Rengen
(ordered especially
for you!)

Form 9-014

UNITED STATES
DEPARTMENT OF THE INTERIOR
GEOLOGICAL SURVEY
WASHINGTON 25, D. C.

STREAMFLOW GENERALIZATION IN AN AREA OF THE CENTRAL UNITED STATES

I. C. James, II

Open-File Report

Lawrence, Kansas

March 1968

☆ U.S. GOVERNMENT PRINTING OFFICE: 1962—O-632210

UNITED STATES

DEPARTMENT OF THE INTERIOR

Robert J. Dingman, District Chief

STREAMFLOW GENERALIZATION

IN AN AREA OF THE CENTRAL UNITED STATES

I. C. James, II
Hydraulic Engineer

Open-File Report

Lawrence, Kansas

March 1968

CONTENTS

ILLUSTRATIONS

TABLES

SYMBOLS

A = Drainage area, in sq mi.

Aa = Alluvial area, in sq mi.

a = Constant in regression equation.

a (subscript) = Annual discharge.

$b_1, b_2, b_3 \ldots b_n$ = Regression coefficients.

Br = Bifurcation ratio.

Ch = Channel entrenchment depth.

D_p = Duration-curve index, i.e., flow exceeded p percent of days, in cfs.

d (subscript) = Duration in days.

El = Basin elevation, in ft, MSL.

Ev = Average annual evaporation from lakes and reservoirs, in inches.

F = Forested area, in percent of total drainage area, and increased by 1.00 percent.

G = Coefficient of skew.

Gr = Streambed grain size, in mm.

h (subscript) = Duration, in hours.

L = Main-channel length, in miles.

Lca = Mean flow length, in miles.

$M_{d,t}$ = Minimum annual d-day average flow expected on the average once each t years, in cfs.

Mr = Meander ratio.

n (subscript) = Number of month starting with January as 1.

P = Mean annual precipitation, in inches

$P_{t,h}$ = Annual maximum h hours precipitation, in inches, expected on average once in t years.

p (subscript) = Percentage of time during which flow is equalled or exceeded.

Q_a = Mean discharge, in cfs.

qm = Map value of mean discharge, in cubic feet per second per square mile.

Q_n = Mean monthly discharge for month n, in cfs.

Qp_t = Annual peak momentary discharge, in cfs, expected on average once in t years.

Rlm = Map value of ratio of 1 percent duration to mean flow.

SD = Standard deviation of flow, in cfs.

Si = Soils infiltration index, in inches.

Sl = Slope, in feet per mile.

Sn = Mean annual snowfall, in inches.

SE = Standard error, in percent.

SR = Serial correlation coefficient.

SW = Streambed width, in feet.

t (subscript) = Average recurrence interval, in years.

T_1 = Mean minimum January temperature, in °F.

T_7 = Mean maximum July temperature, in °F.

v = Map value of variability index, in log units.

$V_{d,t}$ = Maximum annual d-day average flow expected on average once each t years, in cfs.

Vw = Valley width, in miles

W = Basin width, in miles

X_1, X_2, $X_3 \ldots X_n$ = Basin characteristics.

Y_1, Y_2, $Y_3 \ldots Y_n$ = Streamflow characteristics

STREAMFLOW GENERALIZATION IN AN AREA OF THE CENTRAL UNITED STATES

By

Ivan C. James, II

ABSTRACT

Streamflow records of a sample of 41 natural flow stations in an area of the Central Plains were analyzed by electronic computer for the statistics of annual high-flow and low-flow volumes, flow duration, and monthly- and annual-flow discharges. Peak discharges and high- and low-flow volumes for different durations were computed at selected recurrence intervals from statistics of the distributions of the annual values. These dependent variables, discharges and statistics of discharges, were regressed against independent variables of topographic, geomorphic, climatic, and mapped-flow characteristics. The results provide a base for investigating the adequacy of the stream-gaging program and a set of relationships for estimating natural flow characteristics at ungaged sites in the region under study. Comparisons were made of the results of regressions which used only topographic and climatic independent variables with those which also used mapped-flow characteristics as independent variables. The standard error of estimate of the

low-flow characteristics was quite large, indicating that for this discharge range, other methods of obtaining this information may be more accurate or that a quantification of the effect of other factors such as geology on low flow is probably needed to explain part of the present error. Annual and monthly mean flows were defined with standard errors of 18 to 49 percent. Flood peaks and high-flow volumes were defined with standard errors of about 50 percent and 30 percent, respectively, where a significant portion, the residual error, is probably due to sampling error in the dependent variable caused by a short period of record.

INTRODUCTION

An important requirement for sustaining an advanced civilization in a semi-arid environment often has proved to be a water supply sufficient for the development of agriculture and industry. Little was known until recently about the frequency and time distribution of streamflows in the central United States, although the first collection of discharge records in the central plains region began in the 1890's.

Some of these earlier records were published as monthly mean discharges based on daily-stage observations and three or four discharge measurements per year. The information was intended to be used only for the stream on which it was collected and was not considered to have regional value. As water-resources development became more intensive, knowledge of flows on smaller streams was needed, and the stream-gaging system which obtained point information at only a limited number of sites was inadequate to meet the needs. Economics prevented the collection of data at all stream sites of interest, and hydrologists began using analytical methods to transfer streamflow information from one site to another by simple correlation of records, drainage area adjustments, and graphical regression of flow characteristics on physical and climatic characteristics. Thus were developed the needs for a hydrologic-gaging program and a method of generalizing the information in such a manner that it could be used at any site.

Of the methods for generalizing streamflow, multiple regression seems to hold the most promise. The purpose of this study has been to determine the degree of success obtainable by regression as a streamflow generalization tool in an area of the central United States, and to determine the adequacy of the present stream-gaging program, using multiple regression analysis to define natural streamflow characteristics in general.

Symbols are defined on pages 3 and 4. Descriptions of independent variables are in appendix I on pages 58 and 59. A glossary of terms as used in this report is included on page 60.

ACKNOWLEDGMENTS

This report was prepared as a part of a study of streamflow generalization in four pilot areas--Potomac River basin, Louisiana, Central Valley in California, and the area of this report, the Central United States--under the supervision of M. A. Benson and D. M. Thomas, Research Hydrologist and Hydraulic Engineer, Hydrologic Studies Section, Washington, D. C. The report was prepared in the Lawrence, Kansas, district office under the immediate supervision of L. W. Furness, Chief, Hydrologic Studies Section. Special acknowledgment is given to M. A. Benson who initiated this project, supervised its execution, and reviewed the report.

The four computer programs used in analyzing data for this report were written in the Washington office of the U.S. Geological Survey, and computations were performed on the Survey's Burroughs B-220 digital computer.

DESCRIPTION OF STUDY AREA

The study area is located in the Great Plains and Central Lowlands provinces of the Interior Plains. The streamflow sites for which data were analyzed for this report are shown in figure 1 and are mostly in the lower Missouri River basin in Kansas, eastern Colorado, southern Nebraska, and western Missouri. A few sites in the Arkansas River basin bordering on the lower Missouri River basin also were used.

The land surface in the area is flat or gently rolling. The climate of the study area is typical of large continents in middle latitudes. Average annual rainfall ranges from 16 in. in the western part of the study area to 39 in. in the southeastern part. In general, about three-fourths of the annual precipitation falls in the 6-month growing season, April to September. Late May and early June is usually the wettest period of the year although some of the heaviest daily storms have occurred in September as a result of moist gulf air masses from hurricanes meeting cool fronts from the arctic regions. The mean annual temperature for Kansas is 55°F and is slightly less for the study area. Temperatures in Kansas have ranged from 121°F to -40°F. Weather systems in this region can change quite rapidly giving the area what has been called a "vigorous" climate.

