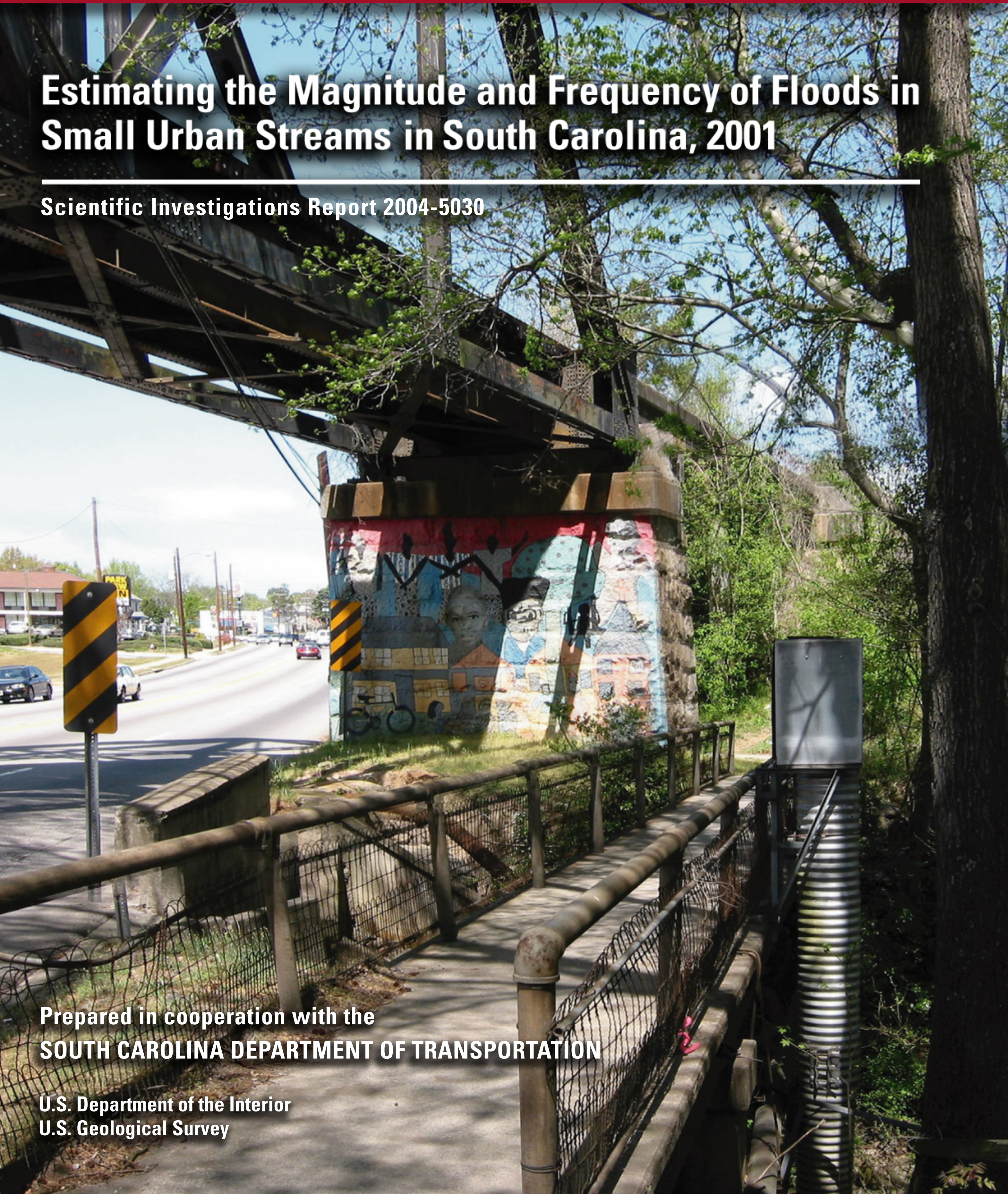
USGS
science for a changing world

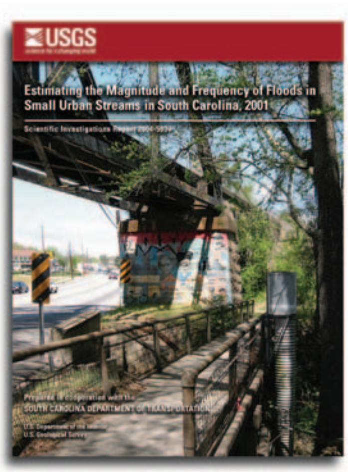


Estimating the Magnitude and Frequency of Floods in Small Urban Streams in South Carolina, 2001

Scientific Investigations Report 2004-5030

Prepared in cooperation with the
SOUTH CAROLINA DEPARTMENT OF TRANSPORTATION

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



COVER PHOTO: U.S. Geological Survey streamflow-gaging station 02162093, Smith Branch at Columbia, S.C.

Photograph by James R. Douglas

Estimating the Magnitude and Frequency of Floods in Small Urban Streams in South Carolina, 2001

By Toby D. Feaster and Wladmir B. Guimaraes

Prepared in cooperation with the
South Carolina Department of Transportation

Scientific Investigations Report 2004–5030

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

U.S. Department of the Interior
Gale A. Norton, Secretary

U.S. Geological Survey
Charles G. Groat, Director

U.S. Geological Survey, Reston, Virginia: 2004
For sale by U.S. Geological Survey, Information Services
Box 25286, Denver Federal Center
Denver, CO 80225

For more information about the USGS and its products:
Telephone: 1-888-ASK-USGS
World Wide Web: http://www.usgs.gov/

Suggested citation:
Feaster, T.D., and Guimaraes, W.B., 2004, Estimating the magnitude and frequency of floods in small urban streams in South Carolina, 2001: U.S. Geological Survey Scientific Investigations Report 2004–5030, accessed May 18, 2004, at *http://pubs.water.usgs.gov/sir2004-5030*

Contents

Figures

Tables

Conversion Factors, Datums, and Abbreviations and Acronyms

Multiply	By	To obtain
	Length	
mile (mi)	1.609	kilometer (km)
	Area	
square mile (mi^2)	2.590	square kilometer (km^2)
	Flow rate	
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)

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

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

Abbreviations and acronyms

BDF	basin development factor
DA	drainage area
GIS	Geographic Information System
GLS	generalized least squares
IA	impervious area
L	length
NWS	National Weather Service
OLS	ordinary least squares
PRESS	prediction error sum of squares
R^2	coefficient of determination
ROI	region of influence
S	slope
SCDOT	South Carolina Department of Transportation
USGS	U.S. Geological Survey
VIF	variance inflation factor

Estimating the Magnitude and Frequency of Floods in Small Urban Streams in South Carolina, 2001

By Toby D. Feaster and Wladmir B. Guimaraes

Abstract

The magnitude and frequency of floods at 20 streamflow-gaging stations on small, unregulated urban streams in or near South Carolina were estimated by fitting the measured water-year peak flows to a log-Pearson Type-III distribution. The period of record (through September 30, 2001) for the measured water-year peak flows ranged from 11 to 25 years with a mean and median length of 16 years. The drainage areas of the streamflow-gaging stations ranged from 0.18 to 41 square miles.

Based on the flood-frequency estimates from the 20 streamflow-gaging stations (13 in South Carolina; 4 in North Carolina; and 3 in Georgia), generalized least-squares regression was used to develop regional regression equations. These equations can be used to estimate the 2-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year recurrence-interval flows for small urban streams in the Piedmont, upper Coastal Plain, and lower Coastal Plain physiographic provinces of South Carolina. The most significant explanatory variables from this analysis were main-channel length, percent impervious area, and basin development factor. Mean standard errors of prediction for the regression equations ranged from -25 to 33 percent for the 10-year recurrence-interval flows and from -35 to 54 percent for the 100-year recurrence-interval flows.

The U.S. Geological Survey has developed a Geographic Information System application called StreamStats that makes the process of computing streamflow statistics at ungaged sites faster and more consistent than manual methods. This application was developed in the Massachusetts District and ongoing work is being done in other districts to develop a similar application using streamflow statistics relative to those respective States. Considering the future possibility of implementing StreamStats in South Carolina, an alternative set of regional regression equations was developed using only main channel length and impervious area. This was done because no digital coverages are currently available for basin development factor and, therefore, it could not be included in the StreamStats application. The average mean standard error of prediction for the alternative equations was 2 to 5 percent larger than the standard errors for the equations that contained basin development factor.

For the urban streamflow-gaging stations in South Carolina, measured water-year peak flows were compared with those from an earlier urban flood-frequency investigation. The peak flows from the earlier investigation were computed using a rainfall-runoff model. At many of the sites, graphical comparisons indicated that the variance of the measured data was much less than the variance of the simulated data. Several statistical tests were applied to compare the variances and the means of the measured and simulated data for each site. The results indicated that the variances were significantly different for 11 of the 13 South Carolina streamflow-gaging stations. For one streamflow-gaging station, the test for normality, which is one of the assumptions of the data when comparing variances, indicated that neither the measured data nor the simulated data were distributed normally; therefore, the test for differences in the variances was not used for that streamflow-gaging station. Another statistical test was used to test for statistically significant differences in the means of the measured and simulated data. The results indicated that for 5 of the 13 urban streamflow-gaging stations in South Carolina there was a statistically significant difference in the means of the two data sets.

For comparison purposes and to test the hypothesis that there may have been climatic differences between the period in which the measured peak-flow data were measured and the period for which historic rainfall data were used to compute the simulated peak flows, 16 rural streamflow-gaging stations with long-term records were reviewed using similar techniques as those used for the measured and simulated data at the urban streamflow-gaging stations. For the rural sites, the period from 1985 to 2001 was compared with the data measured from the beginning of record to 1984. Plots of the two periods at each rural site indicated no significant difference in the data. The statistical test for comparison of variances was applied to the 16 rural streamflow-gaging stations, and the results showed that there was no statistically significant difference in the variances at 14 of the 16 streamflow-gaging stations. The statistical comparisons of the means for the two periods at the rural streamflow-gaging stations showed that there was no statistically significant difference at 12 of the 16 streamflow-gaging stations. Based on these comparisons, the differences between the measured and simulated urban water-year peak flows cannot be completely explained by climatic differences between the periods of record.

Introduction

Knowledge of flood characteristics of streams is needed for the design of roadway drainage structures, the establishment of flood-insurance rates, and for other uses by urban planners and engineers. Because urbanization can produce significant changes in the flood-frequency characteristics of streams, rural basin flood-frequency relations are not applicable to urban streams.

Urban flood-frequency equations were developed by Bohman (1992) for small urban streams in South Carolina using simulated peak-flow data from rainfall-runoff models. Recognizing the importance of measured data for comparison and verification of these equations, the U.S. Geological Survey (USGS), in cooperation with the South Carolina Department of Transportation (SCDOT), continued to collect data at many of the urban streamflow-gaging stations that were established during Bohman's investigation. Comparisons of the measured and simulated peak-flow data indicated that there was enough of a significant difference in the two data sets to warrant updating the urban flood-frequency estimates using only the measured data. This investigation documents these data comparisons and updates the urban flood-frequency equations for South Carolina using measured data collected through the 2001 water year[1].

There are several ways to continue improving the understanding of urban flood-frequency in South Carolina and to increase the confidence in future statistical analyses of water-year peak flows. Hereafter in this report, "peak flow" refers to the maximum peak for the water year. One way is to expand the database used for estimating the magnitude and frequency of floods on small urban streams by continuing to collect streamflow data at existing urban streamflow-gaging stations, which will increase the length of record used in the analysis. Additionally, as funding is available and where appropriate, other streamflow-gaging stations used in the previous urban flood-frequency investigation could be considered for reactivation. It also may be worthwhile to review the geographical coverage of the urban streamflow network and consider the benefits of activating additional new urban streamflow-gaging stations, which will not only improve the geographical coverage of the State, but also increase the number of streamflow-gaging stations in the database. An extended monitoring network and database is likely to provide more accurate flood-frequency equations for use in design and planning.

Purpose and Scope

This report describes the comparison of peak flows measured through water year 2001 with peak flows computed by using a rainfall-runoff model during the previous urban flood-frequency investigation (Bohman, 1992). The flood-frequency estimates for 20 streamflow stations were updated using measured peak-flow data. Methods are presented for predicting the magnitude and frequency of floods in South Carolina at ungaged urban basins in the Piedmont, upper Coastal Plain, and lower Coastal Plain physiographic provinces. Statistics describing the uncertainty in the prediction equations are presented and the limitations of the equations also are discussed.

Previous Investigations

Speer and Gamble (1964) documented the earliest investigation of flood frequency of streams in South Carolina. They presented methods for estimating the magnitude of floods for selected recurrence intervals for rural streams in the South Atlantic slope basin, which extends from the James River in Virginia to the Savannah River along the South Carolina-Georgia State boundary. Whetstone (1982) used multiple regression analyses to define the relation between flows and basin characteristics at recurrence intervals of 2, 5, 10, 25, 50, and 100 years for unregulated rural streams with drainage areas greater than 1.0 square mile (mi^2). Sauer and others (1983) used data from 269 gaged basins in 56 cities in 31 states to develop flood-frequency relations for urban watersheds in the United States. Frequencies of peak flows were regionalized by Guimaraes and Bohman (1991) using generalized least-squares regression methods to define the relation of magnitude and frequency of flows to various basin characteristics on ungaged rural streams that were not affected significantly by regulation.

Bohman (1992) described methods for determining peak-flow frequency relations, flood hydrographs, average basin lag times, and runoff volumes associated with a given peak flow for ungaged urban basins by using data from 34 streamflow-gaging stations in 15 cities in South Carolina, North Carolina, and Georgia. A rainfall-runoff model was calibrated for 23 urban drainage basins in South Carolina. The model was then used to synthesize from 50 to 70 annual peaks, depending on the length of the long-term rainfall data from nearby National Weather Service stations. The logarithms of these peaks were fitted to a Pearson Type-III distribution to determine the frequency of peak discharges having recurrence intervals of 2, 5, 10, 25, 50, 100, and 500 years at each streamflow-gaging station. The final step in analyzing these data was to develop regression equations that could be used to predict the magnitude and frequency of floods at ungaged urban sites in South Carolina. Detailed descriptions of the rainfall-runoff model calibration, the long-term simulation, and the regression analyses are provided in Bohman's (1992) report.

Feaster and Tasker (2002) used generalized least-squares regression to develop a set of predictive equations that can be used to estimate flows at the 2-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year recurrence intervals for rural ungaged basins in the

[1]A water year is the 12-month period from October 1 to September 30 and is designated by the calendar year in which it ends. Thus, the 12-month period ending September 30, 2001, is the 2001 water year.

Blue Ridge, Piedmont, and Coastal Plain physiographic provinces of South Carolina. In addition, a region-of-influence (ROI) method also was developed to interactively estimate the recurrence-interval flows for rural ungaged basins. The predictive abilities of the regional regression equations were compared with the ROI methods for each physiographic province in South Carolina. The ROI method performed systematically better only in the Blue Ridge, which limits its usefulness only to that province.

Peak-Flow Data

The peak flows collected at USGS streamflow-gaging stations are the empirical basis for estimating specific recurrence-interval flows for this investigation. As recommended in Bulletin 17B (Hydrology Subcommittee of the Interagency Advisory Committee on Water Data, 1982), only streamflow-gaging stations with at least 10 years of measured peak flows were used to develop flood-frequency estimates. Of the 34 streamflow-gaging stations used in the Bohman (1992) investigation, streamflow-gaging stations with sufficient lengths of record to be included in this investigation included 16 of the 23 South Carolina streamflow-gaging stations, 3 of the 7 Georgia streamflow-gaging stations, and all 4 North Carolina streamflow-gaging stations (table 1; fig. 1). After reviewing the peak-flow data, data from 3 of the 16 South Carolina streamflow-gaging stations were excluded from the analysis for reasons explained later.

At continuous-record streamflow-gaging stations, the water-surface elevation, or stage, of the stream is recorded at fixed intervals typically ranging from 5 to 60 minutes. At crest-stage, partial-record streamflow-gaging stations, only the crest (highest) stages that occur between site visits, usually 6 to 8 weeks, are recorded. An attempt is made to measure streamflow throughout the range of recorded stages. If this is possible, a relation between stage and streamflow is developed for the gaged site. Using this stage-streamflow relation, or rating, streamflows for recorded stages are estimated. Because stream channels are dynamic, periodic streamflow measurements are made to verify that the hydraulic conditions at the site remain stable. If the measurements indicate that conditions have changed, additional data are collected and used to make adjustments to the stage-streamflow relation. At some crest-stage sites, indirect flow-computation methods are used to develop a theoretical rating. This method has been used extensively to compute streamflows for small drainage areas, which are typical of urban streams (Bodhaine, 1968).

Initial reviews of the peak-flow data for the South Carolina streamflow-gaging stations included comparing the peak flows

Table 1. Streamflow-gaging stations in South Carolina, North Carolina, and Georgia with 10 or more years of record used in the flood-frequency analysis for small urban streams in South Carolina.

[mi^2, square miles]

Station number (fig. 1)	Station name	Location			Drainage area (mi^2)	Period of record
		Latitude	Longitude	Description		
South Carolina						
02110740	Midway Swash at Myrtle Beach, S.C.	33° 39'44"	78° 55'25"	Horry County, at culvert on U.S. Highway 17	0.80	1987–2001
02131130	Gully Branch at Florence, S.C.	34° 53'00"	79° 46'12"	Florence County, at culvert on Cherokee Road	1.92	1985–2001
02135518	Turkey Creek at Sumter, S.C.	33° 55'13"	80° 19'43"	Sumter County, at culvert on East Liberty Street	2.20	1987–2001
02145940	Little Dutchman Creek tributary at Rock Hill, S.C.	34° 58'34"	81° 01'02"	York County, at culvert on Celanese Road	3.50	1986–97
02159785	Fairforest Creek tributary at Spartanburg, S.C.	34° 57'10"	81° 57'57"	Spartanburg County, at culvert on Secondary Road 485	.52	1987–2001
02162093	Smith Branch at Columbia, S.C.	34° 01'38"	81° 02'31"	Richland County, at culvert on North Main Street	5.49	1977–2001
02164011	Brushy Creek (Reedy River tributary) at Greenville, S.C.	34° 49'25"	82° 24'26"	Greenville County, at culvert on Grove Road	3.02	1985–2001

Table 1. Streamflow-gaging stations in South Carolina, North Carolina, and Georgia with 10 or more years of record used in the flood-frequency analysis for small urban streams in South Carolina.—Continued

[mi^2, square miles]

Station number (fig. 1)	Station name	Location			Drainage area (mi^2)	Period of record
		Latitude	Longitude	Description		
			South Carolina (Continued)			
02166975	Sample Branch at Greenwood, S.C.	34° 12'56"	82° 09'20"	Greenwood County, at culvert on U.S. Highway 178 bypass	1.16	1986–2001
02167020	Crane Creek tributary at Columbia, S.C.	34° 03'02"	81° 02'05"	Richland County, at culvert on Carola Street	.28	1986–2001
02168845	Saluda River tributary at Columbia, S.C.	34° 02'26"	81° 08'29"	Richland County, at culvert on Bush River Road	.45	1986–96
02169568	Pen Branch at Columbia, S.C.	34° 00'46"	80° 58'56"	Richland County, at culvert on Brentwood Street	2.26	1986–2001
02173491	Hess Branch at Orangeburg, S.C.	33° 30'12"	80° 52'41"	Orangeburg County, at culvert on Middleton Road	.45	1987–2001
02176380	Coosawhatchie River tributary at Allendale, S.C.	32° 59'53"	81° 19'01"	Allendale County, at culvert on Secondary Road 129	2.06	1986–2001
			North Carolina			
02146300	Irwin Creek near Charlotte, N.C.	35° 11'52"	80° 54'16"	Mecklenburg County, on left bank at city of Charlotte sewage-disposal plant	30.5	1963–77
02146500	Little Sugar Creek near Charlotte, N.C.	35° 09'13"	80° 51'18"	Mecklenburg County, on right bank upstream from bridge on Tyvola Road at city of Charlotte sewage-disposal plant	41.0	1962–77
02146600	McAlpine Creek at Sardis Road near Charlotte, N.C.	35° 08'16"	80° 45'03"	Mecklenburg County, near left bank on downstream end of bridge pier at Sardis Road (Secondary Road 3356)	38.3	1962–77
02146700	McMullen Creek at Sharon View Road near Charlotte, N.C.	35° 08'27"	80° 49'12"	Mecklenburg County, on left bank downstream of culvert wingwall at Sharon View Road (Secondary Road 3673)	6.98	1963–77
			Georgia			
02196760	Rocky Creek tributary at Augusta, Ga.	33° 27'07"	82° 02'57"	Richmond County, at culvert on U.S. Highways 78 and 278	1.56	1979–96
02203543	Wilshire Canal near Savannah, Ga.	31° 59'27"	81° 08'15"	Chatham County, at culvert on Tibet Avenue	.95	1979–96
02203544	Wilshire Canal tributary near Savannah, Ga.	31° 58'25"	81° 08'20"	Chatham County, at culvert on Windsor Road	.18	1979–96

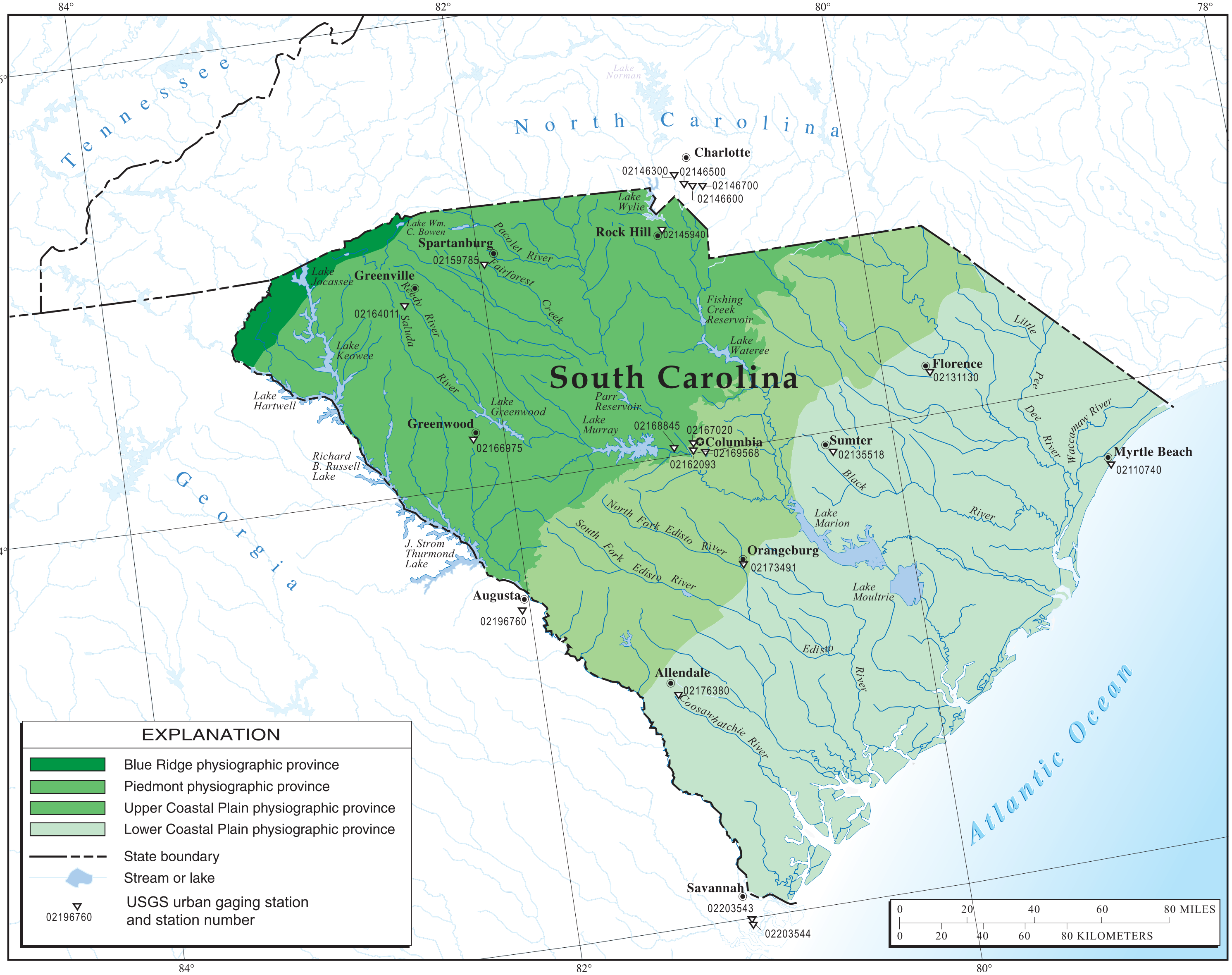


Figure 1. Location of streamflow-gaging stations with 10 or more years of record in urban areas of the Piedmont and upper and lower Coastal Plain physiographic provinces of South Carolina, Georgia, and North Carolina.

listed in the respective station peak-flow file with those in the station's water-year analysis file. Part of this review included a visual inspection of the peak-flow data plotted by water year. Along with the visual inspection, a Kendall tau statistic was used to assess the homogeneity of the record at each streamflow-gaging station.

According to Rantz and others (1982), a rating should not be extended more than two times the maximum measured flow unless no other options are available. Therefore, at streamflow-gaging stations with established ratings, the flow-measurement files were reviewed to determine if the ratings had been excessively extended. The peak flows were plotted against water year along with the maximum measured flow and the flow equal to two times the maximum measured flow. This plot was used to review peaks that may have been estimated from an excessive stage-flow rating extension, but not updated if and when the rating was later defined by greater flow measurements.

From these initial reviews, three of the South Carolina streamflow-gaging stations were excluded from the analysis: station 02160325 (Brushy Creek near Greenville, S.C.), station 02169505 (Rocky Branch at Columbia, S.C.), and station 02173495 (Sunnyside Canal at Orangeburg, S.C.). At station 02160325 (Brushy Creek near Greenville), 8 of the 15 peak flows exceeded the station rating by more than two times the maximum measured flow. At station 02169505 (Rocky Branch at Columbia), the stage-flow relation is incomplete at high stages because physical conditions prevent the flow from being computed. At station 02173495 (Sunnyside Canal at Orangeburg), a trend was detected in the measured data. An exposed pipeline located downstream from this gage often catches debris, which may produce backwater at the gage. As a result of the data trend and problems noted during gage inspections, this station was excluded from the analysis.

Graphical Comparisons of Measured and Simulated Peak Flows

For each of the 13 South Carolina streamflow-gaging stations (fig. 1; table 1), a comparison was made between the measured peak-flow data and the simulated peak-flow data generated by using a rainfall-runoff model (Bohman, 1992). A graphical comparison was made for each streamflow-gaging station by plotting the two sets of data together by water year (fig. 2). At most of the 13 streamflow-gaging stations, the plots show a considerable difference in the variance of the measured and simulated peaks. Note in figure 2 for streamflow-gaging stations 02159785, 02166975, 02173491, and 02176380 that two sets of simulated data are given. Bohman (1992) noted that the rainfall-runoff model showed little sensitivity of volumes and peaks to the evaporation data sets used to synthesize long-term hydrographs but was sensitive to the long-term rainfall record chosen for a basin. Usually, data from the closest long-term rainfall and evaporation stations were used to synthesize the long-term hydrographs. Bohman (1992) noted that "*Even in cases where both long-term rainfall stations seemed to be located in physiographically and meteorologically similar settings, substantially different results were obtained when each rainfall-data set was applied to the calibrated basin models.*" Therefore, for study basins located between rainfall stations where such disparity in results occurred, the flood-frequency estimates were interpolated by weighting the results inversely proportional to the distance between the site and the two rainfall stations.

In the previous investigation by Bohman (1992), five National Weather Service rainfall stations were used in the synthesis of long-term flood-hydrograph data, with periods of record ranging from 49 to 89 years (table 2).

Table 2. National Weather Service rainfall stations used in Bohman's (1992) synthesis of long-term flood-hydrograph data.

