Factors Related to the Joint Probability of Flooding on Paired Streams

By G.F. KOLTUN and J.M. SHERWOOD

U.S. GEOLOGICAL SURVEY

Water-Resources Investigations Report 98–4238

Prepared in cooperation with the

Ohio Departent of Transportation

and the

U.S. Department of Transportation,
Federal Highway Administration

Columbus, Ohio
July, 1998

U.S. DEPARTMENT OF THE INTERIOR
BRUCE BABBITT, Secretary

U.S. GEOLOGICAL SURVEY
Charles G. Groat, Director

For addtional information write to:

District Chief
U.S. Geological Survey
975 West Third Avenue
Columbus, Ohio 43212-3192

Copies of this report can be purchased from:

U.S. Geological Survey
Branch of Information Services
Box 25286, Building 810
Denver Federal Center
Denver, CO 80225-0286

1. Report No.	2.Government Accession No.	3. Recipient's Catalog No.
FHWA/OH-98/009		

4. Title and Subtitle

FACTORS RELATED TO THE JOINT PROBABILITY OF FLOODING ON PAIRED STREAMS

5. Report Date

July, 1998

6. Performing Organization Code

7. Author(s)

G.F. Koltun, J.M. Sherwood

8. Performing Organization Report No.

9. Performing Organization Name and Address

U.S. Geological Survey
Water Resources Division
975 West Third Avenue
Columbus, OH 43212

10. Work Unit No. (TRAIS)

11. Contract or Grant No.

State Job No. 14579(0)

12. Sponsoring Agency Name and Address

Ohio Department of Transportation
25 S. Front St.
Columbus, OH 43215

13. Type of Report and Period Covered

Final Report

14. Sponsoring Agency Code

15. Supplementary Notes

Prepared in cooperation with the U.S. Department of Transportation, Federal Highway Administration

16. Abstract

Factors related to the joint probability of flooding on paired streams were investigated. Stream pairs were considered to have flooded jointly at the design-year flood threshold (corresponding to the 2-, 10-, 25-, or 50-year instantaneous peak streamflow) if peak streamflows at both streams in the pair were observed or predicted to have equalled or exceeded the threshold on a given calendar day. Daily mean streamflow data were used as a surrogate for instantaneous peak streamflow data to determine which flood thresholds were equalled or exceeded on any given day. Instantaneous peak streamflow data, when available, were used preferentially to assess flood-threshold exceedance.

Observed joint probabilities of flooding were computed as the ratios of the number of days when streamflows at both streams concurrently equalled or exceeded their flood thresholds (events) to the number of days where streamflows at either stream equalled or exceeded its flood threshold (trials). Logistic regression equations for estimating the joint probability of flooding at the 2-year flood threshold were developed based on event-trial ratio and basin characteristic data.

Distance between drainage area centroids, the ratio of the smaller to larger drainage area, mean drainage area, and the centroid angle adjusted 30 degrees were the characteristics most closely associated with the joint probability of flooding on paired streams in Ohio. In general, the joint probability of flooding decreased with an increase in centroid distance and increased with increases in drainage area ratio, mean drainage area, and centroid angle adjusted 30 degrees.

Contingency tables were constructed and analyzed to provide information about the bivariate distribution of floods on paired streams.

17. Key Words	18. Distribution Statement
Floods, Probability, Regression models	No Restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified		

CONTENTS

FIGURES

TABLES

CONVERSION FACTORS

Multiply	By	To obtain
Length		
meter (m)	3.281	foot
decameter	32.81	foot
kilometer (km)	0.6214	mile
Area		
square kilometer (km^2)	247.1	acre
square kilometer (km^2)	0.3861	square mile
Volume		
cubic meter (m^3)	35.31	cubic foot
Flow rate		
cubic meter per second (m^3/s)	35.31	cubic foot per second
Slope		
meters per kilometer (m/km)	5.27983	foot per mile
Storage		
cubic meters per square kilometer (m^3/km^2)	0.0021	acre-foot per square mile

Factors Related to the Joint Probability Of Flooding on Paired Streams

By G.F. Koltun *and* J.M. Sherwood

ABSTRACT

The factors related to the joint probability of flooding on paired streams were investigated and quantified to provide information to aid in the design of hydraulic structures where the joint probability of flooding is an element of the design criteria. Stream pairs were considered to have flooded jointly at the design-year flood threshold (corresponding to the 2-, 10-, 25-, or 50-year instantaneous peak streamflow) if peak streamflows at both streams in the pair were observed or predicted to have equaled or exceeded the threshold on a given calendar day. Daily mean streamflow data were used as a substitute for instantaneous peak streamflow data to determine which flood thresholds were equaled or exceeded on any given day. Instantaneous peak streamflow data, when available, were used preferentially to assess flood-threshold exceedance.

Daily mean streamflow data for each stream were paired with concurrent daily mean streamflow data at the other streams. Observed probabilities of joint flooding, determined for the 2-, 10-, 25-, and 50-year flood thresholds, were computed as the ratios of the total number of days when streamflows at both streams concurrently equaled or exceeded their flood thresholds (events) to the total number of days where streamflows at either stream equaled or exceeded its flood threshold (trials).

A combination of correlation analyses, graphical analyses, and logistic-regression analyses were used to identify and quantify factors associated with the observed probabilities of joint flooding (event-trial ratios). The analyses indicated that the distance between drainage area centroids, the ratio of the smaller to larger drainage area, the mean drainage area, and the centroid angle adjusted 30 degrees were the basin characteristics most closely associated with the joint probability of flooding on paired streams in Ohio. In general, the analyses indicated that the joint probability of flooding decreases with an increase in centroid distance and increases with increases in drainage area ratio, mean drainage area, and centroid angle adjusted 30 degrees.

Logistic-regression equations were developed, which can be used to estimate the probability that streamflows at two streams jointly equal or exceed the 2-year flood threshold given that the streamflow at one of the two streams equals or exceeds the 2-year flood threshold. The logistic-regression equations are applicable to stream pairs in Ohio (and border areas of adjacent states) that are unregulated, free of significant urban influences, and have characteristics similar to those of the 304 gaged stream pairs used in the logistic-regression analyses.

Contingency tables were constructed and analyzed to provide information about the bivariate distribution of floods on paired streams. The contingency tables showed that the percentage of trials in which both streams in the pair concurrently flood at identical recurrence-interval ranges generally increased as centroid distances decreased and was greatest for stream pairs with adjusted centroid angles greater than or equal to

60 degrees and drainage area ratios greater than or equal to 0.01. Also, as centroid distance increased, streamflow at one stream in the pair was more likely to be in a less than 2-year recurrence-interval range when streamflow at the second stream was in a 2-year or greater recurrence-interval range.

INTRODUCTION

The flood elevation at a hydraulic structure (such as a culvert or bridge) on a tributary stream may be affected by backwater caused by flooding on a receiving stream. To determine an appropriate hydraulic design at a given location, one must consider the probability of floods of similar magnitude occurring concurrently on the two streams. Because site-specific data are rarely available for directly determining the **joint probability of flooding**[1], designers commonly assume that both the tributary and receiving streams flood concurrently at similar design-year magnitudes. This approach may result in an overly conservative design that can inflate the cost of the hydraulic structure.

A similar concern confronts those who may need to determine the probability of overtopping floods occurring concurrently at two or more hydraulic structures along a critical length of roadway or railway (to determine the probability of the link being closed). Greatly different probabilities may be estimated depending on the number of stream crossings and assumptions about whether the streams at the hydraulic structures flood independently of one another. Again, because of limited availability of **streamflow** information, direct determinations are frequently difficult or impossible to make.

Both concerns described above require assumptions about the statistical dependence of floods at two or more locations. A review of the literature indicated that design problems such as the two previously mentioned occasionally have been addressed by costly site-specific studies (Dyhouse, 1985; Fricke and others, 1983; Morris and Wilson, 1987; Raynal and Salas, 1987; Ribeny, 1971). No studies were found however that addressed the need for empirically based guidelines for evaluating the statistical dependence of floods at two or more locations. For lack of such guidelines, designers are frequently forced to depend on their own judgement and intuition to assess the statistical dependence of floods, and they commonly assume that the occurrence of floods at two or more locations is either completely dependent or completely independent. In reality, stream response will generally be somewhere between these two extremes.

The Ohio Department of Transportation (ODOT) has established general guidelines for evaluating the joint probability of flooding on **confluent streams** (Ohio Department of Transportation, 1992), but more information is needed to support or improve those guidelines. Consequently, the U.S. Geological Survey, in cooperation with the Ohio Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration, conducted a study of factors related to the joint probability of flooding on paired streams. The objectives of the study were to (1) provide information to aid decisions on when **paired streams** can, on average, be considered to react independently during floods (2) identify factors most closely associated with concurrency of flooding on paired streams and quantify those relations, and (3) determine the bivariate frequency distribution of **discretized streamflow data** for paired streams with various **drainage areas** and proximity and, if possible, draw generalizations from those data.

This report summarizes the methods of data retrieval and analysis and the significant findings of the study. The analyses were based upon **instantaneous peak** and **daily mean streamflow** data and basin-characteristic data for 30 streamflow-gaging stations in Ohio and in border areas of adjacent states.

Study approach

The most common U.S. Geological Survey (USGS) methods of streamflow-data processing consist of the computation of an instantaneous streamflow time series with a fixed time interval generally ranging from 5 minutes to 1 hour. The instantaneous streamflow time series is then used to compute a daily mean streamflow time series. The daily mean streamflow data as well as instantaneous **annual peak streamflow** and **partial peak streamflow** data are stored permanently in readily accessible electronic data bases. After daily mean streamflow values have been computed and checked, the fixed-interval instantaneous streamflow time-series data (or the raw data required to compute them) are usually archived off-line because of the large size of the files containing those data. Thus, the fixed-interval instantaneous streamflow time-series data are relatively inaccessible for long periods of record, whereas daily mean streamflow time series and instantaneous peak streamflow (annual peak and partial peak) data are readily accessible for long periods of record.

Because of site-to-site differences in basin lagtime and hydrograph shape, the fixed-interval instantaneous streamflow data are better suited to addressing the issue of joint flooding than are daily mean streamflow data. However, because the fixed-interval instantaneous streamflow data are relatively inaccessible, the approach used to explore the joint probability of flooding on paired streams was based

[1]Terms defined in the glossary are in **bold print** at their first significant use in the main body of the report.

predominantly on daily mean and instantaneous peak streamflow data.

The observed probabilities of joint flooding on paired streams were based on analyses involving the application of **flood thresholds** (discussed in detail later) to recorded streamflow data. A combination of correlation analyses, graphical analyses, and logistic-regression analyses were used to identify and quantify factors most closely associated with the observed probabilities of joint flooding.

Contingency tables were constructed from streamflow data discretized on the basis of estimated ranges in recurrence interval to provide information on the bivariate frequency distribution of the discretized streamflows.

Selection and retrieval of streamflow data

The 30 streamflow-gaging stations selected for use in this study were required to have at least 10 years of continuous record of unregulated flow and were selected so as to represent a variety of drainage-area sizes and geographic locations within Ohio and border areas of adjacent states. An additional criterion for selection of the gaging stations was that there be at least one 10-year period of overlapping unregulated streamflow record with each of the other 29 stations. A special effort was made to select pairs of stations that were on confluent streams; however, that was not a requirement for selection.

The initial data base for the station-selection process consisted of about 300 streamflow-gaging stations in Ohio and numerous stations in neighboring states. From these more than 300 stations, 74 stations were identified that had at least 25 years of continuous *and* concurrent record of unregulated streamflow. Of these 74 stations, 30 were selected for further analyses (table 1).

The 30 stations have from 27 to 80 years of continuous record of unregulated streamflow. Twenty-three of the stations are in Ohio, one is in Pennsylvania, three are in West Virginia, one is in Kentucky, one is in Indiana, and one is in Michigan. Drainage areas for the 30 stations range from 3.50 to 16,395 km^2. Thirteen stations (identified by footnotes in table 1) are on streams that are confluent with streams on which a second station is located. Seventeen stations are on streams in basins sharing a common drainage divide, some of which may also be confluent. Locations of the 30 streamflow-gaging stations and the drainage-area boundary for each streamflow-gaging station are shown in figure 1.

After the 30 streamflow-gaging stations were selected, basin-characteristics data and all daily mean and instantaneous peak streamflow data through **water year** 1993 were retrieved from the WATSTORE (Hutchinson, 1975) and ADAPS (Dempster, 1990) computer data bases. Latitudes, longitudes, drainage areas, and periods of record for the 30 streamflow-gaging stations are listed in table 1. Individual and concurrent lengths of record of the daily mean streamflow data for the 30 streamflow-gaging stations are illustrated in figure 2. The minimum, median, and maximum record lengths of the 30 streamflow-gaging stations are 27, 54, and 80 years, respectively. The minimum, median, and maximum *concurrent* (not necessarily continuous) record lengths for all pairs of streams are 26, 42, and 73 years, respectively.

Three rigorous quality-assurance checks were performed to ensure validity and applicability of data for the subsequent analyses. First, a regression analysis was performed for each station to relate daily mean streamflows to instantaneous peak streamflows occurring on the same days. Data for all days having relatively large **regression residuals** or high influence (as determined from the Cook's D statistic) were checked by verifying the data against original records. Second, ratios of instantaneous peaks to daily mean streamflows were computed for each station for all days where both types of streamflow data were available. Data for all days having relatively high or low ratios also were checked against the original records. Third, data for stations having revised records were checked to ensure that both the daily mean and instantaneous peak streamflow data had been revised. In general, most data were found to be accurate and consistent; however, some corrections were required.

FACTORS RELATED TO JOINT PROBABILITY OF FLOODING

The following sections of the report describe the analytical techniques used and the significant findings of the study. The methodology is based on two assumptions: (1) the observed **event-trial ratio** (discussed in detail later) is a valid measure of the joint probability of flooding on paired streams and (2) the joint probability of flooding can be related to selected basin characteristics.

Determination of flood thresholds

As previously mentioned, the daily mean streamflow time series was used as a substitute for the fixed-interval instantaneous streamflow time series. Consequently, data for each stream were analyzed to develop a relation between daily mean streamflow and instantaneous peak streamflow for the day. This was done by pairing daily mean streamflow data (obtained from the USGS daily-values data base) with annual-peak and partial-peak instantaneous streamflow data (obtained from the USGS peak-flow data base) occurring on the same day. Ordinary least-square regression methods were used to develop equations relating daily mean streamflows to instantaneous peak streamflows for each stream. The regression equations were used to estimate the corresponding daily mean streamflows that would occur on days when the instantaneous peak streamflow equaled selected flood thresholds. Flood thresholds used in this study were the instantaneous peak streamflows with recurrence intervals of

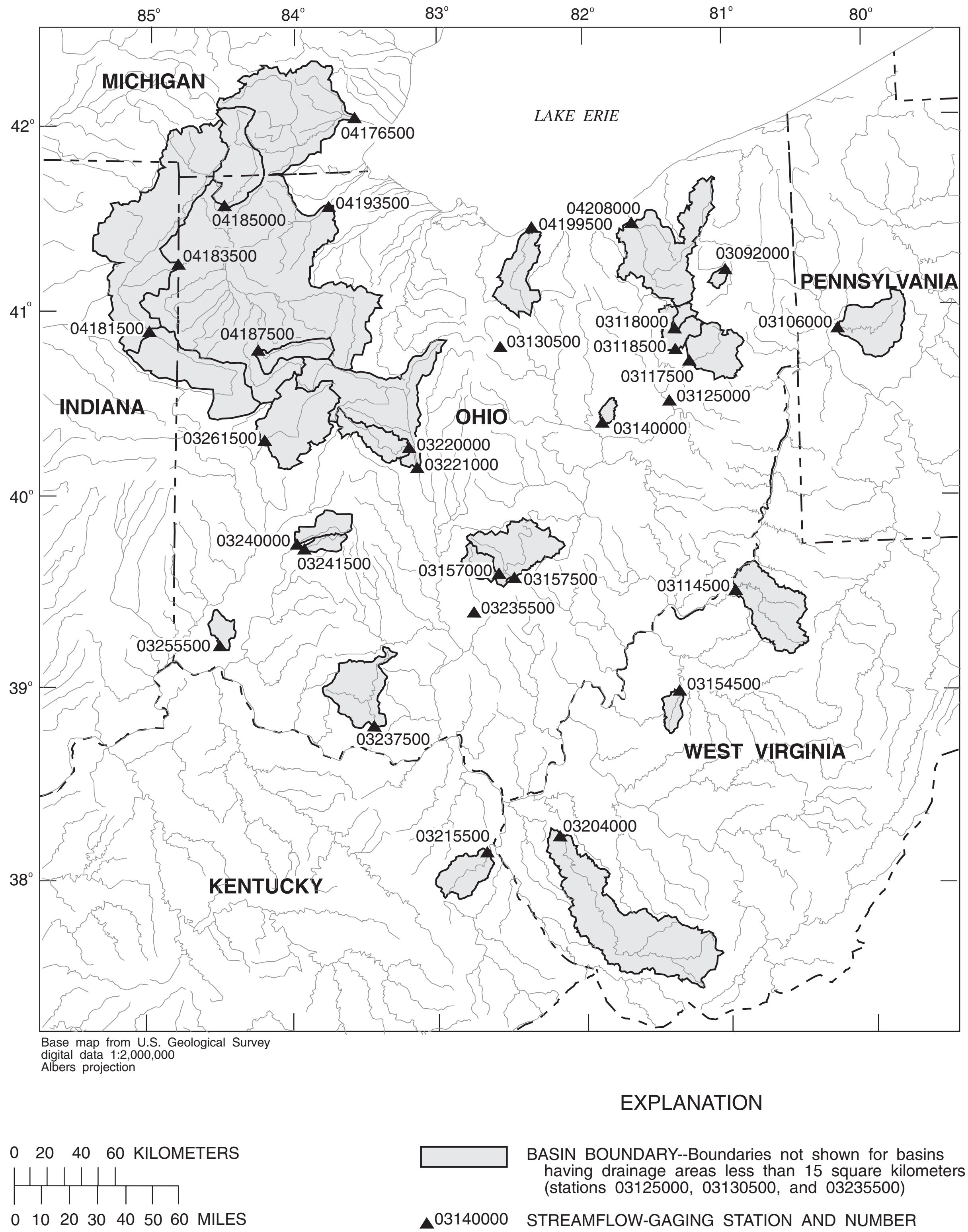

Figure 1. Locations of the 30 streamflow-gaging stations and their drainage basins.