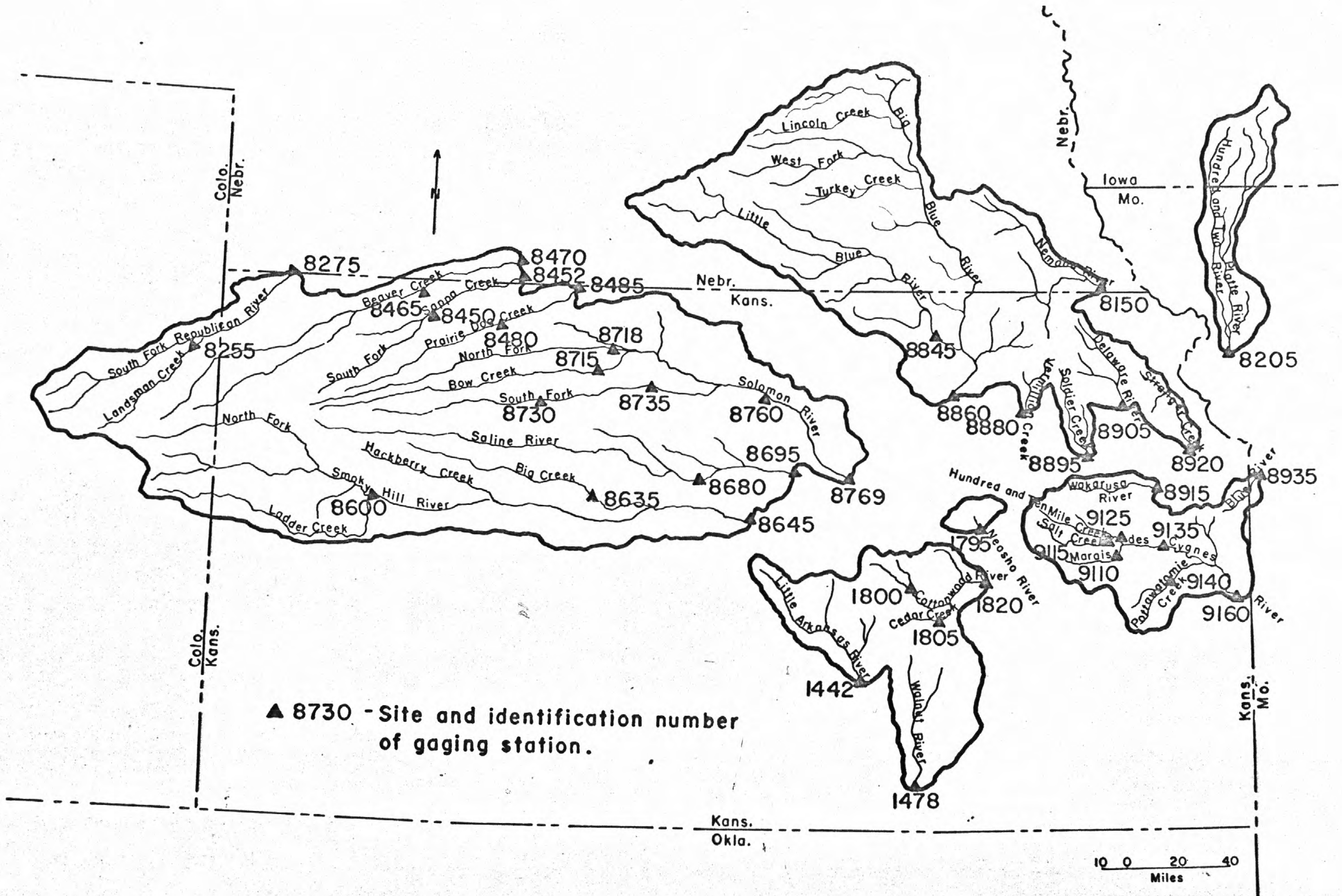

Figure 1.--Locations of gaging stations for which streamflow records were analyzed.

Stream discharge in the study area is normally high in May, June, and early July, decreasing rapidly to very low flow in late July, August, September, and October except for response to intermittent rains. Usually a killing frost in October causes an increase of base flow, and moderate low-flow discharges continue with ice effect and some snowmelt runoff until the first heavy spring rains. The 90-percent duration of flow is zero for about a third of the stations used in the study, and the 10-year, 7-day low flow is zero for about half the stations.

METHOD OF ANALYSIS

The analysis consisted of collecting characteristics of natural streamflow and independent basin variables at 41 streamflow sites. Streamflow characteristics were regressed against the selected independent variables to define equations which may be used to estimate streamflow characteristics at any unregulated stream site in the study area.

Selection of Streamflow Records for Analysis

Water-resources development in the study area has interrupted the collection of natural streamflow records at many sites. This leaves only a moderate number of long records which are not affected by regulation, diversion, or man-made change in the natural hydrology of the basin. The stations selected for analysis contain the longest records of streamflow not appreciably affected by regulation or diversion. Changes in land use with time, man-made changes in the hydraulic characteristics of the natural channel, and changes in small reservoir or farm pond storage usually were considered as not materially affecting the streamflow. Selection of records was further based upon achieving a distribution of independent variables within a subregion. This usually involved a trade-off such as when records from smaller drainage areas were used with a resultant loss in accuracy of streamflow characteristics derived from the short records available on these smaller drainage areas.

The length of record used ranged from 12 to 61 years and averaged 31 years. The period of unregulated record was used at each station with some records starting in the 1895 water year and many ending with the 1963 or 1964 water year. No attempt was made to select streamflow records covering a standard base period nor to extend records to a base period.

Collection of Dependent Variables

Automatic data-processing equipment was used to compute the streamflow characteristics used as dependent variables. Three computer programs were used in this analysis. The first program used daily discharges stored on magnetic tape to compute percent of time selected discharges were exceeded and the lowest and highest mean discharges in each year for durations of 1, 3, 7, 14, 30, 60, 90, 150, 183, and 274 days.

The second program fitted a Pearson type III distribution (defined by mean, standard deviation, and skew) to the annual peaks and to annual highest mean discharges of selected durations, and from these statistics computed the highest mean discharges for the selected durations at specific recurrence intervals. Annual low-flow discharges of selected durations were also computed by this program but could not be accurately fitted by the Pearson type III distribution if there were many values of zero flow. For consistency, all low-flow frequency curves were plotted manually.

The third program used daily discharges on magnetic tape, as input, to compute the means, standard deviations, skew coefficients, and serial correlation coefficients of the monthly and annual discharges.

Selection of Independent Variables

The 25 independent variables selected for analysis can be classed as either topographic, geomorphic, climatic, or mapped-flow characteristics. The distinctions between topographic and geomorphic terms are often arbitrary, and accordingly are combined herein under the general term "physical characteristics". Of the 25 independent variables tried, 15 can be classified as physical, seven as climatic, and three as mapped-flow characteristics. Table 1 gives the maximum, minimum, and median of the values of the independent variables for the 41 stations used in this report. Appendix I gives a definition or description of the method of collection of the independent variables.

Physical Characteristics

The topographic variables selected for this study are indexes of catchment size, shape, drainage efficiency, and location. The most common index of catchment size is drainage area. Main channel length and basin width were also used together as an alternate index of this same factor. For low flows, the emphasis should be on ground-water storage capacity, and alluvial area was investigated as the factor. Slope, mean-flow length, and meander ratio are basin topographic characteristics which were investigated as indexes of drainage efficiency. Forested area was examined as a variable because it may significantly alter the rate of direct runoff and affect the transpiration loss in the basin. Although not used in the final equations, basin elevation has been used in other study areas both as an index of precipitation anomoly and as an index of the variations in factors which affect streamflow such as slope, evaporation, and temperature.

Table 1.--Range in magnitude of independent variables at stations used in this report.