Station number	Location	Number of years of record	Period of record
320800081120050	Savannah, Ga.	89	1898–1987
332200081580050	Augusta, Ga.	72	1902–73
340000081030001	Columbia, S.C.	53	1901–53
345000082240001	Greenville-Spartanburg, S.C.	49	1918–71
351400080560001	Charlotte, N.C.	68	1901–69

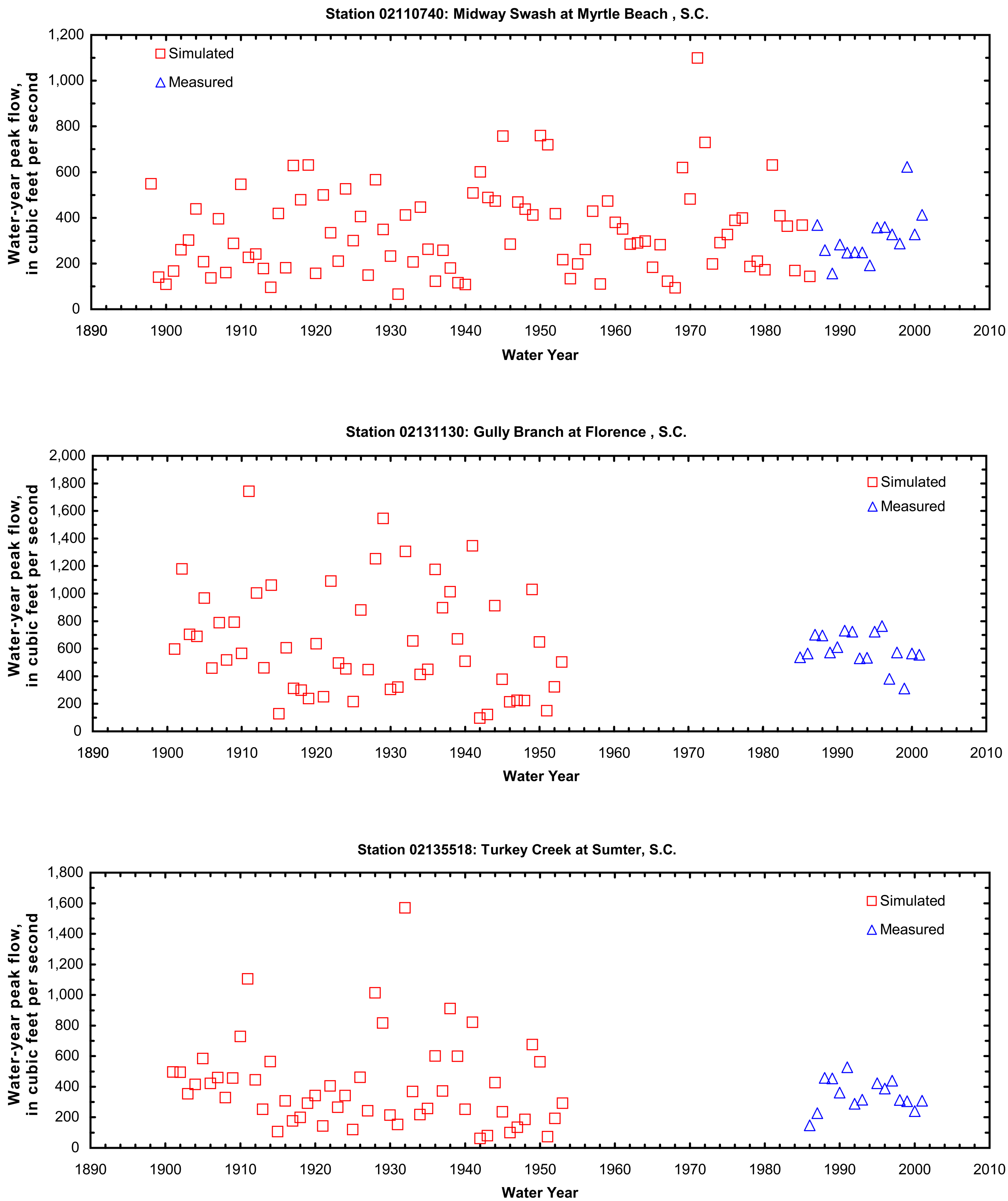


Figure 2. Comparison of measured and simulated peak flows at 13 urban streamflow-gaging stations in South Carolina.

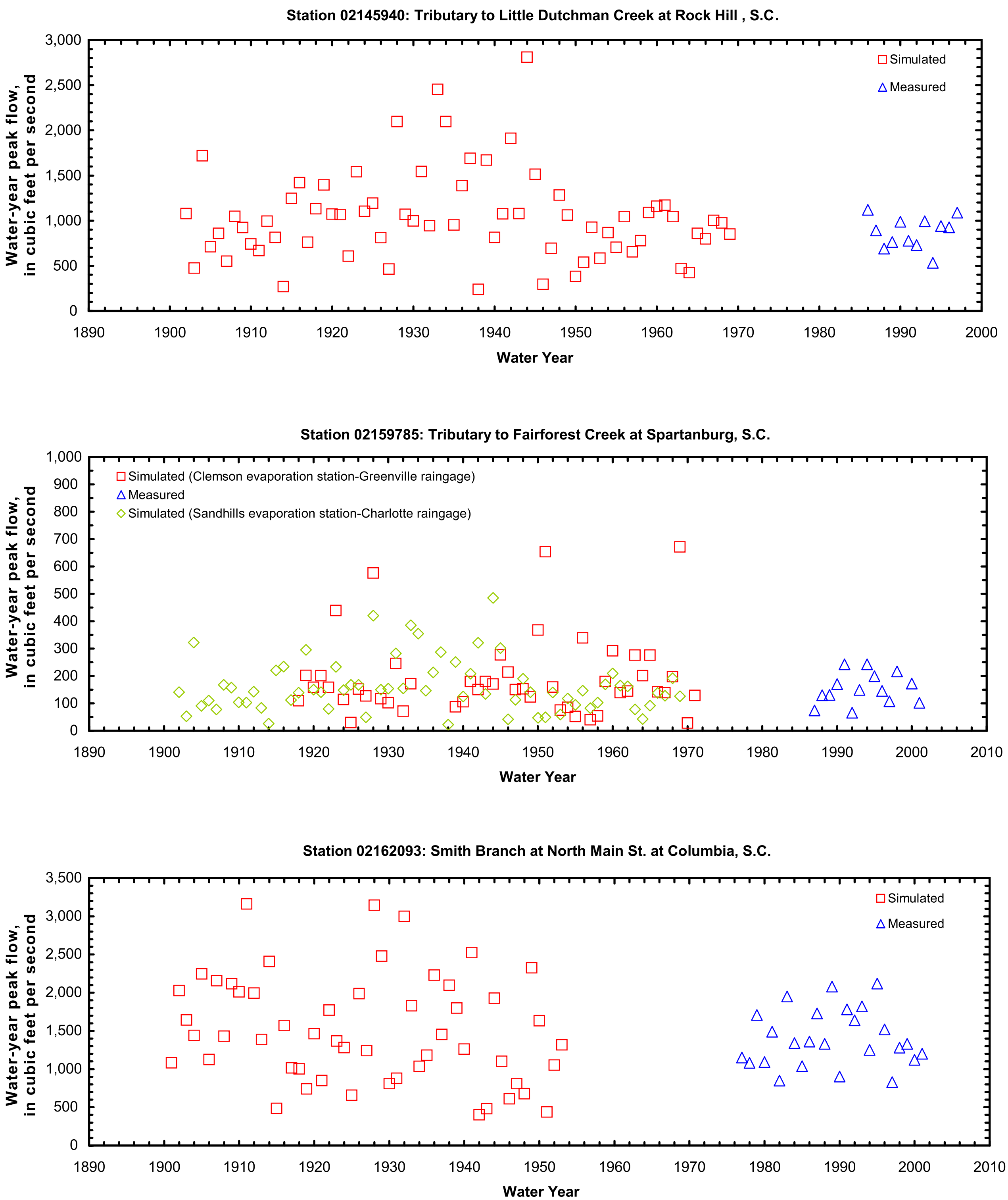
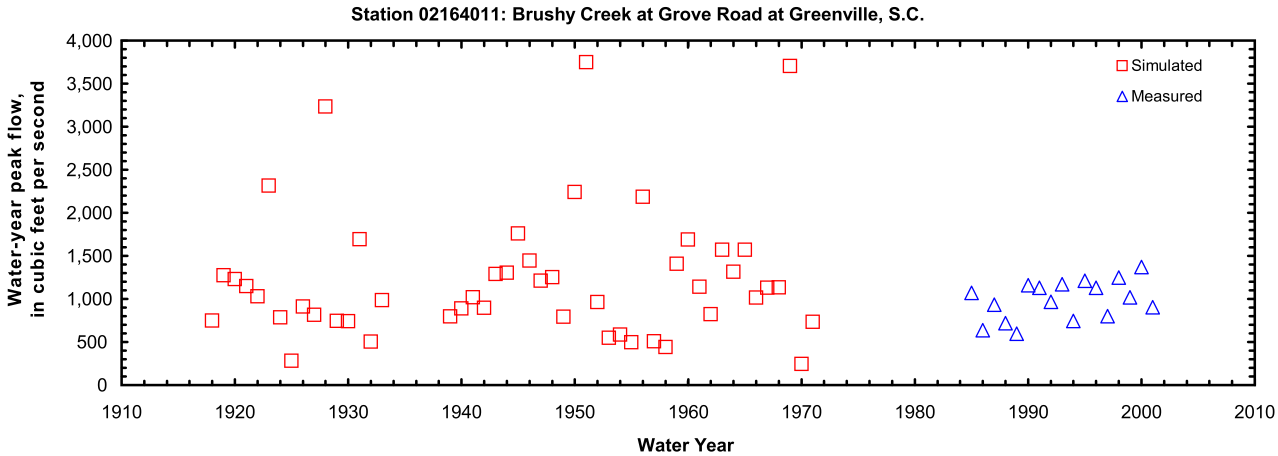
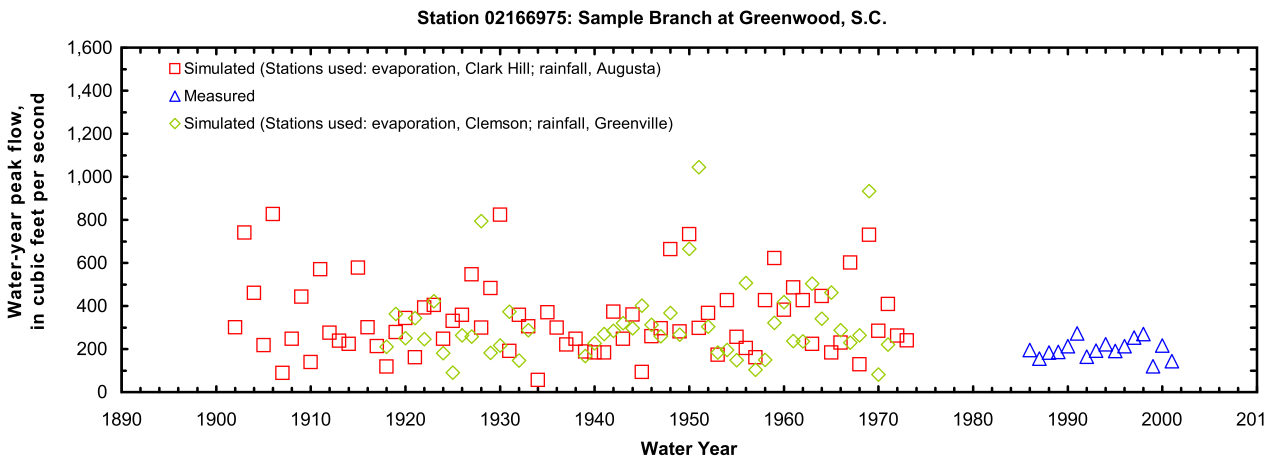


Figure 2. (Continued) Comparison of measured and simulated peak flows at 13 urban streamflow-gaging stations in South Carolina.





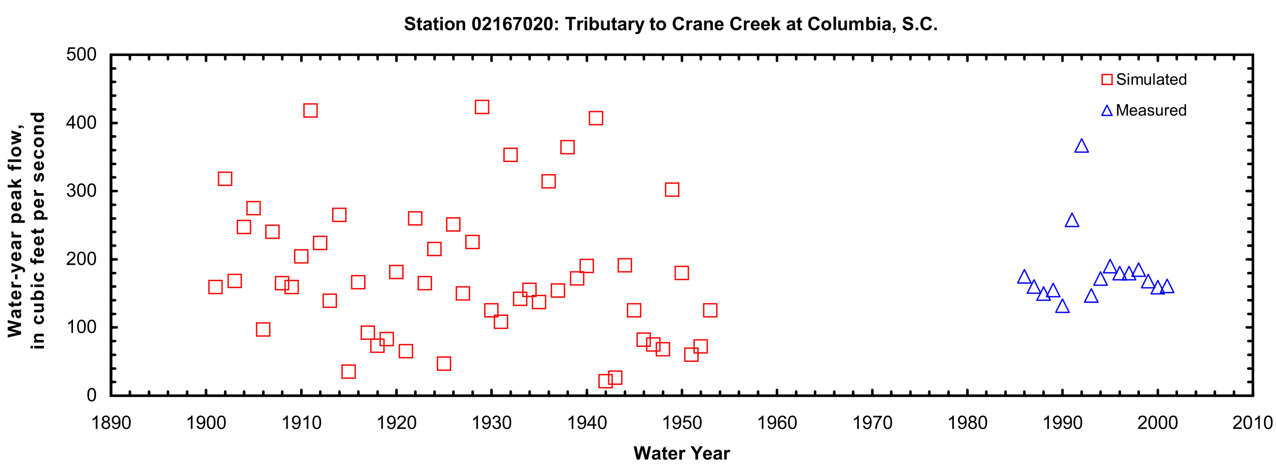
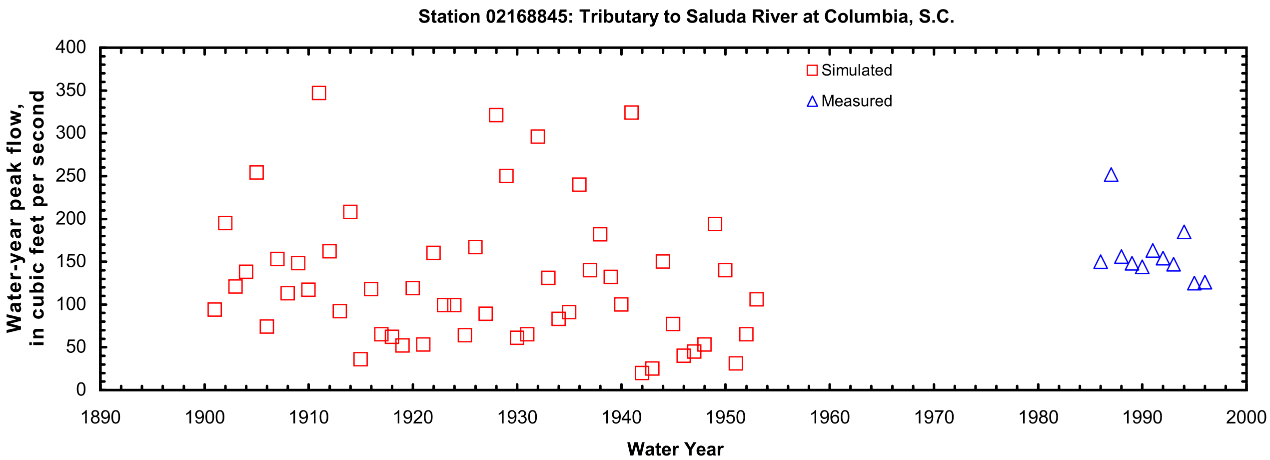
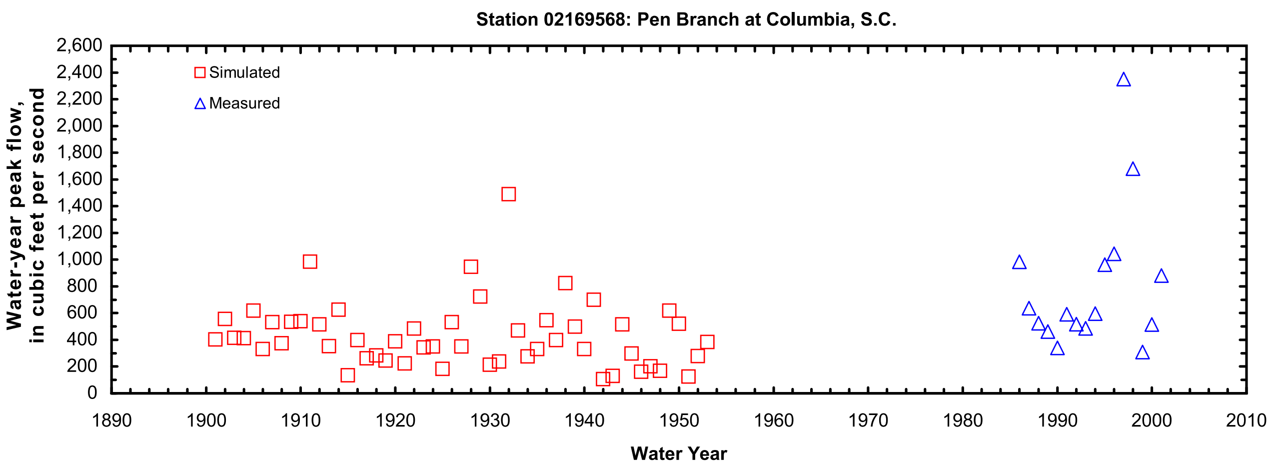


Figure 2. (Continued) Comparison of measured and simulated peak flows at 13 urban streamflow-gaging stations in South Carolina.





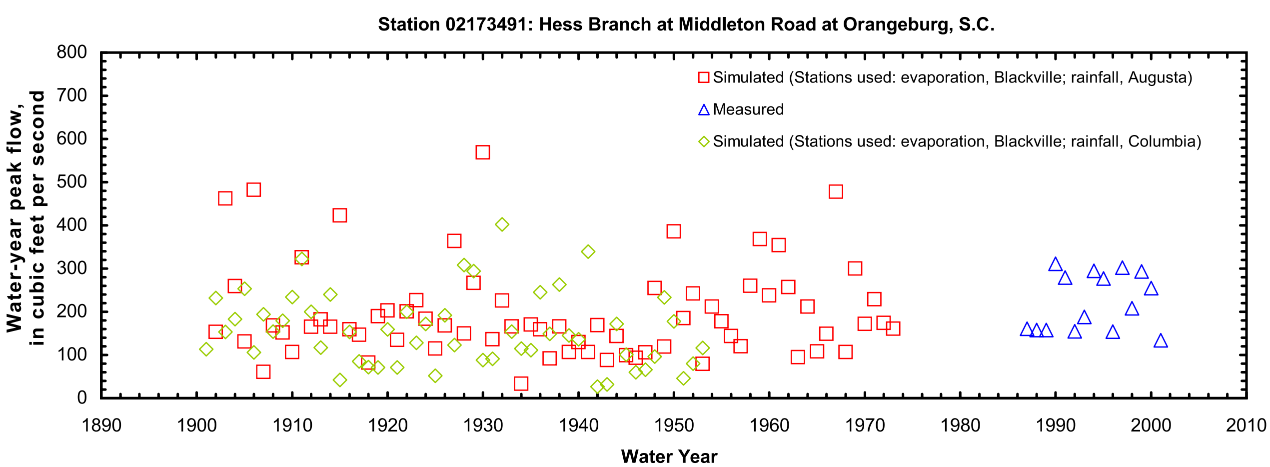


Figure 2. (Continued) Comparison of measured and simulated peak flows at 13 urban streamflow-gaging stations in South Carolina.

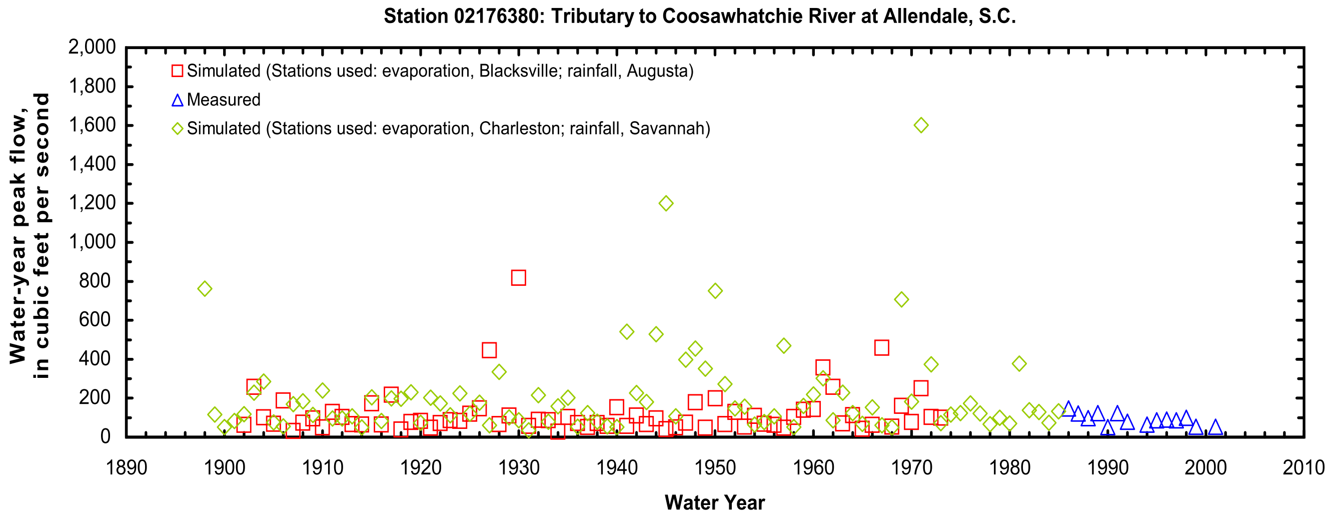


Figure 2. (Continued) Comparison of measured and simulated peak flows at 13 urban streamflow-gaging stations in South Carolina.

To determine if the variance differences between the measured peak-flow data and the simulated peak-flow data could be related to climatic differences in the collection periods of measured data and at the long-term raingages, a review was made of 16 rural streamflow-gaging stations that were included in the South Carolina rural flood-frequency analysis (Feaster and Tasker, 2002). These streamflow-gaging stations all have systematic record lengths through water year 2001 ranging from 42 to 74 years (table 3; fig. 3).

The peak-flow data for the rural streamflow-gaging stations were plotted for two periods—from the beginning of the record to 1984 and from 1985 to 2001 (fig. 4). The 1985–2001 period was chosen because 1985 was the earliest beginning year for 12 of the 13 South Carolina urban streamflow-gaging stations. Station 02162093 (Smith Branch at Columbia, S.C.) is the only continuous-record streamflow-gaging station of the 13 South Carolina urban streamflow-gaging stations with the period of record beginning in 1977. The plots of the peak flows at the rural streamflow-gaging stations do not show significant differences between the two periods, which suggests that the differences in the variances of the simulated and measured data at the 13 South Carolina urban streamflow-gaging stations are not related to climatic differences.

Table 3. Rural streamflow-gaging stations that were used to compare climatic conditions with similar periods at the 13 South Carolina urban streamflow-gaging stations, by physiographic province.

Station number (fig. 3)	Station name	Drainage area (square miles)	Number of years of record	Period of record
		Piedmont		
02147500	Rocky Creek at Great Falls, S.C.	194	45	1952–2001
02154500	North Pacolet River at Fingerville, S.C.	116	71	1931–2001
02160000	Fairforest Creek near Union, S.C.	183	60	1940–2001
02162500	Saluda River near Greenville, S.C.	295	57	1942–2001
02163500	Saluda River near Ware Shoals, S.C.	581	63	1939–2001
02165000	Reedy River near Ware Shoals, S.C.	236	61	1940–2001
02192500	Little River near Mt. Carmel, S.C.	217	59	1940–2001
02196000	Stevens Creek near Modoc, S.C.	545	58	1940–2001
		Upper Coastal Plain		
02130900	Black Creek near McBee, S.C.	108	42	1960–2001
02132500	Little Pee Dee River near Dillon, S.C.	524	61	1940–2001
02173000	South Fork Edisto River near Denmark, S.C.	720	69	1932–2001
02173500	North Fork Edisto River at Orangeburg, S.C.	683	63	1939–2001
02174000	Edisto River near Branchville, S.C.	1,720	56	1946–2001
		Lower Coastal Plain		
02110500	Waccamaw River near Longs, S.C.	1,110	51	1951–2001
02136000	Black River at Kingstree, S.C.	1,250	74	1928–2001
02176500	Coosawhatchie River near Hampton, S.C.	203	51	1952–2001

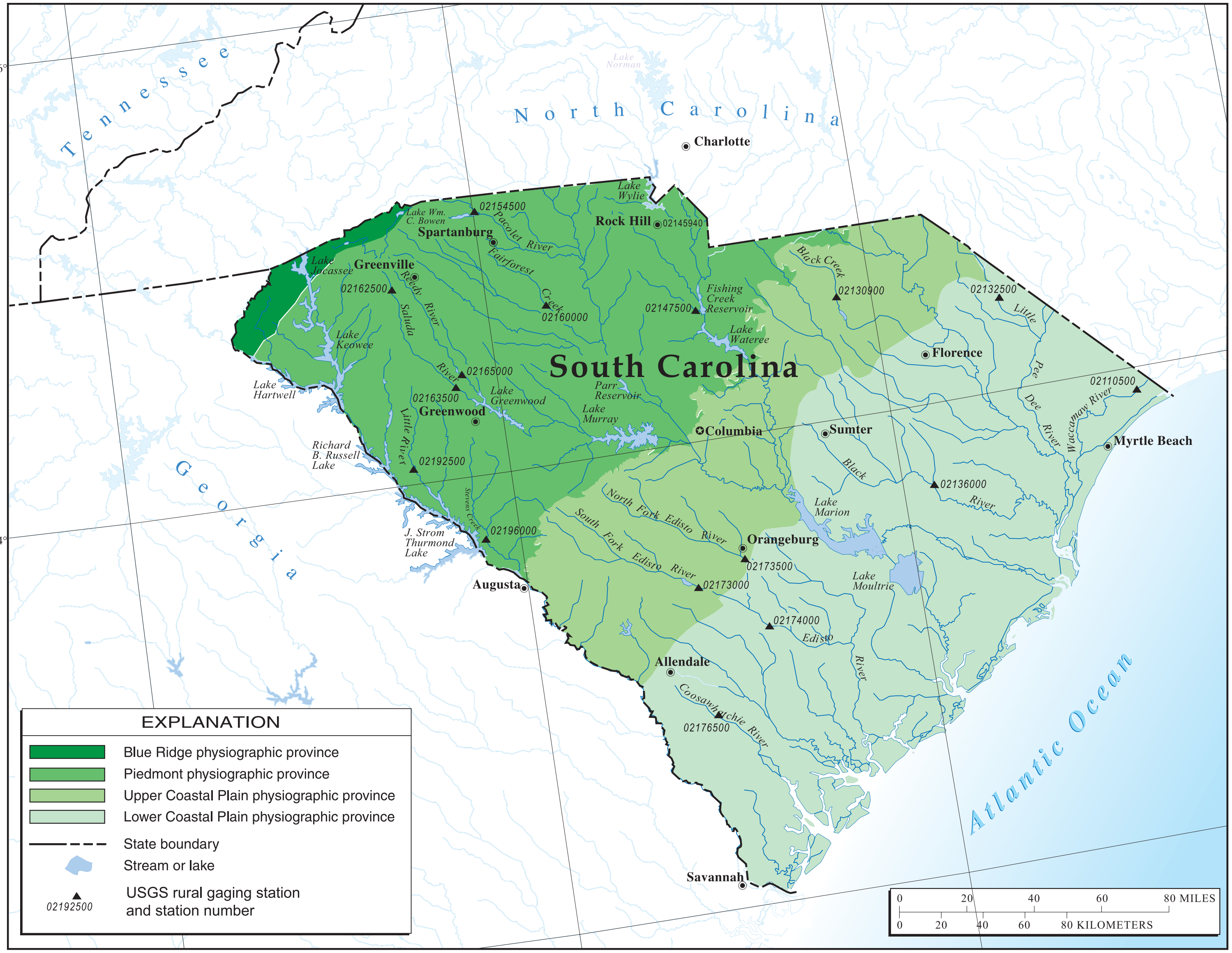


Figure 3. Location of cities, rural streamflow-gaging stations used for comparison with urban streamflow-gaging stations, and physiographic provinces in South Carolina.