Table 1. Station numbers, names, latitudes, longitudes, drainage areas, and observed data available for the 30 study sites
[km^2, square kilometers]

Station number	Station name	Latitude	Longitude	Drainage area (km^2)	Observed streamflow data available: Peak flow: Number of water years	Peak flow: Period	Daily mean: Number of water years	Daily mean: Period[1]
03092000	Kale Creek near Pricetown, Ohio	41°08'23"	80°59'43"	56.7	52	1942-93	53	0541-0993
03106000	Connoquenessing Creek near Zelienople, Pennsylvania	40°49'01"	80°14'33"	922	78	1916-93	74	1019-0993
03114500	Middle Island Creek at Little, West Virginia	39°28'30"	80°59'50"	1,186	76	1875 1916-22 1926-93	73	1015-0920 1025-0993
03117500	Sandy Creek at Waynesburg, Ohio[2]	40°40'21"	81°15'36"	655	55	1939-93	55	1238-0993
03118000	Middle Branch Nimishillen Creek at Canton, Ohio[2,3]	40°50'29"	81°21'14"	112	52	1942-93	52	1041-0993
03118500	Nimishillen Creek at North Industry, Ohio[2,3]	40°44'03"	81°21'08"	453	72	1922-93	72	1021-0993
03125000	Home Creek near New Philadelphia, Ohio	40°28'06"	81°24'10"	4.25	43	1937-79	44	1236-1279
03130500	Touby Run at Mansfield, Ohio	40°45'53"	82°32'43"	14.1	33	1947-78 1987	32	1046-0978
03140000	Mill Creek near Coshocton, Ohio	40°21'46"	81°51'45"	70.4	57	1937-93	57	1136-0993
03154500	Reedy Creek near Reedy, West Virginia	38°57'40"	81°23'25"	206	27	1952-78	27	1051-0978
03157000	Clear Creek near Rockbridge, Ohio[2]	39°35'18"	82°34'43"	231	54	1940-93	54	1039-0993
03157500	Hocking River at Enterprise, Ohio[2]	39°33'54"	82°28'29"	1,189	63	1907 1932-93	63	0531-0993
03204000	Guyandotte River at Branchland, West Virginia	38°13'15"	82°12'10"	3,170	60	1907 1916-22 1929-80	54	1015-0917 0429-0180
03215500	Blaine Creek at Yatesville, Kentucky	38°08'40"	82°41'05"	562	52	1916-20 1938-84	41	1015-0918 0438-0975
03220000	Mill Creek near Bellepoint, Ohio[2]	40°14'54"	83°10'26"	461	52	1913 1943-93	50	1043-0993
03221000	Scioto River below O'Shaughnessy Dam near Dublin, Ohio[2,3]	40°08'36"	83°07'14"	2,538	73	1913 1922-94	73	0421-0993
03235500	Tar Hollow Creek at Tar Hollow State Park, Ohio	39°23'22"	82°45'03"	3.50	32	1947-78	33	1046-1078
03237500	Ohio Brush Creek near West Union, Ohio	38°48'13"	83°25'16"	1,002	62	1927-35 1941-93	62	1026-0935 1040-0993
03240000	Little Miami River near Oldtown, Ohio[2]	39°44'54"	83°55'53"	334	41	1953-93	42	0852-0993
03241500	Massies Creek at Wilberforce, Ohio[2]	39°43'22"	83°52'58"	164	41	1953-93	42	0952-0993
03255500	Mill Creek at Reading, Ohio	39°13'14"	84°26'49"	189	53	1939-91	53	1038-0439
03261500	Great Miami River at Sidney, Ohio[3]	40°17'13"	84°09'00"	1,401	81	1913-93	80	0214-0993
04176500	River Raisin near Monroe, Michigan[3]	41°57'38"	83°31'52"	2,699	56	1938-93	56	1037-0993
04181500	Saint Marys River at Decatur, Indiana[2,3]	40°50'55"	84°56'16"	1,608	62	1932-93	47	0447-0993
04183500	Maumee River at Antwerp, Ohio[2]	41°11'56"	84°44'40"	5,514	71	1912-82	58	1021-0935 0439-0482
04185000	Tiffin River at Stryker, Ohio[2]	41°30'16"	84°25'47"	1,062	62	1913 1922-28 1937 1941-93	60	1021-0928 1040-0993
04187500	Ottawa River at Allentown, Ohio[3]	40°45'18"	84°11'41"	414	52	1924-35 1939 1943-81	53	1023-1235 0943-0382
04193500	Maumee River at Waterville, Ohio[2,3]	41°30'00"	83°42'46"	16,395	73	1900-01 1913 1922-36 1939-93	73	1198-1201 1021-0935 0339-0993
04199500	Vermilion River near Vermilion, Ohio	41°22'55"	82°19'01"	679	32	1950-81	32	0450-0981
04208000	Cuyahoga River at Independence, Ohio[3]	41°23'43"	81°37'48"	1,831	65	1922-23 1928-36 1940-93	65	1021-0523 1027-1235 0440-0993

[1]Period for daily mean observed data available is from beginning month and year to ending month and year.
[2]Station is on a stream that is confluent with a stream where a second station is located.
[3]Station is on a stream whose basin shares a common topographic divide with the basin of a second station on a nonconfluent stream.

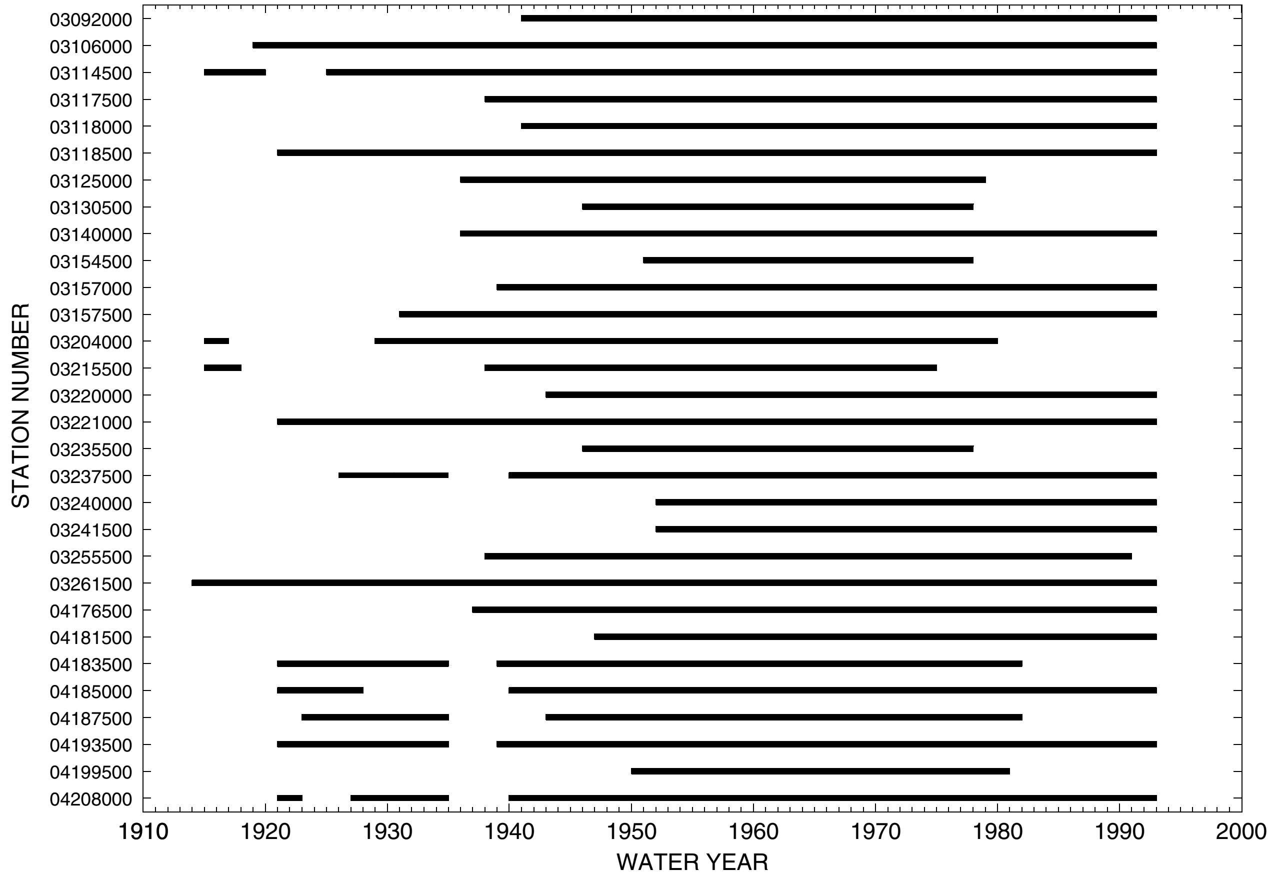

Figure 2. Individual and concurrent lengths of daily mean streamflow record for the 30 study sites in Ohio and border areas of adjacent states.

2, 10, 25, and 50 years as determined by the USGS and published in Choquette (1987), Flippo (1977), Glatfelter (1984), Holtschlag and Croskey (1984), Koltun and Roberts (1990), or Runner (1980).

The equations developed for 24 of the 30 streamflow-gaging stations are listed in table 2. The relations between daily mean streamflow and instantaneous peak streamflow could not be linearized readily for 6 of the 30 stations, so in those cases the daily mean streamflows occurring in association with instantaneous peak streamflows having recurrence intervals of 2, 10, 25, and 50 years were determined graphically. For one of the 30 stations (station number 03125000), the relation between daily mean and instantaneous-peak flows was found to vary with season; consequently, separate equations were developed for the warm season (May-September) and the cool season (October-April).

As shown in table 2, the ratio of instantaneous peak streamflows to daily mean streamflows is generally close to 1.0 for streams having large drainage areas (> 1,500 km^2), and the **R^2 values** for such streams also are relatively close to 1.0. For streams having smaller drainage areas (< 1,000 km^2), the ratio of instantaneous peak streamflows to daily mean streamflows is larger and the R^2 values are generally smaller. Scatterplots (not shown) of daily mean streamflows and (same day) instantaneous peak streamflows for streams having smaller drainage areas show more variability and a greater tendency to be nonlinear than did plots for streams with larger drainage areas.

Figure 3 shows the ratio of the 2-year instantaneous peak streamflow to the corresponding daily mean streamflow (determined either from regression or graphical analyses) for each of the 30 streamflow-gaging stations, plotted as a function of drainage area. The figure shows that ratios are close to 1.00 for large drainage areas and generally increase with decreasing drainage area. Larger drainage areas generally have longer basin lagtimes with broad flat hydrographs; consequently, the mean streamflow for a 24-hour period may

Table 2. Flood threshold data for instantanteous peak and daily mean streamflows at study sites in Ohio and border areas of adjacent states

[km^2, square kilometers; m^3/s, cubic meters per second; DV, daily mean streamflow; PK, instantaneous peak streamflow; W, warm season (May-September) equation; C, cool season (October-April) equation; na, not applicable]

Station number	Drainage area (km^2)	Equation	R-square	Instantaneous peak streamflow (m^3/s) for indicated recurrence interval, in years				Daily mean streamflow (m^3/s) estimated to occur in association with instantaneous peak streamflow with indicated recurrence interval, in years				Instantaneous peak streamflow divided by daily mean streamflow for indicated recurrence interval, in years			
				2	10	25	50	2	10	25	50	2	10	25	50
03092000	56.7	DV=0.420PK+4.05	0.75	31.2	59.5	76.7	90.9	17.1	29.0	36.3	42.2	1.82	2.05	2.12	2.15
03106000	922	DV=0.823PK-10.3	.88	228	360	428	481	178	286	342	386	1.29	1.26	1.25	1.25
03114500	1,186	DV=0.897PK-28.9	.86	377	564	657	725	309	477	560	621	1.22	1.18	1.17	1.17
03117500	655	DV=0.783PK+5.10	.94	97.7	174	219	255	81.6	141	177	205	1.20	1.23	1.24	1.25
03118000	112	DV=0.701PK+1.00	.92	18.7	35.7	45.9	54.1	14.1	26.0	33.2	38.9	1.33	1.37	1.38	1.39
03118500	453	DV=0.691PK-2.15	.78	88.9	155	188	214	59.3	105	128	146	1.50	1.48	1.47	1.47
03125000 W	4.25	DV=0.118PK+0.183	.40	3.43	7.76	10.1	11.8	0.59	1.10	1.37	1.58	5.84	7.07	7.35	7.49
03125000 C	4.25	DV=0.196PK+0.292	.39	3.43	7.76	10.1	11.8	.96	1.81	2.27	2.61	3.56	4.28	4.45	4.53
03130500	14.1	DV=0.233PK+0.348	.44	11.4	22.9	29.7	34.8	3.01	5.69	7.28	8.46	3.79	4.03	4.09	4.12
03140000	70.4	DV=0.231PK+5.21	.71	37.7	105	155	200	13.9	29.5	41.1	51.4	2.71	3.57	3.78	3.89
03154500	206	DV=0.554PK+3.26	.61	103	150	170	184	60.4	86.4	97.5	105	1.71	1.74	1.74	1.75
03157000	231	DV's from graph	na	73.1	152	206	252	40.9	86.0	105	117	1.79	1.76	1.95	2.15
03157500	1,189	DV=0.818PK+4.30	.97	196	428	592	736	165	354	488	607	1.19	1.21	1.21	1.21
03204000	3,170	DV=0.941PK-23.0	.97	623	1102	1351	1538	563	1014	1248	1424	1.11	1.09	1.08	1.08
03215500	562	DV=0.786PK-0.368	.91	167	320	411	481	131	251	322	378	1.28	1.27	1.27	1.27
03220000	461	DV's from graph	na	127	222	276	317	104	181	221	246	1.22	1.23	1.24	1.29
03221000	2,538	DV=0.810PK+29.3	.97	360	663	833	963	321	566	704	809	1.12	1.17	1.18	1.19
03235500	3.50	DV's from graph	na	3.20	9.43	14.1	18.2	.88	1.82	2.18	2.40	3.62	5.20	6.47	7.57
03237500	1,002	DV=0.552PK-5.04	.76	583	1003	1226	1396	317	548	672	766	1.84	1.83	1.83	1.82
03240000	334	DV's from graph	na	78.4	207	295	371	61.5	145	168	181	1.28	1.42	1.76	2.05
03241500	164	DV's from graph	na	44.5	117	164	203	35.0	78.2	91.9	99	1.27	1.50	1.78	2.06
03255500	189	DV=0.538PK-4.05	.53	93.7	142	165	182	46.4	72.1	84.6	94	2.02	1.96	1.95	1.94
03261500	1,401	DV=0.805PK+8.55	.95	193	385	498	592	164	319	410	485	1.18	1.21	1.22	1.22
04176500	2,699	DV=0.944PK-3.31	.98	178	326	396	447	165	304	371	419	1.08	1.07	1.07	1.07
04181500	1,608	DV=0.965PK-0.106	.99	158	272	323	360	152	263	311	347	1.04	1.04	1.04	1.04
04183500	5,514	DV=1.008PK-10.1	.99	399	598	700	773	392	592	695	769	1.02	1.01	1.01	1.01
04185000	1,062	DV=0.943PK+1.46	.99	97.7	169	203	228	93.6	161	193	216	1.04	1.05	1.05	1.05
04187500	414	DV=0.827PK-1.25	.92	87.8	149	176	195	71.4	122	144	160	1.23	1.22	1.22	1.22
04193500	16,395	DV=0.942PK+4.81	.97	1436	2302	2756	3115	1357	2174	2601	2939	1.06	1.06	1.06	1.06
04199500	679	DV's from graph	na	167	323	419	501	130	248	316	368	1.28	1.30	1.33	1.36
04208000	1,831	DV=0.749PK+7.79	.91	243	385	462	524	190	296	354	400	1.28	1.30	1.31	1.31