Characteristics	Symbol	Maximum	Median	Minimum
Physical				
Drainage area	A	9,100	1,250	110
Main channel length	L	381	115	14.1
Basin width	W	34.3	8.09	2.93
Alluvial area	Aa	934	103	3.70
Slope	Sl	17.5	7.50	2.22
Mean flow length	Lca	204	58.0	9.6
Meander ratio	Mr	1.71	1.42	1.02
Forrested area	F	17.6	2.28	1.00
Basin elevation	El	4,215	1,475	885
Bifurcation ratio	Br	5.78	4.57	4.06
Streambed width	SW	250	45	10
Channel entrenchment depth	Ch	54	21	5
Valley width	Vw	3.0	1.0	.3
Streambed grain size	Gr	75	.78	.008
Soil infiltration index	Si	7.6	2.90	1.90
Climatic				
Mean annual precipitation	P	38.76	26.0	15.80
2-year 24-hour precipitation	$P_{2,24}$	3.70	2.91	2.12
100-year 24-hour precipitation	$P_{100,24}$	8.16	6.61	5.04
Mean minimum January temperature	T_1	23	17	14
Mean maximum July temperature	T_7	95.9	93.5	90.4
Mean annual snowfall	Sn	32.2	20.6	13.9
Average annual evaporation	Ev	64	55	40
Mapped flow				
Mean discharge	qm	.574	.170	.007
Variability index	v	1.20	.80	.40
Ratio of 1 percent duration to mean flow	Rlm	19.9	17.2	11.7

Geomorphic variables usually are considered to be both a result and a cause of hydrologic variations. Drainage density is an excellent example. A high drainage density is a result of high precipitation and low infiltration capacity and also is a cause of rapid peaking capabilities of a stream. Drainage densities are extremely tedious to collect; consequently, <u>bifurcation ratio</u> was substituted as a variable indicating degree of drainage development. <u>Streambed width</u> and <u>channel entrenchment depth</u> were tried as variables representing the effect of discharge on channel morphology. <u>Valley width</u> and <u>streambed grain size</u> were studied as possibly related to the rate of ground-water drainage from the alluvium. <u>Soil infiltration index</u> may be related to basin differences in recharge and storm runoff.

Climatic Characteristics

The selected climatic variables described the amount and distribution of precipitation and the water losses to evaporation. Mean annual precipitation was used as an indication of the total supply of water available to appear as streamflow. It is also highly indicative of soil moisture and storm precipitation. The 2-year 24-hour precipitation and 100-year 24-hour precipitation that were tried are measures of the amount of storm rainfall available to contribute to flood peaks. Average minimum January temperature (fig. 2) has been used in other parts of the United States as a factor relating to the potential for snowpack accumulation. On the average, in this study area, runoff events resulting from snowpack melt are minor in comparison to other runoff events and the effects of this variable may not be obvious. Average maximum July temperature (fig. 3) is another factor the influence of which is not fully understood but probably is related to the potential for evapotranspiration. Average annual snowfall (fig. 4), in areas where snow accounts for a significant portion of the total precipitation, is related to the magnitude of the spring runoff. In areas of lesser snowfall, the term may represent the temporal distribution of rainfall or may serve as a loss factor in areas where a large part of the snowfall is sublimated or evaporated. Average annual evaporation is expected to be an index representing losses.

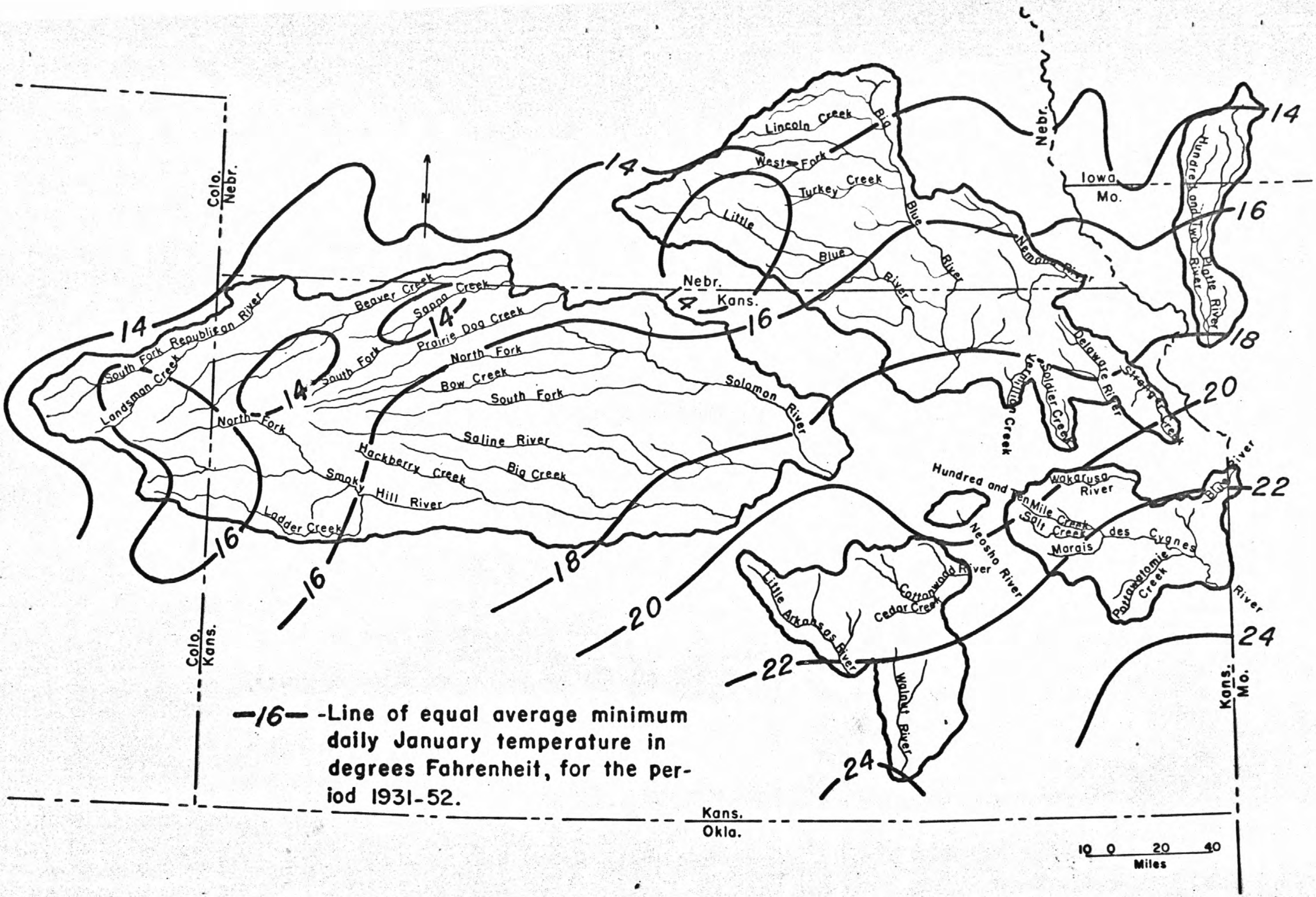

Figure 2.--Average minimum daily January temperatures in the Central United States study area. (Modified after U.S. Department of Commerce, Weather Bureau, 1959a-e)