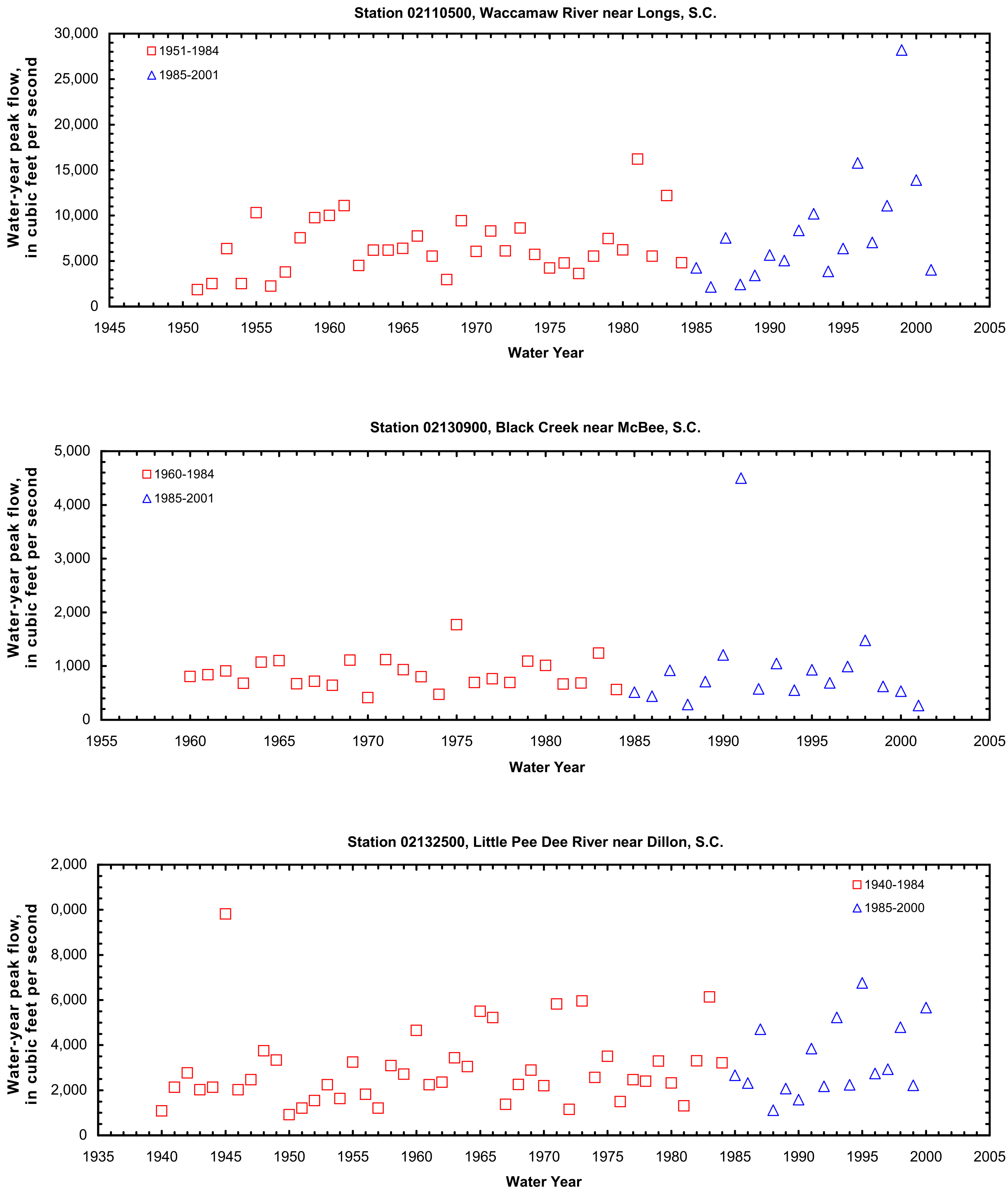
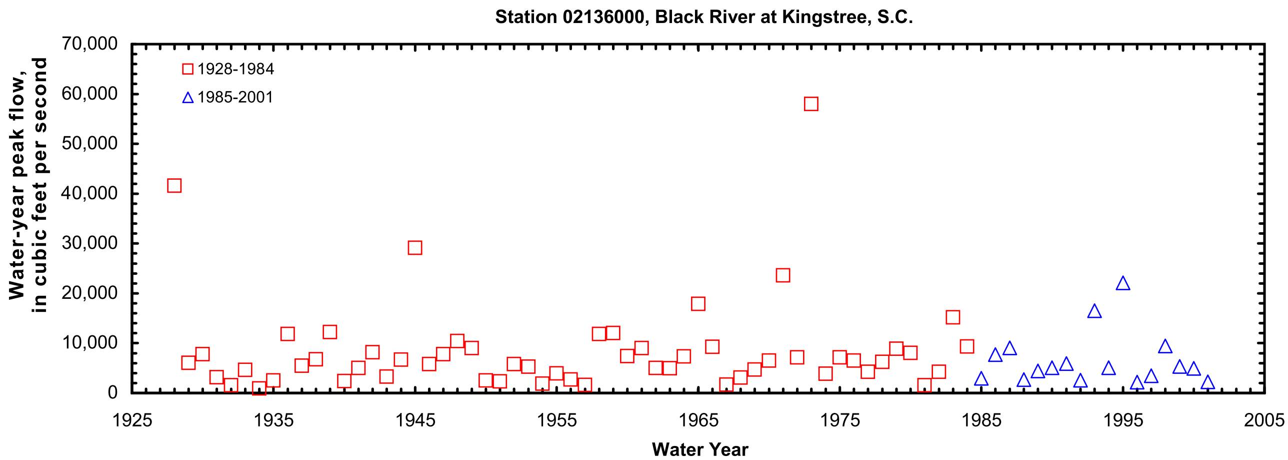
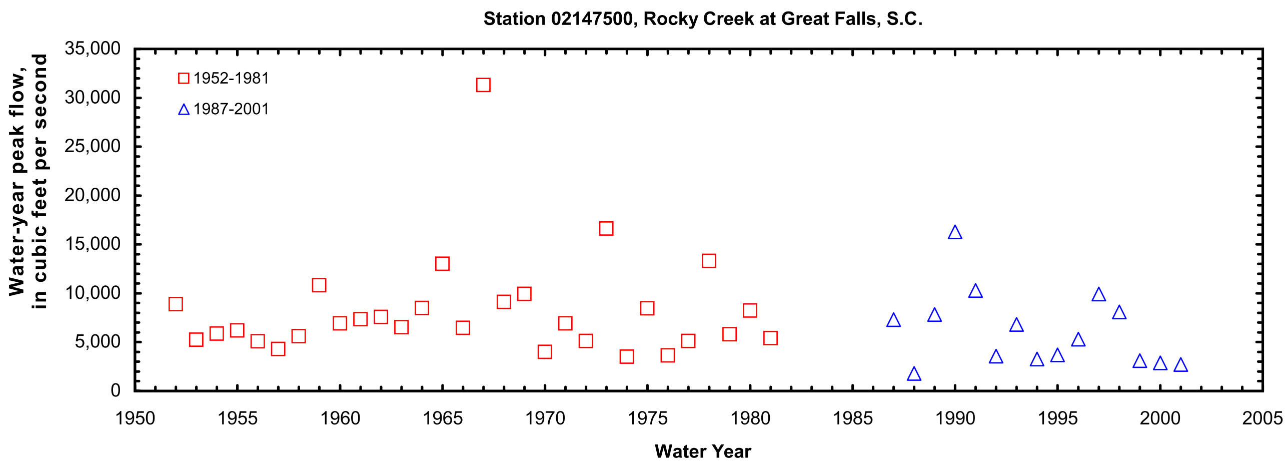


Figure 4. Comparison of two peak-flow periods (beginning of record to 1984 and 1985–2001) at 16 rural streamflow-gaging stations in South Carolina.





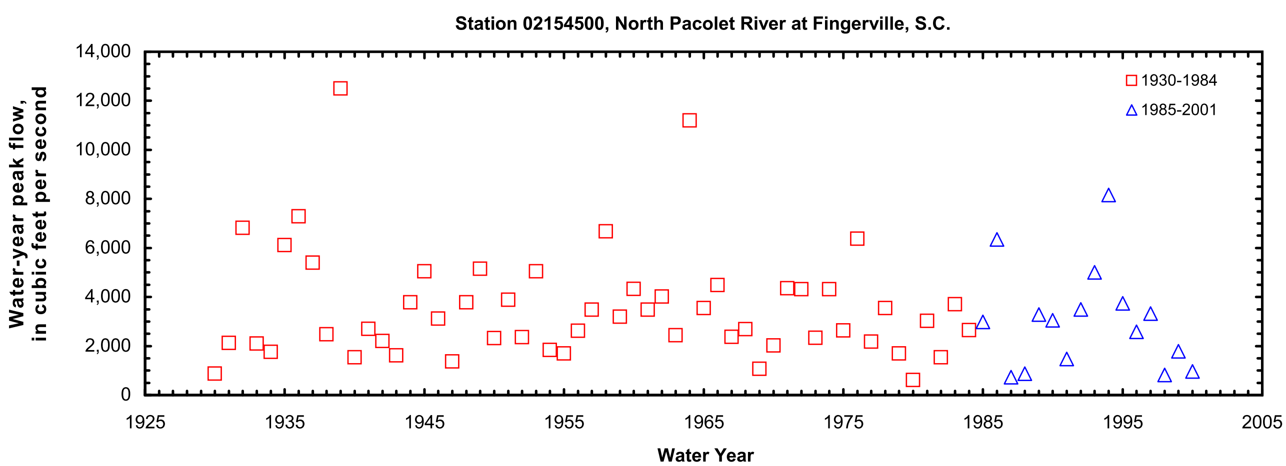
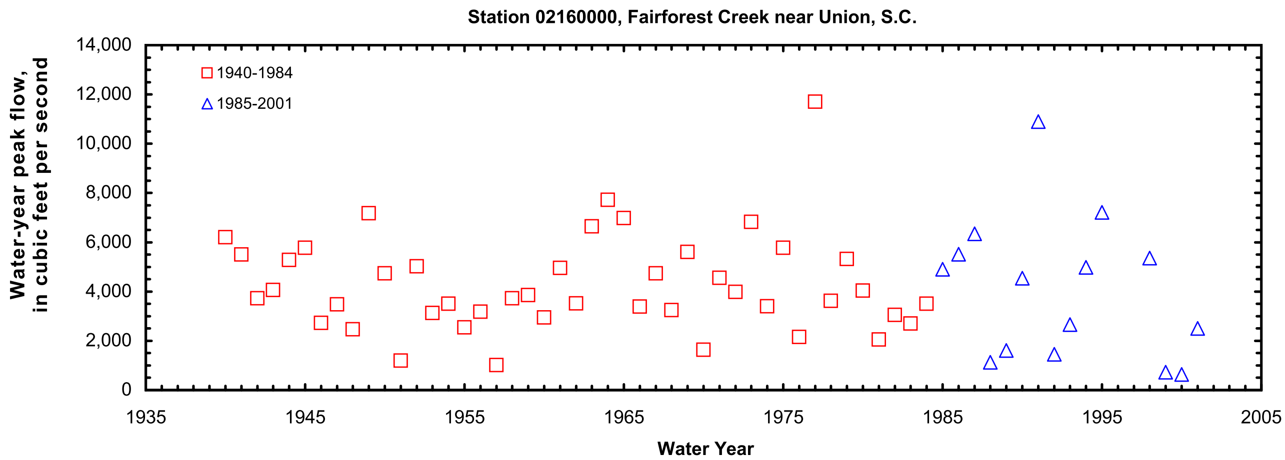
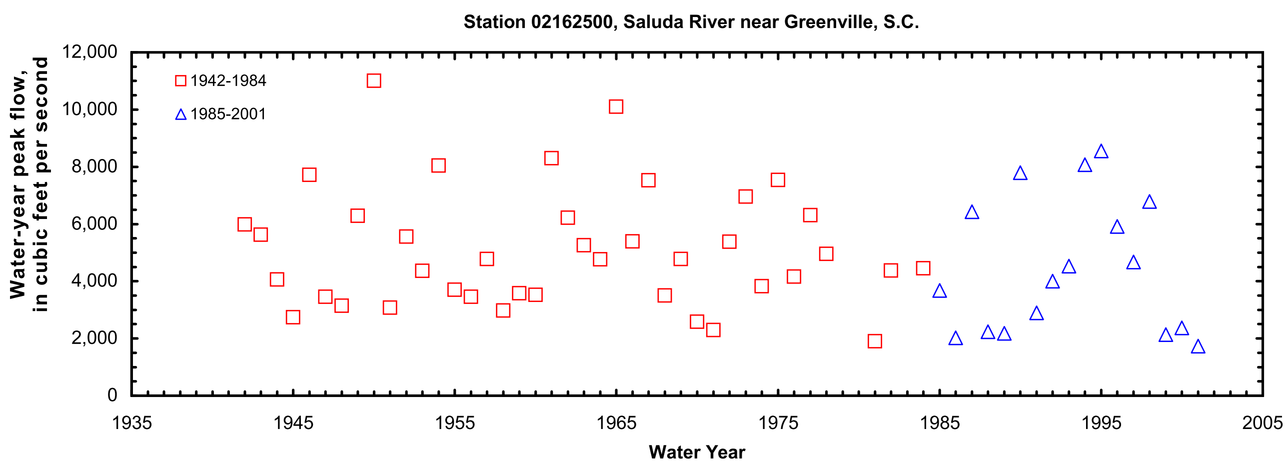


Figure 4. (Continued) Comparison of two peak-flow periods (beginning of record to 1984 and 1985–2001) at 16 rural streamflow-gaging stations in South Carolina.





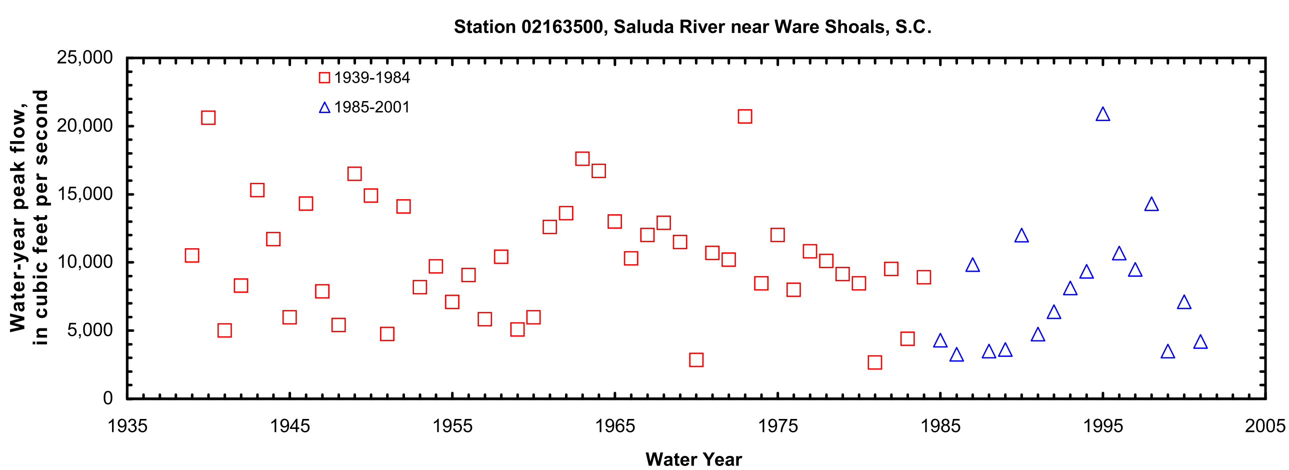
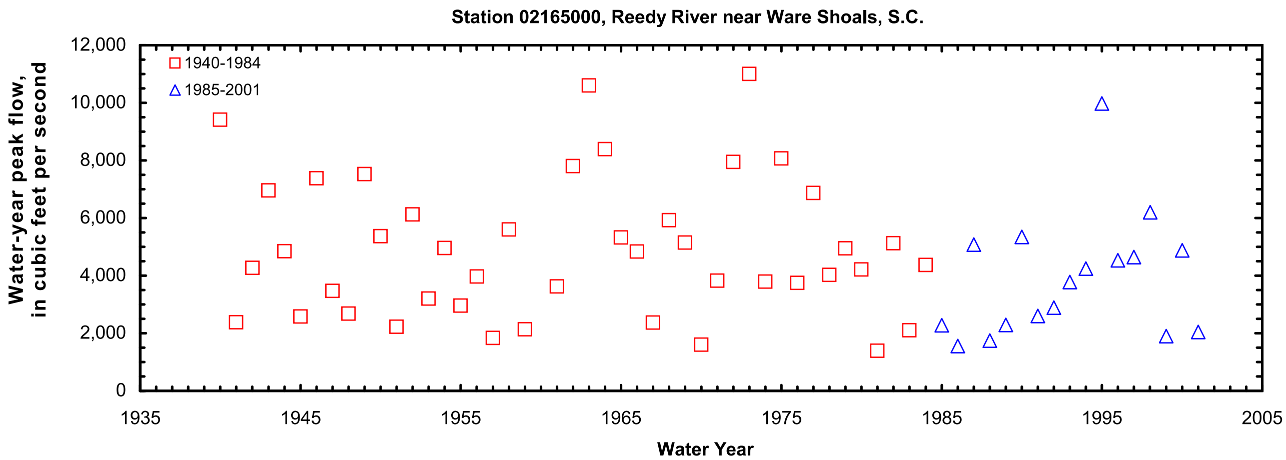
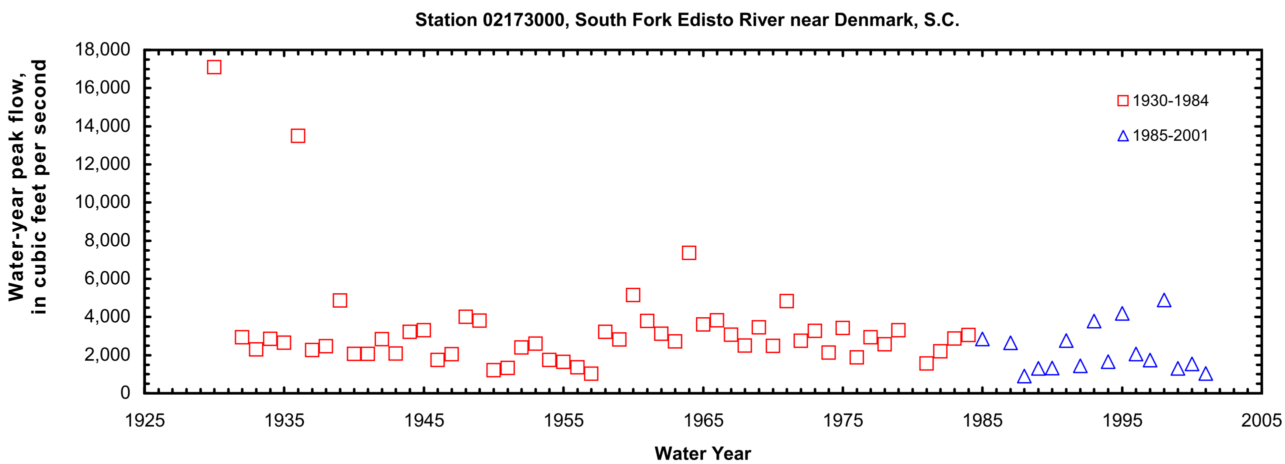


Figure 4. (Continued) Comparison of two peak-flow periods (beginning of record to 1984 and 1985–2001) at 16 rural streamflow-gaging stations in South Carolina.





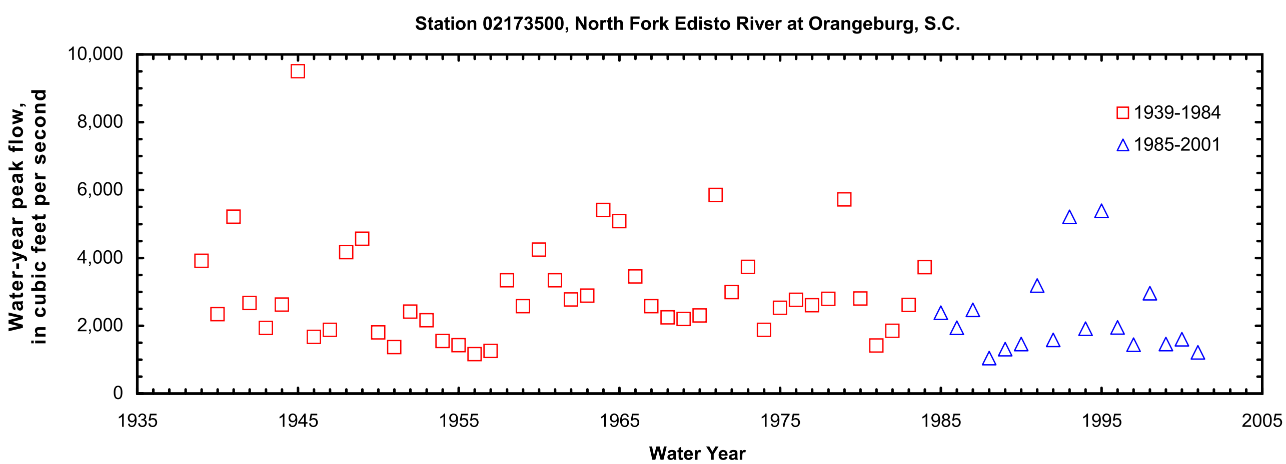
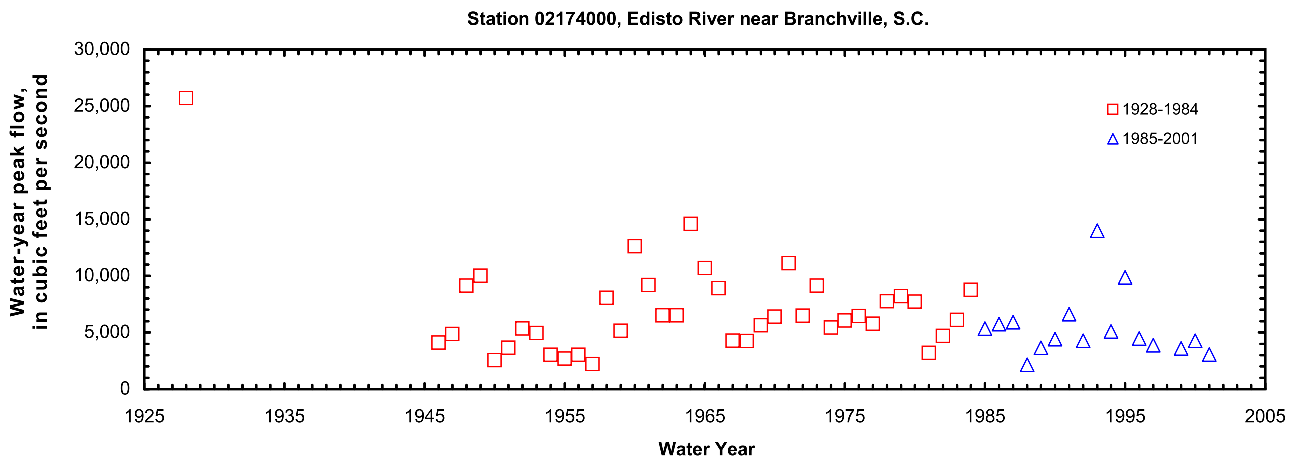
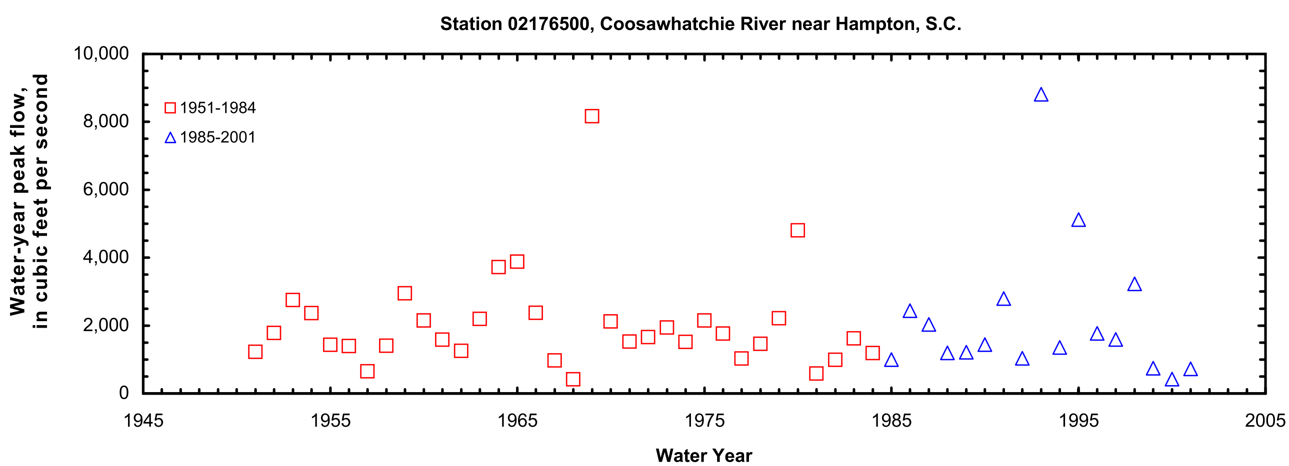


Figure 4. (Continued) Comparison of two peak-flow periods (beginning of record to 1984 and 1985–2001) at 16 rural streamflow-gaging stations in South Carolina.





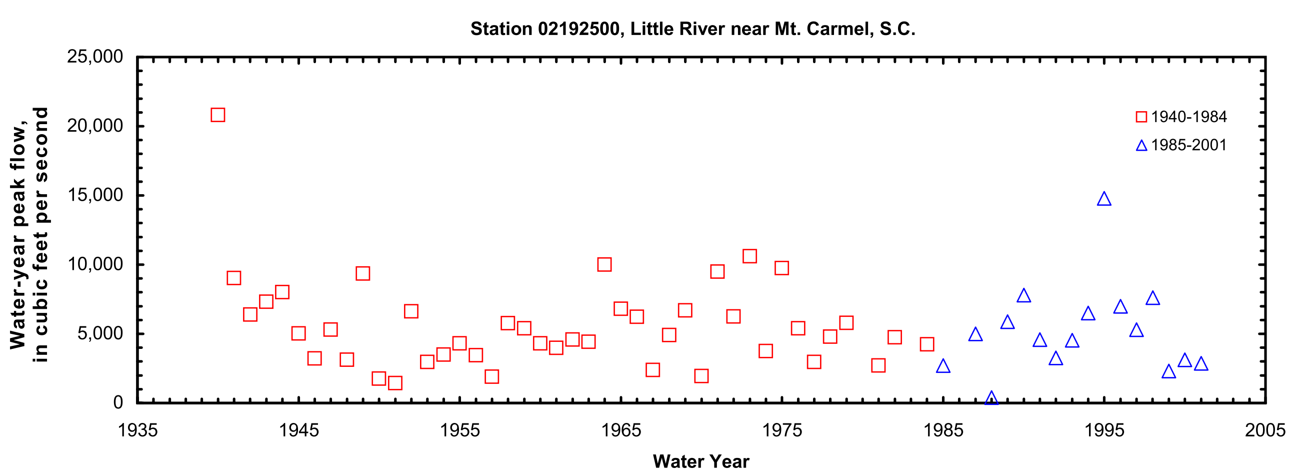


Figure 4. (Continued) Comparison of two peak-flow periods (beginning of record to 1984 and 1985–2001) at 16 rural streamflow-gaging stations in South Carolina.

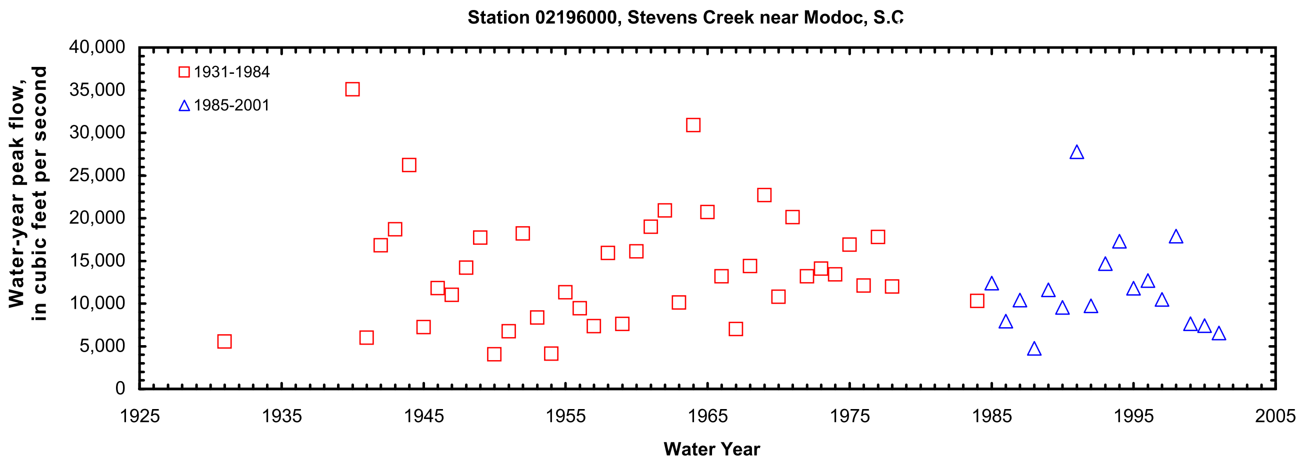


Figure 4. (Continued) Comparison of two peak-flow periods (beginning of record to 1984 and 1985–2001) at 16 rural streamflow-gaging stations in South Carolina.