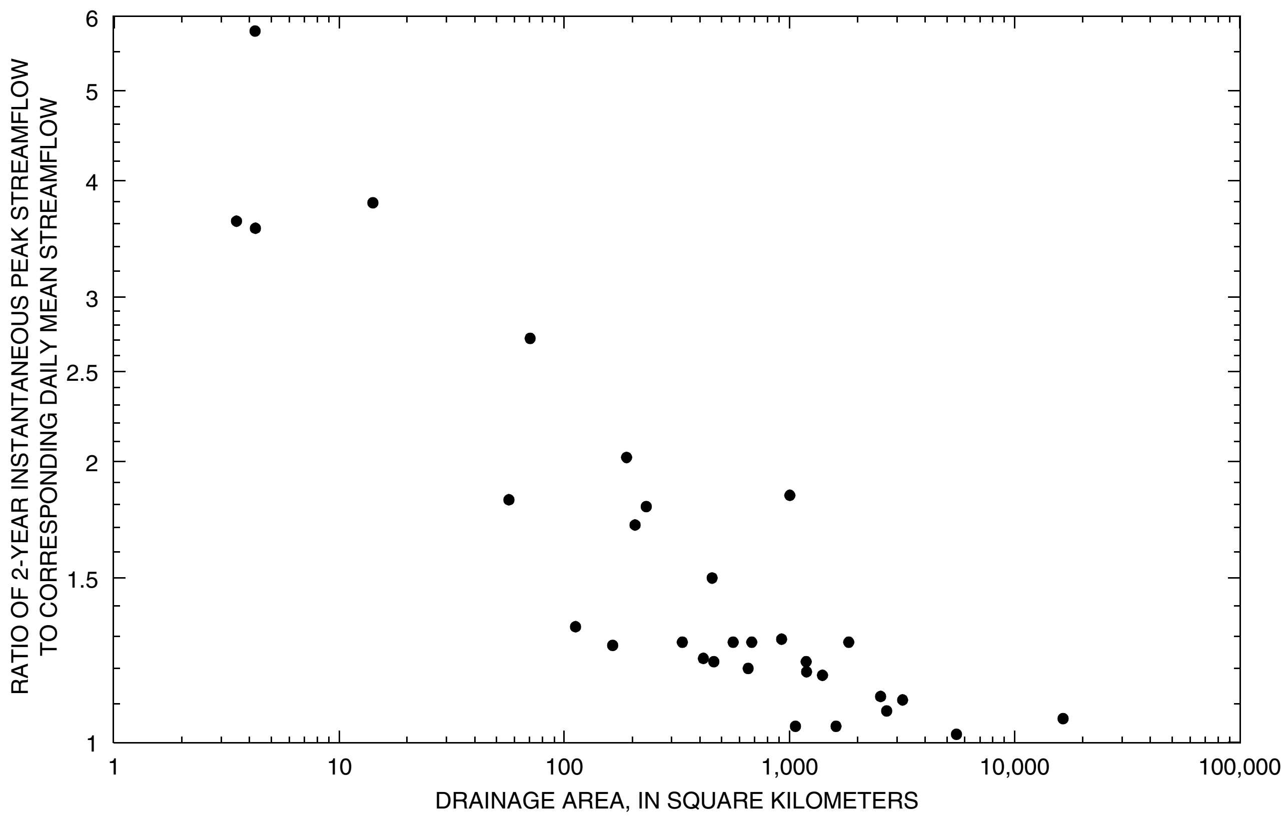

Figure 3. Ratios of 2-year instantaneous peak streamflow to the corresponding daily mean streamflow plotted as a function of drainage area, for 30 study sites in Ohio and border areas of adjacent states.

not be much smaller than the peak streamflow for the same period. In contrast, smaller drainage areas generally have shorter basin lagtimes with more peaked hydrographs, frequently resulting in a mean streamflow for a 24-hour period that is considerably smaller than the peak streamflow for the same period.

Observations of daily mean streamflow were flagged if, for a given day, an instantaneous peak streamflow from the peak-flow data set equaled or exceeded the 2-, 10-, 25-, or 50-year flood threshold. For days not having instantaneous peak streamflow data, observations of daily mean streamflow were flagged if, for a given day, the daily mean streamflow from the daily mean data set equaled or exceeded the daily mean streamflows estimated to be associated with 2-, 10-, 25-, and 50-year flood thresholds. Thus, the peak-flow data set was used to determine directly whether the 2-, 10-, 25-, or 50-year flood thresholds were equaled or exceeded, whereas the daily mean data set was used to estimate whether the 2-, 10-, 25-, or 50-year flood thresholds were equaled or exceeded. In cases where daily mean streamflow data and instantaneous peak streamflow data both were available for the same day, the instantaneous peak streamflow data were used to determine whether the flood threshold was equaled or exceeded. For example, if the daily mean streamflow data indicated an exceedance of the 10-year flood threshold and the instantaneous peak streamflow data indicated an exceedance of the 2-year flood threshold, then the day was flagged as exceeding only the 2-year flood threshold (because the instantaneous peak streamflow data take precedence).

A separate flag was set for each flood-threshold recurrence interval equaled or exceeded so that data for each day could be filtered and categorized as falling within a discrete recurrence-interval range. For example, a streamflow value that exceeded the 2-year flood threshold level but not the 10-year flood threshold level was categorized as falling in the 2- to 10-year recurrence-interval range. Streamflow data that were categorized and analyzed as a function of the bounding range of recurrence interval are subsequently referred to as discretized streamflow data.

The final flagged data sets were derived predominantly from the daily mean streamflow data with some adjustments based on the instantaneous peak streamflow data. The minimum, maximum, and mean number of days when the 2-, 10-, 25-, and 50-year flood thresholds were equaled or exceeded at the 30 streams are listed in table 3.

Table 3. Minimum, maximum, and mean number of days when the 2-, 10-, 25-, and 50-year flood thresholds were equaled or exceeded at the 30 study sites in Ohio and border areas of adjacent states

Flood-threshold recurrence interval	Number of days when flood threshold was equaled or exceeded		
	Minimum	Maximum	Mean
2-year	25	182	74.6
10-year	3	22	8.1
25-year	0	6	2.6
50-year	0	4	1.4

Determination of stream pairs for further analysis

The daily mean streamflow data from each gaging station were paired with concurrent daily mean streamflow data from the other 29 stations. This procedure was repeated for all combinations of stream pairs, yielding an initial data set consisting of 435 stream pairs. To minimize the effects of serial correlation on subsequent analyses, stream pairs were excluded from the data set if two conditions were present: (1) the paired streams were situated so that flow passing one streamflow-gaging station ultimately passed the second streamflow-gaging station, and (2) the ratio of the smaller to larger drainage area was greater than 0.1. It was arbitrarily assumed that serial correlation effects would be minimal for stream pairs in which the ratio of their drainage areas (smaller drainage area divided by the larger drainage area) is less than 0.1. Five stream pairs met these exclusion criteria, resulting in a final data set containing information on 430 stream pairs. Of the 430 stream pairs, 3 unique stream pairs are on confluent streams and 14 stream pairs are on streams in basins sharing a common topographic divide.

Application of flood-threshold filters

For a given recurrence interval, the paired daily mean streamflow data were filtered to remove observations for which flows at both streams were less than their respective flood thresholds (if the threshold was equaled or exceeded at either stream in the pair, then the paired data were retained). If, on a given day, the flood threshold was equaled or exceeded at at least one stream in the pair, a “**trial**” was said to have occurred. An “**event**” was said to have occurred on the same day if flood thresholds were equaled or exceeded at both streams in the pair. The minimum, maximum, and mean frequencies of trials and events determined for the 2-, 10-, 25-, and 50-year flood thresholds for the 430 stream pairs are listed in table 4. The statistics listed in table 4 illustrate that many trials and events were available for analysis of streamflows equal to or exceeding the 2-year flood threshold, whereas far fewer trials and events were available for analyses of streamflows equal to or exceeding the 10-, 25-, and 50-year flood thresholds.

Table 4. Minimum, maximum, and mean frequencies of trials and events determined for the 2-, 10-, 25-, and 50-year flood thresholds for the 430 site pairs in Ohio and border areas of adjacent states

Flood-threshold recurrence interval	Number of trials			Number of events		
	Minimum	Maximum	Mean	Minimum	Maximum	Mean
2-year	25	250	108.0	0	85	8.1
10-year	4	35	12.2	0	9	0.5
25-year	0	10	3.9	0	2	.1
50-year	0	6	2.2	0	2	.1

Event-trial ratios as a measure of the joint probability of flooding

Event-trial ratios determined for the 2-, 10-, 25-, and 50-year flood thresholds were used as measures of the joint probability of flooding at the respective recurrence-interval levels for the 430 stream pairs. These data were analyzed in concert with factors hypothesized to be most closely associated with the joint flooding phenomenon. Examples of factors thought to be important include drainage-area ratios, distance and angle between drainage area centroids, and season. Multiple flood-threshold levels were used to provide further insight into the relations between flood magnitude, basin characteristics, and the joint probability of flooding.

Computations of event-trial ratios

The joint probability of flooding at the 2-, 10-, 25-, and 50-year flood-threshold level was computed as the ratio of the number of times when the respective flood threshold was concurrently equaled or exceeded at both streams in the pair (events) to the number of times when the respective flood threshold was equaled or exceeded at at least one stream in the pair (trials). Because of the nature of the data analyzed, floods at paired sites were considered to be concurrent if they occur on the same calendar day (irrespective of the actual time of day that the peak flows occurred). For example, if the 10-year flood threshold was concurrently equaled or exceeded at both streams in the pair on 2 days (2 events) and the same flood threshold was equaled or exceeded at one or both of the streams in the pair on a total of 8 days (8 trials), then the computed event-trial ratio (the observed joint probability of flooding) at the 10-year flood-threshold level would be 0.25 (2/8). If flood thesholds were exceeded at both streams in the pair on consecutive days, each day was counted as a separate event. In order to facilitate further discussions, event-trial ratios determined for an **N-year flood threshold** (where N = 2, 10, 25 or 50) may be referred to simply as the **N-year event-trial ratio**.

Scatterplots of concurrent daily mean streamflows for 4 of the 430 stream pairs (figs. 4a-d) will be used to illustrate the logic and some nuances associated with the computation of the event-trial ratios. Figure 4a is a scatterplot of concurrent daily mean streamflow data for stations 03157500 and 04183500, one in which the event-trial ratios were very small. The event-trial ratio for the 2-year flood threshold was determined by counting the total number of points that equaled or exceeded the 2-year flood threshold at *both* streams (all points in the dark shaded rectangle) and dividing by the total number of points that equaled or exceeded the 2-year flood threshold at *either* stream (all points in the dark shaded rectangle or in the two lightly shaded rectangles).The unshaded rectangle in the lower left corner (formed by the axes of the plot and the dotted lines representing the respective 2-year flood thresholds for the two streams) contains no plotted points because that area represents the condition in which the daily mean streamflows at both streams were less than the respective 2-year flood thresholds (and thus were filtered from the working data set).

Because instantaneous peak streamflow data (when available) were used preferentially to assess flood-threshold exceedance, the event-trial ratios were in some cases modified from those that would be determined by use of the daily mean streamflow plots alone. In figure 4a, there are two observations in which the 2-year flood threshold was exceeded concurrently at both streams on the basis of daily mean streamflow data. For one of those two observations, however, the instantaneous peak streamflow data indicated that the 2-year flood threshold was not exceeded, so the total number of events was revised to one instead of two. Use of available instantaneous peak streamflow data also resulted in the number of trials being revised from 205 to 202; consequently, the 2-year event-trial ratio was determined to be 0.005 (1/202). The 10-, 25-, and 50-year event-trial ratios are 0.00 because there were no instances in which the 10-, 25-, or 50-year flood thresholds were equaled or exceeded concurrently at both streams, yet each of these flood thresholds was equaled or exceeded at least once at one or more sites in the pair.

Streamflows that equaled or exceeded a flood threshold for a given recurrence interval were also classified as exceeding all flood thresholds of lower recurrence interval. In the example shown in figure 4a, the 50-year flood threshold was exceeded on one day at station 03157500. That same observation was also counted as an exceedance of the 2-, 10-, and 25-year threshold in analyses involving those respective thresholds.

Figure 4b shows a scatterplot of concurrent daily mean streamflow data for stream pair 03118500 and 04187500, in which the event-trial ratio is fairly small for the 2-year (0.09) and 10-year (0.11) flood thresholds, but considerably larger for the 25-year (0.25) and 50-year (0.50) flood thresholds. This plot illustrates the effect that a large flood can have on the computed event-trial ratios for thresholds of higher recurrence interval. The data point in the upper right corner of figure 4b shown to have exceeded the 50-year flood threshold at both streams is for the flood of January 22, 1959 (the second highest flood of record for Ohio in terms of severity and areal extent). Without that data point, the event-trial ratios for the 25-year and 50-year flood thresholds would have been 0.00.

Figure 4b further illustrates how available instantaneous peak streamflow data can affect the computed event-trial ratios. In this case, the instantaneous peak streamflow data indicated that the 50-year flood threshold was exceeded on a second day at one of the two streams. Consequently the number of trials was increased by 1, making the event-trial ratio 0.5 (1/2) instead of the 1.0 (1/1) ratio that would be determined from the daily mean streamflow data alone. Because of the small number of higher recurrence interval floods in the systematic record on which to base the computations, the event-trial ratios computed for the higher recurrence interval flood thresholds are extremely variable and highly influenced by the inclusion or exclusion of even a single event or trial.

Figure 4c is a scatterplot of concurrent daily mean streamflow data for stream pair 04176500 and 04183500, in which the computed event-trial ratios for the 2-year (0.27), 10-year (0.25), and 25-year (0.20) flood thresholds are relatively high and of similar magnitude. The 50-year flood threshold was never equaled or exceeded at either stream, and consequently the event-trial ratio for the 50-year flood threshold was treated as a **missing** value, thereby reducing the effective number of observations at that threshold level.

Figure 4d is a scatterplot of concurrent daily mean streamflow data for stream pair 03240000 and 03241500, in which the event-trial ratios for all four flood thresholds are relatively high but dissimilar in magnitude. An event-trial ratio of 1.00 was computed for the 50-year flood threshold because there were two events and two trials (based on instantaneous peak streamflow data). A value of 1.00 (indicating absolute certainty) most likely overestimates the true joint probability of flooding at the 50-year flood threshold. The event-trial ratios determined for the 2-year (0.53),

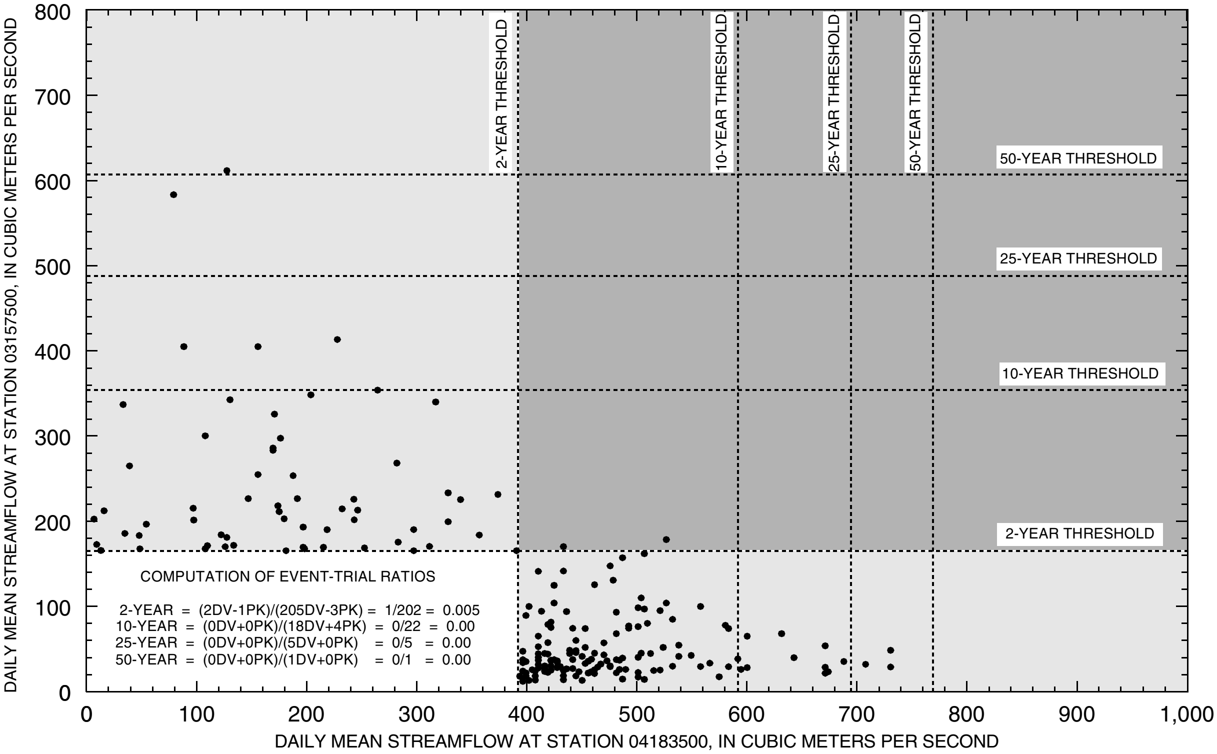

Figure 4a. Concurrent daily mean streamflows for stations 03157500 and 04183500.

