleaf Pine (*Pinus Palustris Mill.*) in the Southeastern Coastal Plain, USA. Dendrochronologia 27 (1):31-43.

Hughes, D.A., E. Kapangaziwiri, and T. Sawunyama, 2010. Hydrological Model Uncertainty Assessment in Southern Africa. Journal of Hydrology 387(3-4):221-232.

Hughes, D.A. and V. Smakhtin, 1996. Daily Flow Time Series Patching rr Extension: A Spatial Interpolation Approach Based on Flow Duration Curves. Hydrological Sciences Journal 41(6):851-872.

Krause, P., D.P. Boyle, and F. Base, 2005. Comparison of Different Efficiency Criteria for Hydrological Model Assessment. Advances in Geosciences 5:89-97.

Kroll, C., J. Luz, B. Allen, and R.M. Vogel, 2004. Developing a Watershed Characteristics Database to Improve Low Streamflow Prediction. Journal of Hydrologic Engineering 9(2):116-125.

Lagarias, J.C., J.A. Reeds, M.H. Wright, and P.E. Wright, 1998. Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. SIAM Journal of Optimization 9(1):112-147.

McIntyre, N., H. Lee, and H. Wheater, 2005. Ensemble Predictions of Runoff in Ungauged Catchments. Water Resources Research 41(12):1-14.

Mohamoud, Y.M., 2008. Prediction of Daily Flow Duration Curves and Streamflow for Ungauged Catchments Using Regional Flow Duration Curves. Hydrological Sciences Journal 53(August):706-724.

Nash, J.E. and J.V. Sutcliffe, 1970. River Flow Forecasting Through Conceptual Models—Part I: A Discussion of Principles. Journal of Hydrology 10(3):282-290.

Niadas, I.A., 2005. Regional Flow Duration Curve Estimation in Small Ungauged Catchments Using Instantaneous Flow Measurements and a Censored Data Approach. Journal of Hydrology 314:48-66.

Ogden, F. and D. Dawdy, 2003. Peak Discharge Scaling in Small Hortonian Watershed. Journal of Hydrologic Engineering 8 (2):64-73.

Omernik, J.M. and R.G. Bailey, 1997. Distinguishing Between Watersheds and Ecoregions. Journal of the American Water Resources Association 33(5):935-949.

Patil, S. and M. Stieglitz, 2012. Controls on Hydrologic Similarity: Role of Nearby Gauged Catchments for Prediction at an Ungauged Catchment. Hydrology and Earth System Sciences 16:551-562.

Pechlivanidis, I.G., N.R. McIntyre, and H.S. Wheater, 2010. Calibration of the Semi-Distributed PDM Rainfall–Runoff Model in the Upper Lee Catchment, UK. Journal of Hydrology 386(1-4):198-209.

Perrin, C., C. Michel, and V. Andréassian, 2003. Improvement of a Parsimonious Model for Streamflow Simulation, Journal of Hydrology 279(1-4):275-289.

Qi, S., G. Sun, Y. Wang, S.G. McNulty, and J.M. Myers, 2009. Streamflow Response to Climate and Landuse Changes in a Coastal Watershed in North Carolina. Trans ASABE 52(3):739-749.

Rao, A.R. and V.V. Srinivas, 2008. Regionalization of Watersheds: An Approach Based on Cluster Analysis (Vol. 58). Water Science and Technology Library, 58, Springer Science + Business Media B.V., New York.

Reusser, D.E., T. Blume, B. Schaefli, and E. Zehe, 2009. Analysing the Temporal Dynamics of Model Performance for Hydrological Models. Hydrology and Earth System Sciences 13(7):999-1018.

Richter, D.D., 1980. Prescribed Fire: Effects of Water Quality and Nutrient Cycling in Forested Watersheds of the Santee Experimental Forest in South Carolina. Ph.D. Dissertation, Duke University, Durham, North Carolina, 194 pp.

Schilling, K.E. and C.F. Wolter, 2005. Estimation of Streamflow, Base Flow, and Nitrate-Nitrogen Loads in Iowa Using Multiple Linear Regression Models. Journal of the American Water Resources Association 41(6):1333-1346.