Statistical Comparisons of Measured and Simulated Peak Flows

Along with the graphical comparisons of the simulated and measured peak flows at the 13 South Carolina urban streamflow-gaging stations, several statistical tests also were used to compare the data. The statistical analyses and computations were made using procedures defined by the SAS Institute, Inc. (1990). The statistical analyses were performed for a *p*-value of 0.05. The *p*-value, also known as the level of significance, is the probability of obtaining the computed test statistic, or one even less likely, when the null hypothesis is true. The null hypothesis is what is assumed to be true about the data until indicated otherwise. It usually states the "null" situation—no difference between groups (Helsel and Hirsch, 1995). In the case of equal variances, the null hypothesis is that the variances between the measured and simulated data are equal. Thus for *p*-values of less than 0.05, the null hypothesis is rejected, and it is reported that there is a statistically significant difference in the variances at the 0.05 level. In other words, there is less than a 5-percent chance that the variances are equal.

Kendall Tau Trend Analysis

One of the assumptions used in the flood-frequency analysis is that the watershed of a streamflow-gaging station does not change significantly through the data-collection period; as such, the peak-flow characteristics at each streamflow-gaging station are homogeneous or do not significantly change over time. The Kendall tau statistical test was chosen to assess the homogeneity of the record at each streamflow-gaging station. The Kendall tau trend analysis was used to determine if a trend exists in the data by measuring the correlation of the peak flow and years (time). The Kendall tau test is based on a ranking system and measures the strength of the monotonic relation between two variables. In a monotonic relation, successive values in a sequence either consistently increase or decrease but do not oscillate in relative value. Being rank based, the Kendall tau statistical test is resistant to the effect of a small number of unusual values (Helsel and Hirsch, 1995).

Measures of monotonic correlation (τ) are characterized by being dimensionless and scaled between positive one and negative one. When the two variables are not correlated, τ equals zero. When one variable increases with the increase of the other variable, τ is a positive number, and when the two variables vary in opposite directions, τ is negative. When one variable is a measure of time or location, correlation becomes a test for temporal or spatial trend. The significance of the correlation is evaluated by forming a null hypothesis that the coefficient is zero against the alternative that it is nonzero, then computing the probability of rejecting the null hypothesis (Helsel and Hirsch, 1995). The probability of rejecting the null hypothesis was computed to a 0.05 level of significance.

Results of the trend analysis of the relation between peak flows and time are shown in table 4. Both the measured and simulated data were analyzed. A trend in the simulated data would indicate a significant change in rainfall patterns over time. A trend in the measured data, in the absence of a climatic trend, would indicate a change in urbanization over time. For the 0.05 level of significance, there were no statistically significant trends in either the measured or the simulated data (table 4), although two streamflow-gaging stations were considered borderline cases with *p*-values at the 0.05 level (simulated data at station 02135518 and measured data at station 02176380).

Table 4. Summary of statistical trends in water-year peak flows for 13 urban streamflow-gaging stations in South Carolina.

Station number (fig. 1)	Period of record	Kendall-tau value	*p*-value
Simulated data			
02110740	1898–1986	0.05	0.53
02131130	1901–53	-.16	.09
02135518	1901–53	-.19	.05
02145940	1902–69	-.05	.57
02159785	1918–71	.01	.90
02162093	1902–53	-.18	.06
02164011	1918–71	.03	.74
02166975	1902–74	.01	.92
02167020	1901–53	-.17	.07
02168845	1901–53	-.14	.13
02169568	1901–53	-.18	.06
02173491	1902–74	.01	.92
02176380	1903–73	.02	.86
Measured data			
02110740	1987–2001	0.35	0.07
02131130	1985–2001	-.14	.43
02135518	1986–2001	-.10	.59
02145940	1986–97	-.02	.94
02159785	1987–2001	.13	.51
02162093	1977–2001	.01	.94
02164011	1985–2001	.30	.09
02166975	1986–2001	.13	.50
02167020	1986–2001	.06	.75
02168845	1986–96	-.34	.11
02169568	1986–2001	.13	.50
02173491	1986–2001	.13	.50
02176380	1986–2001	-.37	.05

F-test for Equality of Variances

The *F*-test was chosen to test the equality of variances for the simulated and measured peak-flow data at the 0.05 level of significance. A random variable that consists of the ratio of two sample variances has an *F* distribution if the two samples are independent and from normal populations with equal population variances (Iman and Conover, 1983). Therefore, before testing for equal variances, a univariate procedure was used to determine if all distributions were normal according to the Shapiro-Wilks statistic (SAS Institute, Inc., 1990). All peak-flow data were transformed to logarithmic units before conducting the statistical analysis and computations. For 8 of the 13 urban streamflow-gaging stations, the Shapiro-Wilks statistic showed that the logarithms of the measured and simulated peak flows were normally distributed (table 5). For seven of the eight streamflow-gaging stations with normal distributions, the *F*-test indicated that there was a statistically significant difference in the variances of the measured and simulated data. For station 02169568, the *F*-test indicated that there was no statistically significant difference in the variances at the 0.05 level of significance. At three of the streamflow-gaging stations where the Shapiro-Wilks statistical test indicated that the peaks were not normally distributed at the 0.05 level of significance, the *p*-values were not much lower than 0.05 (0.04, 0.02, and 0.01). Although the *F*-test is not technically correct for data sets that are not normally distributed, the *F*-test was still used as an indicator for the three streamflow-gaging stations and showed that a statistically significant difference in the variances of the measured and simulated data occurred at all three streamflow-gaging stations. For station 02167020, the Shapiro-Wilks statistical test indicated that neither the measured nor the simulated data were normally distributed; therefore, the *F*-test was not used to compare the variances. It is clear from the plot of the measured and simulated peak-flow data, however, that there is a significant difference in the variances (fig. 2).

For comparison purposes, similar statistics were computed for the two periods at the 16 rural streamflow-gaging stations listed in table 3 and shown in figures 3 and 4. The Shapiro-Wilks statistic indicated that the logarithms of the peak-flow data at 14 of the 16 streamflow-gaging stations were distributed normally for the two periods (table 6). The Shapiro-Wilks statistic indicated that the peak-flow data at station 02192500 for the 1985–2001 period were not distributed normally. A series of peak flows at a streamflow-gaging station may include low or high outliers, which are data points that depart significantly from the range of the remaining data. Based on the flood-frequency analysis at station 02192500, the 1988 peak flow was a low outlier. Therefore, the Shapiro-Wilks test was conducted with the 1988 peak flow excluded, and the test statistic indicated

Table 5. Results of comparison testing of measured and simulated peak-flow data at 13 urban streamflow-gaging stations in South Carolina based on the Shapiro-Wilks test for normality and the *F*-test for equal variances.

[Values in parentheses are *p*-values at the 0.05 level of significance; <, less than]

Station number (fig. 1)	Are the logarithms of the measured data normally distributed?	Are the logarithms of the simulated data normally distributed?	Is there a statistically significant difference in variances?
Piedmont			
02145940	yes (0.46)	yes (0.15)	yes (0.005)
02159785	yes (0.54)	[a]yes, yes (0.18, 0.09)	[a]yes, no (0.04, 0.08)
02164011	yes (0.27)	yes (0.50)	yes (0.001)
02166975	yes (0.73)	[a]yes, yes (0.34, 0.28)	[a]yes, yes (0.001, 0.001)
02168845	no (0.04)	yes (0.69)	[b]yes (0.0002)
Upper Coastal Plain			
02162093	yes (0.64)	yes (0.12)	yes (0.001)
02167020	no (0.002)	no (0.03)	did not compute
02169568	yes (0.31)	yes (0.71)	no (1.00)
02173491	no (0.02)	[a]yes, yes (0.25, 0.30)	[a, b]yes, no (0.04, 0.01))
Lower Coastal Plain			
02110740	yes (0.80)	yes (0.31)	yes (0.02)
02131130	no (0.01)	yes (0.18)	[b]yes (<0.0001)
02135518	yes (0.32)	yes (0.92)	yes (0.003)
02176380	yes (0.35)	[a]no, no (0.003, 0.01)	[a, b]yes, yes (0.02, 0.002)

[a]Two sets of peak flows were synthesized using rainfall data from two different long-term rainfall gages.

[b]Because the Shapiro-Wilks test indicated that one of the data sets was not normally distributed, the *F*-test is not technically valid but was still used as an indicator.

that the logarithms of the peak data were distributed normally. In addition, the Shapiro-Wilks statistic indicated that the logarithms of the peak-flow data at station 02173000 from the beginning of the record to water year 1984 were not distributed normally. Based on the flood-frequency analysis at station 02173000, there were two large peaks that exceeded the high-outlier threshold. When the Shapiro-Wilks test was conducted excluding those two peaks, the logarithms of the data from the beginning of the record to 1984 were distributed normally. Therefore, the *F*-test also was conducted on stations 02192500 and 02173000. The *F*-test indicated that there was no statistically significant difference in the variances for the 1985–2001 period or for the beginning of record to 1984 period for 14 of the 16 rural streamflow-gaging stations. For the 1985–2001 period at station 02130900, the log-Pearson Type-III analysis indicated that the 1991 peak was a high outlier. Consequently, the *F*-test was run excluding the 1991 peak, and the results indicated that there was no statistically significant difference in the period from 1985 to 2001 and the beginning of record to 1984. The *F*-test indicated that there was a statistically significant difference in the variances in the peak-flow data for the two periods at station 02160000. However, the plot of the data in figure 4 indicates that the variances do not appear to be drastically different.

Overall, the *F*-test indicated that there is a statistically significant difference in the variances at 11 of the 13 urban streamflow-gaging stations. The *F*-test indicated no significant difference at one of the urban streamflow-gaging stations. The remaining streamflow-gaging station was not analyzed because the Shapiro-Wilks statistic indicated that the data were not distributed normally. However, for the rural streamflow-gaging stations, the *F*-test indicated that there were no statistically significant differences in the variances at 14 of the 16 streamflow-gaging stations. At station 02130900, a statistically significant difference in the variances of the two periods was present; however, when analyzed without the 1991 peak, which was a high-outlier in the log-Pearson analysis, there was no significant difference in the variances. Consequently, this comparison suggests that differences in the variances between the measured and simulated peak-flow data at the urban streamflow-gaging stations cannot be solely attributed to changes in climatic conditions.

Table 6. Results of comparison testing of peak-flow data from 16 rural streamflow-gaging stations in South Carolina for two periods (beginning of record to 1984 and 1985–2001) based on the Shapiro-Wilks test for normality and the *F*-test for equal variances.

[mi^2, square miles; Values in parentheses are *p*-values at the 0.05 level of significance]

Station number (fig. 3)	Drainage area (mi^2)	Period of measured data	Are the logarithms of the data from 1985 to 2001 normally distributed?	Are the logarithms of the data from the beginning of record to 1984 normally distributed?	Is there a statistically significant difference in variances?
Piedmont					
02147500	194	1952–2001	yes (0.69)	yes (0.07)	no (0.20)
02154500	116	1931–2001	yes (0.28)	yes (0.92)	no (0.19)
02160000	183	1940–2001	yes (0.26)	yes (0.28)	yes (0.002)
02162500	295	1942–2001	yes (0.12)	yes (1.00)	no (0.11)
02163500	581	1939–2001	yes (0.25)	yes (0.09)	no (0.31)
02165000	236	1940–2001	yes (0.51)	yes (0.60)	no (0.92)
02192500	217	1940–2001	no (0.02)	yes (0.87)	no (0.052)
02196000	545	1940–2001	yes (0.99)	yes (0.70)	no (0.43)
Upper Coastal Plain					
02130900	108	1960–2001	yes (0.17)	yes (0.72)	yes (0.002)
02132500	524	1940–2001	yes (0.63)	yes (0.57)	no (0.94)
02173000	720	1932–2001	yes (0.46)	no (0.0004)	no (0.89)
02173500	683	1939–2001	yes (0.12)	yes (0.62)	no (0.78)
02174000	1,720	1946–2001	yes (0.44)	yes (0.81)	no (0.57)
Lower Coastal Plain					
02110500	1,110	1951–2001	yes (0.98)	yes (0.55)	no (0.14)
02136000	1,250	1928–2001	yes (0.29)	yes (0.37)	no (0.39)
02176500	203	1952–2001	yes (0.91)	yes (0.58)	no (0.23)

Wilcoxon Rank Sum Test

The Wilcoxon rank sum test, also know as the Mann-Whitney or Wilcoxon-Mann-Whitney test, was used to check for statistically significant differences in the means of the measured and simulated peak-flow data. The Wilcoxon test is similar to the *t*-test except that it is applied to the ranks of the data rather than to the data values. In addition, the *t*-test assumes that both groups of data are distributed normally and that the variances are the same (Helsel and Hirsch, 1995). The Wilcoxon test makes no such assumptions about how the data are distributed nor does it require that the groups have the same variance. The statistical procedure used to perform the computations was NPAR1WAY (SAS Institute, Inc., 1990).

For the urban streamflow-gaging stations in South Carolina, the Wilcoxon test indicated that there was no statistically significant difference in the mean values of the measured and simulated peak-flow data at 8 of the 13 streamflow-gaging stations (about 62 percent) at the 0.05 level of significance (table 7). As can be seen in table 7, there appears to be no bias with respect to physiographic province or drainage-area size. The logarithmic data also are presented in box plots (fig. 5). Although there is no statistically significant difference in the mean values at 8 of the 13 urban streamflow-gaging stations, once again it is clear that there is a significant difference in the variance of the two groups.

For comparison with the urban streamflow-gaging stations, the 16 rural streamflow-gaging stations used in the comparison of variances also were analyzed using the Wilcoxon rank sum test. As with the comparison of variances, the period from 1985 to 2001 was compared with the period from the beginning of record to 1984. The Wilcoxon test indicated that 12 of the 16 rural streamflow-gaging stations had no statistically significant difference in mean peak-flow values at the 0.05 level of significance (table 8; fig. 6). Of the four streamflow-gaging stations that had a significant difference between the periods, station 02174000 had no significant difference when analyzed without the historic peak that occurred in water year 1928. Consequently, 81 percent of the rural streamflow-gaging stations had no statistically significant difference in mean peak flows from 1985 to 2001 and from the beginning of station record to 1984.

Overall, the results of the Wilcoxon test of the peak-flow data for the urban and rural streamflow-gaging stations were somewhat similar, with 62 percent of the urban streamflow-gaging stations and 81 percent of the rural streamflow-gaging stations showing no statistically significant difference in the mean values. At 31 percent of the urban streamflow-gaging stations, the mean peak flows for the measured data were greater than the mean peak flows for the simulated data, with the average difference being about 18 percent. For 25 percent of the rural streamflow-gaging stations, the mean peak flows for 1984–2001 were greater than the mean peak flows for the beginning of record to the 1984 period, with the average difference being about 12 percent. Consequently, with respect to mean peak flows, the simulated and measured data at the urban streamflow-gaging stations had similar bias as the rural streamflow-gaging stations for the two periods compared. Four of the rural streamflow-gaging stations had peak-flow values that were high outliers based on the log-Pearson analysis—stations 02130900, 02173000, 02173500, and 02174000. The high outliers were excluded in the computations of the mean peak-flow values at these streamflow-gaging stations.

Table 7. Results of comparison testing of measured and simulated peak-flow data from 13 urban streamflow-gaging stations in South Carolina based on the Wilcoxon rank sum test.

[Values in parentheses are *p*-values at the 0.05 level of significance]

Station number (fig. 1)	Drainage area (in square miles)	Does the Wilcoxon test indicate a statistically significant difference in the means?
Piedmont		
02145940	3.50	no (0.14)
02159785	.52	[a]no, no (0.32, 0.39)
02164011	3.02	no (0.22)
02166975	1.16	[a]yes, yes (0.0001, 0.001)
02168845	.45	yes (0.01)
Upper Coastal Plain		
02162093	5.49	no (0.35)
02167020	.28	no (0.27)
02169568	2.26	yes (0.0004)
02173491	.45	[a]yes, yes (0.0498, 0.002)
Lower Coastal Plain		
02110740	0.80	no (0.43)
02131130	1.92	no (0.36)
02135518	2.20	no (0.46)
02176380	2.06	[a]no, yes (0.49, 0.006)

[a]Two sets of peak flows were synthesized for this station using rainfall data from two different long-term rainfall gages.

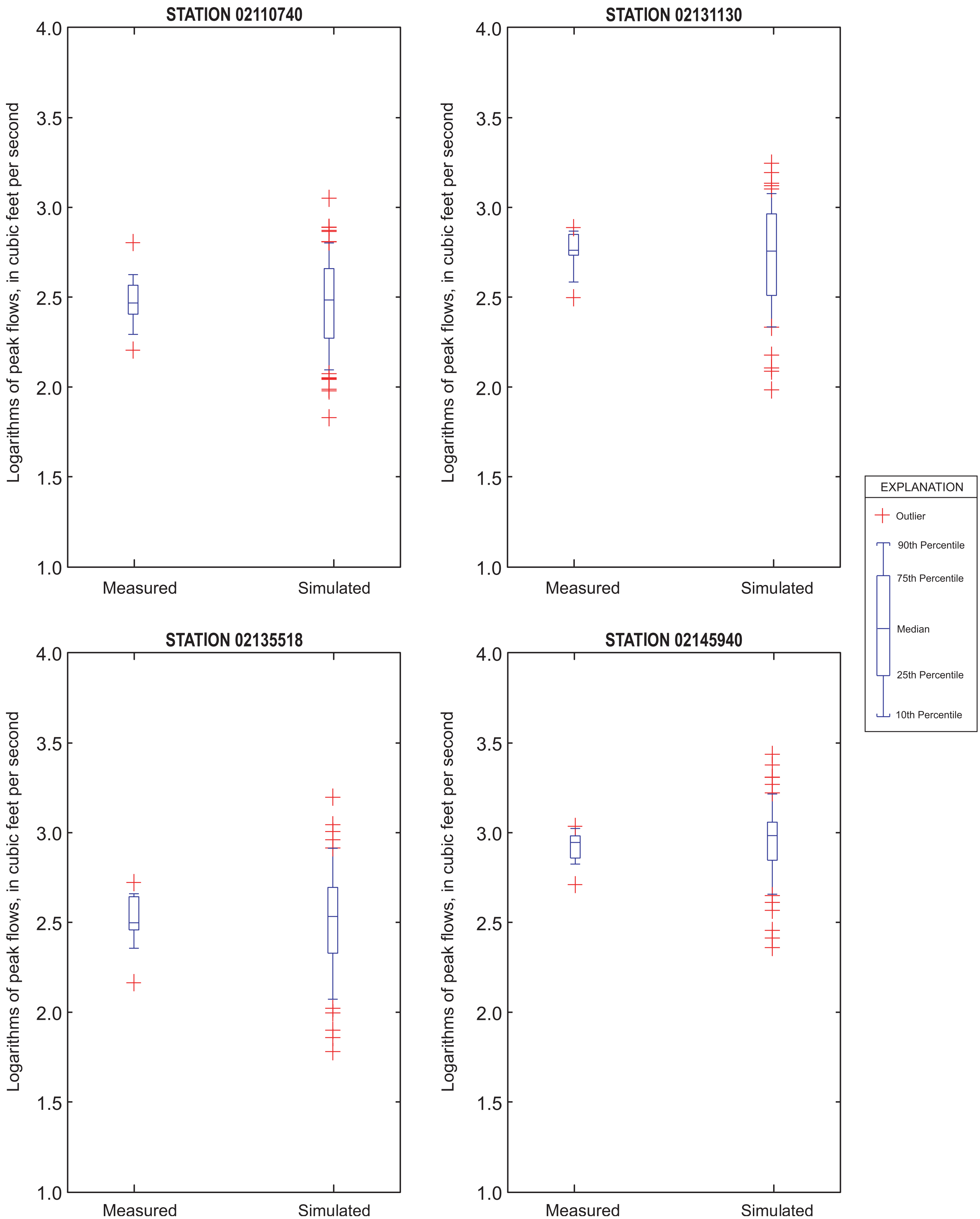


Figure 5. Box plots of the logarithms of the measured and simulated peak-flow data at 13 urban streamflow-gaging stations in South Carolina.

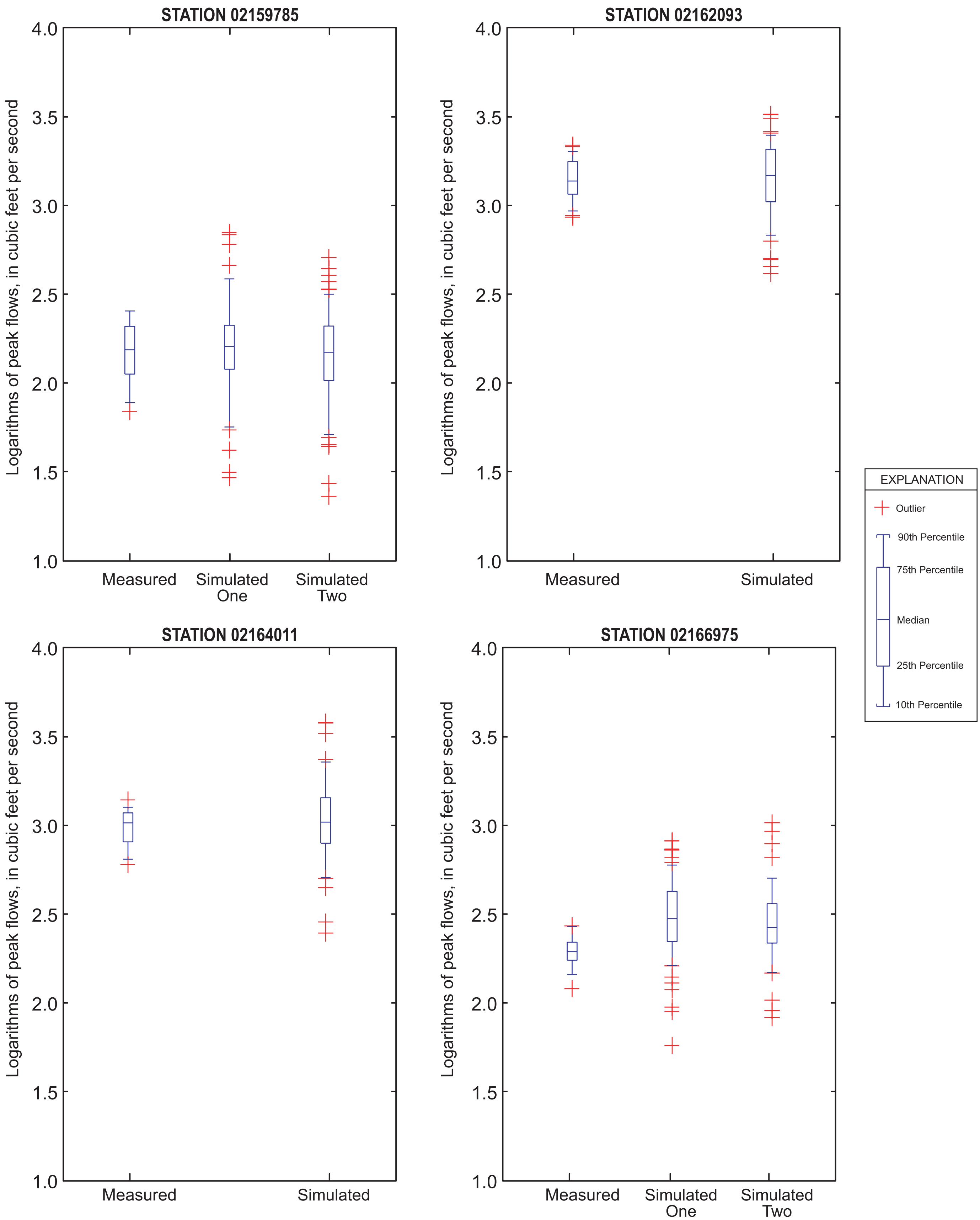


Figure 5. (Continued) Box plots of the logarithms of the measured and simulated peak-flow data at 13 urban streamflow-gaging stations in South Carolina.

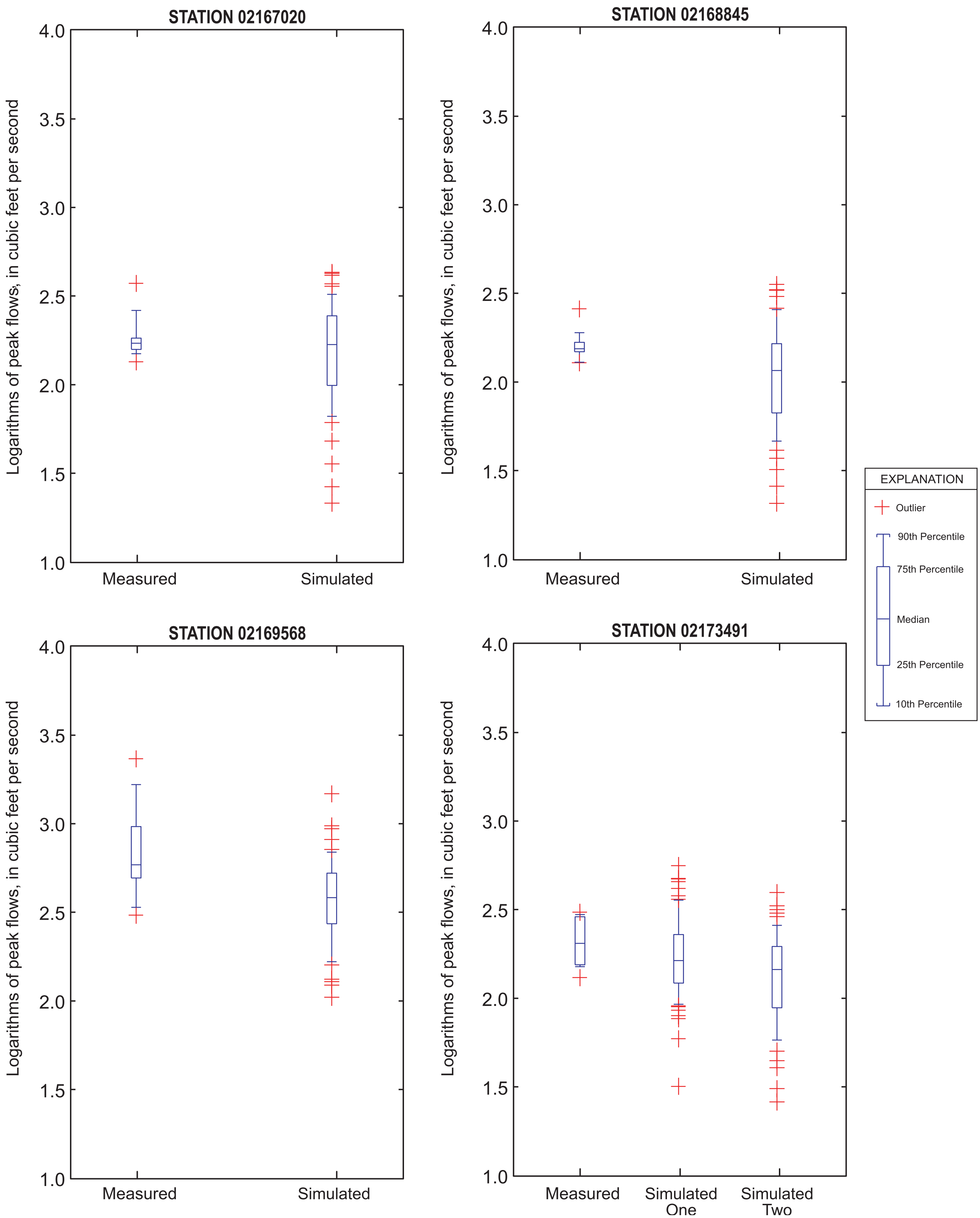


Figure 5. (Continued) Box plots of the logarithms of the measured and simulated peak-flow data at 13 urban streamflow-gaging stations in South Carolina.