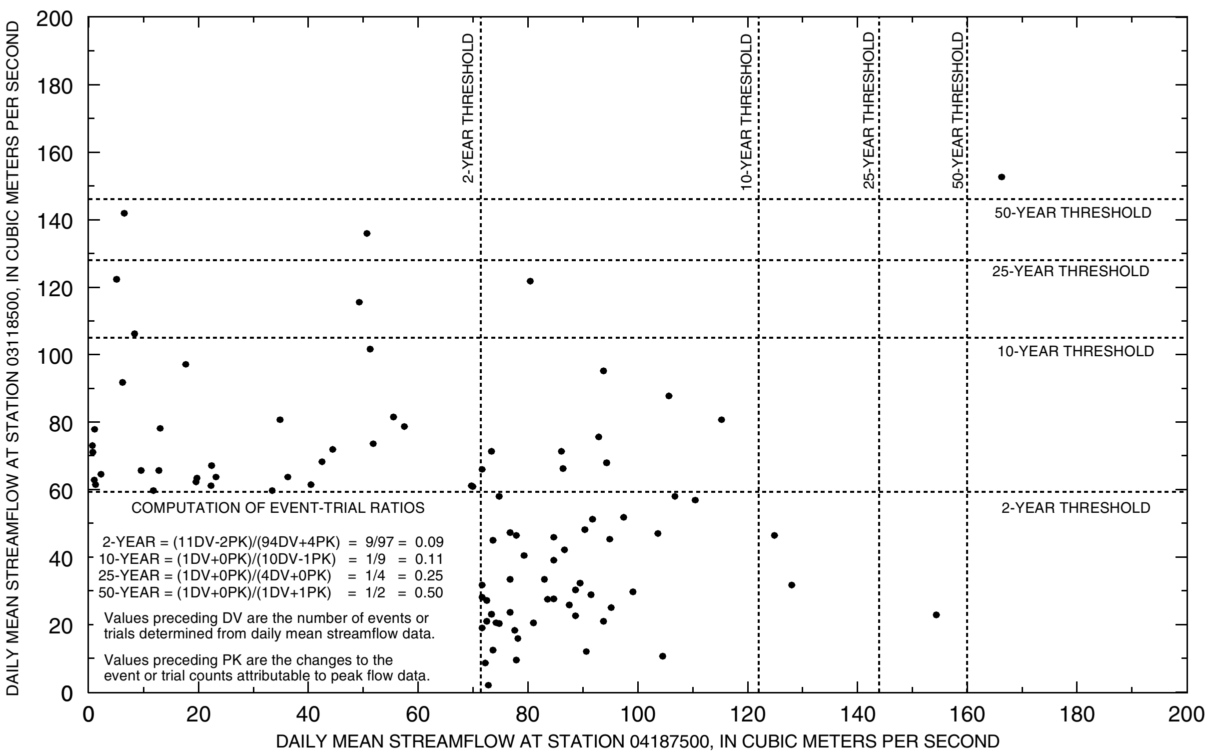

Figure 4b. Concurrent daily mean streamflows for stations 03118500 and 04187500.

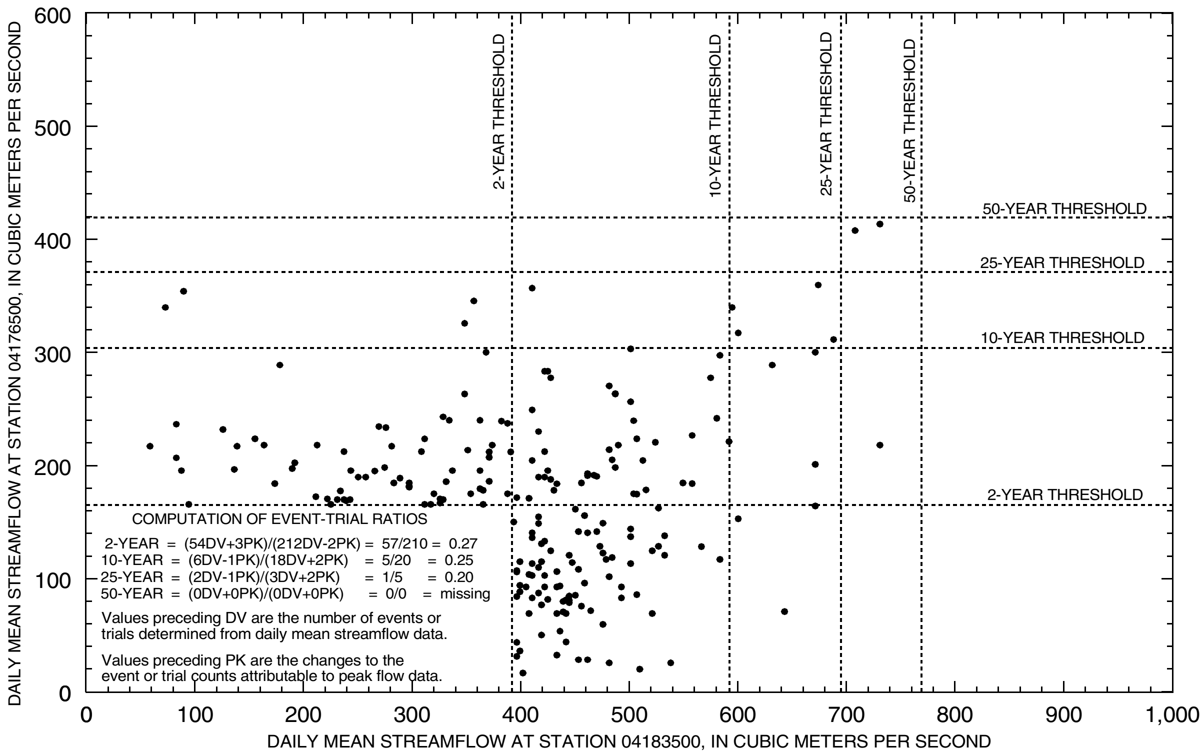

Figure 4c. Concurrent daily mean streamflows for stations 04176500 and 04183500.

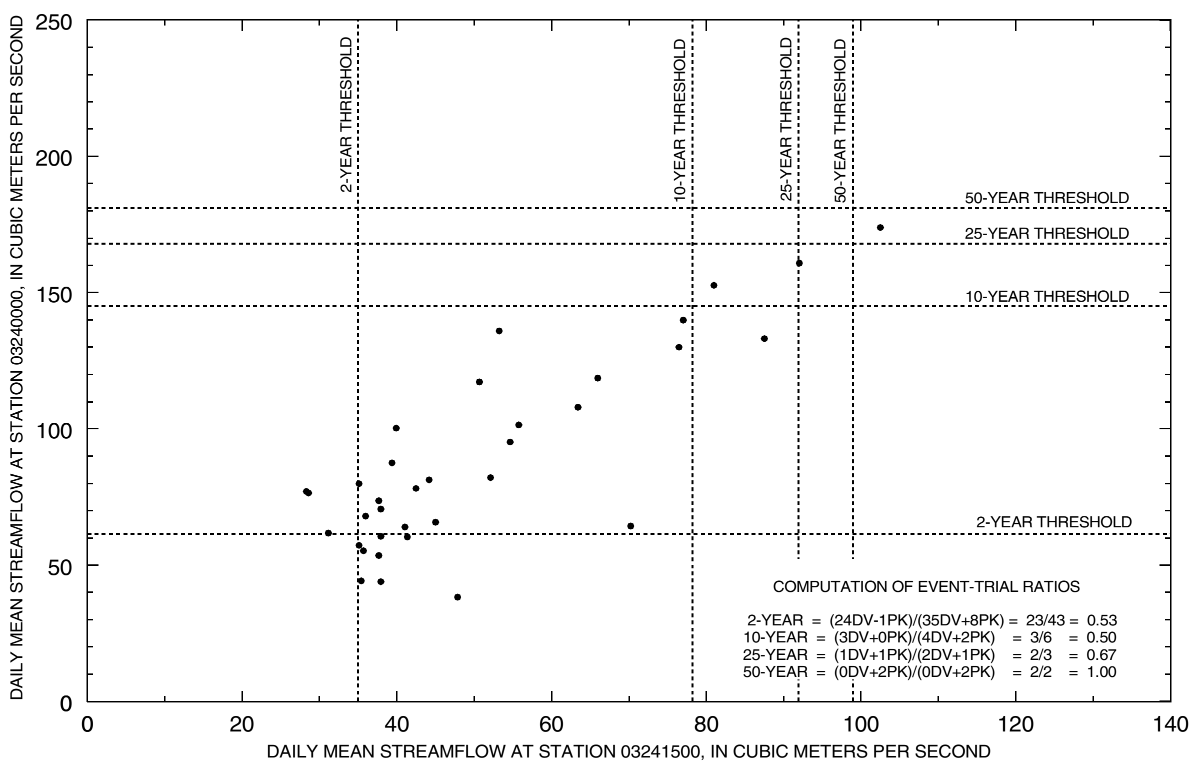

Figure 4d. Concurrent daily mean streamflows for stations 03240000 and 03241500.

10-year (0.50), and 25-year (0.67) flood thresholds might, in this case, be more reasonable values of the joint probability of flooding at the 50-year flood-threshold level.

Characteristics of event-trial ratios

Summary statistics for **non-missing** event-trial ratios determined for the 2-, 10-, 25-, and 50-year flood thresholds are listed in table 5. All 430 stream pairs had at least one trial in the 2- and 10-year flood threshold analyses, whereas no trials occurred in the 25- and 50-year flood threshold analyses for 25 and 55 stream pairs, respectively. Also evident in table 5 is a decrease in the mean and an increase in the standard deviation of event-trial ratios with an increase in recurrence interval.

Summary statistics of non-zero event-trial ratios determined for the 2-, 10-, 25-, and 50-year flood thresholds are listed in table 6. As is evident in the table, zero-valued event-trial ratios occurred for all four recurrence intervals, but the number of zero values increased considerably as recurrence interval increased. With respect to non-missing values, 11 percent of the 2-year event-trial ratios were equal to zero, whereas 68, 86, and 90 percent of the respective 10-, 25-, and 50-year event-trial ratios were equal to zero.

Histograms of event-trial ratios determined for the 2-, 10-, 25-, and 50-year flood thresholds are shown in figure 5. Sixty-six percent or more of the N-year event-trial ratios (where N = 2, 10, 25 or 50) fell in a range from 0.00 to 0.10. More than 88 percent of event-trial ratios in that range equaled zero for the 10-year flood threshold and 100 percent of the event-trial ratios in that range equaled zero for the 25-, and 50-year flood thresholds.

Data analyzed for the 25-, and 50-year flood thresholds have a greater proportion of high event-trial ratios than do data analyzed for the 2- and 10-year flood thresholds. This outcome is in part an artifact of the small numbers of events and trials in the systematic record for the 25- and 50-year flood thresholds. For example, if a stream pair has two trials for the 50-year flood threshold, the possible event-trail ratio values are 1.00 (2/2), 0.50 (1/2), or 0.00 (0/2). If instead there are three trials, then the possible event-trail

Table 5. Summary statistics of non-missing event-trial ratios determined for the 2-, 10-, 25-, and 50-year flood thresholds

["Missing" refers to the condition where the flood threshold being considered was not exceeded at either site in the pair]

Flood-threshold recurrence interval	Number of observations			Statistic for non-missing event-trial ratios			
	Total	Missing	Non-missing	Mean	Standard Deviation	Minimum	Maximum
2-year	430	0	430	0.081	0.083	0	0.535
10-year	430	0	430	.051	.087	0	.500
25-year	430	25	405	.042	.117	0	.667
50-year	430	55	375	.043	.153	0	1.000

Table 6. Summary statistics of non-zero event-trial ratios determined for the 2-, 10-, 25-, and 50-year flood thresholds

["Zero" refers to the condition where the flood threshold being considered was exceeded at at least one site in the pair but was never concurrently exceeded at both sites in the pair]

Flood-threshold recurrence interval	Number of observations			Statistic for non-zero event-trial ratios			
	Non-missing	Zero	Non-zero	Mean	Standard Deviation	Minimum	Maximum
2-year	430	46	384	0.091	0.083	0.005	0.535
10-year	430	292	138	.158	.081	.036	.500
25-year	405	348	57	.301	.141	.143	.667
50-year	375	339	36	.450	.247	.200	1.000

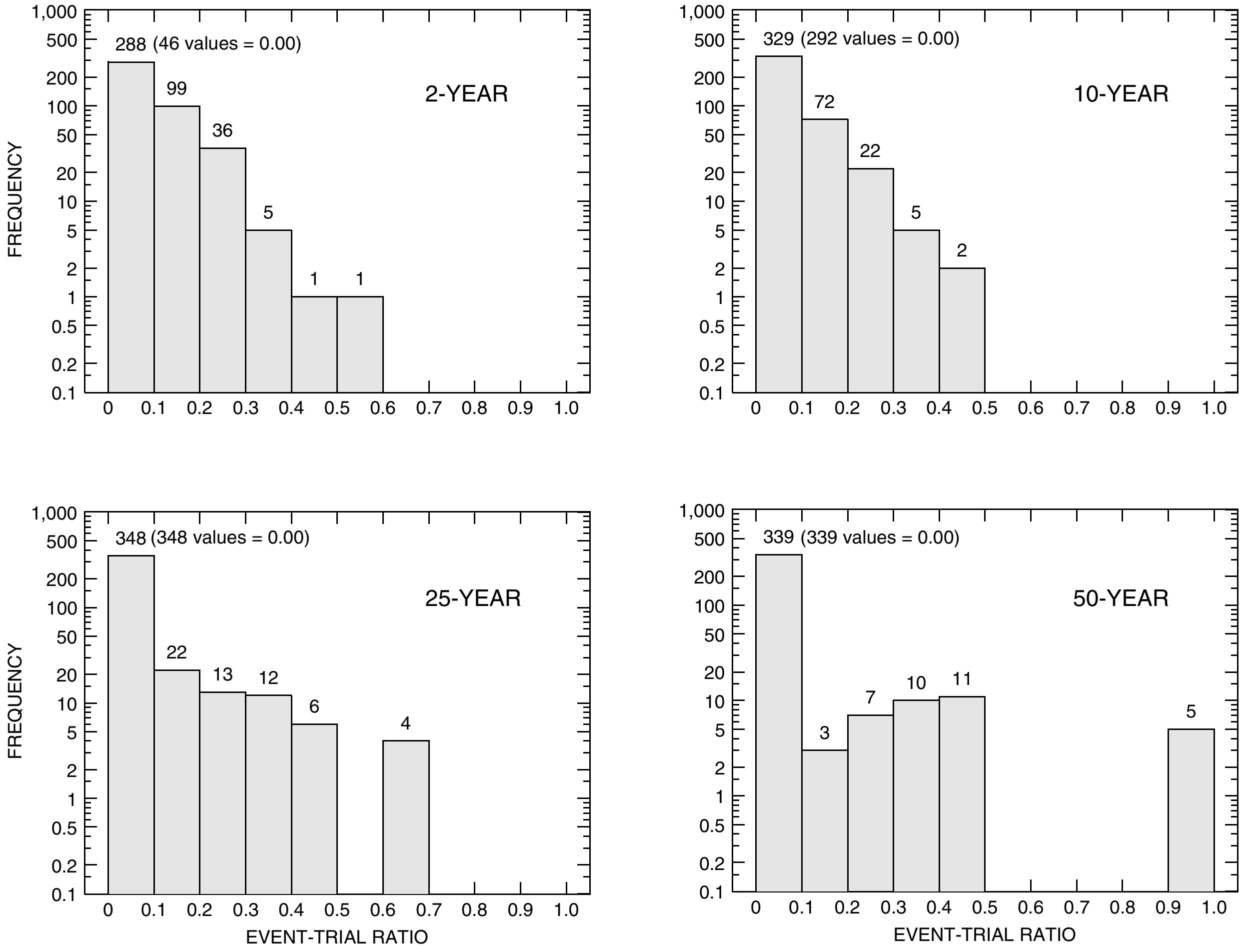

Figure 5. Frequency of event-trial ratios determined for the 2-, 10-, 25-, and 50-year flood thresholds among 30 sites in Ohio and border areas of adjacent states.

ratio values are 1.00 (3/3), 0.67 (2/3), 0.33 (1/3), or 0.00 (0/3), and so forth. From this example, it is clear that the occurrence of as few as one or two events can result in large event-trial ratios.