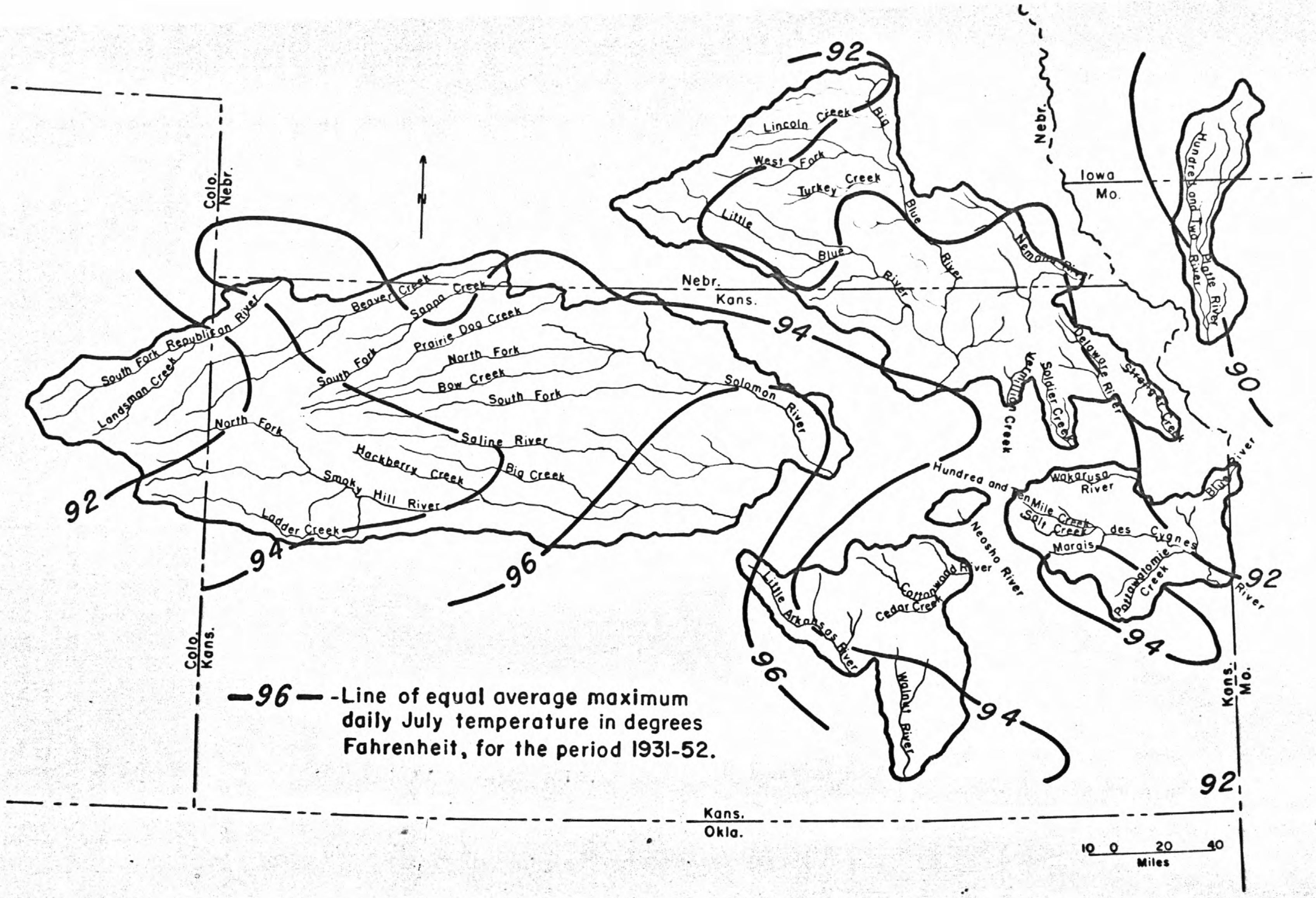

Figure 3.--Average maximum daily July temperatures in the Central United States study area. (Modified after U.S. Department of Commerce, Weather Bureau, 1959a-e)

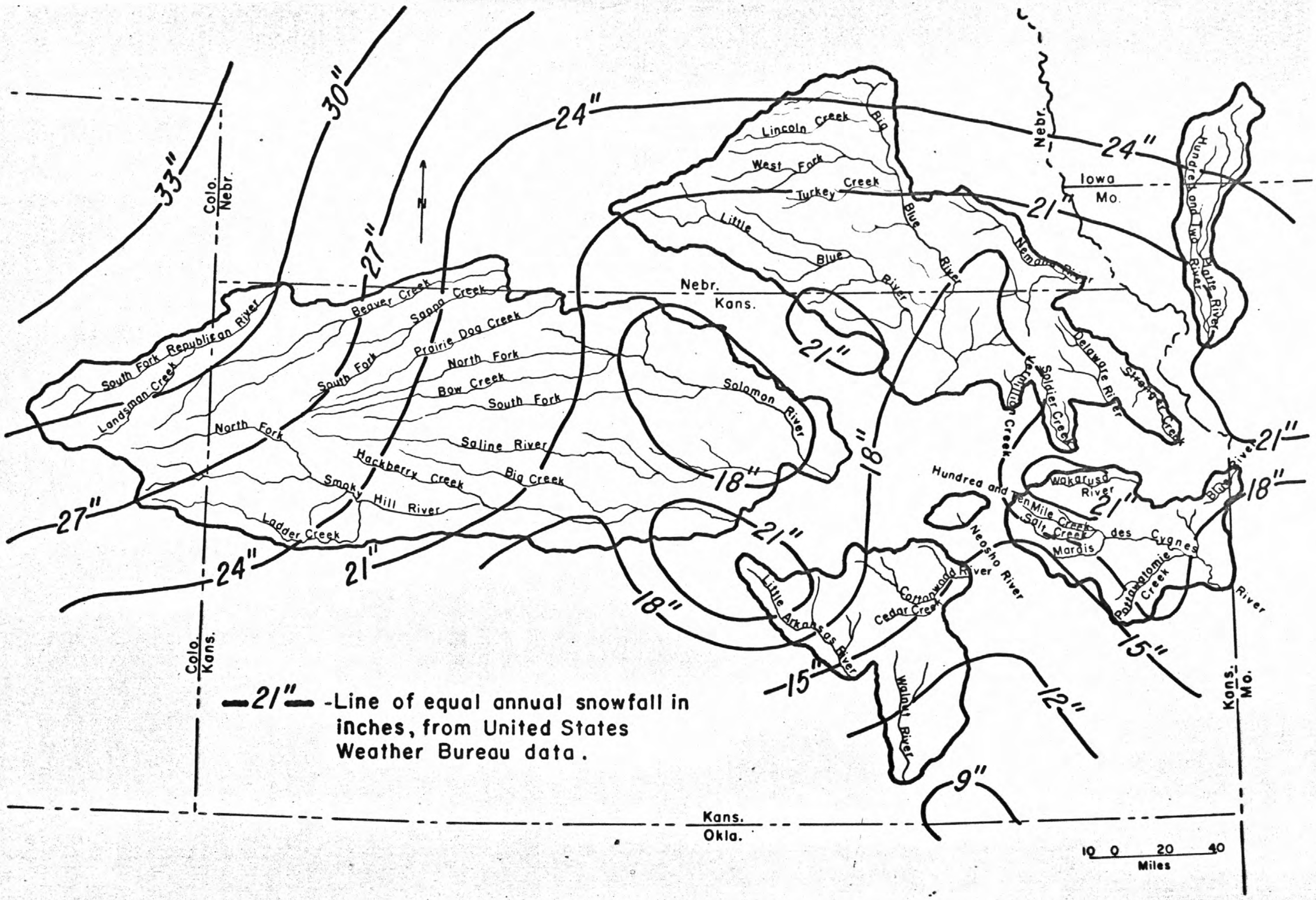

Figure 4.--Average annual snowfall in the Central United States study area. (Modified after Flora, 1948)

Mapped-Flow Characteristics

The regional variations of three hydrologic factors as defined by isopleths on maps that were described by Furness (1966) were used Mean discharge is the residual of annual precipitation after evaporation, transpiration, and underflow at the gaging sites. It includes all flows--high, medium, and low. Since it is a mean value, it can be expected to be more closely related to the higher discharge range which contributes the greatest volume of water. Since it is a summation of all flows, it is related to all flows, though not uniquely. It is determined by averaging the isopleths over the basin weighted by area. Variability index is the standard deviation of daily discharges, in logarithmic units, and has been adjusted to that of a common sized basin because it is influenced by basin size. It is a resultant of many physical and climatic factors, including infiltration, transmissibility, and the frequency of streamflow-producing precipitation. It would be expected to be negatively correlated with low flows and possibly positively correlated with high flows and peaks. Maps of mean discharge and variability index are shown in figures 5 and 6, respectively. Ratio of 1 percent duration to mean flow (fig. 7) is a measure of the slope of the upper portion of the duration curve. In the eastern part of the study area where mean flow for a station can be between the 14-percent and 17-percent points on the duration curve, the factor is almost entirely dependent on the upper-end duration curve slope. In the western section of the study area where the duration curve may deviate strongly from a straight line, mean flow for a station may be anywhere between the 5-percent and 30-percent points on the duration curve, and the ratio will be influenced by many hydrologic characteristics of the basin.