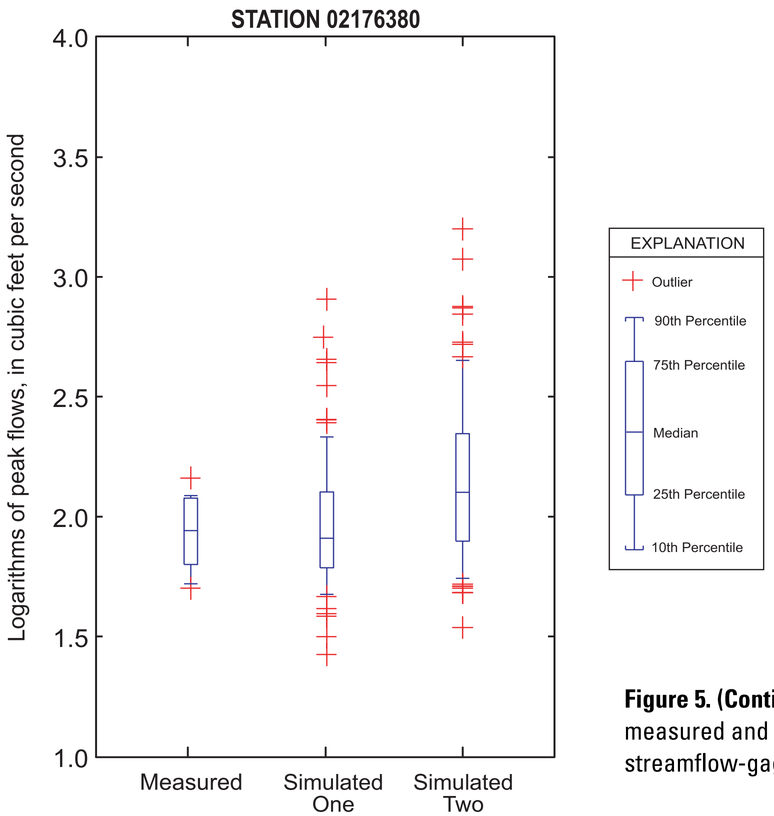


Figure 5. (Continued) Box plots of the logarithms of the measured and simulated peak-flow data at 13 urban streamflow-gaging stations in South Carolina.

Table 8. Results of comparison testing of peak-flow data from 16 rural streamflow-gaging stations in South Carolina for two periods (beginning of record to 1984 and 1985–2001) based on the Wilcoxon rank sum test.

[Values in parentheses are *p*-values at the 0.05 level of significance]

Station number (fig. 3)	Drainage area (in square miles)	Does the Wilcoxon test indicate a statistically significant difference in the means?
	Piedmont	
02147500	194	no (0.055)
02154500	116	no (0.17)
02160000	183	no (0.27)
02162500	295	no (0.14)
02163500	581	yes (0.02)
02165000	236	no (0.053)
02192500	217	no (0.36)
02196000	545	no (0.10)
	Upper Coastal Plain	
02130900	108	no (0.15)
02132500	524	no (0.20)
02173000	720	yes (0.01)
02173500	683	yes (0.01)
02174000	1,720	[a]yes, no (0.04, 0.052)
	Lower Coastal Plain	
02110500	1,110	no (0.34)
02136000	1,250	no (0.23)
02176500	203	no (0.31)

[a]Analytical results exclude the 1928 peak, which was a high outlier.

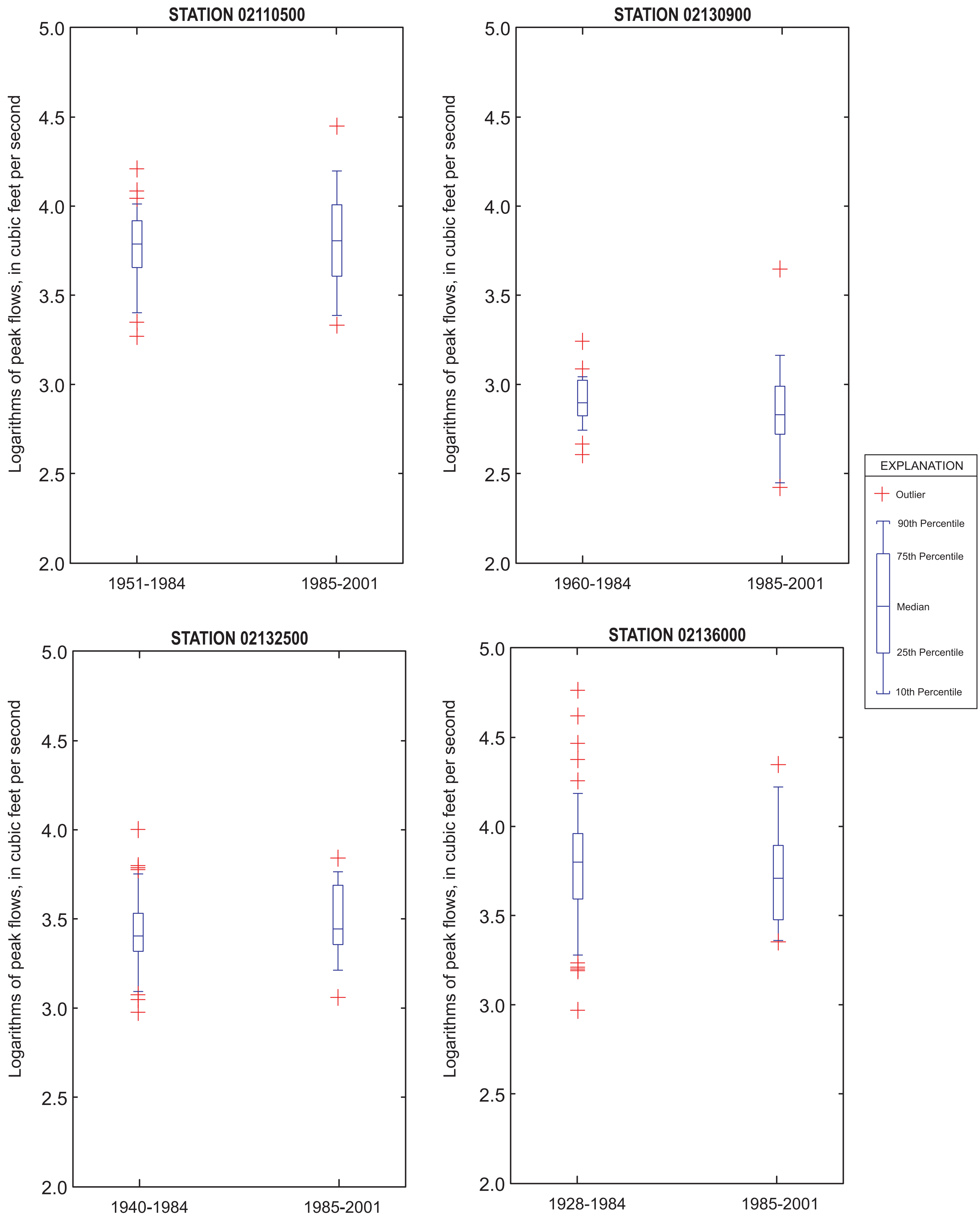


Figure 6. Box plots of the logarithms for two peak-flow periods (beginning of record to 1984 and 1985–2001) at 16 rural streamflow-gaging stations in South Carolina.

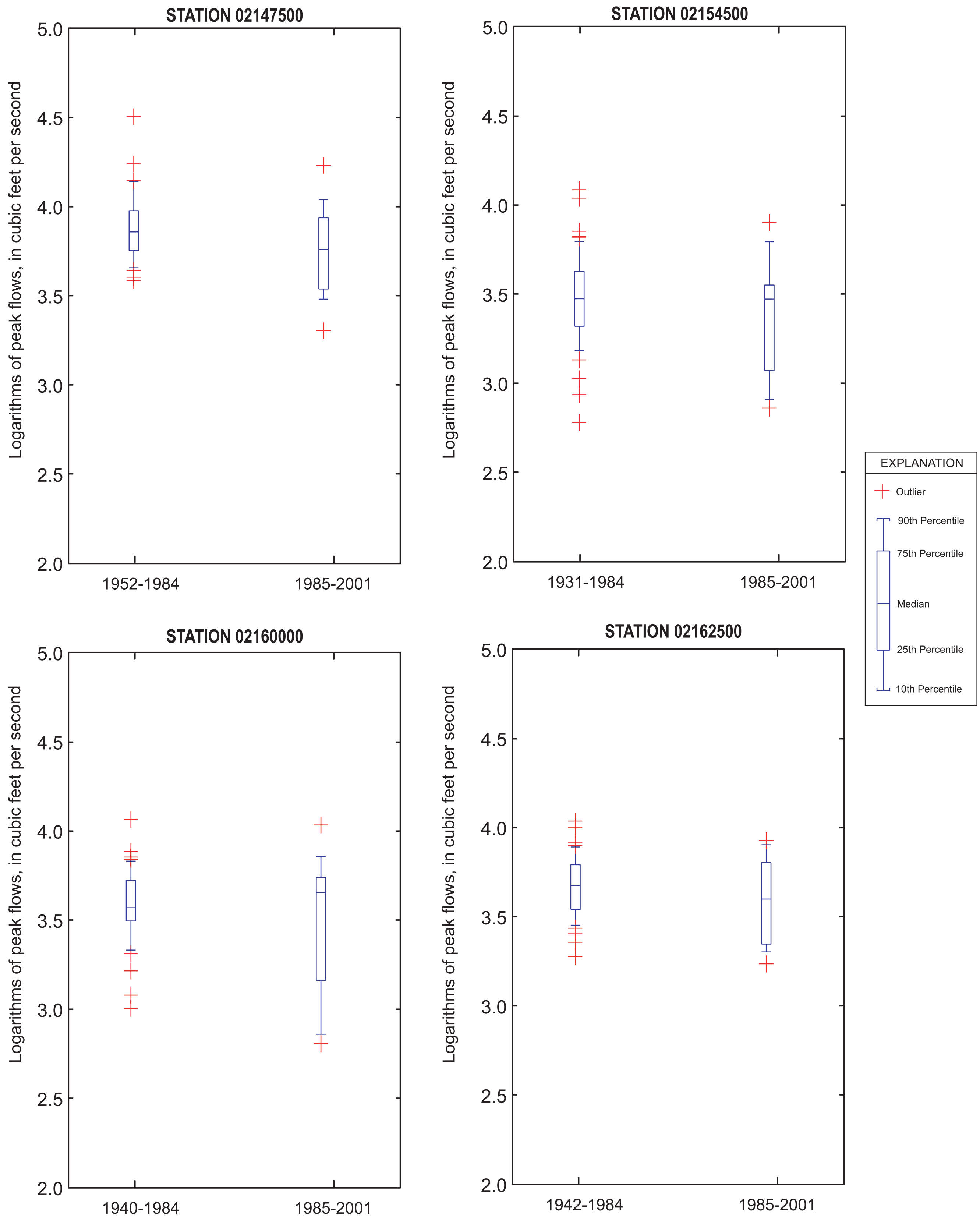


Figure 6. (Continued) Box plots of the logarithms for two peak-flow periods (beginning of record to 1984 and 1985–2001) at 16 rural streamflow-gaging stations in South Carolina.

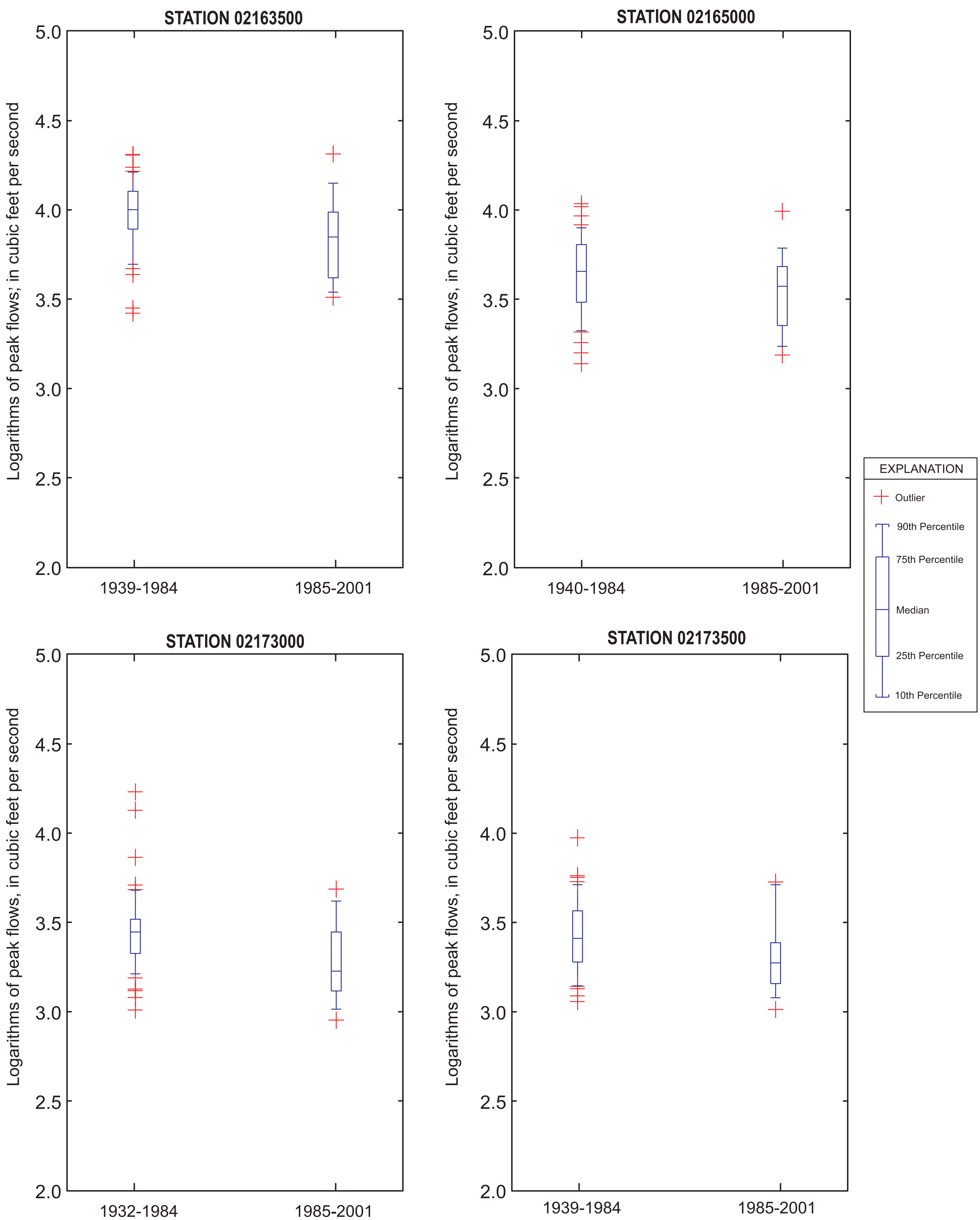


Figure 6. (Continued) Box plots of the logarithms for two peak-flow periods (beginning of record to 1984 and 1985–2001) at 16 rural streamflow-gaging stations in South Carolina.

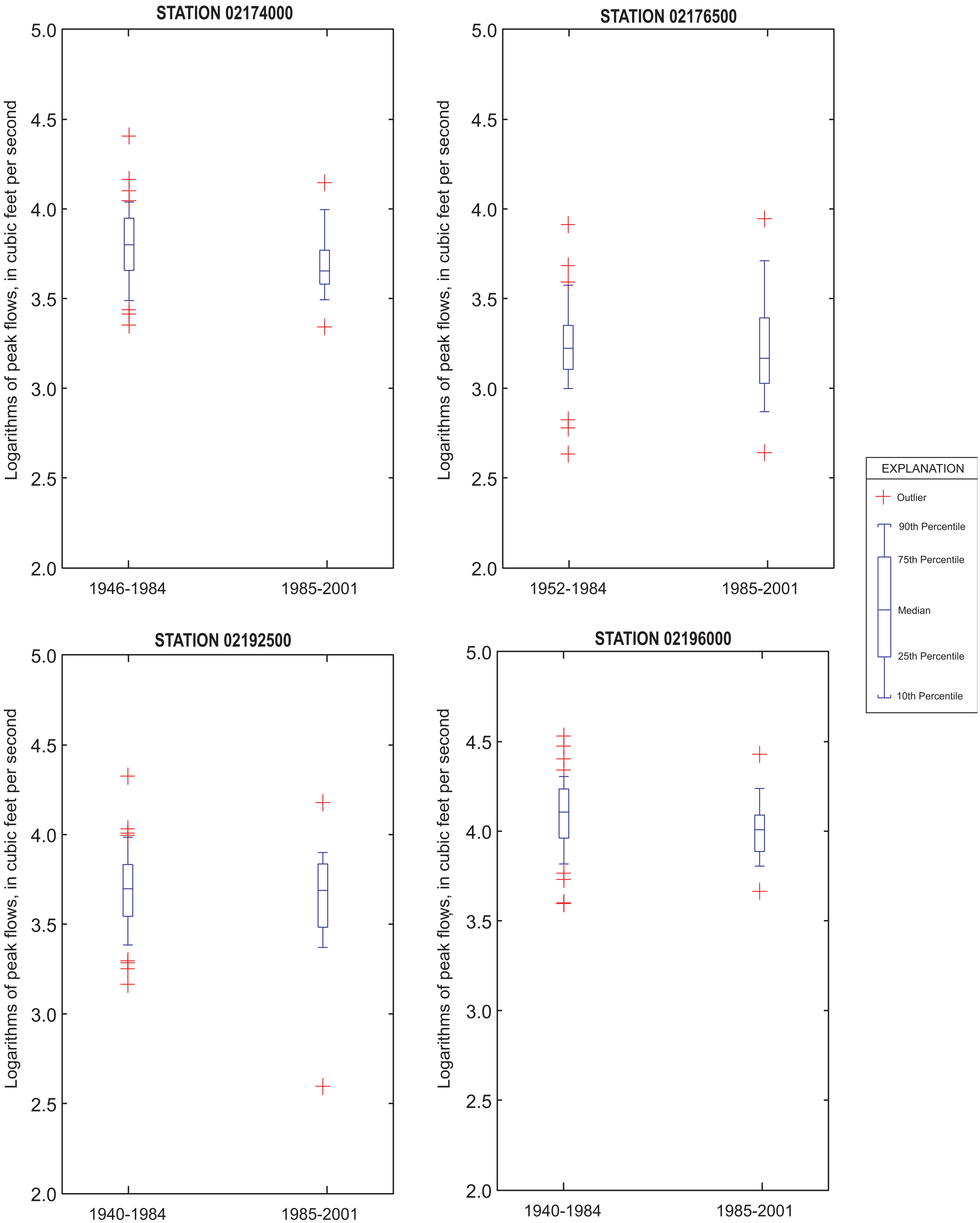


Figure 6. (Continued) Box plots of the logarithms for two peak-flow periods (beginning of record to 1984 and 1985–2001) at 16 rural streamflow-gaging stations in South Carolina.

Estimation of Flood Magnitude and Frequency at Gaged Sites

A frequency analysis of peak flows at a streamflow-gaging station provides an estimate of the flood magnitude and frequency at that specific site. The estimates typically are presented as a set of exceedance probabilities or, alternatively, recurrence intervals along with the associated flows. Exceedance probability is defined as the probability of exceeding a specified flow in a 1-year period and is expressed as a decimal fraction less than 1.0 or as a percentage less than 100. A flow with an exceedance probability of 0.01 has a 1-percent chance of being exceeded in any given year. Recurrence interval is defined as the number of years, on average, during which the specified flow is expected to be exceeded one time. A flow with a 100-year recurrence interval is one that, on average, will be exceeded once every 100 years. However, a flood with a 100-year recurrence can occur more frequently than once every 100 years and could occur more than once in a given year. Recurrence interval and exceedance probability are mathematically inverse; therefore, a flow with an exceedance probability of 0.01 has a recurrence interval of 1/0.01 or 100 years.

Flood Frequency

Flood-frequency estimates at gaged sites can be computed by fitting the peak flows to a known statistical distribution. This investigation followed the guidelines and computational methods described in Bulletin 17B of the Hydrology Subcommittee of the Interagency Advisory Committee on Water Data (1982). The flood-frequency estimates were computed by fitting the logarithms (base 10) of the peak flows to a Pearson Type-III distribution. The equation for fitting the log-Pearson Type-III distribution to a series of peak flows is as follows:

$$(\log Q_T = \bar{X} + KS), \tag{1}$$

where

Q_T is the T-year recurrence-interval flow, in cubic feet per second,

$\bar{X}$ is the mean of the log-transformed peak flow,

K is a factor dependent on recurrence interval and the skew coefficient of the log-transformed peak flow, and

S is the standard deviation of the log-transformed peak flow.

A series of peak flows at a streamflow-gaging station may include low or high outliers, which are data points that depart significantly from the trend of the remaining data. The streamflow-gaging station record also may include information about maximum peak flows that occurred outside the period of regularly collected, or systematic, record. These peak flows are known as historic peaks and are often the maximum peak flows known to have occurred during an extended period of time beyond the period of collected record. Bulletin 17B (1982) provides guidelines for detecting and interpreting low and high outliers and historic data points and provides computational methods for making appropriate corrections to the distribution to account for their presence. In some cases, low or high outliers may be excluded from the record; therefore, the number of systematic peaks may not be equal to the number of years in the period of record.

The USGS computer program PEAKFQ (W.O. Thomas and others, U.S. Geological Survey, written commun., January 1998) was used to compute the relation between flood magnitude and probability of occurrence at each of the 20 urban streamflow-gaging stations in South Carolina (13), North Carolina (4), and Georgia (3; fig. 1). PEAKFQ includes the features described in Bulletin 17B (1982) but requires the user to exercise judgment when providing data on historic peaks, specifying screening levels for outliers, and interpreting the appropriateness of the resultant frequency curve to the measured data set. In the previous urban flood-frequency investigation (Bohman, 1992) and in a similar investigation in North Carolina (Robbins and Pope, 1996), the authors used the flood-frequency estimates from Sauer and others (1983) for the four streamflow-gaging stations in Charlotte, N.C. (02146300, 02146500, 02146600, and 02146700; fig. 1). Because there were no recommended or generally accepted procedures available for estimating skew coefficients for urban areas, Sauer and others (1983) defined an average skew value for each city or metropolitan area. The assigned city skew coefficients were then weighted with skew coefficients computed from actual flood-peak records according to published guidelines (Water Resources Council, 1977). For consistency with the other streamflow-gaging stations in this investigation, the flood-frequency estimates for the four Charlotte, N.C., streamflow-gaging stations were computed using their respective station skew (table 9).

Comparison of Selected Recurrence-Interval Flows Computed From Measured Data and Simulated Data

One purpose of this investigation was to compare the selected recurrence-interval (T-year) flows computed by using the rainfall-runoff model data from the previous investigation (Bohman, 1992) with those computed from the measured data. If the comparison showed that there was no statistically significant difference in the T-year flows, the data sets could be combined resulting in a longer period of record for the frequency analysis. Based on comparisons of the measured and simulated peak flows presented earlier in this report, it was obvious that the data sets should not be combined. As shown in equation 1, the standard deviation of the peak flows is an important part of the equation for computing the T-year recurrence-interval flows. Because the variance at each streamflow-gaging station is directly related to the standard deviation, it is apparent that the simulated data will predict T-year recurrence-interval flows

Table 9. Flood-frequency statistics for measured data from 20 urban streamflow-gaging stations in South Carolina, North Carolina, and Georgia.

[L, lower Coastal Plain; U, upper Coastal Plain; P, Piedmont]

Station number (fig. 1)	Physio-graphic province	Drainage area (in square miles)	Period of record	Statistical data for water-year peak flows			Recurrence-interval flow (in cubic feet per second)		
				Mean (log)	Standard deviation (log)	Skew of logarithms	2-year	25-year	100-year
				Town of Allendale, S.C.					
02176380	L	2.06	1986–2001	1.972	0.196	0.807	88.3	231	348
				City of Augusta, Ga.					
02196760	U	1.56	1979–96	2.554	0.194	0.530	345	844	1,200
				City of Columbia, S.C.					
02162093	U	5.49	1977–2001	3.131	0.117	-0.075	1,360	2,150	2,490
02167020	U	.28	1986–2001	2.250	.105	2.069	165	305	430
02168845	P	.45	1986–96	2.194	.084	1.585	149	237	299
02169568	U	2.26	1986–2001	2.840	.239	.757	646	2,060	3,340
				City of Charlotte, N.C.					
02146300	P	30.5	1963–77	3.493	0.194	1.011	2,890	7,770	12,100
02146500	P	41.0	1962–77	3.636	.161	.158	4,320	8,490	10,800
02146600	P	38.3	1962–77	3.431	.188	-.509	2,800	5,310	7,230
02146700	P	6.98	1963–77	2.967	.153	-.431	950	1,620	1,870
				City of Florence, S.C.					
02131130	L	1.92	1985–2001	2.772	0.083	-0.857	608	777	817
				City of Greenville, S.C.					
02164011	P	3.02	1985–2001	2.983	0.107	-0.614	986	1,400	1,530
				City of Greenwood, S.C.					
02166975	P	1.16	1986–2001	2.290	0.097	-0.454	198	278	304
				City of Myrtle Beach, S.C.					
02110740	L	.80	1987–2001	2.474	0.142	0.222	294	542	674
				City of Orangeburg, S.C.					
02173491	U	.45	1987–2001	2.327	0.134	-0.064	213	362	430
				City of Rock Hill, S.C.					
02145940	P	3.50	1986–97	2.940	0.076	-0.278	879	1,160	1,260
				City of Savannah, Ga.					
02203543	L	.95	1979–96	2.445	0.178	1.409	254	666	1,060
02203544	L	.18	1979–96	1.911	.108	-.409	82.8	121	134
				City of Spartanburg, S.C.					
02159785	P	.52	1987–2001	2.155	0.177	-0.510	148	270	316
				City of Sumter, S.C.					
02135518	L	2.20	1986–2001	2.537	0.155	-1.302	370	538	588

greater than those computed from the measured data. Several comparisons were made to evaluate the magnitude of these differences.

A comparison was made of the 2-, 25-, and 100-year recurrence-interval flows computed from the simulated data with those computed from the measured data. The three streamflow-gaging stations in Georgia (02196760, 02203543, and 02203544; fig. 1) were included in these comparisons because the recurrence-interval flows computed from the simulated data were readily available (Bohman, 1992). The Georgia streamflow-gaging stations were not included in earlier comparisons of peak flows because the modeling for these streamflow-gaging stations was done by the USGS Georgia District for the Bohman (1992) investigation, and the data were not readily available. The percent difference between the measured flows and simulated flows was computed as follows:

$$\text{Percent difference} = ((\text{simulated}Q_T - \text{measured}Q_T) / \text{measured}Q_T) \times 100. \qquad (2)$$

where

simulated Q_T is the T-year recurrence-interval flow computed from the simulated peak-flow data, in cubic feet per second, and

measured Q_T is the T-year recurrence-interval flow computed from the measured peak-flow data, in cubic feet per second.

The result from the percent difference computation represents the percentage by which the measured value would have to be changed to obtain the simulated value. Consequently, a positive percent difference indicates the simulated value is greater than the measured value, and a negative percent difference indicates the simulated value is less than the measured value.

As shown in figure 7, the 2-year recurrence-interval flows compared well with the data, remaining relatively close to the line of equality throughout the range of values. The percent differences ranged from -41.5 to 45.5 percent, with a mean difference of -1.5 percent and a median difference of -0.4 percent (table 10). For the 25-year recurrence interval, the percent differences ranged from -53.9 to 158 percent, with a mean difference of 56.4 percent and a median difference of 56.3 percent. For the 100-year recurrence interval, the percent differences ranged from -62.3 to 222 percent, with a mean difference of 78.1 percent and a median difference of 72.9 percent. As the percent differences indicate, the recurrence-interval flows computed from the simulated data in most cases are greater than those computed from the measured data. The mean and median percent differences were about the same, indicating that there is no significant skew in the data and that it is fairly uniform about the mean.

According to the values shown in figure 7, as the recurrence interval increased, the data plotted farther away from the line of equality indicating that the recurrence-interval flows computed from the simulated peaks were increasingly greater than those computed from the measured peaks. Two streamflow-gaging stations in the upper Coastal Plain had greater recurrence-interval flows from the measured data than from the simulated data. Station 02169568, Pen Branch at Columbia, S.C., had differences of -41.5, -53.9, and -62.3 percent for the 2-, 25-, and 100-year recurrence-interval flows, respectively. Station 02196760, Rocky Creek tributary at Augusta, Ga., had differences of -18.8, -9.4, and -9.2 percent for the 2-, 25-, and 100-year recurrence-interval flows, respectively (fig. 1; table 10).

For the 16 rural streamflow-gaging stations with long-term records, the 2-, 25-, and 100-year recurrence-interval flows for the 1985–2001 period were compared with flows from the total period of record (table 11; fig. 8). A positive percent difference indicates that the 1985–2001 Q_T value is greater than the total period of record Q_T value, and a negative percent difference indicates that the 1985–2001 Q_T value is less than the total period of record Q_T value. The differences for the 2-year recurrence interval ranged from -23.6 to 13.0 percent, with a mean difference of -11.6 percent and a median difference of -14.3 percent. Overall, the 2-year flows for the 1985–2001 period were less than those for the entire period of record. This can probably be attributed to the severe drought that occurred in South Carolina during 1998–2001. The differences for the 25-year recurrence-interval flows ranged from -26.1 to 44.3 percent, with the mean and median differences of 0.5 and -3.9 percent, respectively. The differences for the 100-year recurrence-interval flows ranged from -31.4 to 77.9 percent, with the mean and median differences of 7.0 and 1.0 percent, respectively. The plots in figure 8 indicate that the data are well distributed about the line of equality and that there is no significant deviation between the recurrence-interval flows for the 1985–2001 period and those for the total period of record.