Figure 6 consists of scatterplots of the 2-year event-trial ratio (on the abscissa) and the 10-, 25-, and 50-year event-trial ratios (on the ordinates) for each stream pair. A spline curve is drawn on each scatterplot to show general trends in the relation between the 2-year event-trial ratios and event-trial ratios for thresholds of higher recurrence interval. The cubic spline interpolation used to determined the curves minimizes a linear combination of the sum of squares of the residuals of fit and the integral of the square of the second derivative (Reinsch, 1967). A smoothing factor of 65 was used for each plot. The spline curves show that the 2-year event-trial ratios and the 10-, 25-, and 50-year event-trial ratios tend to increase together and, their average slopes indicate that, for a given stream pairing, the 2-year event-trial ratio tends to be larger than the corresponding 10-, 25-, and 50-year event-trial ratios. It is apparent, however, that the relation between the 2-year event-trial ratio and event-trial ratios for thresholds of higher recurrence interval become considerably less linear and more scattered as threshold recurrence intervals increase.

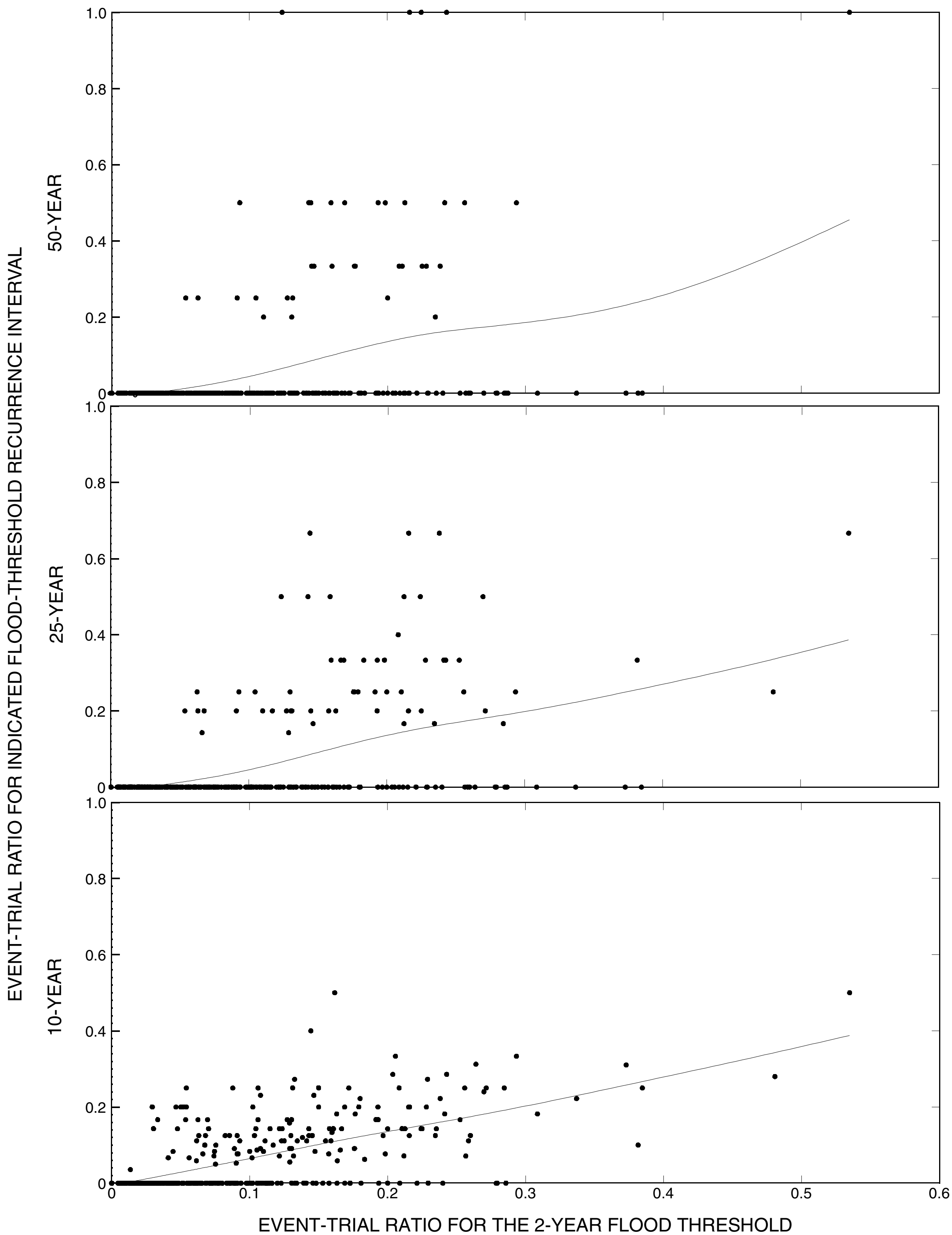

Figure 6. Relation between 2-year event-trial ratios and 10-, 25-, and 50-year event-trial ratios for the 30 study sites in Ohio and border areas of adjacent states.

Probabilities of joint flooding as a function of basin characteristics

Correlation analyses, graphical analyses, and logistic-regression analyses were used in combination to identify and quantify factors most closely associated with the observed probabilities of joint flooding (event-trial ratios) determined for the 2-, 10-, 25-, and 50-year flood thresholds. The following is a list of the basin characteristics that were determined for each stream pair and tested in these analyses:

- distance between centroids of the drainage areas of the paired streams
- drainage area ratio (smaller drainage area divided by the larger drainage area)
- difference between drainage areas of the paired streams
- mean drainage area of the paired streams
- mean main-channel length of the paired streams
- angle between a line connecting centroids of the drainage areas of the paired streams and true north
- distance between centroids divided by the mean drainage area
- distance between centroids divided by the mean main-channel length
- distance between gaging-station locations
- main-channel slope ratio and difference
- drainage area storage ratio and difference
- mean annual precipitation ratio and difference
- decimal latitude ratio and difference
- ratio of the N-year flood magnitudes, where N = 2, 5, 10, 25, 50, and 100

These basin characteristics (and assorted arithmetic transformations thereof) were in many cases derived from basin characteristics that were found to be statistically significant in previous studies of the relations between flood characteristics and basin characteristics in Ohio and elsewhere (Jennings and others, 1994; Koltun and Roberts, 1990; Thomas and Benson, 1970).

The correlation analyses, graphical analyses, and logistic-regression analyses discussed in the next sections of this report indicate that, in general, the basin characteristics most closely associated with the joint probability of flooding of paired streams in Ohio are the following:

CD **Centroid distance** (in kilometers) - The distance between the centroids of the drainage areas of the paired streams. The locations of the centroids may be determined by use of a geographic information system (GIS), as was done for this study, or they may be determined graphically as follows:

1. Draw a straight line across the drainage area in any direction so that approximately half of the area within the drainage area boundary lies on either side of the line.
2. Determine the area of each half of the drainage area (by planimeter, digitizer, grid, or other method).
3. Adjust the location of the line across the drainage area until the areas on both sides of the line are equal (this may take several iterations).
4. Repeat steps 1, 2, and 3 with a second line drawn across the drainage area approximately perpendicular to the first line.
5. The intersection of the two lines is the centroid of the drainage area.

DAR **Drainage area ratio** (on a scale from 0 to 1) - The drainage area (contributing surface runoff) of the smaller of the paired streams divided by the drainage area of the larger of the paired streams, as determined from U.S. Geological Survey 7.5-minute topographic quadrangle maps.

MDA **Mean drainage area** (in square kilometers) - The mean of the drainage areas of the paired streams (sum of the two drainage areas divided by 2), as determined from U.S. Geological Survey 7.5-minute topographic quadrangle maps.

CA **Centroid angle** (in degrees) - The angle between a line connecting the centroids of the drainage areas of the paired streams and a zero-degree reference line coincident with true north. A negative angle indicates that the centroid of the more northerly drainage area of the pair lies west of the centroid of the more southerly drainage area (-89 degrees is almost directly west), a positive angle indicates that the centroid of the more northerly drainage area lies to the east of the centroid of the more southerly drainage area (89 degrees is almost directly east), and drainage areas whose centroids lie directly north and south of each other are assigned a centroid angle of 0. Thus, centroid angles range from -90 to 90 degrees.

Scatterplots of centroid angles against event-trial ratios (not shown) indicated that the relation was not linear. In order to linearize the relation for use in subsequent analyses, it was necessary to take the absolute value of the centroid angle and then further adjust it by addition or subtraction of a constant. Various angular adjustments to the zero-degree reference line were tested to determine an optimum adjustment angle. Figure 7 is a scatterplot showing the correlations (as measured by **Spearman's rho**) between the 2-, 10-, 25-, and 50-year event-trial ratios and adjusted centroid angle as a function of the angle of adjustment. The plot shows that for data analyzed at most flood threshold levels, the highest correlation coincides with an adjustment of about 30 degrees. This outcome indicates that basin pairs with centroids aligned in a west-southwest (240^{o}) to east-northeasterly (60^{o}) direction tend to have a higher joint probability of flooding than those aligned in other direc-

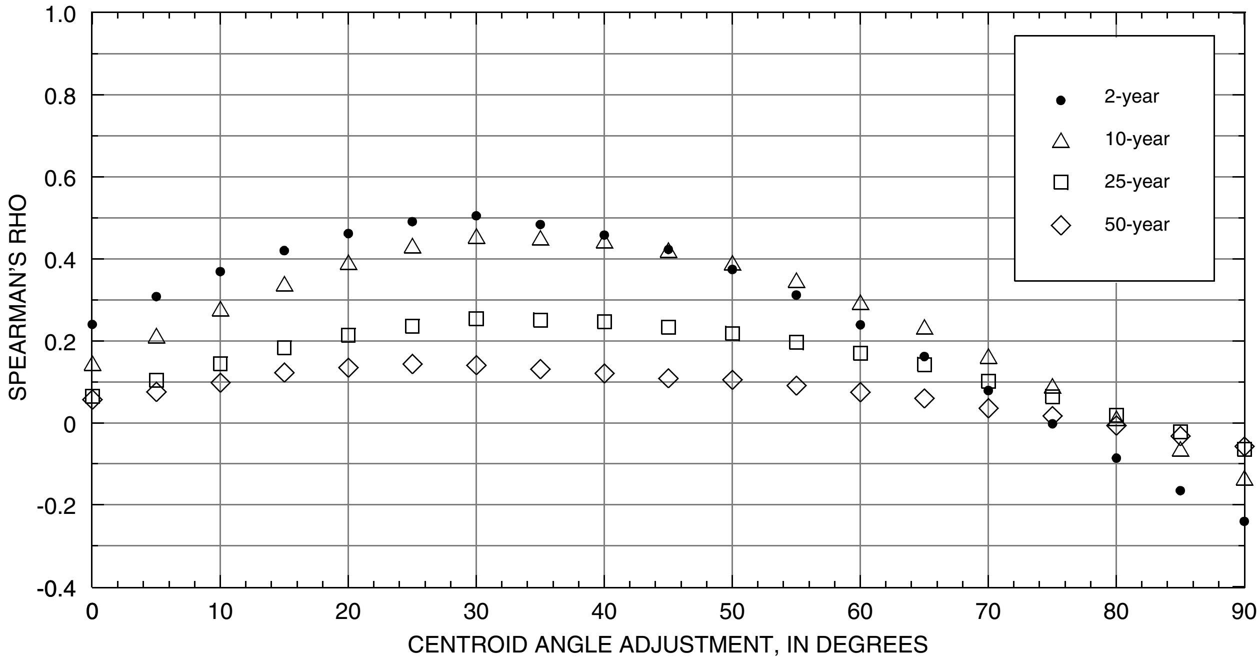

Figure 7. Correlation between adjusted centroid angle and the 2-, 10-, 25-, and 50-year event-trial ratio as a function of centroid angle adjustment for the 30 study sites in Ohio and border areas of adjacent states.

tions. This result is consistent with the findings of Huff and Angel (1992), who found that moderately heavy rainstorms in the Midwest are oriented most frequently from west-southwest to east-northeast or southwest to northeast. As a result of this finding, a 30-degree adjustment to the centroid angle was made and used in subsequent analyses. The **adjusted centroid angle**, hereafter referred to as *ACA30*, was determined as follows and as illustrated in figure 8:

$$ACA30 = \text{abs}(CA+30^{o}), \text{ where } CA \leq 60^{o} \qquad (1)$$

and

$$ACA30 = \text{abs}(CA-150^{o}), \text{ where } CA > 60^{o}, \qquad (2)$$

where abs() is the absolute value operator.

Correlation analyses

Correlation analyses were used to explore associations between various basin characteristics and event-trial ratios determined for the 2-, 10-, 25-, and 50-year flood thresholds. The analyses indicate that centroid distance (*CD*), drainage area ratio (*DAR*), mean drainage area (*MDA*), and adjusted centroid angle (*ACA30*) are the basin characteristics most highly correlated with the event-trial ratios.

Figure 9 shows bar graphs representing the correlation (as measured by Spearman's rho) between selected basin characteristics (*CD*, *DAR*, *MDA*, and *ACA30*) and event-trial ratios grouped by centroid distance and flood-threshold recurrence interval. The graphs show correlation coefficients for observations grouped into four ranges of centroid distance: 0-80 km, >80-160 km, >160-507 km, and 0-507 km (centroid distances for the 430 stream pairs range from 9.46 to 507 km).

Some patterns are evident from examining the bar graphs for the full data set (centroid distances from 0 to 507 km, shown in the right-hand column) in figure 9:

- correlation coefficients generally decrease in absolute value with an increase in recurrence interval
- signs and relative magnitudes of the correlation coefficients are fairly consistent for all four recurrence intervals

These patterns indicate that the basin characteristics have similar associations with the event-trial ratios regardless of flood-threshold recurrence interval. However, the

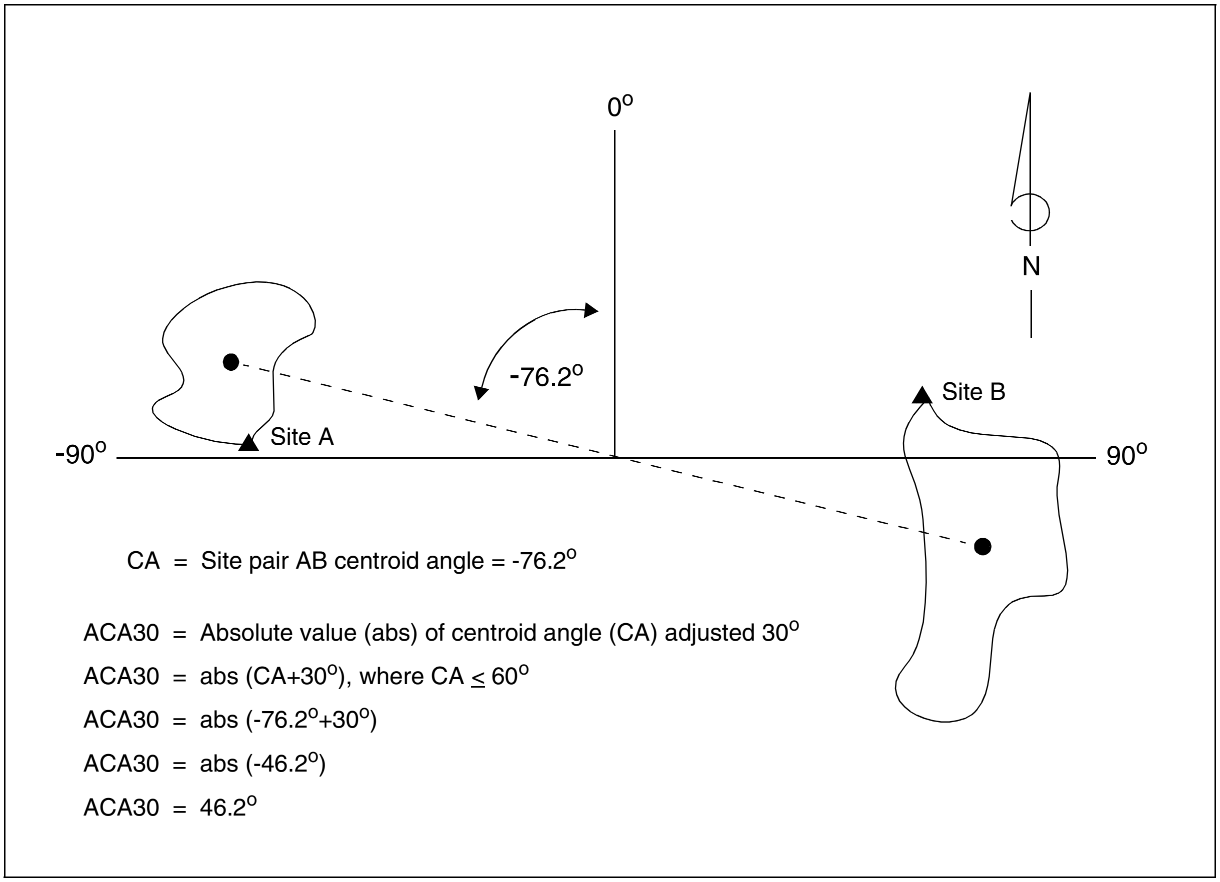

Figure 8. Diagram showing measurement of centroid angle (CA) and computation of absolute value of centroid angle adjusted 30 degrees (ACA30) for hypothetical site pair AB.

strengths of those associations diminish with increasing recurrence interval.