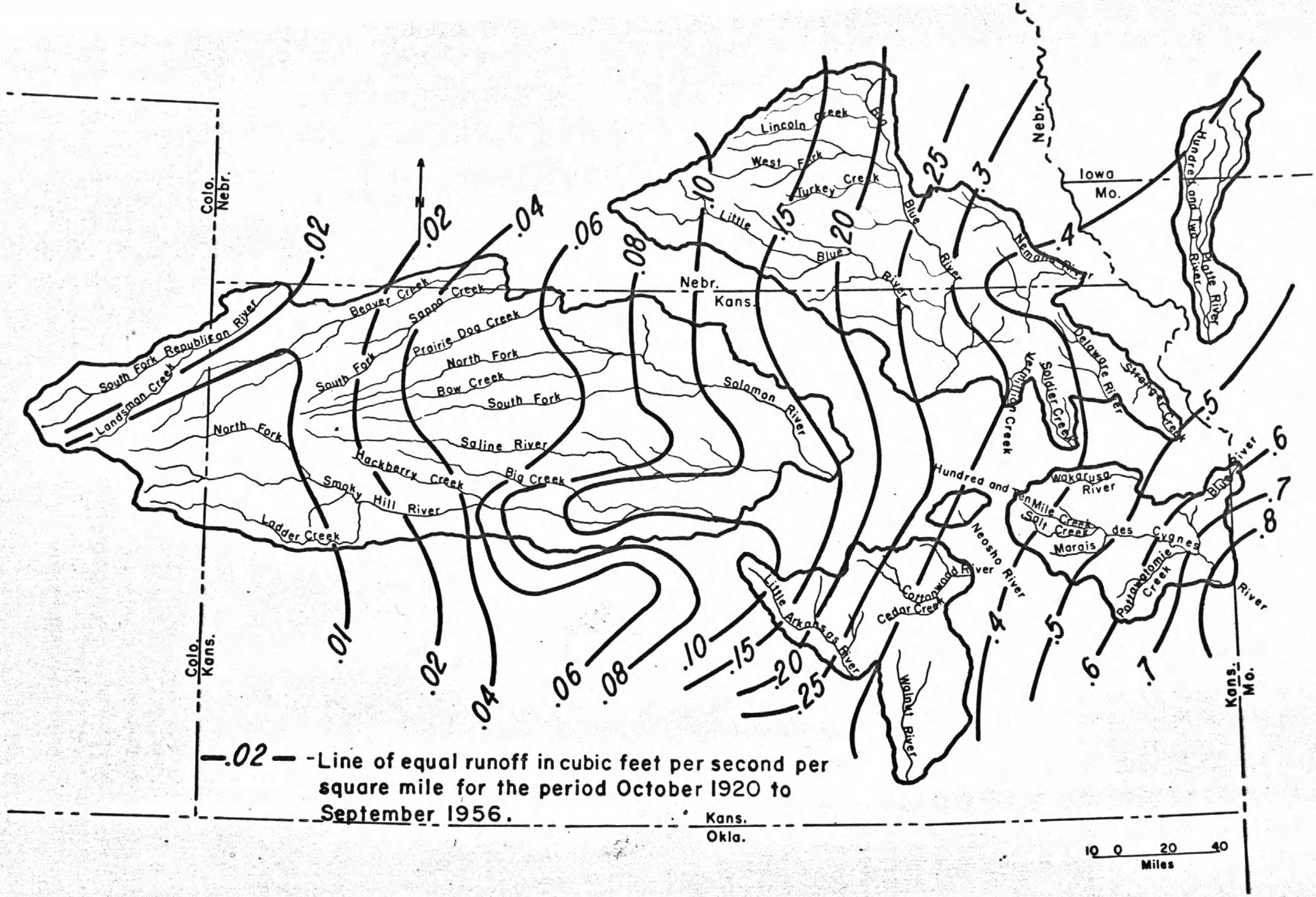

Figure 5.--Mean discharge in the Central United States study area. (Modified after Furness and others,1966)

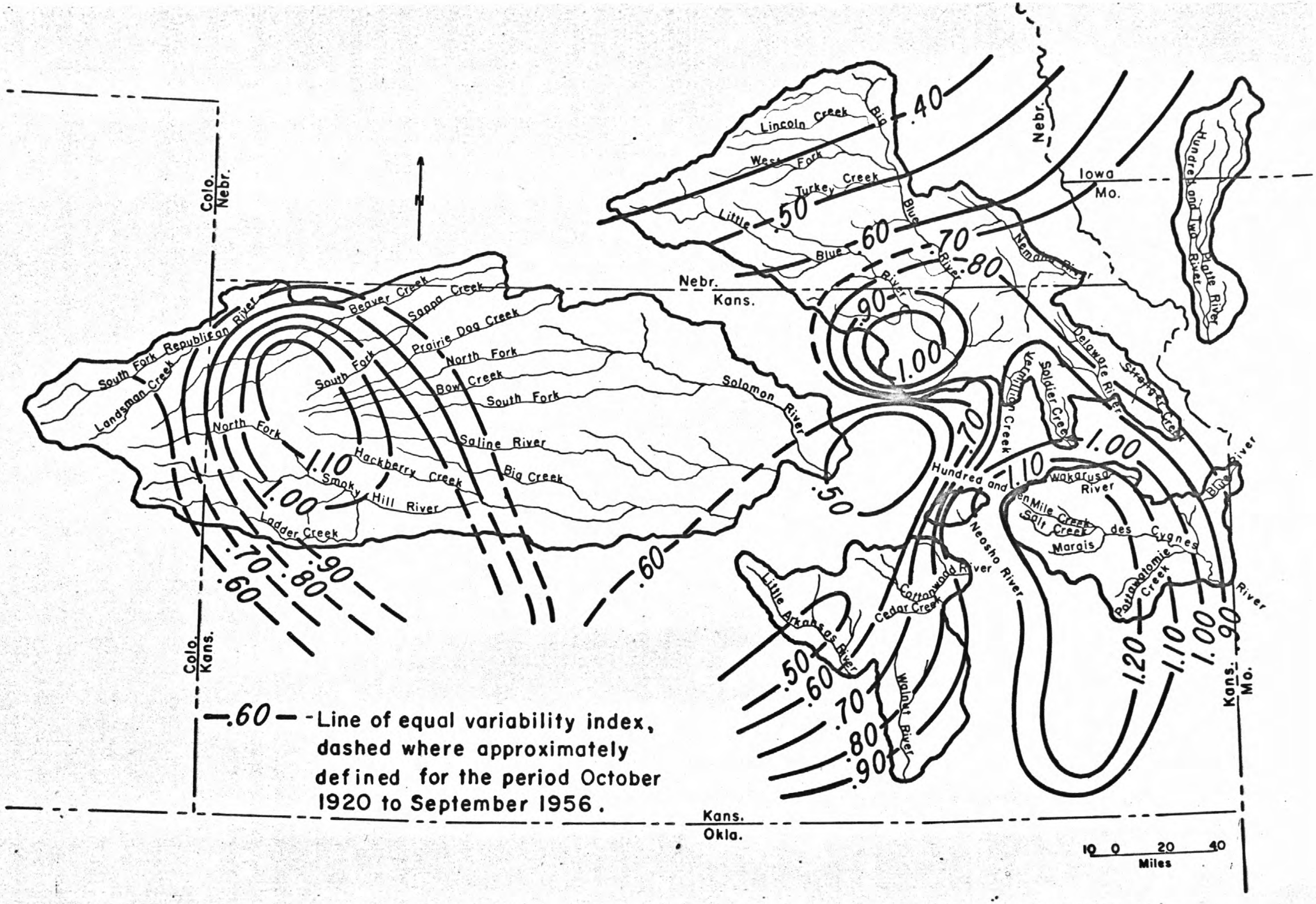

Figure 6.--Variability index for basins of 500 square miles in the Central United States study area. (Modified after Furness and others, 1966)