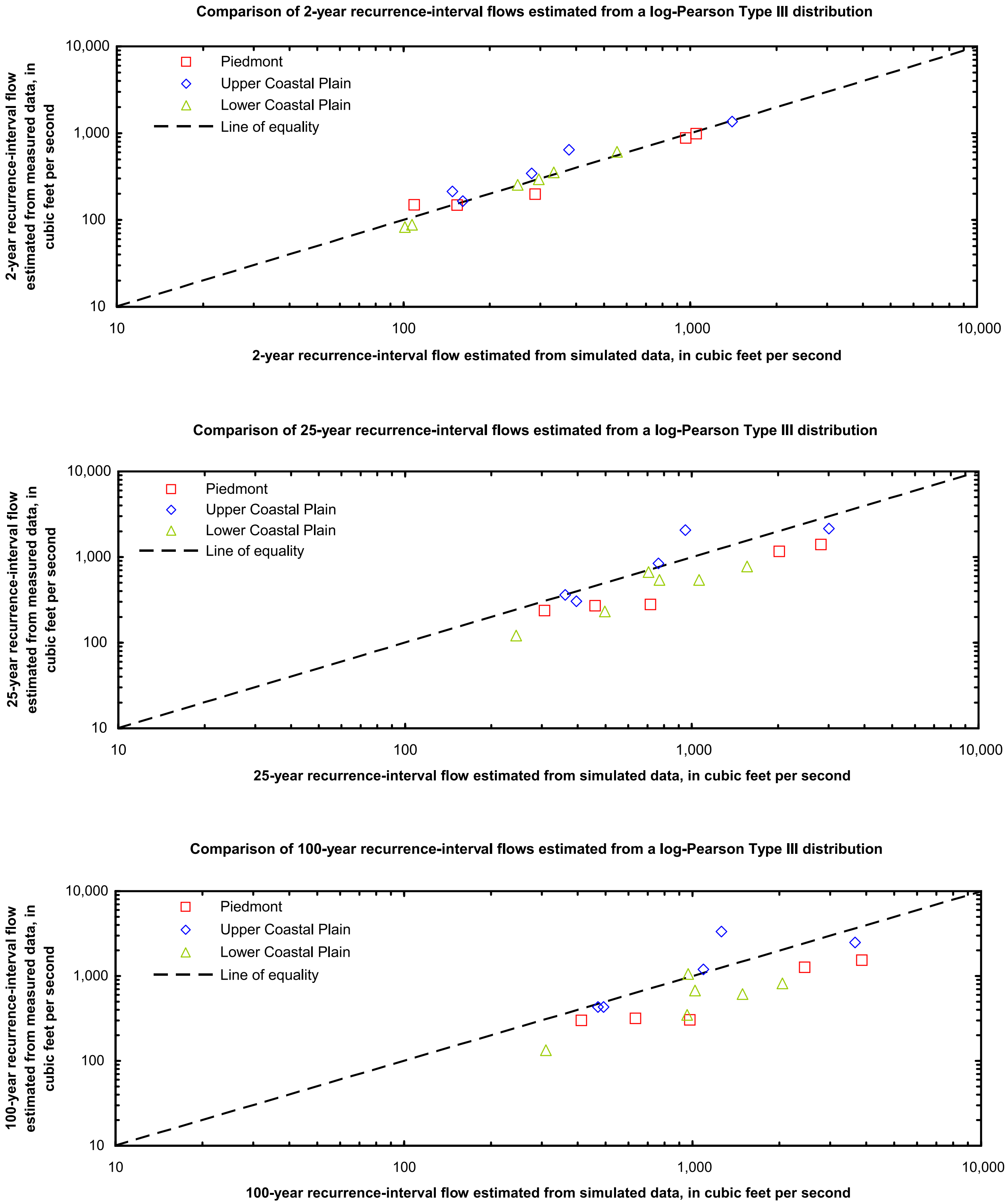


Figure 7. Comparison of the 2-, 25-, and 100-year recurrence-interval flows computed from the measured peak flows with those computed from the simulated peak flows.

Table 10. Comparison of 2-, 25-, and 100-year flows computed from simulated peak flows (Bohman, 1992) with those computed from measured peak flows.

[ft^3/s, cubic feet per second; yr, year]

Station number (fig. 1)	Flows computed from simulated peak flows (ft^3/s)			Period of record	Flows computed from measured peak flows (ft^3/s)			Percent difference		
	Recurrence interval				Recurrence interval			Recurrence interval		
	2-yr	25-yr	100-yr		2-yr	25-yr	100-yr	2-yr	25-yr	100-yr
Piedmont										
02145940	966	2,020	2,450	1986–1997	879	1,160	1,260	9.9	74.1	94.4
02159785	154	460	635	1987–2001	148	270	316	4.1	70.4	101
02164011	1,050	2,820	3,860	1985–2001	986	1,400	1,530	6.5	101	152
02166975	288	718	980	1986–2001	198	278	304	45.5	158	222
02168845	109	307	412	1986–96	149	237	299	-26.8	29.5	37.8
Upper Coastal Plain										
02162093	1,400	3,010	3,660	1977–2001	1,360	2,150	2,490	2.9	40.0	47.0
02167020	161	396	491	1986–2001	165	305	430	-2.4	29.8	14.2
02169568	378	949	1,260	1986–2001	646	2,060	3,340	-41.5	-53.9	-62.3
02173491	148	362	470	1987–2001	213	362	430	-30.5	0.0	9.3
02196760	280	765	1,090	1979–96	345	844	1,200	-18.8	-9.4	-9.2
Lower Coastal Plain										
02110740	296	771	1,020	1987–2001	294	542	674	0.7	42.3	51.3
02131130	555	1,560	2,050	1985–2001	608	777	817	-8.7	101	151
02135518	334	1,060	1,490	1987–2001	353	539	615	-5.4	96.7	142
02176380	107	497	958	1986–2001	88.3	231	348	21.2	115	175
02203543	250	706	965	1979–96	254	666	1,060	-1.6	6.0	-9.0
02203544	101	244	310	1979–96	83.0	121	134	21.7	102	131
							Mean:	-1.5	56.4	78.1
							Median:	-0.4	56.3	72.9

Table 11. Comparison of 2-, 25-, and 100-year recurrence-interval flows from two periods (period of record and 1985–2001) at 16 rural streamflow-gaging stations in South Carolina.

[ft^3/s, cubic feet per second; yr, year]

Station number (fig. 3)	Flows computed from peak flows for the total period of record (ft^3/s)			Period of record	Flows computed from peak flows for 1985–2001 (ft^3/s)			Percent difference		
	Recurrence interval				Recurrence interval			Recurrence interval		
	2-yr	25-yr	100-yr		2-yr	25-yr	100-yr	2-yr	25-yr	100-yr
Piedmont										
02147500	6,340	17,200	24,500	1952–2001	5,150	15,600	22,700	-18.8	-9.3	-7.4
02154500	2,930	8,390	11,600	1931–2001	2,430	8,390	12,500	-17.1	0.0	7.8
02160000	3,850	9,010	11,300	1940–2001	3,070	13,000	20,100	-20.3	44.3	77.9
02162500	4,490	9,680	12,300	1942–2001	3,880	10,300	14,200	-13.6	6.4	15.4
02163500	8,850	19,700	24,700	1939–2001	6,760	18,800	26,700	-23.6	-4.6	8.1
02165000	4,160	10,100	13,300	1940–2001	3,380	8,630	11,900	-18.8	-14.6	-10.5
02192500	4,800	12,300	16,700	1940–2001	4,650	11,900	16,400	-3.1	-3.2	-1.8
02196000	12,300	27,400	35,000	1940–2001	10,800	22,900	29,600	-12.2	-16.4	-15.4
Upper Coastal Plain										
02130900	754	1,970	2,870	1960–2001	706	2,600	4,290	-6.4	32.0	49.5
02132500	2,610	6,670	9,250	1940–2001	2,950	7,140	9,600	13.0	7.0	3.8
02173000	2,470	6,040	8,360	1932–2001	1,930	4,870	6,700	-21.9	-19.4	-19.9
02173500	2,460	6,360	9,070	1939–2001	1,960	4,870	6,830	-20.3	-23.4	-24.7
02174000	5,610	13,500	18,300	1946–2001	4,770	11,000	14,900	-15.0	-18.5	-18.6
Lower Coastal Plain										
02110500	6,010	16,500	23,200	1951–2001	6,370	22,500	35,100	6.0	36.4	51.3
02136000	5,540	24,100	41,100	1928–2001	5,040	17,800	28,200	-9.0	-26.1	-31.4
02176500	1,660	5,380	8,100	1951–2001	1,580	6,320	10,300	-4.8	17.5	27.2
							Mean:	-11.6	0.5	7.0
							Median:	-14.3	-3.9	1.0

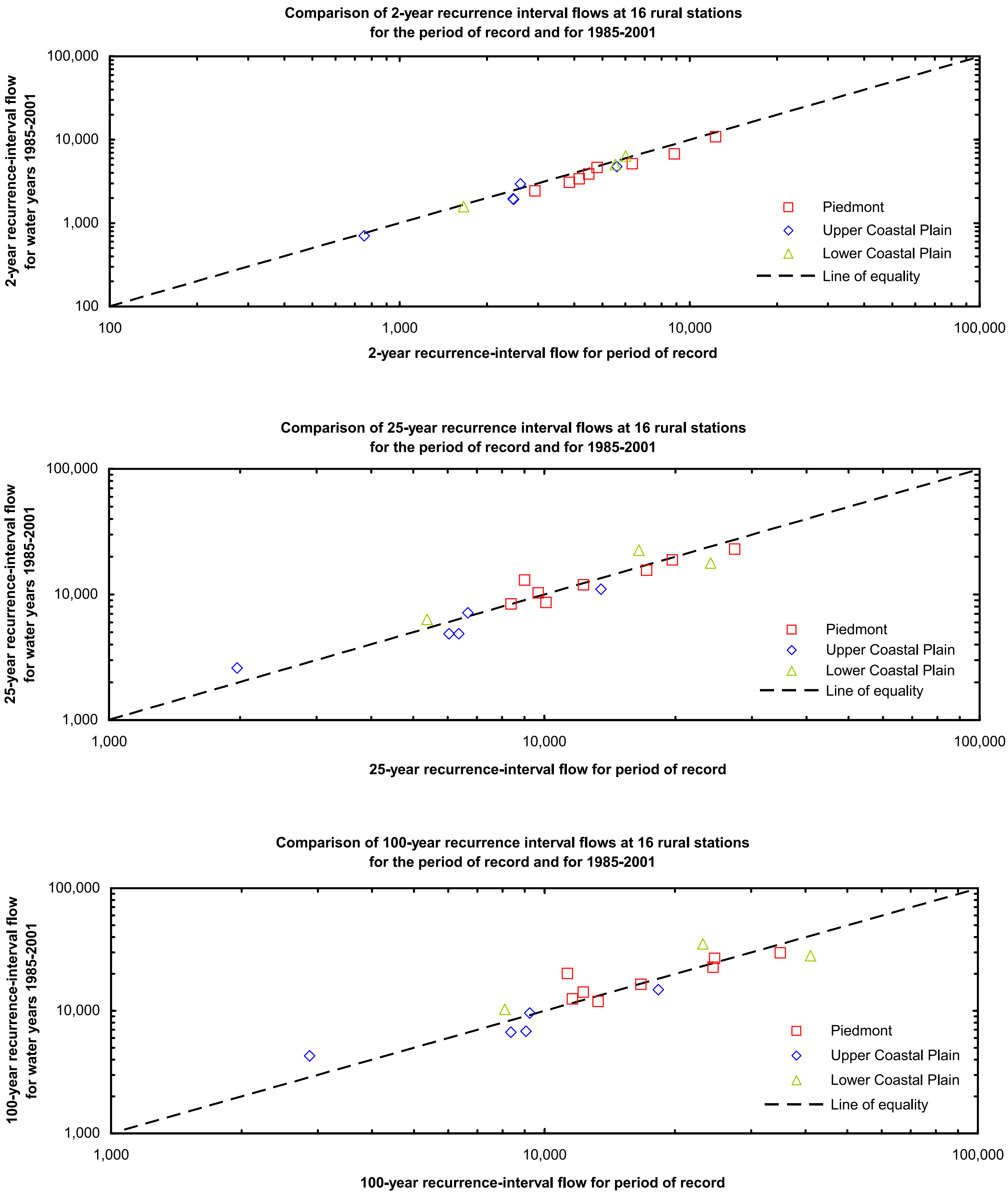


Figure 8. Comparison of 2-, 25-, and 100-year recurrence-interval flows from two peak-flow periods (period of record and 1985–2001) at 16 rural streamflow-gaging stations in South Carolina.

Conclusions From Comparisons of Measured and Simulated Data

Comparisons of the measured and simulated data at the urban streamflow-gaging stations and of data from two periods at the rural streamflow-gaging stations indicate that there is no evidence of significant climatic trends in the peak-flow data. The comparisons also indicate a significant difference in the variances of the simulated and measured urban data that was not evident in the two periods used to compare the rural data. The Wilcoxon rank sum test results for the simulated and measured urban data were similar to the results for the rural data for the two periods analyzed, which suggests that the average simulations from the rainfall-runoff model were similar to the average measured data. However, the results from comparisons of recurrence-interval flows computed from the simulated urban data with those computed from the measured urban data indicate that the mean difference for the 25- and 100-year recurrence-interval flows were 56 and 78 percent, respectively (table 10). The mean difference for the 2-year recurrence-interval was only -1.5 percent. For the rural streamflow-gaging stations, the mean differences in the 25- and 100-year recurrence-interval flows for the two periods were 0.5 and 7.0 percent, respectively (table 11).

Revisiting the calibration of and assumptions for the rainfall-runoff model used in the Bohman (1992) investigation is beyond the scope of this report. However, several possibilities have been considered regarding the significant differences between the simulated peak-flow data and the measured peak-flow data.

The data-collection period for the Bohman (1992) investigation was from 1983 to 1990, but 1986 was considered a major period of drought in South Carolina (Waters, 2003). The maximum peak flows from a data report documenting the hydrographs used to calibrate the rainfall-runoff model (Logan and others, 1995) were compared with the log-Pearson results computed for the current investigation for the 13 urban streamflow-gaging stations in South Carolina. Six of the 13 streamflow-gaging stations had maximum peak flows with recurrence intervals less than 2 years. Four of the 13 streamflow-gaging stations had maximum peak flows with recurrence intervals between 2 and 5 years, and the three remaining streamflow-gaging stations had maximum peak flows with recurrence intervals of approximately 6, 12, and 18 years. Consequently, for about half of the South Carolina streamflow-gaging stations included in the current investigation, the rainfall-runoff models from the previous investigation were calibrated to hydrographs with peak flows less than a 2-year recurrence interval, and 77 percent were calibrated with peak flows less than a 5-year recurrence interval. It is interesting to note that the maximum peak flows at station 02169568 (Pen Branch at Columbia, S.C.; fig. 1) had the longest recurrence interval of approximately 18 years. As previously noted, station 02169568 is one of only two streamflow-gaging stations where the measured data were consistently greater than the simulated data. Although a few "large" flood events did occur during the 4- or 5-year data-collection period, a total of 30 to 45 events were utilized to calibrate the model parameters for each basin. Though the larger events undoubtedly carried more weight in the subjective calibration process, the results are probably most representative of flooding that could be expected to occur many times each year.

A related factor that may account for some of the differences between the simulated and measured data is storage. The drainage areas of the 13 urban streamflow-gaging stations in South Carolina range from 0.28 to 5.49 mi^2 (table 1). It would not be unreasonable to assume that the temporary storage characteristics for these small basins would be significantly different between small and large storm events. Small storm events can be expected to move water through the basin without significant storage; however, large storm events likely cause more water to be stored behind roadways and upstream from clogged culverts and storm drains. This tends to reduce the peaks and decrease the variance. However, because the rainfall-runoff models were calibrated to small storm events, the routing parameters in the models will be biased toward little or no storage. Therefore, for large storm events (longer recurrence intervals), the models will route the flows downstream as if for a small flood with little or no storage, resulting in greater simulated flows than those actually measured.

A third possibility for the significant difference between the measured and simulated peak-flow data is the rainfall-runoff modeling assumption of spatially uniform rainfall. The rainfall-runoff models were calibrated based on storm hydrographs and rainfall collected at each streamflow-gaging station. The calibrated models then were used to synthesize long-term flow data at a streamflow-gaging station by using long-term rainfall data from a National Weather Service (NWS) rainfall station (table 2) with the assumption that rainfall patterns are similar at the two locations. Station 02159785 was one of several streamflow-gaging stations where the long-term synthesized streamflow data were generated by using long-term rainfall data from two NWS stations and then weighting the results to obtain the final synthesized record. As shown in figure 2, the peak flow at times was significantly different depending on which long-term rainfall-gaging station was used in the rainfall-runoff modeling. During an investigation in Mecklenburg County, North Carolina, from July 1995 through June 1997, precipitation data were collected at 46 sites (Robinson and others, 1998). The results of the investigation indicated that the distribution of annual rainfall in parts of Mecklenburg County (528 mi^2) ranged from 35 to 50 inches. In addition, the distribution of recurrence intervals for a 24-hour rainfall duration in the city of Charlotte for the storm of August 26–27, 1995, based on 24 rainfall-gaging stations, ranged from less than 2 years to more than 100 years (Hershfield, 1961; Hazell and Bales, 1997). Consequently, the investigation showed that rainfall patterns can be significantly different over the same time period in a single county. For the Bohman (1992) investigation, simulated annual peak flows were synthesized using the largest three to five rainfall events (in total inches) each year. Unlike the calibration phase, no attempt was made to verify the areal uniformity of the events that were selected for long-term simulation. Point rainfall extremes, as shown in the 1995 Charlotte example, can vary greatly and the assumption that these large rainfall amounts occurred evenly over the entire basin may be erroneous and could result in unusually high peaks, which would further translate to overestimation of flood magnitude and frequency values.

Inman (1997) compared the 2-, 25-, and 100-year recurrence-interval peak flows computed from measured data with urban flood-frequency estimating equations for Georgia (Inman, 1995) and found that

> "*The flood-frequency data computed from the statewide regression equations are higher than the flood-frequency data computed from observed data for the 2-year flood at 15 urban stations and are equal at one urban station; higher for the 25-year flood at 20 stations and equal at one station; and higher for the 100-year flood at 22 stations * * *. Therefore, the peak flows computed with the statewide estimating equations generally are higher than those computed using the observed data.*"

The mean residuals in the Inman (1997) investigation for the 2-, 25-, and 100-year floods estimated from the regional regression equations were 13.5, 19.9, and 22.4 percent higher, respectively, than the mean residuals of the corresponding flows computed from the same 20 years of simulated annual peak flows. Although the *t*-test indicated that the differences were statistically significant in all cases, the differences were within the range of standard error of prediction for the statewide regression equations (Inman, 1995); therefore, despite the apparent bias, the Georgia urban regression equations were not updated using observed data.

Estimation of Flood Magnitude and Frequency at Ungaged Sites

For small urban streams for which gaged data are not available, regionalization methods can be used to transfer flood-characteristic information from gaged to ungaged sites. Regionalization methods typically define relations between flood-frequency characteristics and explanatory drainage basin variables for gaged streams that have similar characteristics in a specific class or region. For this investigation, ordinary least-squares regression was used as an exploratory tool to select the preliminary regression model by using numerous combinations from a set of explanatory variables. A qualitative explanatory variable, which indicated location by physiographic province, also was tested. After the initial model was chosen, generalized least-squares regression was used to define a set of predictive equations relating peak flows for the 2-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year recurrence intervals to selected basin characteristics.

Urban Basin Characteristics

Several urban basin characteristics describing basin size and shape, indices of urban development and channel improvement, climate, and rural or background flood characteristics were included in this investigation (table 12). All of the basin characteristics used in the preliminary regional analysis were obtained from the database used by Bohman (1992) in the previous urban flood-frequency investigation except for the rural peak-flow characteristics, which were computed with regression equations developed by Feaster and Tasker (2002).

Characteristics describing basin size and shape included drainage area (DA), measured in square miles and determined from USGS topographic maps, and(or) storm-sewer maps obtained from city engineering or public works departments (Logan and others, 1995). Main-channel length (L) was measured in miles from the streamflow-gaging station upstream along the channel to the basin divide. Channel slope (S) was measured in feet per mile and computed as the difference in elevation between the 10- and 85-percent points along the stream channel divided by the length between those two points.

Urban development and channel improvement were characterized by percent impervious area (IA) and basin development factor (BDF). Impervious area is a dimensionless value determined by overlaying a grid on basin maps, delineating the impervious areas, such as roads, parking lots, and rooftops, and determining the percent of grid cells that constitute areas impervious to the infiltration of rainfall. According to Cochran (1963), a minimum of 200 points, or grid intersections, for each area or subbasin will provide a confidence level of 0.10. Grid intersections over points on buildings, streets, and parking lots were counted as impervious surface points. Grid intersections located over forests, lawns, unpaved industrial yards, and so forth, were treated as pervious surface points. An estimate of the percent of total impervious area was determined by dividing the impervious points by the total number of grid intersections. Three counts of at least 200 points for each subbasin were obtained with different orientations of the grid network, and the results were averaged for the final value.

The BDF can be determined by using the methods described in the following excerpt from Sauer and others (1983, p. 8).

> *"The most significant index of urbanization that resulted from this study is a basin development factor (BDF), which provides a measure of the efficiency of the drainage system. This parameter * * * can be easily determined from drainage maps and field inspections of the drainage basin. The basin is first divided into thirds * * *. Then, within each third, four aspects of the drainage system are evaluated and each assigned a code as follows:*
>
> 1. *Channel improvements.—If channel improvements such as straightening, enlarging, deepening, and clearing are prevalent for the main drainage channels and principal tributaries*

Table 12. Selected basin characteristics for the 20 urban streams in South Carolina, North Carolina, and Georgia

[DA, drainage area; mi^2, square miles; L, main-channel length; mi, miles; S, slope; ft/mi, feet per mile; $L/S^{0.5}$, a ratio, where L and S have been defined previously; IA, impervious area; BDF, basin development factor; $RI_{2,2}$, 2-year, 2-hour rainfall amount; in., inches; RQ, rural equivalent peak flow for 2-, 5-, 10-, 25-, 50-, and 100-year flows; ft^3/s, cubic feet per second]

Station number (fig. 1)	DA (mi^2)	L (mi)	S (ft/mi)	$L/S^{0.5}$	IA (percent)	BDF	$RI_{2,2}$ (in.)	RQ2 (ft^3/s)	RQ5 (ft^3/s)	RQ10 (ft^3/s)	RQ25 (ft^3/s)	RQ50 (ft^3/s)	RQ100 (ft^3/s)
02110740	0.80	1.40	9.2	0.462	23.0	7	2.24	[a]DA was outside of rural-regression analysis limitations.					
02131130	1.92	1.97	29.1	.365	31.0	9	2.24	79.7	148	207	299	380	471
02135518	2.20	1.79	20.2	.398	25.0	8	2.19	87.2	162	226	326	414	513
02145940	3.50	2.90	46.8	.424	19.0	6	1.95	260	448	589	784	939	1,100
02146300	30.5	11.2	13.7	3.026	20.0	9	1.90	1,490	2,450	3,210	4,320	5,270	6,310
02146500	41.0	11.0	13.1	3.039	22.0	9	1.90	1,830	2,990	3,900	5,220	6,360	7,590
02146600	38.3	8.72	12.2	2.496	10.0	7	1.90	1,740	2,860	3,730	5,000	6,090	7,270
02146700	6.98	5.20	20.9	1.137	12.0	9	1.90	528	902	1,210	1,670	2,070	2,510
02159785	.52	1.75	112	.165	14.0	0	2.00	71.2	128	172	233	282	334
02162093	5.49	3.37	30.6	.609	34.0	8	2.10	76.6	117	146	186	217	251
02164011	3.02	3.01	48.7	.431	34.0	6	2.15	235	407	536	714	856	1,000
02166975	1.16	1.63	47.2	.237	24.0	6	2.04	123	217	288	388	468	552
02167020	.28	.69	196	.049	40.0	8	2.10	[a]DA was outside of rural-regression analysis limitations.					
02168845	.45	1.07	120	.098	23.0	9	2.10	64.6	117	156	212	257	305
02169568	2.26	2.30	55.5	.309	29.0	10	2.10	[a]DA was outside of rural-regression analysis limitations.					
02173491	.45	1.17	82.0	.129	29.0	7	2.20	[a]DA was outside of rural-regression analysis limitations.					
02176380	2.06	2.02	24.3	.410	13.0	2	2.28	83.5	155	217	313	397	492
02196760	1.56	2.07	111	.196	23.0	8	2.16	[a]DA was outside of rural-regression analysis limitations.					
02203543	.95	1.78	13.0	.494	29.7	9	2.56	[a]DA was outside of rural-regression analysis limitations.					
02203544	.18	.51	29.9	.093	25.9	9	2.56	[a]DA was outside of rural-regression analysis limitations.					

[a]For more details, see "Limitations" section on p. 24 of Feaster and Tasker (2002).

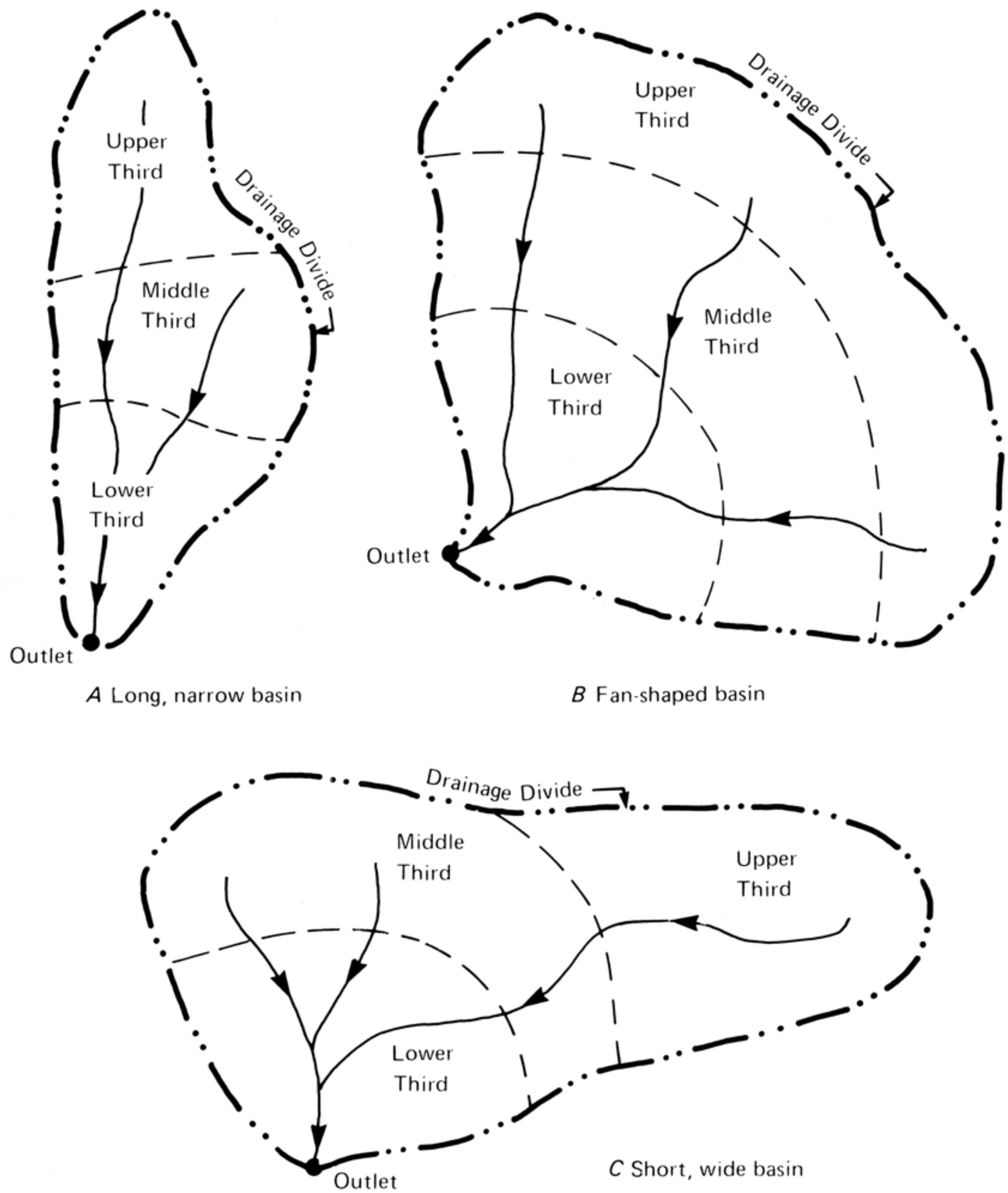


Figure 9. Schematic of typical drainage basin shapes and subdivision into basin thirds for determination of basin development factor (BDF; from Sauer and others, 1983, fig. 2, p. 7).