Figure 9 also shows several patterns indicating that as centroid distance increases, there is

- a slight decrease in the magnitudes of the correlation coefficients for *DAR*,
- a large increase and change in sign of the correlation coefficients for *ACA30*, and
- a moderate decrease and change in sign of the correlation coefficients for *MDA*.

The patterns of correlation indicate that for drainage areas whose centroids are relatively close to each other (0-80 km), the association between drainage area ratios and event-trial ratios is stronger and the association between adjusted centroid angle and event-trial ratios is weaker than comparable associations for drainage areas whose centroids are separated by greater distances. The patterns observed for *MDA* are more difficult to interpret. One possible explanation is that large drainage areas whose centroids are relatively distant are less likely to flood concurrently owing to (1) the lower probability of concurrently being affected by convective storms (because of the storms generally limited areal extent) and (2) greater differences in timing of runoff associated with a frontal storm passing through drainage areas that are more distant.

Graphical analyses

Graphical analyses were used to further explore the associations between various basin characteristics and event-trial ratios determined for the 2-, 10-, 25-, and 50-year flood thresholds. Scatterplots of 2-year event-trial ratios and selected basin characteristics (*CD*, *DAR*, *MDA*, and *ACA30*) for four ranges of centroid distance are shown in figures 10 and 11. The scatterplots are shown only for the 2-year data because the relations between the 2-year event-trial ratios and the selected basin characteristics are much better defined than those for the higher recurrence intervals (10-, 25-, and 50-years). The plots shown in figures 10 and 11 help illustrate the relations identified in the correlation analyses; how-

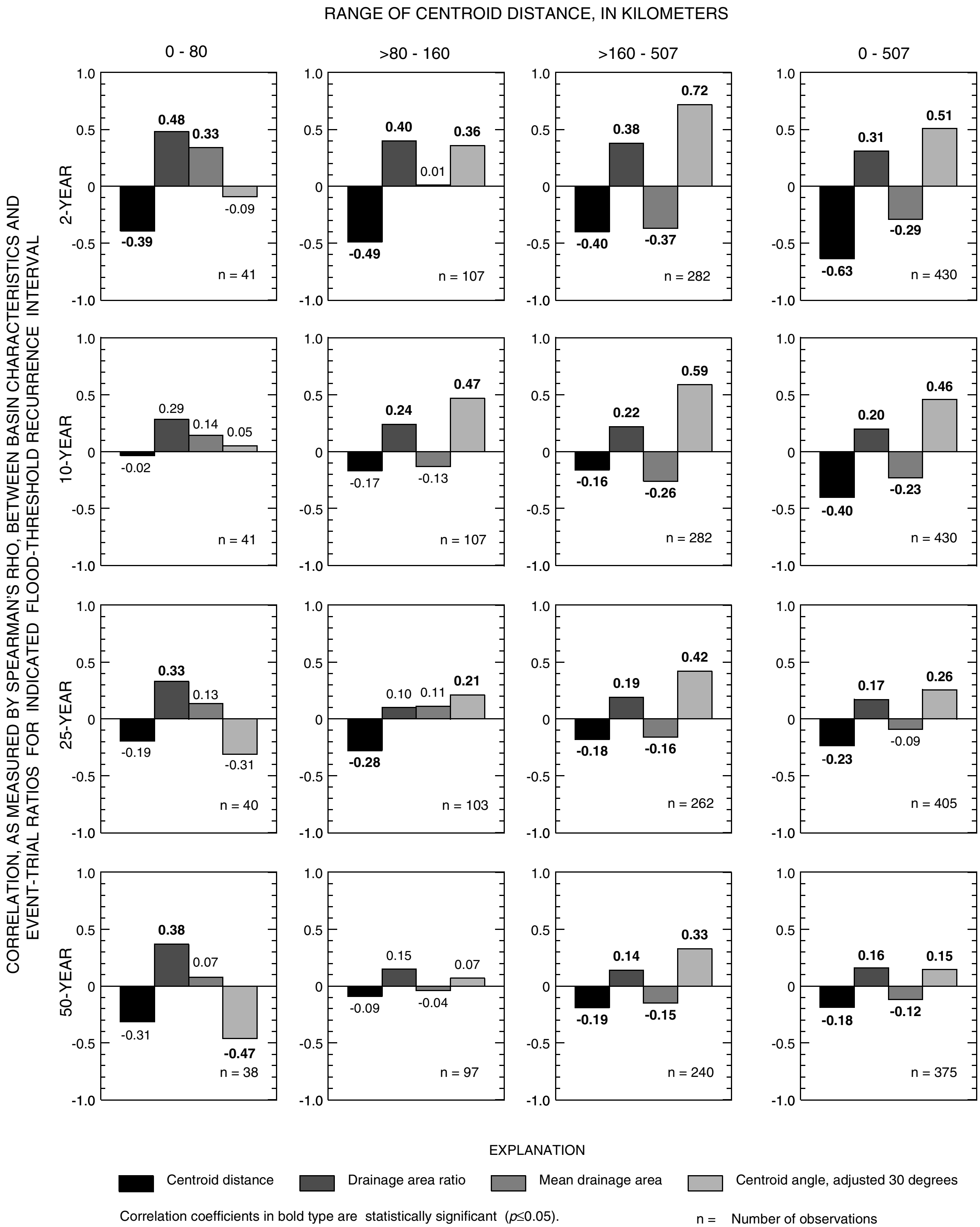

Figure 9. Correlation between event-trial ratios and selected basin characteristics grouped by centroid distance and flood-threshold recurrence interval.

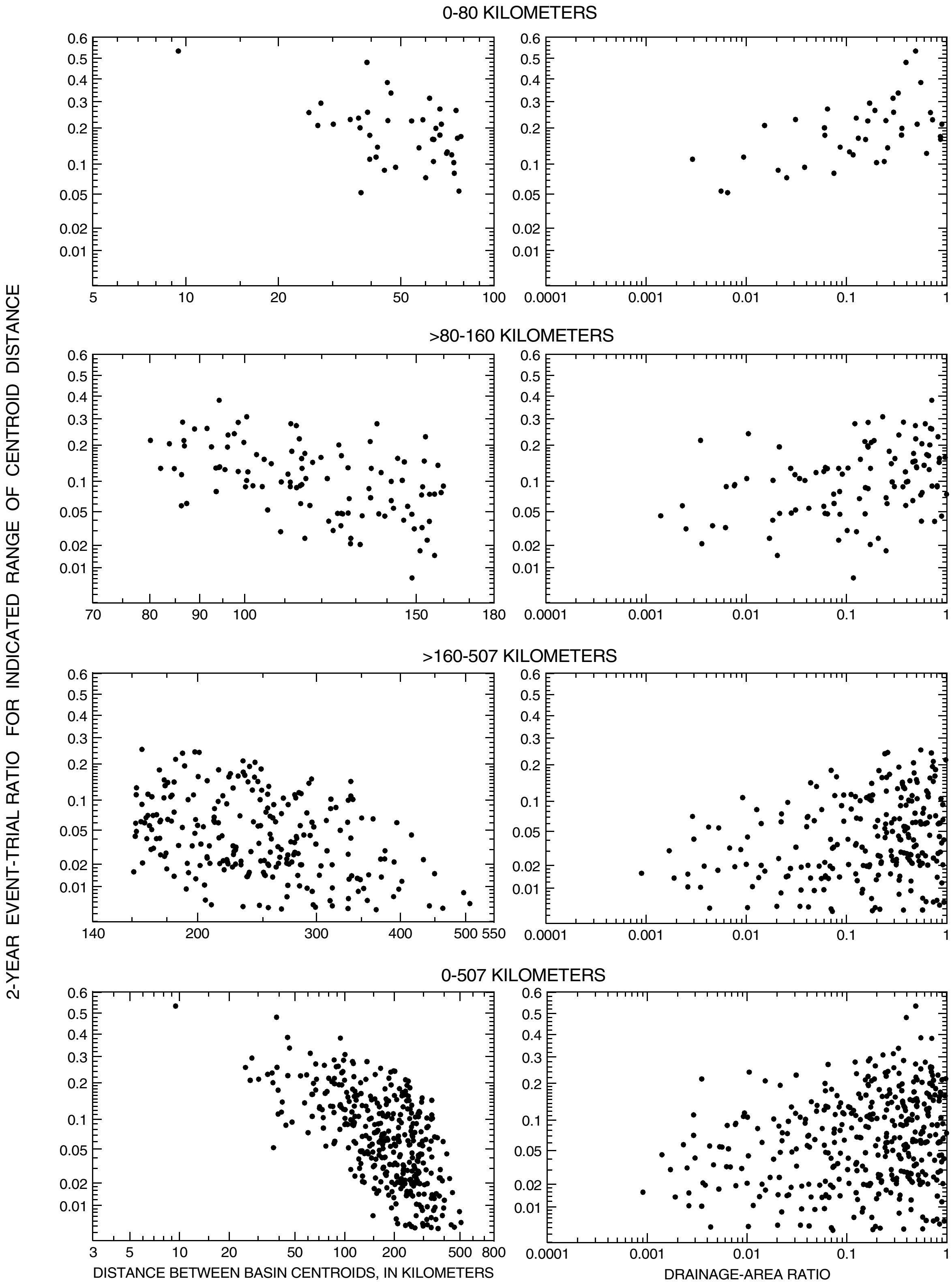

Figure 10. Event-trial ratios for the 2-year flood threshold plotted as a function of centroid distance and drainage-area ratio, for selected ranges of centroid distance among 30 sites in Ohio and border areas of adjacent states.

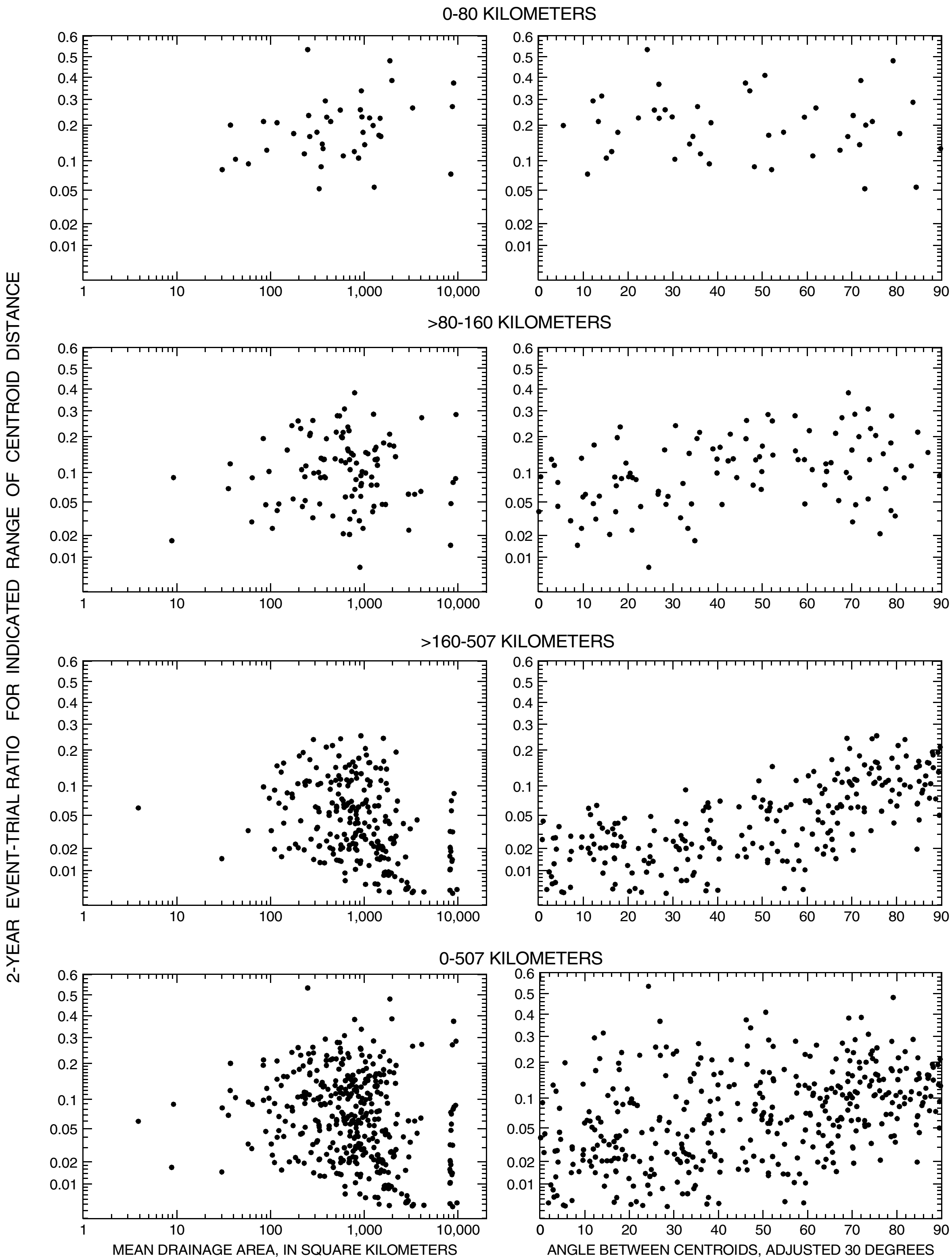

Figure 11. Event-trial ratios for the 2-year flood threshold plotted as a function of mean drainage area and adjusted centroid angle, for selected ranges of centroid distance among 30 sites in Ohio and border areas of adjacent states.

ever, they also provide added information about variability and linearity of the relations.

Logistic-regression analyses

Factors closely associated with the joint probability of flooding on paired streams have been identified and discussed in the previous sections. However, little direct information has been provided with which to estimate those probabilities. To that end, logistic-regression analyses were used in an attempt to develop empirical equations relating selected basin characteristics to the joint probability of flooding on paired streams in Ohio.

Logistic regression is a statistical procedure for analyzing probability data (or binary data) as a response variable. It is similar to ordinary least-squares regression in that the analysis results in the development of an equation for estimating values of the response variable (the logistic transformation of probability) from values of one or more explanatory variables. In this case, logistic-regression analysis was used to model the probability that the flood magnitudes equaled or exceeded the N-year flood-threshold level (where N = 2, 10, 25, and 50) concurrently at both streams in a pair, given that the flood magnitude at at least one of the streams exceeded the N-year flood-threshold level.

An event-trial form of the logistic-regression model was used, where the response variable was the logistic transformation of the ratio of events to trials. As discussed earlier, a trial is defined as the condition where the flood magnitude at at least one stream in a pair equaled or exceeded the flood threshold, and an event is defined as the condition where the flood magnitudes at both streams in the pair equaled or exceeded their flood thresholds concurrently (on the same day).

Logistic-regression analyses were performed using the logistic transformation of observed event-trial ratios as the response variable and the basin characteristics listed on page 16 as the explanatory variables. Numerous combinations and transformations of the basin characteristics were tested as potential explanatory variables. The combination of explanatory variables that provided the best fit between the observed and predicted response variable was ultimately selected subject to the following constraints: (1) the inclusion of each explanatory variable was hydrologically valid, (2) the regression coefficients associated with each explanatory variable were statistically significant at the 5-percent level, and (3) regression diagnostics did not reveal any problems that would potentially invalidate the model.

The Statistical Analysis System (SAS Institute, 1989) was used to perform the logistic-regression analyses. The general form of the linear logistic-regression equation is

$$\ln(P_N/(1-P_N)) = a + b(B) + c(C) + d(D) + e(E) + \ldots, \quad (3)$$

where

$\ln(P_N/(1-P_N))$ is the logistic transformation of P_N,

P_N is the event-trial ratio for the N-year flood threshold (the observed joint probability of flooding),

N is the recurrence interval, in years (2, 10, 25, or 50),

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

B, C, D, E are basin characteristics (explanatory variables).

More detailed information on logistic regression can be found in Collett (1991) and Hosmer and Lemeshow (1989).