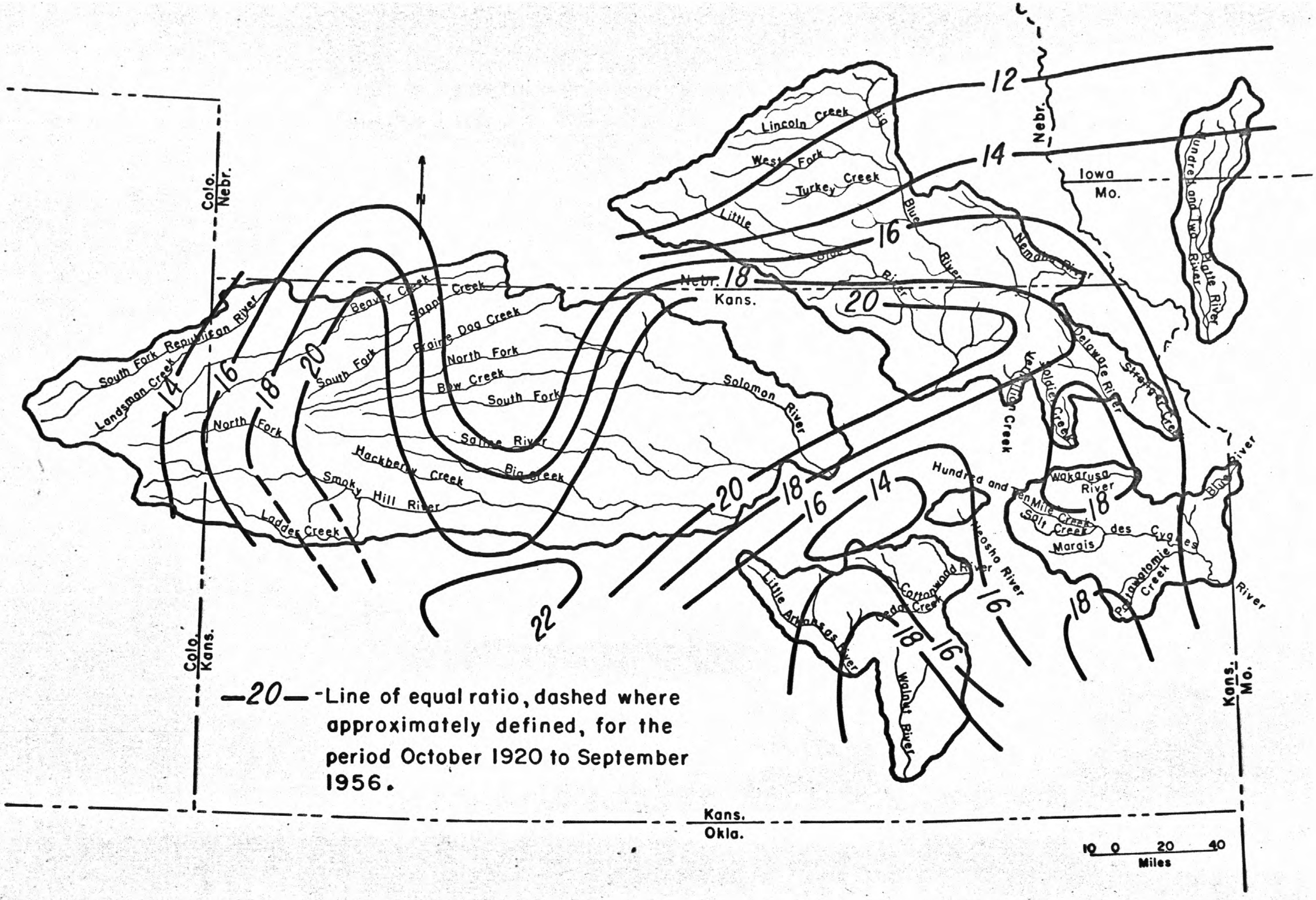

Figure 7.--Ratio of 1-percent duration to mean of total flow for basins of 500 square miles in the Central United States study area. (Modified after Furness and others, 1966)

RESULTS

Multiple regression is used in data analysis to obtain the best fit of a set of observations of independent and dependent variables by an equation of the form:

$$Y = a + b_1X_1 + b_2X_2 + . . . + b_nX_n$$

where:

Y is a dependent variable such as a streamflow characteristic,

X_1, X_2,...X_n are independent variables such as basin and climatic characteristics,

a is the regression constant, and

b_1, b_2,...b_n are regression coefficients.

The Y's and X's are known data, and the a's and b's are constants which are computed with the criterion that the sum of squares of residuals of the relationship be minimized. In practice, it has been found that a logarithmic transformation will linearize the relationships of many hydrologic variables. This has been done for all the data so that the resulting equation has the form:

$$\text{Log } Y = \log a + b_1 \log X_1 + b_2 \log X_2 + . . . + b_n \log X_n$$

by taking antilogs we get the equivalent form:

$$Y = aX_1^{b_1} X_2^{b_2} . . . X_n^{b_n}.$$

The calculations involved in solving for the constants are quite extensive and have been programed on a digital computer. The program used in this study has an option which will successively delete the least significant independent variable from the original set of specified variables until only one independent variable remains. When this option was used, all succeeding equations following the one in which all independent variables were significant at the 0.05 level are included in the table of regression results so that the reduction in standard error of estimate of the relationship for a particular variable can be determined. By use of the 0.05 significance level, we can be at least 95 percent confident that the included independent variables are effective in explaining variations of the dependent variable.

The results are divided into two groups, those in which only basin and climatic characteristics were used as independent variables, such as in the similar studies in the Potomac River basin, Louisiana, and the Central Valley in California; and those in which mapped-flow variables also were used such as has previously been done in Kansas and reported in the Kansas Water Resources Board's technical report series. It is not the object of this report to determine whether or not mapped-flow terms should be used as independent variables but rather to present the results both with and without them.

Drainage area, annual precipitation, annual snowfall, main channel length, and soils infiltration index were found to be some of the most important basin and climatic variables affecting streamflow. Mean discharge and variability index were the most important mapped-flow variables explaining variations in streamflow. Alluvial area proved to be a more effective variable than drainage area in many low-flow relationships.

In some of the regression relationships developed, high simple correlation of the independent variables used in the equation gave unreasonable results, or widely varying b coefficients, as successive variables were dropped. Table 2 is a matrix of the simple correlation coefficients of the independent variables. It shows that P, $P_{2,24}$, $P_{100,24}$, F, El, qm, and Sn are quite highly correlated. For example, going across row 19 to column 5, we find that P and El have a -0.99 correlation coefficient. If more than one of the highly correlated variables appeared at one time in an equation, the results were used only if they seemed hydrologically reasonable. Another set of independent variables which were highly related were A, Aa, L, and Lca.

Table 2.--Matrix of simple correlation coefficients of transformed independent variables.