Schroedl, S., 2010. Feature Selection Based on Interaction (Mutual) Information. http://www.mathworks.com/matlabcentral/fileexchange/26981-feature-selection-based-on-interaction-information, *accessed* September18, 2012.

Segura, C. and J. Pitlick, 2010. Scaling Frequency of Channel-Forming Flows in Snowmelt-Dominated Streams. Water Resources Research 46:W06524.

Shu, C. and T.B.M.J. Ouarda, 2012. Improved Methods for Daily Streamflow Estimates at Ungauged Sites. Water Resources Research 48(2):1-15.

Singh, V.P. and D.K. Frevert, 2002a. Mathematical Models of Large Watershed Hydrology. Water Resources Publications, LLC, Highlands Ranch, Colorado.

Singh, V.P. and D.K. Frevert, 2002b. Mathematical Models of Small Watershed Hydrology and Applications. Water Resources Publications, LLC, Highlands Ranch, Colorado.

Singh, V.P. and D.K. Frevert, 2006. Watershed Models. CRC Press, Boca Raton, Florida.

Sivapalan, M., K. Takeuchi, S.W. Franks, V.K. Gupta, H. Karambiri, V. Lakshmi, X. Liang, J.J. McDonnell, E.M. Mendiondo, P.E. O'Connell, T. Oki, J.W. Pomeroy, D. Schertzer, S. Uhlenbrook, and E. Zehe, 2003. IAHS Decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an Exciting Future for the Hydrological Sciences. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques 48(6):857-880.

Smakhtin, V.Y., 1999. Generation of Natural Daily Flow Time-Series in Regulated Rivers Using a Non-Linear Spatial Interpolation Technique. October, 15(4):311-323.

Smakhtin, V.Y., D. Hughes, and E. Creuse-Naudin, 1997. Regionalization of Daily Flow Characteristics in Part of the Eastern Cape, South Africa. Hydrological Sciences Journal 42(6):919-936.

Smakhtin, V.Y. and B. Masse, 2000. Continuous Daily Hydrograph Simulation Using Duration Curves of a Precipitation Index. Hydrological Processes 1100(August 1999):1083-1100.

Ssegane, H., 2011. In Search of Causal Variables for Watershed Classification and Daily Streamflow Prediction at Ungauged Watersheds. Ph.D. Dissertation, Department of Biological and Agricultural Engineering, University of Georgia, Athens, Georgia.

USEPA, 2008. Handbook for Developing Watershed Plans to Restore and Protect Our Waters. U.S. Environmental Protection Agency (USEPA), Office of Water Nonpoint Source Control Branch. DIANE Publishing, Washington, D.C.

U.S. Geological Survey, U.S. Dept. of Interior, 2000. The National Flood Frequency Program - Methods for Estimating Flood Magnitude and Frequency in Rural and Urban Areas in South Carolina. USGF Fact Sheet 001-00, January, 2000.

Verdin, K.L. and B. Worstell, 2008. A Fully Distributed Implementation of Mean Annual Stream Flow Regional Regression Equations. Journal of the American Water Resources Association 44 (6):1537-1547.

Vogel, R.M. and N.M. Fennessey, 1994. Flow-Duration Curves .2. New Interpretation and Confidence-Intervals. Journal of Water Resources Planning and Management-ASCE 120(4):485-504.

Vogel, R.M. and A. Sankarasubramanian, 2000. Spatial Scaling Properties of Annual Streamflow in the United States. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques 45 (3):465-476.

Wagener, T. and A. Montanari, 2011. Convergence of Approaches Toward Reducing Uncertainty in Predictions in Ungauged Basins. Water Resources Research 47:WO6301.

Wolock, D.M., T.C. Winter, and G. McMahon, 2004. Delineation and Evaluation of Hydrologic-Landscape Regions in the United States Using Geographic Information System Tools and Multivariate Statistical Analyses. Environmental Management 34 (Supplement 1):S71-S88.

Young, C.E., 1967. Streamflow - An Important Factor in Forest Management in the Coastal Plain. South. Lumberman, Christmas Issue 215(2680):109-110.

Zhu, Y. and R.L. Day, 2009. Regression Modeling of Streamflow, Baseflow, and Runoff Using Geographic Information Systems. Journal of Environmental Management 90(2):946-953.