Initial logistic-regression analyses were conducted using event-trial ratios determined for the 2-year flood thresholds (P_2) as the response variable because of the strength of its association with several basin characteristics in the correlation and graphical analyses. A preliminary logistic-regression equation (based on the full data set of 430 stream pairs) was developed in which the basin characteristics *CD*, *DAR*, *MDA* and *ACA30* were found to be statistically significant (p-values <0.05). Model fit was evaluated on the basis of comparison of the -2 log-likelihood criterion and by plotting observed values of P_2 against values of P_2 estimated by use of the equations. Some improvement in fit was subsequently achieved by use of natural logarithmic transformations of *CD*, *DAR*, *MDA*, and *ACA30*.

All equations were checked for parametrical bias by plotting **standardized deviance residuals** against each of the explanatory variables and the **linear predictor**. For an equation to be free of parametrical bias, the standardized deviance residuals should not be autocorrelated and should exhibit uniform variance throughout the range of explanatory variables and the linear predictor. Residual plots prepared for the initial regression model showed indications of bias with respect to the linear predictor for the total data set (centroid distance = 0-507 km) and bias with respect to the linear predictor and all of the explanatory variables (*CD*, *DAR*, *MDA*, and *ACA30*) for stream pairs with centroid distances in the 0-80 km range. Because the relations between P_2 and *CD*, *DAR*, *MDA*, and *ACA30* vary as a function of centroid distance, the data set was divided into subsets based on range of centroid distance, and the subsets were tested in further logistic-regression analyses. A logistic-regression equation based on the 304 stream pairs with centroid distances in the 0-260 km range provided negligible parametrical bias for stream pairs in the >80-260 km range; however, parametrical bias was still present for stream pairs in the 0-80 km range. A second logistic-regression equation based on the 41 stream pairs with centroid distances in the 0-80 km range provided negligible parametrical bias for stream pairs in the 0-80 km range.

The following equations, determined by means of logistic regression analyses, can be used to estimate the joint probability of flooding at the 2-year flood threshold level on unregulated paired streams in Ohio with centroid distances in the range of 0-260 km:

$$\hat{P}_2 = \frac{A}{(1 + A)} \quad (4)$$

where A is equal to the logistic transformation of the 2-year event-trial ratio and,

for stream pairs with $0 < CD \leq 80$ km,

$$A = 0.98(CD)^{-1.00}(DAR)^{0.25}(MDA)^{0.29}(ACA30)^{0.30}\ ; \quad (5)$$

for stream pairs with $80 < CD \leq 260$ km,

$$A = 0.89(CD)^{-1.00}(DAR)^{0.25}(MDA)^{0.15}(ACA30)^{0.61}\ ; \quad (6)$$

and where

$\hat{P}_2$ is the estimated joint probability of flooding at the 2-year flood threshold (estimated 2-year event-trial ratio),

CD is the centroid distance (in kilometers),

DAR is the drainage area ratio (unitless, on a scale from zero to one),

MDA is the mean drainage area (in square kilometers), and

ACA30 is the centroid angle adjusted 30 degrees (in degrees).

Figure 12 shows scatterplots of the 2-year event-trial ratios (the observed probabilities of joint flooding) and estimated probabilities of joint flooding determined from equations 4-6. Summary statistics of the observed and estimated 2-year event-trial ratios, as well as selected statistics describing model fit, are listed in table 7. The minimum, maximum, and mean values of the estimated probabilities are of comparable magnitudes to the minimum, maximum, and mean values of the observed probabilities for both ranges of centroid distance. The table also indicates that the magnitudes of the average absolute, average positive, and average negative simple residuals (observed minus estimated values of P_2) are comparable for both ranges of centroid distance.

Logistic regression analyses were also performed using the 10-, 25-, and 50-year event-trial ratios as response variables; however, satisfactory equations could not be developed because of the large number of zero-valued event-trial ratios and the extreme variability of the few non-zero event-trial ratios present in those data sets. Those analyses did indicate, however, that most of explanatory variables that were statistically significant in the 2-year flood-threshold regression analyses were also influential in the logistic regression analyses for the 10-, 25-, and 50-year flood-threshold data sets. Additionally, the signs of the regression coefficients in these higher recurrence interval analyses were unchanged from the 2-year logistic-regression analyses, and the magnitudes of the regression coefficients were comparable. Although not conclusive, these results indicate that factors associated with the joint probability of flooding on

Table 7. Summary statistics of observed and estimated 2-year event-trial ratios

[km, kilometers]

Observed event-trial ratio			Equation number for estimating event-trial ratios	Range of centroid distance (km) for equation	Estimated event-trial ratio			Mean of absolute values of observed minus estimated values	Mean of negative values of observed minus estimated values	Mean of positive values of observed minus estimated values
Minimum	Maximum	Mean			Minimum	Maximum	Mean			
0.052	0.535	0.199	5	0-80	0.058	0.557	0.200	0.039	-0.043	0.036
.000	.535	.103	6	0-260	.001	.557	.100	.040	-.032	.052

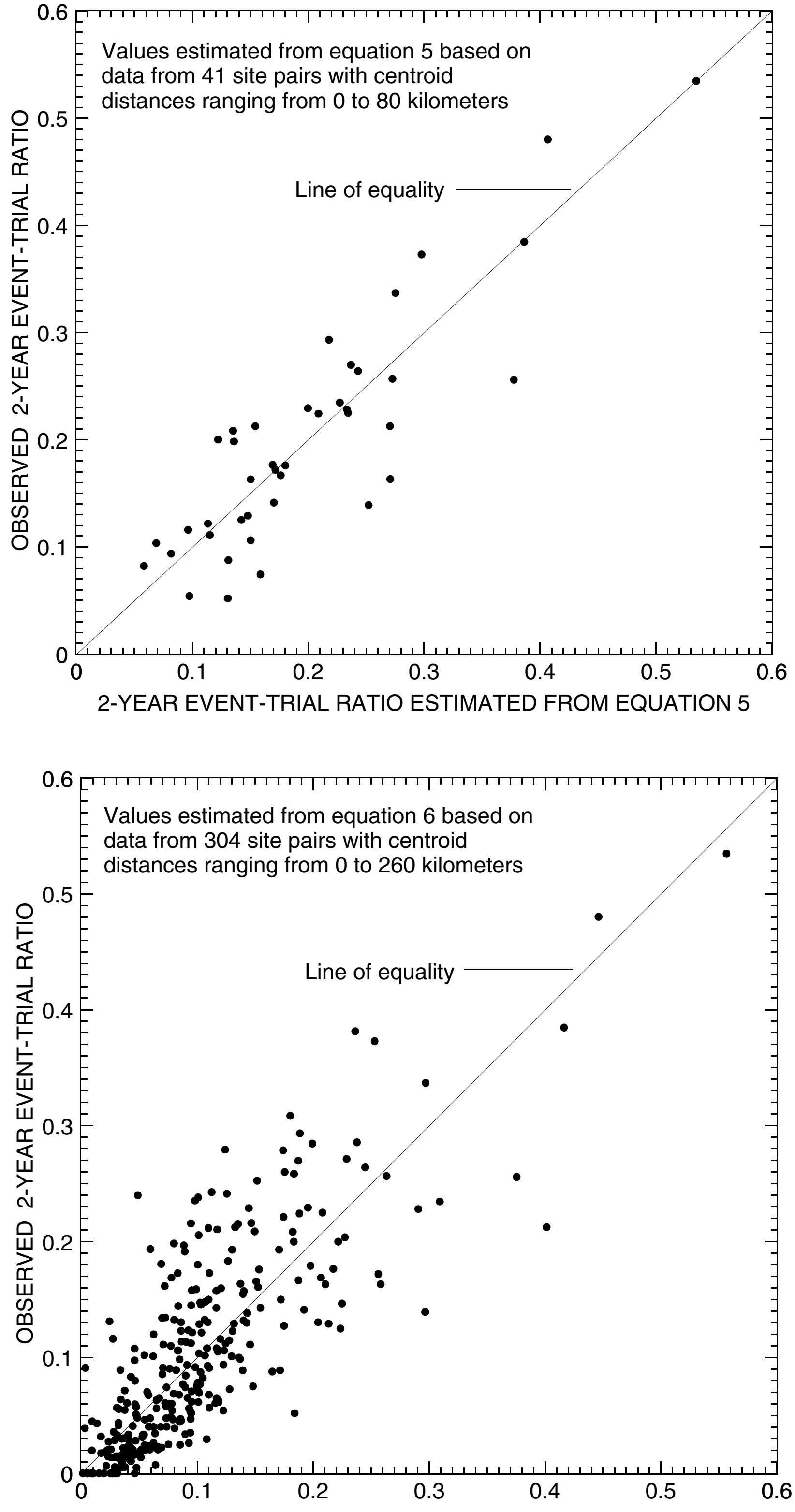

Figure 12. Observed 2-year event-trial ratios plotted as a function of the estimated 2-year event-trial ratios for 30 sites in Ohio and border areas of adjacent states.

paired streams at the 2-year flood-threshold level are similarly associated with the joint probability of flooding on paired streams at higher recurrence interval levels.

Association between season and the joint probability of flooding

The association between season and the joint probability of flooding was not tested directly in the logistic-regression analyses; however, it was explored in separate analyses. The frequency of annual peaks and annual plus partial peaks occurring in each calendar month was determined for each stream and for an aggregate of all 30 streams. When data for all streams were analyzed in aggregate, the month of March was found to have the greatest numbers of both annual and partial peaks and October, the smallest numbers. Stream-specific frequency analyses showed that the largest number of annual peaks occurred in March at two-thirds of the streams. At the remaining streams, the largest number of annual peaks occurred in January, February, or April. The fewest number of annual peaks occurred in October at half of the streams, with the remaining streams having the fewest number of annual peaks in May, June, July, August, September, November, or December.

Streams whose drainage areas are greater than 2,500 km^2 generally had few or no annual peaks between June and November, whereas the number of annual peaks between June and November was slightly greater for streams whose drainage areas are less than 2,500 km^2. A few streams with drainage areas less than about 450 km^2 had several annual peaks in July and August. These results can be explained by the fact that convective storms, characterized by local intense rainfall, are most common in Ohio during the summer months, whereas frontal storms, characterized by widespread moderate rainfall, are most common during the late winter and spring months. Convective storms and frontal storms can cause both large flood events on small streams, with most annual peak streamflows resulting from frontal storms. However, because of the limited areal extent of convective storms, large floods on large streams are caused almost exclusively by frontal storms. A consequence of this finding is that some seasonality might be expected in the joint probability of flooding when the stream pairing involves one stream with a large drainage area and a second stream with a small drainage area.

The frequency data were also examined for areal trends, but none were evident.

Bivariate flood-frequency distributions

The previous analyses dealt with the probability of two streams having concurrent floods of a magnitude that equaled or exceeded some flood threshold level, given that at least one of the two streams had a flood that equaled or exceeded the same flood threshold level. In those analyses, there was no consideration as to how much the flood threshold was exceeded by, but only that it was exceeded. Consequently, contingency tables were prepared to provide more information about the bivariate distribution of floods for paired streams.

The following discrete ranges of flood-threshold recurrence intervals were used to construct the contingency tables: <2 year, 2 to <10 year, 10 to <25 year, 25 to <50 year, and ≥50 year. Contingency tables were developed from the discretized streamflow data for the 430 stream pairs aggregated on three ranges of centroid distance: 0 to 80 km, >80 to 160 km, and > 160 km (table 8).

The contingency tables show, for three ranges of centroid distance, the percentage of trials at each level of aggregation for which concurrent streamflows fell within selected recurrence-interval ranges. For example, one of the cells of the contingency table shows the percentage of time that one stream in the pair had a daily mean streamflow (or instantaneous peak streamflow) corresponding to a flood peak with an associated recurrence interval in the 2- to <10-year range when a second stream in the pair had a concurrent daily mean streamflow (or instantaneous peak streamflow) corresponding to a flood peak with an associated recurrence interval in the 10- to <50-year range. Table cells that lie along the main diagonals of the contingency tables (shaded in gray) correspond to the condition where both streams in the pair concurrently flood in identical recurrence-interval ranges. No data are reported for the upper left main-diagonal cells because those cells represent the condition where streamflows at both streams in the pair were smaller than the 2-year flood threshold level (and consequently were filtered from the data set).

Two important conclusions can be drawn from table 8. First, the percentage of trials that correspond to the condition where both streams in the pair concurrently flood at identical recurrence-interval ranges (the shaded main diagonal cells) generally decreased as centroid distances increased. For instance, the main diagonal cells comprise about 18 percent of the trials for stream pairs with centroid distances ≤80 km, whereas the main diagonal cells comprise only about 3 percent of the trials for stream pairs with centroid distances greater than 160 km.

A second conclusion that can be drawn is that as centroid distance increased, streamflow at one stream in the pair was more likely to be in the <2-year recurrence-interval range when streamflow at the second stream was in a 2-year or greater recurrence-interval range. For example, in only about 4.5 percent of the trials did streamflows at both sites concurrently exceed the 2-year flood threshold when centroid distances were greater than 160 km. By comparison, streamflows at both sites in the pair concurrently exceeded the 2-year flood threshold in about 23.8 percent of the trials for site pairs with basin centroids that were within 80 km of one another.

Contingency tables (not shown) were also constructed for stream pair data aggregated on two ranges of the

Table 8. Relative frequencies of days, aggregated by range of centroid distance, in which concurrent streamflows at paired sites fell within selected recurrence-interval ranges given that the 2-year flood threshold was equaled or exceeded at at least one site in the pair

[ND, not determined; RI, recurrence interval; n, number of observations]

RI range for one site in pair (years)	Percentage of total trials for indicated centroid-distance range where sites flooded concurrently at the indicated flood-threshold RI ranges					Row sum (percent)
	RI range for second site in pair (years)					
	<2	2 to <10	10 to <25	25 to <50	≥50	
Centroid distance = 0 to 80 km (n=4,584)						
<2	ND	71.44	3.54	0.41	0.81	76.19
2 to <10		17.28	3.40	0.55	0.72	21.95
10 to <25			0.50	0.37	0.46	1.33
25 to <50				0.07	0.13	0.20
≥50					0.33	0.33
Centroid distance = >80 to 160 km (n=10,236)						
<2	ND	80.09	5.98	1.04	1.55	88.65
2 to <10		7.64	1.96	0.40	0.67	10.67
10 to <25			0.21	0.16	0.19	0.56
25 to <50				0.04	0.03	0.07
≥50					0.05	0.05
Centroid distance = >160 km (n=32,238)						
<2	ND	85.85	6.84	1.33	1.49	95.50
2 to <10		3.23	0.64	0.17	0.21	4.25
10 to <25			0.05	0.05	0.05	0.15
25 to <50				0.01	0.03	0.04
≥50					0.06	0.06

adjusted centroid angle (*ACA30*) and two ranges of drainage-area ratio (*DAR*). Ranges of *ACA30* used for aggregation were *ACA30* < 60 degrees and *ACA30* ≥ 60 degrees. The percentage of trials that correspond to the conditions where both streams in the pair concurrently flood in identical recurrence-interval ranges (the main diagonal cells) was about 9 percent for stream pairs with *ACA30* values greater than or equal to 60 degrees as compared to 4 percent for stream pairs with *ACA30* values less than 60 degrees. Ranges of *DAR* used for aggregation were *DAR* < 0.01 and *DAR* ≥ 0.01. For this aggregation scheme, the percentage of trials that correspond to the condition where both streams in the pair concurrently flood in identical recurrence-interval ranges was about 6 percent for stream pairs with *DAR* values greater than or equal to 0.01 as compared to 2 percent for stream pairs with *DAR* values less than 0.01.

ESTIMATION OF THE JOINT PROBABILITY OF FLOODING ON UNGAGED PAIRED STREAMS

Equations 4-6 (page 23) can be used to estimate the probability that streamflows at two ungaged streams concurrently equal or exceed the 2-year flood threshold given that the streamflow at at least one of the two streams equals or exceeds the 2-year flood threshold. The equations are applicable to stream pairs in Ohio (and border areas of adjacent states) that are unregulated, free of significant urban influences, and similar in basin characteristics to those of the gaged stream pairs used in the logistic-regression analyses (as summarized in tables 9 and 10). In general, drainage areas having usable storage of less than 49,100 m^3/km^2 are considered to be unregulated; however, the streamflow for a site directly below a large reservoir could be considered to be regulated regardless of the usable storage criterion (Benson, 1962).