	A	L	W	Vw	P	Sl	Ch	SW	T_7	T_1	Ev	Aa	F	Gr	$P_{2,24}$	$P_{100,24}$	Sn	Si	El	qm	v	Br	Lca	Mr	Rlm
A	1.00																								
L	.89	1.00																							
W	.70	.31	1.00																						
Vw	.41	.34	.35	1.00																					
P	-.44	-.62	-.00	.04	1.00																				
Sl	-.26	-.16	-.27	-.34	-.53	1.00																			
Ch	.19	.08	.20	.44	.40	-.45	1.00																		
SW	.43	.16	.66	.31	.07	-.13	.13	1.00																	
T_7	.26	.40	-.14	.24	-.33	.04	.21	-.10	1.00																
T_1	-.42	-.56	-.03	.08	.81	-.34	.35	.03	.03	1.00															
Ev	.28	.44	-.10	.05	-.72	.41	-.17	-.06	.76	-.29	1.00														
Aa	.92	.83	.66	.46	-.37	-.35	.15	.40	.23	-.39	.24	1.00													
F	-.37	-.53	.05	.05	.88	-.51	.34	.12	-.58	.58	-.89	-.31	1.00												
Gr	-.09	-.24	.15	.05	.16	.02	.00	.20	.15	.33	.15	-.09	-.07	1.00											
$P_{2,24}$	-.46	-.66	.03	.04	.99	-.49	.39	.11	-.32	.84	-.67	-.38	.85	.25	1.00										
$P_{100,24}$	-.45	-.64	.01	.06	.98	-.51	.42	.09	-.24	.87	-.63	-.38	.83	.24	1.00	1.00									
Sn	.32	.48	-.06	-.20	-.95	.36	-.52	-.15	-.15	-.84	.21	.26	-.51	-.40	-.80	-.83	1.00								
Si	.16	.30	-.05	.03	-.82	.58	-.42	-.08	.16	-.62	.56	.18	-.69	-.13	-.78	-.79	.59	1.00							
El	.38	.55	-.03	-.07	-.99	.58	-.44	-.11	.36	-.75	.77	.31	-.91	-.09	-.97	-.96	.72	.83	1.00						
qm	-.36	-.53	.04	.14	.98	-.55	.47	.16	-.26	.77	-.70	-.30	.86	.13	.95	.96	-.77	-.80	-.98	1.00					
v	.00	-.01	-.02	.17	.15	-.13	.24	.08	-.03	.09	-.14	-.05	.16	.11	.12	.13	-.13	-.23	-.15	.12	1.00				
Br	.16	.42	-.34	-.29	-.57	.27	-.16	-.33	.26	-.57	.41	.12	-.46	-.33	-.62	-.61	.49	.35	.53	-.54	-.06	1.00			
Lca	.92	.98	.38	.33	-.61	-.18	.09	.19	.36	-.56	.42	.85	-.52	-.20	-.64	-.62	.44	.29	.55	-.54	.01	.40	1.00		
Mr	.37	.60	-.19	.19	-.23	-.20	.27	-.30	.48	-.06	.29	.32	-.30	-.26	-.26	-.22	.12	.07	.21	-.21	.01	.28	.55	1.00	
Rlm	.07	.15	-.18	-.02	.08	-.15	.37	-.25	.43	.09	.17	.08	.00	-.16	.05	.09	-.17	-.34	-.09	.05	.26	.33	.16	.29	1.00

Using Physical and Climatic Characteristics as Variables

Of the 22 physical and climatic characteristics investigated as independent variables for developing regression equations, only nine were found significant at the .05 level in the best of the developed regression relationships. These developed equations are shown in table 3 and in all cases have the lowest standard errors of estimate of any logical combination of independent variables.

Flood peaks were generalized with standard errors of estimate from 43.6 percent for the 10-year flood to 57.8 percent for the 50-year flood. The variables significantly related to flood peaks were drainage area, annual precipitation, main channel length, and annual snowfall.

High-flow volumes were generalized with standard errors of estimate from 26.3 percent for the 10-year 7-day high to 46.3 percent for the 50-year 3-day high. Drainage area, mean annual precipitation, and average annual snowfall were the variables significantly affecting high-flow volumes. High-flow volumes had higher standard deviations than flood peaks, but standard errors of estimate for flood peaks were higher. This could be because drainage area explains more of the variance for high flows than for flood peaks, and flood peaks are more subject to variations in drainage efficiency which is less adequately quantified than other variables used in this study.

Durations of daily flows exceeded 10, 50, and 80 percent of the time were generalized with standard errors of 55.2, 86.0, and 194.9 percent, respectively. The variables effective in explaining the variation of the flow-duration curves were drainage area, mean annual precipitation, and soils infiltration index. The soils infiltration index probably is a measure of the potential for recharge to the ground-water system that feeds the streams during base-flow periods.

Annual flow statistics generalized were the mean, standard deviation, skew coefficient, and first-order serial correlation coefficient. The mean should be influenced by the basin size, precipitation, and losses. Drainage area and annual precipitation were found significant, but neither of the indexes of losses, average maximum-daily July temperature or average annual evaporation, significantly affected the relations. The standard error of estimate of the regression equation for mean annual flow was 28.2 percent. The standard deviation of annual discharges was defined with a standard error of 33.0 percent by a regression equation using drainage area, average annual precipitation, and average annual snowfall. Average annual evaporation, alluvial area, and drainage area were found to affect skew coefficients. Average minimum-daily January temperature, alluvial area, and average maximum-daily July temperatures were significantly related to the first-order serial correlation coefficient of annual discharge. The reduction in standard error of estimate from the standard deviation of either skew or serial correlation is small. This could be attributed to the fact that relatively long records are necessary for good estimates of these statistics, and therefore the station values used as dependent variables may have a significant amount of sampling error variance which is unexplainable by hydrologic characteristics.

Generalization was attempted on 7-day low-flow frequencies. Although drainage area, mean annual precipitation, average annual snowfall, soils infiltration index, and average minimum January temperature were found significant for the 2-year exceedance value, the standard error of estimate was 306 percent. Regression equations for the 10- and 20-year exceedance intervals had standard errors of estimate over 400 percent.

The best estimates of monthly discharge were obtained during the spring and fall with the lowest standard error of estimate of 28.4 percent for the average May discharge. The highest standard error of estimate was 48.9 percent for the average January discharge. Drainage area and mean annual precipitation were significant variables in all monthly regression equations. Other variables which were significant in some of the monthly relationships were soils infiltration index, average minimum January temperature, average annual snowfall, average maximum July temperature, and alluvial area. Standard deviations of monthly discharge were consistently related to drainage area and mean annual precipitation, and sometimes to average annual snowfall and soils infiltration index. Standard errors of estimate of standard deviations ranged from 30.6 to 52.6 percent.

Table 3.--Regression coefficients and standard errors for relationships developed using only physical and climatic characteristics as independent variables.

General equation: $\log Y = a + b_1 \log A + b_2 \log L + b_3 \log P + b_4 \log T_7 + b_5 \log T_1 + b_6 \log Ev + b_7 \log Aa + b_8 \log Sn + b_9 \log Si.$

All regression coefficients are statistically significant at the 1 percent level except those underlined which are significant at the 5 percent level and those followed by the letter d which are nonsignificant at the 5 percent level.

Regression Coefficients

Y	a	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	SE%	SE logs
Qp_{50}	--	--	--	--	--	--	--	--	--	--	91.9	0.359
	7.442	--	--	--	--	--	--	--	-2.213	--	73.5	.296
	7.180	0.329	--	--	--	--	--	--	-2.761	--	61.9	.254
	6.621	.710	$\underline{-0.642}$	--	--	--	--	--	-2.227	--	57.8	.239
Qp_{20}	--	--	--	--	--	--	--	--	--	--	91.6	.351
	1.926	--	--	1.735	--	--	--	--	--	--	72.1	.288
	- .303	.396	--	2.475	--	--	--	--	--	--	55.1	.230
	.742	.780	- .691	1.901	--	--	--	--	--	--	50.5	.211
	3.415	.737	$\underline{-.632}$	$\underline{1.187}$	--	--	--	--	$\underline{-1.269}$	--	47.1	.198
Qp_{10}	--	--	--	--	--	--	--	--	--	--	90.4	.355
	1.563	--	--	1.895	--	--	--	--	--	--	69.8	.280
	- .653	.394	--	2.631	--	--	--	--	--	--	52.7	.219
	.515	.823	- .773	1.989	--	--	--	--	--	--	45.9	.193
	2.703	.788	- .717	1.405	--	--	--	--	$\underline{-1.039}$	--	43.6	.184
Qp_5	--	--	--	--	--	--	--	--	--	--	95.1	.369
	1.163	--	--	2.058	--	--	--	--	--	--	69.3	.281
	-1.081	.399	--	2.804	--	--	--	--	--	--	52.5	.219
	.175	.861	- .832	2.113	--	--	--	--	--	--	44.6	.188