(those that drain directly into the main channel), then a code of 1 is assigned. Any or all of these improvements would qualify for a code of 1. To be considered prevalent, at least 50 percent of the main drainage channels and principal tributaries must be improved to some degree over natural conditions. If channel improvements are not prevalent, then a code of zero is assigned.

2. *Channel linings.—If more than 50 percent of the length of the main drainage channels and principal tributaries has been lined with an impervious material, such as concrete, then a code of 1 is assigned to this aspect. If less than 50 percent of these channels is lined, then a code of zero is assigned. The presence of channel linings would obviously indicate the presence of channel improvements as well. Therefore, this is an added factor and indicates a more highly developed drainage system.*

3. *Storm drains, or storm sewers.—Storm drains are defined as enclosed drainage structures (usually pipes), frequently used on the secondary tributaries where the drainage is received directly from streets or parking lots. Many of these drains empty into open channels; however, in some basins they empty into channels enclosed as box or pipe culverts. When more than 50 percent of the secondary tributaries within a subarea (third) consists of storm drains, then a code of 1 is assigned to this aspect; if less than 50 percent of the secondary tributaries consists of storm drains, then a code of zero is assigned. It should be noted that if 50 percent or more of the main drainage channels and principal tributaries are enclosed, then the aspects of channel improvements and channel linings would also be assigned a code of 1.*
4. *Curb-and-gutter streets.—If more than 50 percent of a subarea (third) is urbanized (covered by residential, commercial, and/or industrial development), and if more than 50 percent of the streets and highways in the subarea are constructed with curbs and gutters, then a code of 1 would be assigned to this aspect. Otherwise, it would receive a code of zero. Drainage from curb-and-gutter streets frequently empties into storm drains.*

The above guidelines for determining the various drainage-system codes are not intended to be precise measurements. A certain amount of subjectivity will necessarily be involved. Field checking should be performed to obtain the best estimate. The basin development factor (BDF) is the sum of the assigned codes; therefore, with three subareas (thirds) per basin, and four drainage aspects to which codes are assigned in each subarea, the maximum value for a fully developed drainage system would be 12. Conversely, if the drainage system were totally undeveloped, then a BDF of zero would result. Such a condition does not necessarily mean that the basin is unaffected by urbanization. In fact, a basin could be partially urbanized, have some impervious area, have some improvement of secondary tributaries, and still have an assigned BDF of zero . . . such a condition still frequently causes peak discharges to increase.

The BDF is a fairly easy index to estimate for an existing urban basin. The 50-percent guideline will usually not be difficult to evaluate because many urban areas tend to use the same design criteria, and therefore have similar drainage aspects, throughout. Also, the BDF is convenient for projecting future development. Obviously, full development and maximum urban effects on peaks would occur when BDF = 12. Projections of full development or intermediate stages of development can usually be obtained from city engineers."

One of the main assumptions in a regional flood-frequency analysis is that the data are from stable basins. The urban streamflow-gaging stations used in the previous investigation were selected partly based on their locations in older, well-established urban areas (L.R. Bohman, U.S. Geological Survey, written commun., 2000). In order to test the assumption of basin stability, seven of the urban streamflow-gaging stations were examined to determine if significant changes had occurred in land use and(or) urbanization. The seven streamflow-gaging stations are listed in the following table.

Station number	Station name
02145940	Little Dutchman Creek tributary at Rock Hill, S.C.
02146100	Manchester Creek tributary at Rock Hill, S.C.
02159785	Fairforest Creek tributary at Spartanburg, S.C.
02160325	Brushy Creek (Reedy River tributary) at Greenville, S.C.
02166975	Sample Branch at Greenwood, S.C.
02168845	Saluda River tributary at Columbia, S.C.
02174240	Middle Pen Branch at Orangeburg, S.C.

Although several of these streamflow-gaging stations were not included in the regional analysis for reasons other than significant basin changes, these were determined to be the best indicator sites or sites that would have the highest probability of significant changes with regard to land use.

The seven urban basins were examined for land-use changes by comparing aerial photographs from 1986 with aerial photographs from 1999. Newly developed areas were delineated on the 1999 photographs. If the newly developed areas were deemed to be significant, the drainage area of the new development was computed and compared to the total drainage area of the basin. The average percent impervious area was then computed based on information provided by the Soil Conservation Service (1986). Of the sites reviewed, the impervious area changes ranged from 0 to 4 percent. Consequently, it was concluded that no significant changes had occurred in the impervious area percentages computed during the previous investigation and that the same values could be used.

Climatic measurements included in the regional analysis were the cumulative 2-hour rainfall corresponding to the 2-year recurrence interval ($RI_{2,2}$) as determined from the NWS, formerly known as the U.S. Weather Bureau (1961). The rural flood characteristics were provided as the rural flood-frequency values for the 2-, 5-, 10-, 25-, 50-, and 100-year recurrence-interval flows. Measures of the rural flood characteristics were derived from regression equations developed by Feaster and

Tasker (2002) for the Blue Ridge, Piedmont, upper Coastal Plain, and lower Coastal Plain physiographic provinces of South Carolina.

Ordinary Least-Squares Analysis

Ordinary least-squares (OLS) regression techniques were used to select the explanatory variables for the preliminary models. In OLS regression, linear relations between the explanatory and response variables are necessary; consequently, variables must often be transformed. For example, the relation between drainage area and peak flow typically is not linear; however, the relation between the logarithms of drainage area and the logarithms of peak flows often is linear. Homoscedasticity (a constant variance in the response variable over the range of the explanatory variables) about the regression line and normality of residuals also are requirements for OLS regression. Transformation of the flow data and certain other variables to logarithms often enhances the homoscedasticity of the data about the regression line. Linearity, homoscedasticity, and normality of residuals were examined in residual plots.

The hydrologic model used in the regression analysis is of the form:

$$Q_T = aA^bB^cC^d\ldots, \qquad (3)$$

where

Q_T is the response variable, a flood magnitude having T-year recurrence interval, in cubic feet per second;

A, B, C are explanatory variables (basin characteristics); and

a, b, c, d are regression coefficients.

If the response and explanatory variables are logarithmically transformed, the hydrologic model has the following linear form:

$$\text{Log } Q_T = \log a + b(\log A) + c(\log B) + d(\log C) + \ldots, \qquad (4)$$

where the variables are previously defined. A combination of the arithmetic and logarithmic relation was used in this investigation because the logarithmic transformation of IA and BDF did not improve the linear relation with Q_T.

Ordinary least-squares regression of all possible subsets was used to determine the best combination of explanatory variables to use in the final regression equations. Cross products, such as $DA \times L$, also were computed using combinations of the explanatory variables and were included in the OLS regression. The use of cross products allows the regression lines to converge or diverge, thereby decreasing or increasing the effect of one variable with the effect of another variable. For example, the percentage of IA may decrease with increasing DA. None of the cross products were found to be statistically significant in the regression. The best combination of the variables was chosen on the basis of Mallow's *Cp* statistic, the adjusted coefficient of determination (R^2), and the statistical significance of the explanatory variables. The best combination of explanatory variables that were consistent for all response variables (the base-10 logarithms of the T-year peak flows; T = 2, 5, 10, 25, 50, 100, 200, and 500) was chosen for the final models and consisted of L, IA, and BDF.

Throughout the years, USGS regionalization studies for flood frequency have resulted in DA being the most significant explanatory variable. For this investigation, L was determined to be the most significant explanatory variable although not drastically different from the statistical significance of DA. Usually, DA and L are strongly related, which was the case in this investigation; therefore, it would not be appropriate to include both variables in the regression model because of multicollinearity, which occurs when the explanatory variables are correlated.

Often regional regression analysis for urban streams results in an equation that includes the equivalent rural recurrence-interval flow. In such cases, the equivalent rural recurrence-interval flow is viewed as an intercept value for the urban recurrence-interval flow as IA and(or) other characteristics related to urbanization approach zero. In this investigation, the equivalent rural recurrence-interval flows were not determined to be part of the most significant basin characteristics. For 7 of the 20 (35 percent) urban streamflow-gaging stations included in this investigation, the DA size was outside the limits used for the rural flood-frequency analysis (Feaster and Tasker, 2002). Consequently, even if the equivalent rural recurrence-interval flows had been determined to be statistically significant, rural values that were computed outside the limits of the rural-regression analysis would be considered unreliable.

Regression diagnostic tools were used to test the adequacy of the OLS regressions. Multicollinearity was measured by the variance inflation factor (VIF). There were no problems with multicollinearity. The influence of individual streamflow-gaging stations was measured by Cook's D and DFFITS (Helsel and Hirsch, 1995). Any streamflow-gaging stations that were noted as having a high influence were further reviewed for potential problems with the data. Regression residuals (for the 100-year peak flows) also were plotted and inspected for geographical patterns of bias. No distinct patterns were apparent in the mapped residuals. In addition, "qualitative variables" for the Piedmont, upper Coastal Plain, and lower Coastal Plain physiographic provinces were included in the regression analysis to detect geographical bias (Feaster and Tasker, 2002). The regression analysis also indicated no statistically significant difference among the three physiographic provinces. Therefore, the urban flood-frequency equations may be used to estimate flood magnitude and frequency in the Piedmont, upper Coastal Plain, and lower Coastal Plain physiographic provinces of South Carolina.

Generalized Least-Squares Analysis

Generalized least-squares (GLS) regression, as described by Stedinger and Tasker (1985), was used to compute the final coefficients and the measures of accuracy for the regression equations using the USGS computer program GLSNET (G.D. Tasker, K.M. Flynn, A.M. Lumb, and W.O. Thomas, Jr., U.S. Geological Survey, written commun., 1995). Stedinger and Tasker (1985) found that GLS regression equations are more accurate and provide a better estimate of the accuracy of the equations than OLS regression equations when streamflow records at streamflow-gaging stations are of different and widely varying lengths and when concurrent flows at different streamflow-gaging stations are correlated. GLS regression techniques give less weight to streamflow-gaging stations that have shorter periods of record than streamflow-gaging stations with longer periods of record. Less weight also is given to streamflow-gaging stations where concurrent peak flows are correlated with other streamflow-gaging stations (Hodgkins, 1999). Table 13 lists the peak-flow regression equations for recurrence intervals of 2, 5, 10, 25, 50, 100, 200, and 500 years that resulted from the GLS regression analysis for South Carolina.

Table 13. South Carolina urban regional regression equations.

[L, main-channel length, in miles; IA, impervious area, in percent; BDF, basin development factor, dimensionless]

Urban flood-recurrence interval (years)	Urban regional regression equations
2	$34.8\ L^{1.40}\ 10^{0.0158IA}\ 10^{0.0319BDF}$
5	$48.8\ L^{1.43}\ 10^{0.0144IA}\ 10^{0.0324BDF}$
10	$57.1\ L^{1.45}\ 10^{0.0138IA}\ 10^{0.0337BDF}$
25	$65.7\ L^{1.47}\ 10^{0.0131IA}\ 10^{0.0356BDF}$
50	$71.0\ L^{1.48}\ 10^{0.0127IA}\ 10^{0.0369BDF}$
100	$75.6\ L^{1.50}\ 10^{0.0124IA}\ 10^{0.0384BDF}$
200	$79.6\ L^{1.51}\ 10^{0.0123IA}\ 10^{0.0398BDF}$
500	$84.2\ L^{1.52}\ 10^{0.0122IA}\ 10^{0.0419BDF}$

Accuracy of the Method

The regional regression analyses for South Carolina resulted in the development of a set of equations that allow the user to estimate the 2-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year recurrence-interval flow at an ungaged, unregulated stream in an urban drainage basin (table 13). When applying these equations, users should not interpret these empirical results as exact. These regression equations are statistical models that should be interpreted and applied within the limits of the data and with the understanding that the results are best-fit estimates with an associated scatter or variance.

One measure of how well the regression equations estimate peak flows at an ungaged site is the standard error of prediction (S_p). The S_p is the square root of the mean square error of prediction (MSEp). The MSEp is the sum of two components—the mean square error resulting from the model, γ^2, and the sampling mean square error, $MSE_{s,i}$, which results from estimating the model parameters from samples of the population. The mean square model error, γ^2, is a characteristic of the model, constant for all sites, and cannot be reduced by additional data collection. The mean square sample error, $MSE_{s,i}$, for a given site, however, depends on the values of the explanatory variables; in this case, L, IA, and BDF were used to develop the flow estimate at the site. Consequently, the sampling error can be reduced by additional data collection at existing streamflow-gaging stations, or by installing new streamflow-gaging stations in the same physiographic province, or by some combination of both. The standard error of prediction for a site, i, is computed as:

$$S_{p,\ i} = (\gamma^2 + MSE_{s,\ i})^{0.5}, \tag{5}$$

(variables previously defined) and varies from site to site. Assuming the explanatory variables for the gaged sites in the regression are a representative sample of all sites in the region, the average accuracy of prediction for the regression model can be determined by computing the average standard error of prediction:

$$S_p = \left\{\gamma^2 + \frac{1}{n}\sum_{i=1}^{n} MSE_{s,\ i}\right\}^{0.5}, \tag{6}$$

where n is the number of observations and all other variables have been defined previously.

The standard error of the model ($SE_{(model)}$) can be converted from log (base 10) units to error percentage by using the transformation formulas,

$$+\ \text{Percent}SE_{(model)} = 100[10^{\gamma} - 1] \tag{7}$$

and

$$-\ \text{Percent}SE_{model} = 100[10^{-\gamma} - 1], \tag{8}$$

where the variables have been defined previously.

Similarly, the average standard error of prediction (S_p) can be transformed to positive or negative error percentage by substituting S_p^2 for γ^2 in equations 7 and 8, respectively. Computation of $S_{p,i}$ for a given ungaged site, i, involves fairly complex matrix algebra; therefore, a computer program that computes the standard error of prediction for any study site has been developed. Details of the GLS regression method and calculation of prediction errors and intervals are provided in the Appendix.

Another overall measure of how well regression equations can be used to estimate flood peaks when applied to ungaged basins is the PRESS (prediction error sum of squares) (Helsel and Hirsch, 1995) statistic. The PRESS statistic is a validation-type statistic. To compute the PRESS statistic, one streamflow-gaging station is removed from the streamflow-gaging stations used to develop the regression equation, then the value of the one removed is predicted. The difference between the predicted value from the regression equation and the measured peak flow at the streamflow-gaging station is computed. The streamflow-gaging station that is removed is then changed and the above process repeated until every streamflow-gaging station has been left out once. The prediction errors are then squared and summed (Helsel and Hirsch, 1995). PRESS/n is analogous to the average variance of prediction, and the square root of PRESS/n is analogous to the average standard error of prediction. Values of the square root of PRESS/n close to the values of the average standard error of prediction provide some measure of validation of the regression equations (Hodgkins, 1999).

A third measure of the overall accuracy of the regression equations is the average equivalent years of record. This measure represents the average number of years of peak-flow data needed to provide an estimate by using log-Pearson Type-III techniques that would be equal in accuracy to an estimate made by using regional methods (table 14). The average equivalent years of record is a function of the accuracy of the regression equations, the recurrence interval, and the average variance and skew of the peak flows at streamflow-gaging stations (Hardison, 1971).

Sensitivity Analysis

The sensitivity of the equations to errors in the explanatory variables, L, IA, and BDF, was evaluated by changing each explanatory variable by 10-percent increments from -50 to +50 percent, while holding the other variables constant. For the base computations, the mean values for each explanatory variable were substituted in the equations for the 2-, 25-, and 100-year recurrence-interval floods. The results are shown in figure 10. For the -50 to +50 percent change in L, the percent change in the peak flows ranged from -62 percent to 84 percent, respectively. For the -50 to +50 percent change in IA, the change in peak flows ranged from -30 to +42 percent, respectively. For the -50 to +50 percent change in BDF, the change in peak flows ranged from -24 to +38 percent, respectively. From these percentages and the graphs shown in figure 10, L is the most significant explanatory variable. From the OLS regression of the 100-year peak flow, R^2 was 0.94 with all three explanatory variables included. When only L was included, the R^2 was 0.83, indicating that 83 percent of the variation in the 100-year peak flow can be explained by L. For IA, the percent change in the peak flows relative to the percent change in the IA decreased when going from the 2-year flood to the 100-year flood. As the sensitivity analysis suggests, it is reasonable to assume that as the magnitude of a flood increases, the effects of the IA decrease.

Limitations

The multiple-regression equations developed in this report for estimating flood magnitude and frequency are applicable to sites on small urban streams in South Carolina that have basin characteristics within the following range: L, 0.51 to 11.2 mi; IA, 10 to 40 percent; and BDF, 0 to 10. Applying the equations to sites on streams having basin characteristics outside of the

Table 14. Accuracy statistics for the urban regional regression equations.

[PRESS, prediction error sum of squares; n, number of sites]

Urban flood-recurrence interval (years)	Mean standard error of prediction (percent)	Average mean standard error of prediction (percent)	$(PRESS/n)^{0.5}$ (percent)	Average $(PRESS/n)^{0.5}$ (percent)	Average equivalent years of record
2	-30 to 44	± 37	-32 to 47	± 39	0.8
5	-27 to 36	± 31	-29 to 40	± 34	1.7
10	-25 to 33	± 29	-27 to 37	± 32	2.8
25	-25 to 33	± 29	-27 to 37	± 32	3.9
50	-27 to 37	± 32	-29 to 41	± 35	4.3
100	-29 to 41	± 35	-32 to 47	± 39	4.2
200	-32 to 47	± 39	-35 to 55	± 45	4.0
500	-36 to 57	± 46	-40 to 68	± 54	3.6

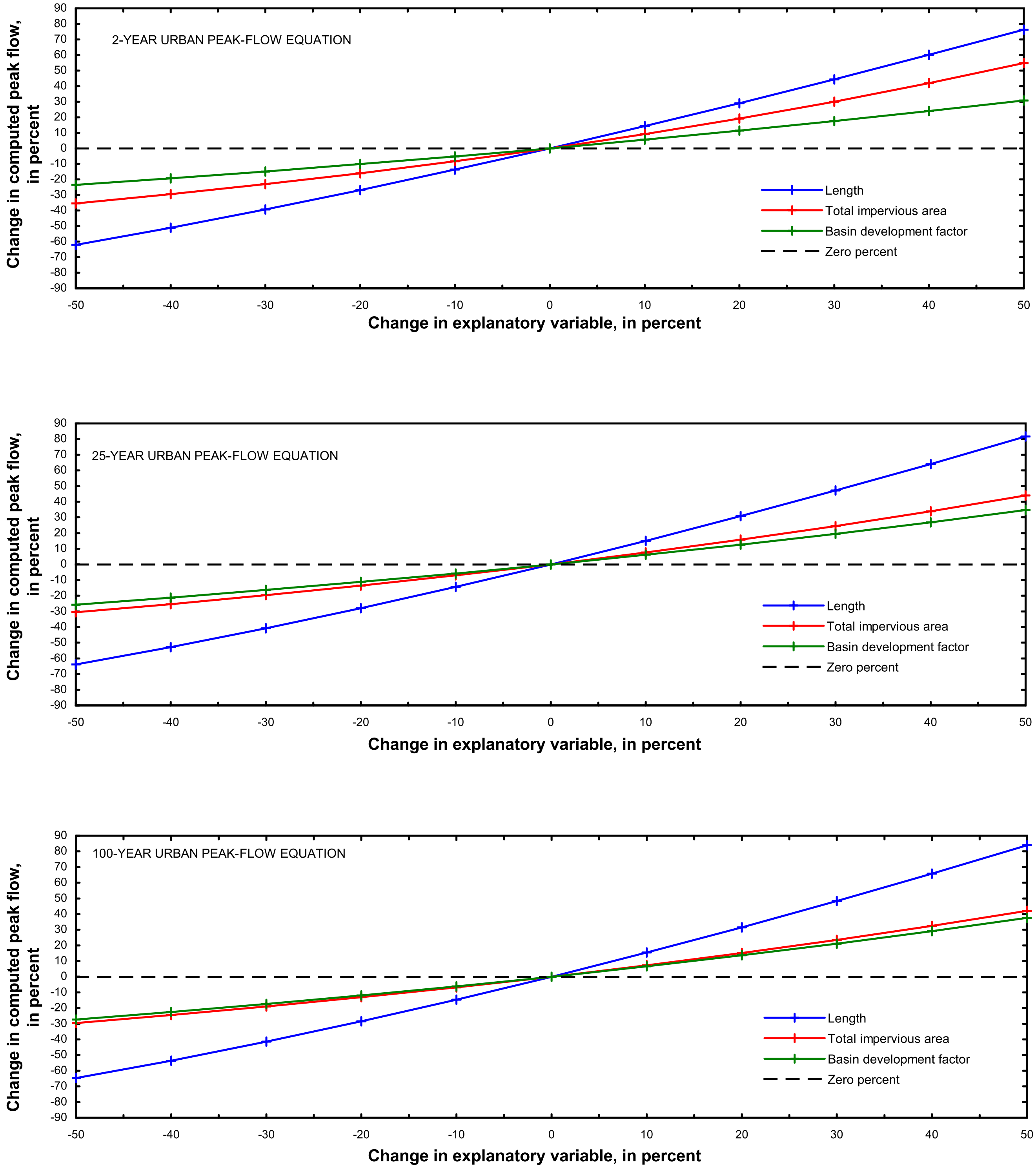


Figure 10. Sensitivity of computed peak flows to changes in the three explanatory variables for selected peak-flow-frequency equations.

range of those used in this investigation may result in prediction errors that are considerably greater than those suggested by the standard error of prediction percentages listed in table 14.

The equations may not apply to urban streams where temporary in-channel storage or detention storage significantly affects the magnitude of peak flows. Detention storage, for this report, is defined as the storage occurring in planned areas, such as ponds upstream from dams, or in unplanned detention areas, such as upstream from highway and railroad embankments. Detention storage typically reduces the peak outflow from the detention area as compared to the peak inflow. The investigation sites chosen probably reflect average storage conditions and negligible permanent storage (Bohman, 1992).

The length of record for the streamflow-gaging stations used in this investigation ranged from 11 to 25 years, with the mean and median being 16 years. Although the Wilcoxon rank sum test for 12 of 16 rural streamflow-gaging stations with long-term record showed that there was no statistically significant difference in the means of the peak-flow data for the periods from the beginning of record to 1984 and 1985–2001, it is important to understand that there is greater uncertainty in the peak-flow predictions at the higher recurrence intervals. Previous investigators (Hydrology Subcommittee of the Interagency Advisory Committee on Water Data, 1982) suggest that a streamflow-gaging station should have at least 10 years of record to warrant statistical analysis; other investigators suggest that estimates of frequencies of floods should be no greater than twice the length of record (Viessman and Lewis, 1996). However, water-resource managers often must tolerate limitations and uncertainty associated with estimating equations because long-term estimates of the magnitude and frequency of floods are needed at locations where no long-term measured data are available.

In some cases, a comparison of the recurrence-interval flows computed for a small urban stream using the equations in table 13 may result in values that are less than those obtained by using the equations for a rural stream of the same size. This most often occurs for higher recurrence-interval flows. Sometimes this condition is caused by time-sampling errors in the data and(or) errors in the statistical model; however, research also indicates that factors such as detention storage and location of urbanization can reduce peak flows (Sauer and others, 1983). For this investigation, 5 of the 20 streamflow-gaging stations had 100-year peak flows that were lower when computed with the rural equation than when computed with the urban equation. The ratio of urban to rural 100-year peak flows for these five sites ranged from 0.76 to 0.99 with an average of 0.83. When this occurs, it is left to the discretion of the user, based on hydrologic judgment and knowledge of the area, to decide which computed peak flow to use.

Program For Computing Flood-Frequency Estimates

A computer program that was developed to estimate the magnitude and frequency of floods at rural ungaged sites in South Carolina (Feaster and Tasker, 2002) was modified for this investigation to allow the user to also compute the magnitude and frequency of floods at small, urban ungaged sites in South Carolina. The program produces an on-screen summary of results and generates an output file containing the results of the frequency estimates for a site. For each site of interest, the output file contains flood-magnitude predictions, standard error of the predictions, and 90-percent prediction intervals for the 2-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year recurrence-interval flows.

The computer program and necessary data files can be downloaded, using an Internet browser to a personal computer from the USGS Web site http://sc.water.usgs.gov/SCFFREQ/. From this site, the compressed (or zipped) file named scff.zip can be downloaded to a designated directory on a personal computer. The following six files then can be extracted: (1) scff.for—the program source code; (2) scff.exe—the executable file; (3) sc.cmn—a common block file used with the source code; and (4) sc115.txt, (5) scroi.cr, and (6) scroi.rl, which are data files for the region-of-influence method. The region-of-influence method is an optional method for computing rural flood-frequency estimates in the Blue Ridge physiographic province of South Carolina (Feaster and Tasker, 2002). Once the files have been extracted, the program can be run by opening a DOS window and typing "scff."

Example Computation

Flood-frequency estimates at specific recurrence intervals can be made for small urban drainage basins located in the Piedmont and upper and lower Coastal Plain physiographic provinces of South Carolina. For the site of interest, determine the L, IA, and BDF as described earlier in this report. Verify that the basin characteristics are within the range of those described in the "Limitations" section of this report and apply the appropriate equation from table 13.

The following example can be used to estimate the 100-year recurrence-interval flow for a site having the following characteristics and located in the upper Coastal Plain of South Carolina: L = 1.79 mi^2, IA = 25.0 percent, BDF = 8. The basin characteristics are within the limitations of the regression analysis; therefore, the 100-year recurrence-interval flow (UQ_{100}) is computed from the following equation in table 13:

$$UQ_{100} = 75.6(L^{1.50})(10^{0.0124IA})(10^{0.0384BDF})$$
$$UQ_{100} = 75.6(1.79^{1.50})(10^{0.0124*25.0})(10^{0.0384*8})$$
$$UQ_{100} = 750 \text{ ft}^3/\text{s}.$$

If the computer program is used, the results will appear on the screen and in an output file as follows:

```
REGIONAL REGRESSION METHOD
Flood frequency estimates for
urban.test
Region: Urban
Main-channel Length:   1.79 miles
Impervious area:  25.00 percent
Basin Development Factor:   8.00
RI       DISCHARGE    - SE (%)    + SE (%)       90% PRED. INTERVAL
           (cfs)
     2          351.0       -28.8         40.4       194.0        636.0
     5          467.0       -25.0         33.4       282.0        773.0
    10          547.0       -23.2         30.2       345.0        868.0
    25          633.0       -23.3         30.4       398.0       1010.0
    50          689.0       -24.8         33.0       418.0       1140.0
   100          750.0       -27.1         37.1       432.0       1300.0
   200          810.0       -27.6         38.0       461.0       1420.0
   500          891.0       -63.1        171.1       156.0       5100.0
```

The abbreviations used in the output file are RI, recurrence interval; SE, standard error; and cfs, cubic feet per second.

Estimation of Flood Magnitude and Frequency at Ungaged Sites—Alternative Equations

Of the three basin characteristics that are included in the regressions equations in table 13, L was determined to be the most statistically significant variable followed by IA and BDF. When BDF was not included in the GLS analysis, the average mean standard error of prediction for the recurrence-interval flows increased from 2 to 5 percent with the average increase being 4 percent. It was concluded that the improvement in the standard error warranted including BDF in the regression equations. However, BDF is probably the most time-consuming characteristic to obtain because it requires a field visit to the basin of interest, whereas, L and IA can often be determined from maps or digital data.

The USGS Massachusetts District has developed a Geographic Information System (GIS) application called StreamStats to make the process of computing streamflow statistics at ungaged sites faster and more consistent than manual methods (Ries and others, 2000). The USGS is currently working on a new prototype of StreamStats in several other States. The new prototype has been designed so that it can be implemented in any state. Currently, no GIS coverages for basin characteristics, such as BDF, are available. Consequently, flood-frequency equations that include BDF would not be compatible with StreamStats. Considering the future possibilities of implementing StreamStats in South Carolina, a second set of regression equations was developed that only includes L and IA (table 15). The accuracy statistics for the alternative equations are shown in table 16. In figure 11, the 3-parameter (L, IA, and BDF) equations and the 2-parameter (L and IA) equations for the 2-, 25-, and 100-year recurrence-interval flows are compared with the observed peak flows at the 20 urban stations included in this investigation.

Table 15. Alternative urban regional regression equations for South Carolina.