Table 9. Ranges of the explanatory variables used in the logistic-regression equations

[CD, centroid distance; DAR, drainage area ratio; MDA, mean drainage area; ACA30, centroid angle adjusted 30 degrees]

Variable	Minimum	Maximum	Units
Equation 5 (0 - 80 kilometers)			
CD	9.46	78.2	kilometers
DAR	0.003	0.899	scale from 0 to 1
MDA	30.5	9000	square kilometers
ACA30	5.52	89.8	degrees
Equation 6 (80 - 260 kilometers)			
CD	9.46	259	kilometers
DAR	0.0002	0.998	scale from 0 to 1
MDA	3.87	9550	square kilometers
ACA30	.0600	89.8	degrees

Table 10. Ranges of four selected basin characterisitics of the 30 study sites in Ohio and border areas of adjacent states

Basin characteristic	Minimum	Maximum	Units
Drainage area	3.50	16,395	square kilometers
Main-channel slope	0.24	26.5	meters per kilometer
Basin storage	.00	4.00	percent
Forested area	.10	96.0	percent

For illustrative purposes, the joint probability of flooding will be estimated for one of the 430 pairs of gaged streams used in this study. The following table summarizes some basic information about the streams being paired:

Characteristic	Name of streamflow-gaging station	
	Maumee River at Antwerp, Ohio	**Tiffin River at Stryker, Ohio**
Station number	04183500	04185000
Station latitude	41°11'56"	41°30'16"
Station longitude	84°44'40"	84°25'47"
Drainage area	5,514 km^2	1,062 km^2
Centroid latitude	41°08'03"	41°46'16"
Centroid longitude	84°47'32"	84°20'12"

The basin-characteristic variables needed for use in the equations are the following:

CD = centroid distance = 75.4 km,

DAR = drainage area ratio (smaller/larger)
= 1,062/5,514 = 0.193,

MDA = mean drainage area
= (1,062 + 5,514)/2 = 3,288 km^2, and

CA = centroid angle = 32°

Because the centroid angle is less than 60°, equation 1 (on page 17) is applied to compute the absolute value of the centroid angle adjusted 30° (*ACA30*):

$$ACA30 = \text{abs}\,(CA+30^{o}), \text{ where } CA \leq 60^{o}$$

$$ACA30 = \text{abs}\,(32^{o}+30^{o})$$

$$ACA30 = \text{abs}\,(62^{o})$$

$$ACA30 = 62^{o}$$

Because the centroid distance is between 0 and 80 km, equation 5 (on page 23) is applied:

$$A = 0.98\,(CD)^{-1.00}\,(DAR)^{0.25}\,(MDA)^{0.29}\,(ACA30)^{0.30}$$

$$A = 0.98\,(75.4)^{-1.00}\,(0.193)^{0.25}\,(3{,}288)^{0.29}\,(62)^{0.30}$$

$$A = 0.311$$

Equation 4 (on page 23) is then used as follows to obtain the estimated probability of both streams concurrently equaling or exceeding their respective 2-year flood thresholds ($\hat{P}_2$):

$$\hat{P}_2 = \frac{A}{(1+A)}$$

$$\hat{P}_2 = \frac{0.311}{(1+0.311)}$$

$$\hat{P}_2 = 0.24$$

The observed 2-year event-trial ratio for this stream pair is 0.26. In comparison to the other stream pairs, which were either confluent or shared a common topographic divide, the observed value (0.26) for the example stream pair ranked 7 from the highest out of 14.

SUMMARY AND CONCLUSIONS

The factors related to the joint probability of flooding on paired streams were investigated and quantified to provide information to aid the design of hydraulic structures where the joint probability of flooding is an element of the design criteria. The data used for the study consisted of instantaneous peak and daily mean streamflow data and basin-characteristics data for 30 streamflow-gaging stations in Ohio and border areas of adjacent states. Drainage areas of the 30 gaging stations ranged from 3.50 to 16,395 km^2, and lengths of daily streamflow record ranged from 27 to 80 years.

Regression and graphical analyses were used to estimate daily mean streamflows from instantaneous peak streamflows for each stream. This was required so that the daily mean streamflow data could be used as a substitute for instantaneous peak streamflow data in order to determine which flood thresholds (corresponding to the 2-, 10-, 25-, or 50-year instantaneous peak streamflow) were equaled or exceeded on any given day.

The daily mean streamflow data for each stream were paired with concurrent daily mean streamflow data at the other 29 streams, yielding 435 stream pairs (5 of which were later dropped from the working data set to minimize serial correlation effects). Observed probabilities of joint flooding, determined for the 2-, 10-, 25-, and 50-year flood thresholds, were computed as the ratios of the total number of days when streamflows at both streams equaled or exceeded their flood thresholds on the same day (events) to the total number of days when streamflows at either stream equaled or exceeded its flood threshold (trials). Instantaneous peak streamflow data, when available, were used preferentially to assess flood-threshold exceedance.

Smooth curves were overlaid on scatterplots showing 2-year event trial ratios against corresponding 10-, 25-, and 50-year event-trial ratio. The curves illustrate that the 2-year event-trial ratios and the 10-, 25-, and 50-year event-trial ratios tend to increase together and that the 2-year event-trial ratio tended to be larger than the corresponding 10-, 25-, and 50-year event-trial ratios.

A combination of correlation analyses, graphical analyses, and logistic-regression analyses were used to identify and quantify factors most closely associated with the observed probabilities of joint flooding (event-trial ratios) determined for the 2-, 10-, 25-, and 50-year flood thresholds. In all three analyses, centroid distance (*CD*), drainage area ratio (*DAR*), mean drainage area (*MDA*), and centroid angle adjusted 30 degrees (*ACA30*) were found to be closely associated with the joint probability of flooding on paired streams in Ohio. These analyses also showed that the observed probabilities of joint flooding tended to decrease with an increase in centroid distance and increase with increases in drainage area ratio, mean drainage area, and centroid angle adjusted 30 degrees.

A series of bar graphs was prepared to show the correlation between selected basin characteristics and event-trial ratios, grouped as a function of centroid distance and flood-threshold recurrence interval. Patterns observed in the bar graphs indicate that, for drainage areas whose centroids are relatively close to each other (0-80 km), the association between drainage area ratios and event-trial ratios is stronger and the association between adjusted centroid angle and event-trial ratios is weaker than comparable associations for drainage areas whose centroids are separated by greater distances.

Logistic-regression equations were developed using the 2-year event-trial ratio as the response variable and using the centroid distance, drainage area ratio, mean drainage area, and centroid angle adjusted 30 degrees as explanatory variables. The equations are applicable to stream pairs in Ohio (and border areas of adjacent states) that are unregulated, free of significant urban influences, and similar in basin characteristics to those of the gaged stream pairs used in the logistic-regression analyses.

Logistic-regression equations could not be developed for the 10-, 25-, and 50-year event-trial ratios, owing to the large number of zero-valued event-trial ratios and the extreme variability of the few non-zero event-trial ratios present in the data set.

The frequency of annual peaks and annual plus partial peaks occurring in each calendar month was determined for each stream and for an aggregate of all 30 streams. The month of March was found to have had the greatest numbers of both annual and partial peaks and October, the smallest numbers. Streams whose drainage areas are greater than 2,500 km^2 generally had few or no annual peaks between June and November, whereas the number of annual peaks between June and November was slightly greater for streams having drainage areas less than 2,500 km^2. Drainage-area-related differences in the seasonal occurrence of annual peak flows are surmised to be associated with the seasonal nature of convective storms as opposed to frontal storms and the relative effects of the two storm types on peak streamflows at sites with large rather than small drainage areas.

Contingency tables were constructed to provide more information about the bivariate distribution of floods on paired streams. Contingency tables aggregated on ranges of centroid distance indicated that the percentage of trials where both streams in the pair concurrently flood at identical recurrence-interval ranges generally decreased as centroid distances increased. The tables also indicated that as centroid distance increased, streamflow at one stream in the pair was more likely to be in the <2-year recurrence-interval range when streamflow at the second stream was in a 2-year or greater recurrence-interval range.

Contingency tables aggregated on two ranges of adjusted centroid angle and two ranges of drainage-area ratio indicated that the percentage of trials that correspond to the conditions where both streams in the pair concurrently flood in identical recurrence-interval ranges was greatest for aggregations of stream pairs with adjusted centroid angles greater than or equal to 60 degrees and drainage area ratios greater than or equal to 0.01.

SUGGESTIONS FOR FURTHER STUDY

Information on the joint probability of flooding is in greatest demand for streams that are confluent and (or) in close proximity to one another. The authors attempted to use as much data as possible from confluent streams; however, only a limited amount of such data are available for Ohio and border areas of adjacent states. Selection of streamflow-gaging stations from a broader geographic area would provide more data from confluent and (or) close proximity streams; however, doing so would likely not appreciably increase the average concurrent length of record and may result in the introduction of other factors that could confound our ability to make accurate predictions at any flood threshold.

The methods employed in this study consider floods to be concurrent if they occurred on the same calendar day. However, designers are generally interested in knowing about flood conditions at the instant in time when the stream of interest reaches its design peak flow. When considering pairs of streams with moderately large drainage areas, probability estimates based on the “same day” measure of flood concurrency are reasonable. When considering pairs of streams that involve one or more streams with smaller drainage areas, probability estimates based on the “same day” measure of flood concurrency may be of marginal utility because of potential differences in hydrograph timing. In this latter case, it would be preferable to use fixed-interval instantaneous streamflow information to assess concurrency of flooding. Given the recent advances in affordable data-storage media, the adoption of a program of systematic storage of fixed-interval instantaneous streamflow information would enable questions such as those considered in this report to be addressed more directly.

REFERENCES CITED

Benson, M.A., 1962, Factors influencing the occurrence of floods in a humid region of diverse terrain: U.S. Geological Survey Water-Supply Paper 1580-B, 64 p.

Choquette, A.F., 1987, Regionalization of peak discharges for streams in Kentucky: U.S. Geological Survey Water-Resources Investigations Report 87-4209, 105 p.

Collett, D., 1991, Modelling binary data: London, Chapman and Hall, 369 p.

Dempster, G.R., compiler, 1990, National water information system user’s manual, v. 2, chapter 3, Automated data processing system: U.S. Geological Survey Open-File Report 90-116 [variously paginated].

Dyhouse, G.R., 1985, Stage-frequency analysis at a major river junction: Journal of Hydraulic Engineering, v. 111, p. 565-583.

Flippo, H.N., Jr., 1977, Floods in Pennsylvania: Pennsylvania Department of Environmental Resources Bulletin 13, 59 p.

Fricke, T.J., Kennedy, M.R., and Wellington, N.B., 1983, The use of rainfall correlation in determining design storms for waterways on a long railway line: Hobart, Australia, Hydrology and Water Research Symposium, Proceedings, p. 215-219.

Glatfelter, D.R., 1984, Techniques for estimating magnitude and frequency of floods on streams in Indiana: U.S. Geological Survey Water-Resources Investigations Report 84-4134, 110 p.

Holtschlag, D.J., and Croskey, H.M., 1984, Statistical models for estimating flow characteristics of Michigan streams: U.S. Geological Survey Water-Resources Investigations Report 84-4207, 80 p.

Hosmer, D.W., and Lemeshow, S., 1989, Applied logistic regression: New York, John Wiley and Sons, 307 p.

Huff, F. A., and Angel, J. R., 1992, Rainfall frequency atlas of the Midwest: Illinois State Water Survey Bulletin 71, 141 p.

Hutchinson, N.E., compiler, 1975, WATSTORE, National water-data storage and retrieval system—User's guide, v. 1: U.S. Geological Survey Open-File Report 75-426, 505 p.

Jennings, M.E., Thomas, W.O., and Riggs, H.C., compilers, 1994, Nationwide summary of U.S. Geological Survey equations for estimating magnitude and frequency of floods for ungaged sites: U.S. Geological Survey Water-Resources Investigations Report 94-4002, 196 p.

Koltun, G.F., and Roberts, J.W., 1990, Techniques for estimating flood-peak discharges for rural, unregulated streams in Ohio: U.S. Geological Survey Water-Resources Investigations Report 89-4126, 68 p.

Morris, C.D., and Wilson, L.C., 1987, Concurrent flooding probabilities. *in* Singh, V.P., ed., Hydrologic frequency modeling: Dordrecht, The Netherlands, D. Reidel Publishing Company, p. 603-614.

Ohio Department of Transportation, 1992, Location and design manual: v. 2, p. 11-2.

Raynal, J.A., and Salas, J.D., 1987, A probabilistic model for flooding downstream of the junction of two rivers, *in* Singh, V.P., ed., Hydrologic frequency modeling: Dordrecht, The Netherlands, D. Reidel Publishing Company, p. 595-602.

Reinsch, C.H., 1967, Smoothing by spline functions: Numerische Mathematik, v. 10, p. 177-183.

Ribeny, F.M., 1971, On the chance of culvert washouts on a long railway line: Sydney, Australia, Institution of Engineers, Hydrology Papers, p. 55-59.

Runner, G.S., 1980, Runoff studies on small drainage areas—Technique for estimating magnitude and frequency of floods in West Virginia: U.S. Geological Survey Open-File Report 80-1218, 44 p.

SAS Institute Inc., 1989, SAS/STAT User's guide, version 6, fourth edition, volume 2: Cary, N.C., 846 p.

Thomas, D.M., and Benson, M.A., 1970, Generalization of streamflow characteristics from drainage-basin characteristics: U. S. Geological Survey Water-Supply Paper 1975, 55 p.

GLOSSARY

Adjusted centroid angle (ACA30) — ACA30 is the absolute value of the centroid angle where the zero-degree reference line has been adjusted counterclockwise by 30 degrees.

Annual peak streamflow — The maximum instantaneous streamflow at a streamflow-gaging station during the water year.

Centroid angle (CA) — The angle (in degrees) between a line connecting the centroids of the drainage areas of a pair of streams and a zero-degree true-north reference line.

Centroid distance (CD) — The distance (in kilometers) between the centroids of the drainage areas of a pair of streams.

Confluent streams — Two streams that flow together.

Daily mean streamflow — Time-weighted mean streamflow for a calendar day.

Discretized streamflow data — Streamflow data categorized into discrete ranges of recurrence interval.

Drainage area — The area of land, measured in a horizontal plane, that contributes surface runoff to a specified location on a stream.

Drainage area ratio (DAR) — The drainage area (contributing surface runoff) of the smaller of a pair of streams divided by the drainage area of the larger of a pair of streams.

Event —the condition when streamflows at both streams in a pair equal or exceed their respective N-year (where N = 2, 10, 25, or 50) flood thresholds on the same calendar day.

Event-trial ratio — The ratio of the total number of days when streamflows at *both* streams in a pair equaled or exceeded their N-year (where N = 2, 10, 25, or 50) flood thresholds on the same day (events) to the total number of days when streamflows at *either* stream in a pair equaled or exceeded their N-year flood threshold (trials). Also referred to as the observed probability of joint flooding.

Flood threshold — Instantaneous peak streamflows with recurrence intervals of 2-, 10-, 25-, and 50-years. Determinations of when flood thresholds were equaled or exceeded was based upon instantaneous peak streamflow data, when available, *or*, when instantaneous peak streamflow data were not available, by comparison with daily mean streamflow values estimated to occur in association with the 2-, 10-, 25-, and 50-year instantaneous peak floods.

Instantaneous streamflow — Streamflow at a particular instant in time.

Instantaneous peak streamflow — The instantaneous streamflow at the peak of a flood event. Annual peak and partial peak streamflows are specific categories of instantaneous peak streamflow.

Joint probability of flooding — The probability that streamflows at both streams in a pair equal or exceed their N-year (where N = 2, 10, 25, or 50) flood thresholds on the same calendar day.

Linear predictor — The estimated value of the linear systematic component of the logistic-regression model.

Mean drainage area (MDA) — The arithmetic mean of the drainage areas of a pair of streams (in square kilometers).

Missing — In the context of event-trial ratios, refers to the condition in which the flood threshold being considered was not exceeded at either stream in a pair. Missing values are not counted as observations and consequently have no effect on statistical measures such as summary statistics. In contrast, an event-trial ratio was 0.00 (a “non-missing” value) if the flood threshold being considered was exceeded at at least one stream in the pair but was never concurrently exceeded at both streams in the pair.

Non-missing — In the context of event-trial ratios, refers to the condition in which the flood threshold being considered was exceeded at either stream in the pair. Values for non-missing event-trial ratios range from 0.00 to 1.00.

N-year event-trial ratio — Event-trial ratios determined for an N-year flood threshold (where N = 2, 10, 25, or 50).

N-year flood threshold — Flood threshold with a recurrence interval of N years (where N = 2, 10, 25 or 50).

Paired streams — Any two streams considered together, regardless of their proximity to one another. Paired streams may or may not be confluent.

Partial peak streamflow — An instantaneous peak streamflow at a streamflow-gaging station during the water year, that is less than the annual peak streamflow yet greater than a base streamflow established for the gaging station.

R^2 value — A value, measured on a scale from 0 to 1, representing the fraction of variance in the dependent variable explained by the regression.

Regression residuals — Observed value minus the corresponding value predicted by a regression equation.

Spearman’s rho — measured on a scale from -1 to 1, a nonparametric measure of the strength of association between two continuous variables.

Standardized deviance residual — a measure of the difference between the observed and fitted values in a logistic regression analysis, standardized to have approximately unit variance.

Streamflow — The flow of water past a point on a stream, expressed in units of volume per time.

Trial — The condition when streamflow at either stream in a pair equaled or exceeded their N-year (where N = 2, 10, 25, or 50) flood threshold.

Water year — A water year is the 12-month period, October 1 through September 30, designated by the calendar year in which it ends. Thus, the 12-month period ending September 30, 1993, is called the "1993 water year."