[L, main-channel length, in miles; IA, impervious area, in percent]

Urban flood-recurrence interval (years)	Alternative urban regional regression equation
2	$41.6\ L^{1.47}\ 10^{0.0213IA}$
5	$58.8\ L^{1.50}\ 10^{0.0198IA}$
10	$69.9\ L^{1.51}\ 10^{0.0192IA}$
25	$82.3\ L^{1.53}\ 10^{0.0187IA}$
50	$90.3\ L^{1.55}\ 10^{0.0185IA}$
100	$97.2\ L^{1.56}\ 10^{0.0185IA}$
200	$103\ L^{1.58}\ 10^{0.0185IA}$
500	$111\ L^{1.60}\ 10^{0.0187IA}$

Table 16. Accuracy statistics for the alternative urban regional regression equations.

[PRESS, prediction error sum of squares; n, number of sites]

Urban flood-recurrence interval (years)	Mean standard error of prediction (percent)	Average mean standard error of prediction (percent)	$(PRESS/n)^{0.5}$ (percent)	Average $(PRESS/n)^{0.5}$ (percent)	Average equivalent years of record
2	-32 to 47	± 39	-33 to 50	± 42	0.7
5	-29 to 41	± 35	-31 to 45	± 38	1.4
10	-28 to 39	± 33	-30 to 43	± 36	2.1
25	-28 to 39	± 34	-30 to 43	± 37	3.0
50	-30 to 42	± 36	-32 to 47	± 39	3.3
100	-32 to 47	± 39	-34 to 52	± 43	3.4
200	-34 to 52	± 43	-37 to 60	± 49	3.3
500	-38 to 62	± 50	-42 to 73	± 57	3.1

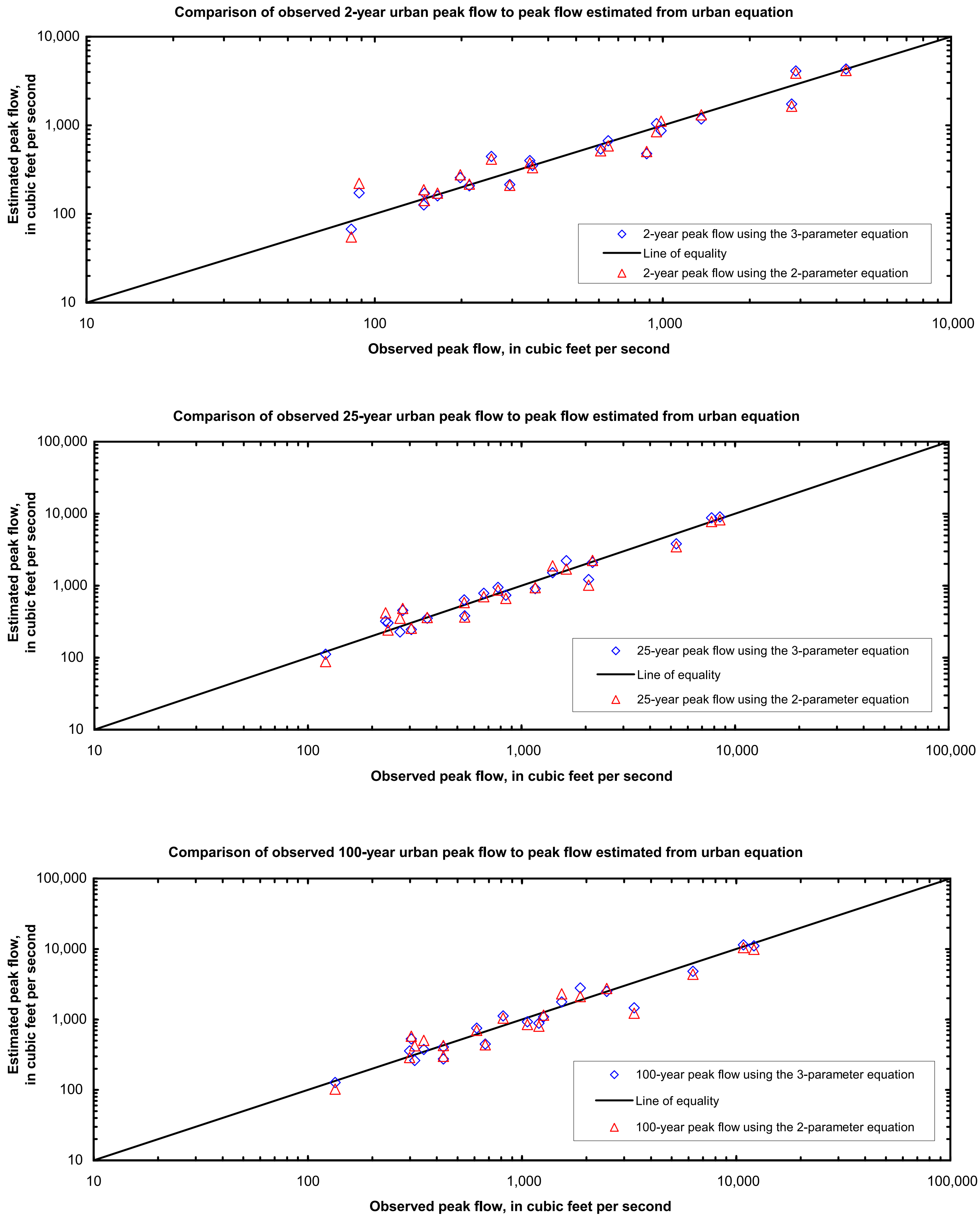


Figure 11. Comparison of observed and estimated peak flows for the 2-, 25-, and 100-year recurrence intervals.

Summary

Knowledge of flood characteristics of streams is needed for designing roadway drainage structures, establishing flood-insurance rates, and for other uses by urban planners and engineers. Because urbanization can produce significant changes in the flood-frequency characteristics of streams, flood-frequency relations for rural streams are not applicable to urban streams. Methods for estimating the magnitude and frequency of floods on small urban streams in South Carolina were developed by using data from 20 streamflow-gaging stations in South Carolina, North Carolina, and Georgia.

The magnitude and frequency of floods at the 20 urban streamflow-gaging stations were estimated by fitting the logarithm of the maximum peak flows to a Pearson Type-III distribution. Ordinary least-squares regression was used to select the explanatory variables for the preliminary statistical models to estimate the 2-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year recurrence-interval floods at ungaged urban streams. The best combination of variables was chosen based on Mallow's *Cp*, the adjusted coefficient of determination, and the statistical significance of the explanatory variables. Based on the results of the preliminary regression, generalized least-squares regression was used to compute the final coefficients and the measures of accuracy for the regional regression equations. The explanatory variables included in the final model were main-channel length, in miles, percent impervious area, and basin development factor.

Several measures of how well the regression equations estimate the specified recurrence-interval floods at an ungaged site were used; these included the standard error of prediction, the PRESS (prediction error sum of squares) statistic, and the average equivalent years of record. The mean standard error of prediction for the urban regional regression equations ranged from -25 to +33 percent for the 10- and 25-year recurrence-interval floods and -36 to +57 percent for the 500-year recurrence-interval flood. The square root of the PRESS/n, where n is the number of streamflow-gaging stations in the regression, is analogous to the average standard error of prediction and provides some measure of validation of the regression equations. For this investigation, the square root of the PRESS/n values ranged from -27 to +37 percent for the 10- and 25-year recurrence-interval floods and -40 to +68 percent for the 500-year recurrence-interval floods. The average equivalent years of record represents the average number of years of peak-flow data needed to provide an estimate using log-Pearson Type-III techniques that would be equal in accuracy to an estimate made by using regional methods. For this investigation, the average equivalent year of record ranged from 0.8 years for the 2-year recurrence-interval floods to 4.3 years for the 50-year recurrence-interval flood.

A comparison was made of measured peak flows with those that were computed using a rainfall-runoff model during a previous urban flood-frequency investigation. From a graphical comparison of these two data sets, an apparent difference in the variances was observed. To statistically compare the data sets, the Kendall tau trend test was used to test for trends over time, the *F*-test was used to test the equality of variances, and the Wilcoxon rank sum test was used to test for differences in the means. From the Kendall tau test, there were no statistically significant trends in either the measured or the simulated data. Results from the *F*-test indicated that there were statistically significant variances in the equality of variances at 85 percent of the South Carolina urban streamflow-gaging stations. The Wilcoxon rank sum test indicated that there was a statistically significant difference in the means at 38 percent of the South Carolina urban streamflow-gaging stations.

To test the hypothesis that climatic differences may have occurred between the period in which the peak-flow data were measured and the historic period in which rainfall data were used to compute the simulated peak flows, similar statistical tests were applied on long-term data collected at 16 rural streamflow-gaging stations in South Carolina. The periods of record at these rural streamflow-gaging stations ranged from 42 to 74 years with a mean length of 58.5 years and a median length of 59.5 years. For the rural sites, the period from 1985 to 2001 was compared with the data measured from the beginning of record to 1984. Plots of the two periods at each rural site indicated that there were no significant differences in the data. The statistical test for comparison of variances was applied to the 16 rural streamflow-gaging stations, and the results indicated that there were no statistically significant differences in the equality of variances at 88 percent of the sites. The statistical comparisons of the means of the two periods for the rural streamflow-gaging stations indicated that there was no statistically significant difference at 75 percent of the streamflow-gaging stations.

For the urban streamflow-gaging stations, results of the Wilcoxon test indicated no statistically significant difference between the average peak-flow simulations from the rainfall-runoff model and the average measured peak-flow data at approximately 60 percent of the urban streamflow-gaging stations. For the rural streamflow-gaging stations, the test indicated no statistically significant difference at 75 percent of the rural streamflow-gaging stations for the two periods compared. This indicates that the rainfall-runoff model reasonably simulated the average peak flows. However, the results from the comparisons of recurrence-interval flows computed from the simulated urban data with those computed from the measured urban data indicated that the mean differences for the 25- and 100-year recurrence-interval flows were 56 and 78 percent, respectively. The difference for the 2-year recurrence-interval was only -1.5 percent. This indicates that the significant differences in the variances of the simulated and modeled data at the urban streamflow-gaging stations strongly affects the estimated recurrence-interval flows, especially at higher recurrence intervals. For the rural streamflow-gaging stations, the differences in the 25- and 100-year recurrence-interval flows for the two periods were 0.5 and 7.0 percent, respectively. These findings indicate that there is no significant climatic trend that explains the differences between the measured and simulated urban peak flows. A more thorough review of the differences would require

revisiting the assumptions and calibrations of the rainfall-runoff models from the previous urban flood-frequency investigation, which is beyond the scope of this investigation.

Selected References

Becker, L.D., 1986, Techniques for estimating flood-peak discharges from urban basins in Missouri: U.S. Geological Survey Water-Resources Investigations Report 86-4322, 38 p.

Bodhaine, G.L., 1968, Measurement of peak discharge at culverts by indirect methods: U.S. Geological Survey Techniques of Water-Resources Investigations book 3, chap. A3, 60 p.

Bohman, L.R., 1992, Determination of flood hydrographs for streams in South Carolina—Volume 2. Estimation of peak-discharge frequency, runoff volumes, and flood hydrographs for urban watersheds: U.S. Geological Survey Water-Resources Investigations Report 92-4040, 79 p.

Cochran, W.G., 1963, Sampling techniques: New York, John Wiley, p. 71–86.

Feaster, T.D., and Tasker, G.D., 2002, Techniques for estimating the magnitude and frequency of floods in rural basins of South Carolina, 1999: U.S. Geological Survey Water-Resources Investigations Report 02-4140, 34 p.

Guimaraes, W.B., and Bohman, L.R., 1991, Techniques for estimating magnitude and frequency of floods in South Carolina, 1988: U.S. Geological Survey Water-Resources Investigations Report 91-157, 174 p.

Hardison, C.H., 1971, Prediction error of regression estimates of streamflow characteristics at ungaged sites: U.S. Geological Survey Professional Paper 750-C, p. 228–236.

Hazell, W.F., and Bales, J.D., 1997, Real-time rainfall measurement in the city of Charlotte and Mecklenburg County, North Carolina: U.S. Geological Survey Fact Sheet FS-052-97, 4 p.

Heath, R.C., 1977, Water resources data for North Carolina, water year 1976: U.S. Geological Survey Water-Data Report NC-76-1, 411 p.

Helsel, R.M, and Hirsch, D.R., 1995, Studies in environmental science 49—Statistical methods in water resources: Amsterdam, Elsevier Science, 529 p.

Hershfield, D.M., 1961, Rainfall frequency atlas of the United States: Washington, D.C., Department of Commerce, Weather Bureau, Technical Paper No. 40, 114 p.

Hodgkins, G.A., 1999, Estimating the magnitude of peak flows for streams in Maine for selected recurrence intervals: U.S. Geological Survey Water-Resources Investigations Report 99-4008, 45 p.

Hydrology Subcommittee of the Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood frequency: U.S. Geological Survey Bulletin 17B, Office of Water Data Collection, Reston, Va., 183 p.

Iman, R.L., and Conover, W.J., 1983, A modern approach to statistics: New York, John Wiley & Sons, 497 p.

Inman, E.J., 1995, Flood-frequency relations for urban streams in Georgia—1994 update: U.S. Geological Survey Water-Resources Investigations Report 95-4017, 27 p.

Inman, E.J., 1997, Comparison of the 2-, 25-, and 100-year recurrence interval floods computed from observed data with the 1995 urban flood-frequency estimating equations for Georgia: U.S. Geological Survey Water-Resources Investigations Report 97-4118, 14 p.

Landers, M.N., and Wilson, K.V., Jr., 1991, Flood characteristics of Mississippi streams: U.S. Geological Survey Water-Resources Investigations Report 91-4037, 82 p.

Logan, S.W., Eckenwiler, M.R., and Bohman, L.R., 1995, Selected hydrologic data for urban watersheds in South Carolina, 1983–90: U.S. Geological Survey Open-File Report 95-351, 434 p.

Ragland, B.C., Walters, D.A., Cartano, G.D., and Taylor, J.E., 2003, Water resources data, North Carolina, water year 2002: U.S. Geological Survey Water-Data Report NC-02-1B, 645 p.

Rantz, S.E., and others, 1982, Measurement and computation of streamflow—Volume 2. Computation of discharge: U.S. Geological Survey Water-Supply Paper 2175, 347 p.

Rasmussen, T.J., and Perry, C.A., 2001, Trends in peak flows of selected streams in Kansas: U.S. Geological Survey Water-Resources Investigations Report 01-4203, 62 p.

Ries, K.G., III, and Friesz, P.J., 2000, Methods for estimating low-flow statistics for Massachusetts streams: U.S. Geological Survey Water-Resources Investigations Report 00-4135, 87 p.

Ries, K.G., III, Steeves, P.A., Freeman, A., and Singh, R., 2000, Obtaining streamflow statistics for Massachusetts streams on the World Wide Web: U.S. Geological Survey Fact Sheet 104-00, 4 p.

Robbins, J.C., and Pope, B.F., 1996, Estimation of flood-frequency characteristics of small urban streams in North Carolina: U.S. Geological Survey Water-Resources Investigations Report 96-4084, 21 p.

Robinson, J.B., Hazell, W.F., and Garrett, R.G., 1998, Precipitation, streamflow, and water-quality data from selected sites in the city of Charlotte and Mecklenburg County, North Carolina, 1995–97: U.S. Geological Survey Open-File Report 98-67, 220 p.

SAS Institute, Inc., 1990, SAS procedures guide: Cary, N.C., 705 p.

Sauer, V.B., Thomas, W.O., Stricker, V.A., and Wilson, K.V., 1983, Flood characteristics of urban watersheds in the United States: U.S. Geological Survey Water-Supply Paper 2207, 63 p.

Soil Conservation Service, 1986, Urban hydrology for small watersheds (2d ed.): U.S. Department of Agriculture, Soil Conservation Service Technical Release No. 55.

Speer, P.R., and Gamble, C.R., 1964, Magnitude and frequency of floods in the United States, Part 2-A, South Atlantic Slope basins, James River to Savannah River: U.S. Geological Survey Water-Supply Paper 1673, 329 p.

Stamey, T.C., and Hess, G.W., 1993, Techniques for estimating magnitude and frequency of floods in rural basins of Georgia: U.S. Geological Survey Water-Resources Investigations Report 93-4016, 75 p.

Stedinger, J.R., and Tasker, G.D., 1985, Regional hydrologic analysis 1. Ordinary, weighted, and generalized least square compared: Water-Resources Research, v. 21, no. 9, p. 1421–1432.

Stedinger, J.R., and Tasker, G.D., 1986, Correction to regional hydrologic analysis 1. Ordinary, weighted, and generalized least square compared: Water-Resources Research, v. 22, no. 5, p. 844.

U.S. Weather Bureau, 1961, Rainfall-frequency atlas of the United States: U.S. Weather Bureau Technical Paper 40, 115 p.

Viessman, W., Jr., and Lewis, G.L., 1996, Introduction to hydrology (4th ed.): New York, Harper Collins College Publ., 760 p.

Water Resources Council, 1977, Guidelines for determining flood flow frequency: U.S. Water Resources Council Bulletin 17A, 26 p.

Waters, K.E., 2003, Ground-water levels in South Carolina—A compilation of historical water-level data: South Carolina Department of Natural Resources Water Resources Report, 18 p.

Whetstone, B.H., 1982, Techniques for estimating magnitude and frequency of floods in South Carolina: U.S. Geological Survey Water-Resources Investigations Report 82-1, 78 p.

Appendix—Generalized Least-Squares Model Description and Assumptions

A streamflow characteristic, such as the logarithm of the 50-year peak flow, is estimated at each gaged site,

$$y_i = \Psi_i + \eta_i \,, \tag{1}$$

where Ψ_i is the true (but unknown) log of the 50-year peak and η is a random error. If y_i is an unbiased estimate of ψ_i, then η (sometimes called time sampling error) has a mean of zero and a variance that is a function of how many years of data are available for the site and the standard deviation of water-year peaks. In addition, there are k basin characteristics, such as log of drainage area, that are measured with negligible error.

Assuming that (within the region defined by the basin characteristics at the n stations) ψ is approximately linearly related to the basin characteristics (x's), then the model formulation can be written as:

$$\Psi_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \ldots + \beta_k x_{ki} + \varepsilon_i \quad \text{(i=1, 2, ..., n; n>k)}, \tag{2}$$

where ε_i is a model error assumed uncorrelated from observation to observation, with mean zero and constant variance, γ^2. Substituting into equation 1,

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \eta + \varepsilon_i \,. \tag{3}$$

In matrix notation:

$$\mathrm{Y} = \mathrm{X}\beta + \upsilon \,, \tag{4}$$

where

$$\mathrm{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{bmatrix} \qquad \mathrm{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \ldots x_{k1} \\ 1 & x_{12} & x_{22} & \ldots x_{k2} \\ \ldots & \ldots & \ldots & \ldots \\ 1 & x_{1n} & x_{2n} & \ldots x_{kn} \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \ldots \\ \beta_k \end{bmatrix} \qquad \upsilon = \begin{bmatrix} \varepsilon_1 + \eta_1 \\ \varepsilon_2 + \eta_2 \\ \ldots \\ \varepsilon_n + \eta_n \end{bmatrix}, \tag{5}$$

where $E[\upsilon]=0$, and $E[\upsilon\upsilon^T]=\Lambda$. Now the GLS estimator of β is:

$$b = (\mathrm{X}^T \Lambda^{-1} \mathrm{X})^{-1} \mathrm{X}^T \Lambda^{-1} \mathrm{Y} \,. \tag{6}$$

The problem with this estimator is that Λ is unknown and must be estimated from the data. In ordinary least-squares regression, Λ is estimated as $\sigma^2 I$, which would be a good estimate if all stations in that region have approximately the same lengths of record, or if the variance of η is small relative to the variance of ε_i at every station in the region.

Because this assumption may be hard to justify, a better estimate of Λ is attempted. Denote this estimated covariance matrix $\hat{\Lambda}$, and the GLS estimator, b, will be referred to as an Estimated Generalized Least Squares (EGLS) estimator.

EGLS Regression

An example illustrates how $\hat{\Lambda}$ is estimated. Suppose that y_i is the log of the 50-year peak flow estimated from m_i years of record and that the water-year peaks follow a log-Pearson Type-III (LPIII) distribution at all sites. Further, to minimize notation, assume that the skew coefficient at all sites is zero. The elements of $\hat{\Lambda}$ would be given by:

$$\lambda_{ij} = \begin{cases} \gamma^2 + \dfrac{\sigma^2_i(1 + 0.5K^2)}{m_i} for(i = j) \\ or \\ \dfrac{\rho_{ij}\sigma_i\sigma_j m_{ij}(1 + 0.5K^2)}{m_i m_j} for(i \neq j) \end{cases} . \tag{7}$$

In this equation, K (LPIII standard deviate for zero skewness and 50-year recurrence interval), m_i (record length at station i), m_j (record length at station j), and m_{ij} (concurrent record length for stations i and j) are known, but σ_i (standard deviation of water-year peaks at station i), ρ_{ij} (cross correlation of water-year peaks at stations i and j), and γ^2 (variance of model error) must be estimated from the data. Furthermore, s_i (the sample estimate of σ_i) cannot be used as an estimate of σ_i without introducing bias, and the use of r_{ij} (sample cross correlations) for ρ_{ij} often causes numerical problems. Therefore, we estimate σ_i and ρ_{ij} as follows.

The standard deviation of water-year peaks, σ_i, is estimated from a regional regression of the form:

$$ln(s_i) = b_0 + b_1x_{1i} + b_2x_{2i} + \ldots + b_kx_{ki} . \tag{8}$$

By estimating the standard deviations, s_i, that enter into equation 7 with equation 8, we are assured that the rows of the Λ matrix are not correlated with the observed dependent variable Y. This quality is necessary for the estimates of β to be unbiased.

The cross correlation coefficient, ρ_{ij}, is estimated by developing an empirical relation between sample cross correlations, r_{ij}, and distance between stations of the form:

$$r_{ij} = \Theta^{\left[\frac{d_{ij}}{\alpha d_{ij} + 1}\right]} . \tag{9}$$

Estimating the cross correlations in this manner assures us that the matrix Λ will be positive definite. Figure 1 below shows a smooth curve with $\Theta = 0.9812$ and $\alpha = 0.00412$ based on data from Illinois. This curve was developed by running the GLSNET program that will be described later.

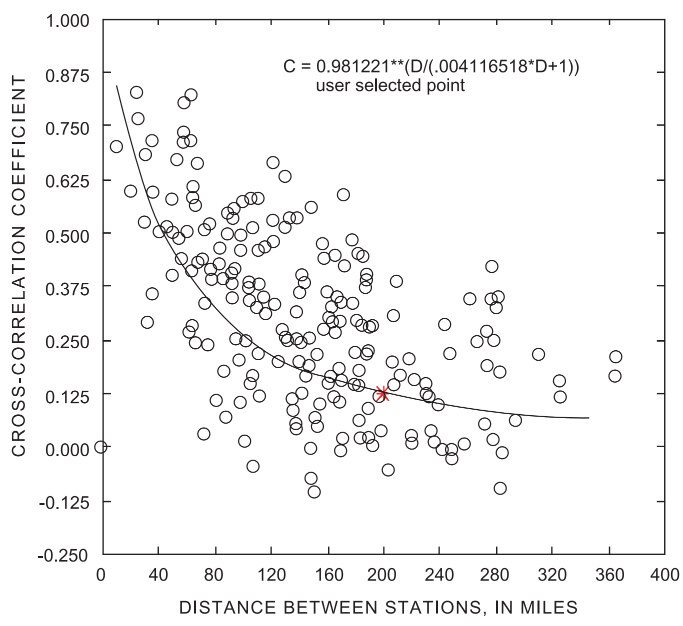


Figure 1. Relation between cross correlation and distance.

Now the only parameters left to find in the EGLS model are the regression coefficients, b, and variance of the model error, γ^2. The model error variance, γ^2, and regression coefficients, b, are found by iteratively searching for the best non-negative solution to the equation:

$$E\{(Y - X\beta)^T \Lambda^{-1}(Y - X\beta)\} = n - k - 1 . \tag{10}$$

The GLSNET/AIDE package leads one through the development of equations 8 and 9 in preparation for the estimation of the GLS regression coefficients.

Reporting Results and Errors

The predicted response at ungaged site k with basin characteristics $x_k = (1, x_{k,1}, x_{k,2}, ..., x_{k,p})$ is:

$$\hat{y}_k = x_k b. \tag{11}$$

The standard error of the prediction in OLS regression is:

$$S(\hat{y}_k) = \{\sigma^2[1 + x_k(X'X)^{-1}x'_k]\}^{0.5} . \tag{12}$$

In GLS regression, the standard error of prediction is:

$$S(\hat{y}_k) = \sqrt{\hat{\gamma}^2 + x_k X' \hat{\Lambda}^{-1} x'_k} . \tag{13}$$

The $S(\hat{y}_k)$ is a function of x and the computed standard error of a prediction in percent also will be a function of x.

Standard Errors in Percent

When a standard error or average prediction error in log units follows a normal distribution, the error may be expressed in percent of the predicted value in cubic feet per second. Denote σ as the standard error in log (base 10) units, S_{cfs} as the standard error in cubic feet per second, and $E(q|x_k)$ as the predicted value of q, in cubic feet per second, given x_k, and $x_k = (1, x_{k,1}, x_{k,2}, ..., x_{k,p})$ is a vector of basin characteristics. The standard error in percent, $S_{percent}$, is given by:

$$S_{percent} = 100\frac{S_{cfs}}{E(q|x_k)} = 100\sqrt{(e^{5.302\sigma^2} - 1)} \tag{14}$$

(Aitcheson and Brown, 1957).

Sometimes it is said in OLS that two-thirds of the points lie within one standard error of estimate of the regression function. This is true for the log unit standard error of estimate, σ, but it generally is not correct for $S_{percent.}$ This is true because the errors in log space are symmetrically distributed under the assumption of normality of the log errors, but the errors in cubic feet per second are skewed. One can, however, calculate +percent and -percent errors with the following formulas:

$$S_{plus} = 100(10^{\sigma} - 1) ; \tag{15}$$

and

$$S_{minus} = 100(10^{-\sigma} - 1) . \tag{16}$$

The three formulas above apply not only to the standard error of estimate for a regression, but they also apply to the standard error of the model, $\hat{\gamma}$, in GLS regression, the average prediction error, and standard error of a prediction in both OLS and GLS.

Average Prediction Error (APE)

One overall measure of how good the regression model is for prediction is the average prediction error (Hardison, 1971), where the average is taken over prediction sites with X variables identical to the observed data. This measure assumes the observed data have been collected at a representative set of sites in the region. It is computed as:

$$APE = \left(\sum_{i=1}^{n} \frac{\hat{\gamma}^2{}_i}{n} + \sum_{i=1}^{n} \frac{x_i (X'\hat{\Lambda}^{-1} X)^{-1} x'_i}{n} \right)^{1/2} . \qquad (17)$$

The first term in the brackets on the right side of equation 17 represents an estimate of the average squared model error for the n sites and the second term inside the brackets is an estimate of the average squared error due to estimating true model parameters from a sample of data.

Prediction Interval

Users of the regression model are probably more interested in a measure of error in a particular prediction rather than an average prediction. A good measure of the error of a particular prediction is the confidence interval of a prediction, or prediction interval. Let x_0 represent the usual row vector of basin characteristics at a prediction site. As usual x_0 is augmented by a 1 as the first element. The predicted value is $\hat{y}_0 = x_0 b$. A 100(1-α) prediction interval would be:

$$\hat{y}_0 - T \le y_0 \le \hat{y}_0 + T , \qquad (18)$$

where

$$T = t_{\frac{\alpha}{2},\, n-p'} \sqrt{(\hat{\gamma}^2{}_0 + x_0 (X'\hat{\Lambda}^{-1} X)^{-1} x'_0)} , \qquad (19)$$

where $t_{\alpha/2,\, n-p'}$ is the critical value from a t-distribution for n–p' degrees of freedom.

If a log transform had been made so that $y_0 = \log_{10}(q_0)$, then the prediction interval would be:

$$10^{\hat{y}_0 - T} \le q_0 \le 10^{\hat{y}_0 + T} . \qquad (20)$$

Selected References to the Appendix

Acreman, M.C., and Wiltshire, S.E., 1987, Identification of regions for regional flood frequency analysis [abs.]: EOS, v. 68, no. 44.

Aitcheson, J., and Brown, J.A.C., 1957, The log-normal distribution: London, Cambridge University Press.

Hardison, C.H., 1971, Prediction error of regression estimates of streamflow characteristics at ungaged sites: U.S. Geological Survey Professional Paper 750-C, p. 228–236.

Estimating the Magnitude and Frequency of Floods in Small Urban Streams in South Carolina

Scientific Investigations Report 2001-5030

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