Table 3.--Continued

Regression Coefficients

Y	a	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	SE%	SE logs
Qp_2	--	--	--	--	--	--	--	--	--	--	105.1	0.398
	0.494	--	--	2.288	--	--	--	--	--	--	73.8	.297
	-1.905	0.426	--	3.085	--	--	--	--	--	--	55.4	.230
	- .698	.870	-0.798	2.422	--	--	--	--	--	--	48.5	.204
$Qp_{1.1}$	--	--	--	--	--	--	--	--	--	--	121.6	.477
	- .281	--	--	2.404	--	--	--	--	--	--	89.6	.350
	-3.194	.517	--	3.372	--	--	--	--	--	--	64.8	.265
	-6.782	.629	--	5.084	--	--	--	--	--	1.869	58.5	.241
$V_{3,50}$	--	--	--	--	--	--	--	--	--	--	125.4	.463
	1.304	--	--	2.013	--	--	--	--	--	--	106.7	.403
	-3.042	.772	--	3.457	--	--	--	--	--	--	49.4	.207
	- .399	.767	--	2.724	--	--	--	--	-1.208	--	46.3	.195
$V_{3,20}$	--	--	--	--	--	--	--	--	--	--	125.8	.457
	.841	--	--	2.239	--	--	--	--	--	--	97.7	.378
	-3.412	.755	--	3.652	--	--	--	--	--	--	39.5	.168
	-1.058	.751	--	2.999	--	--	--	--	-1.076	--	36.4	.155
$V_{3,10}$	--	--	--	--	--	--	--	--	--	--	129.5	.471
	.455	--	--	2.420	--	--	--	--	--	--	97.6	.373
	-3.859	.766	--	3.853	--	--	--	--	--	--	34.1	.146
	-1.630	.762	--	3.234	--	--	--	--	-1.019	--	30.9	.132
$V_{3,2}$	--	--	--	--	--	--	--	--	--	--	146.7	.512
	- .697	--	--	2.863	--	--	--	--	--	--	102.1	.389
	-5.270	.812	--	4.382	--	--	--	--	--	--	31.8	.136
	-3.777	.809	--	3.968	--	--	--	--	- .682	--	30.6	.131

Table 3.--Continued.

Regression Coefficients

Y	a	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	SE%	SE logs
$V_{7,50}$	--	--	--	--	--	--	--	--	--	--	138.1	0.489
	2.431	0.493	--	--	--	--	--	--	--	--	116.1	.429
	-3.807	.875	--	3.615	--	--	--	--	--	--	45.0	.189
	-1.144	.869	--	2.876	--	--	--	--	-1.218	--	41.4	.175
$V_{7,20}$	--	--	--	--	--	--	--	--	--	--	137.7	.488
	.627	--	--	2.228	--	--	--	--	--	--	111.5	.417
	-4.251	.866	--	3.848	--	--	--	--	--	--	35.3	.151
	-1.844	.862	--	3.180	--	--	--	--	-1.100	--	31.6	.135
$V_{7,10}$	--	--	--	--	--	--	--	--	--	--	141.4	.494
	.229	--	--	2.416	--	--	--	--	--	--	109.2	.409
	-4.637	.864	--	4.033	--	--	--	--	--	--	29.9	.128
	-2.491	.860	--	3.438	--	--	--	--	- .981	--	26.3	.113
$V_{7,2}$	--	--	--	--	--	--	--	--	--	--	154.9	.530
	- .883	--	--	2.832	--	--	--	--	--	--	110.9	.417
	-5.881	.888	--	4.492	--	--	--	--	--	--	28.2	.121
	-4.539	.885	--	4.120	--	--	--	--	- .613	--	27.1	.116
D_{10}	--	--	--	--	--	--	--	--	--	--	209.7	.646
	.385	.657	--	--	--	--	--	--	--	--	170.3	.565
	-7.966	1.168	--	4.839	--	--	--	--	--	--	55.2	.229
D_{50}	--	--	--	--	--	--	--	--	--	--	205.5	.637
	- .739	.727	--	--	--	--	--	--	--	--	155.4	.532
	-7.133	1.119	--	3.705	--	--	--	--	--	--	92.8	.359
	-11.259	1.247	--	5.674	--	--	--	--	--	2.149	86.0	.338

Table 3.--Continued.

Regression Coefficients

Y	a	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	SE%	SE logs
D_{80}	--	--	--	--	--	--	--	--	--	--	441.4	0.952
	-2.905	1.150	--	--	--	--	--	--	--	--	287.6	.772
	-9.126	1.531	--	3.605	--	--	--	--	--	--	227.0	.677
	-18.351	1.818	--	8.006	--	--	--	--	--	4.805	194.9	.617
Q_a	--	--	--	--	--	--	--	--	--	--	165.6	.555
	-1.666	--	--	2.712	--	--	--	--	--	--	126.9	.460
	-7.222	.987	--	4.558	--	--	--	--	--	--	28.2	.121
SD_a	--	--	--	--	--	--	--	--	--	--	165.5	.557
	-1.540	--	--	2.570	--	--	--	--	--	--	132.8	.473
	-7.143	.995	--	4.431	--	--	--	--	--	--	36.4	.155
	-4.762	.991	--	3.771	--	--	--	--	-1.088	--	33.0	.141
G_a	--	--	--	--	--	--	--	--	--	--	37.0	.168
	-2.331	--	--	--	--	--	1.578	--	--	--	35.0	.149
	-2.065	--	--	--	--	--	1.316	0.097	--	--	33.0	.141
	-1.956	- .250	--	--	--	--	1.465	.296	--	--	31.2	.133
SR_a	--	--	--	--	--	--	--	--	--	--	17.7	.078
	- .701	--	--	--	--	0.608	--	--	--	--	14.7	.064
	-1.100	--	--	--	--	.819	--	.071	--	--	12.1	.052
	4.587	--	--	--	-2.914	.850	--	.080	--	--	11.6	.050
$M_{7,2}$+.01	--	--	--	--	--	--	--	--	--	--	660.8	1.163
	6.680	--	--	--	--	--	--	--	-4.924	--	598.8	1.080
	5.597	1.362	--	--	--	--	--	--	-7.194	--	368.3	.875
	.800	1.459	--	1.779[d]	--	--	--	--	-5.680	--	368.4	.876
	-9.346	1.746	--	6.420	--	--	--	--	-5.221	4.762	335.2	.836
	2.272	1.646	--	8.284	--	-7.688	--	--	-8.590	4.913	306.0	.798

Table 3.--Continued.

		Regression Coefficients										
Y	a	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	SE%	SE logs
$M_{7,10}$ +.01	--	--	--	--	--	--	--	--	--	--	1059.0	1.251
	-5.603	1.521	--	--	--	--	--	--	--	--	509.7	1.011
	-11.631	1.890	--	3.493	--	--	--	--	--	--	441.9	.952
$M_{7,20}$ +.01	--	--	--	--	--	--	--	--	--	--	712.0	1.113
	-5.136	1.292	--	--	--	--	--	--	--	--	410.5	.921
Q_{10}	--	--	--	--	--	--	--	--	--	--	178.3	.581
	-2.352	--	--	3.090	--	--	--	--	--	--	126.3	.459
	-7.797	.967	--	4.899	--	--	--	--	--	--	34.5	.147
	-24.202	.954	--	5.010	8.266	--	--	--	--	--	33.0	.141
Q_{11}	--	--	--	--	--	--	--	--	--	--	235.3	.692
	-3.967	--	--	4.033	--	--	--	--	--	--	146.5	.509
	-9.884	1.051	--	5.999	--	--	--	--	--	--	45.7	.192
	-12.772	1.141	--	7.377	--	--	--	--	--	1.505	40.2	.170
Q_{12}	--	--	--	--	--	--	--	--	--	--	206.0	.638
	-3.773	--	--	3.824	--	--	--	--	--	--	126.6	.458
	-8.951	.920	--	5.544	--	--	--	--	--	--	47.3	.199
	-11.876	1.011	--	6.940	--	--	--	--	--	1.524	42.0	.177
Q_1	--	--	--	--	--	--	--	--	--	--	230.6	.683
	-4.213	--	--	4.146	--	--	--	--	--	--	135.9	.483
	-9.577	.953	--	5.928	--	--	--	--	--	--	54.5	.227
	-12.755	1.052	--	7.445	--	--	--	--	--	1.656	48.9	.205

Table 3.--Continued.

Regression Coefficients

Y.	a	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	SE%	SE logs
Q_2	--	--	--	--	--	--	--	--	--	--	202.3	0.632
	-3.020	--	--	3.449	--	--	--	--	--	--	140.3	.490
	-8.560	0.984	--	5.290	--	--	--	--	--	--	50.7	.212
	-10.777	1.053	--	6.347	--	--	--	--	--	1.155	48.1	.202
	-10.079	1.041	--	7.170	--	-1.497	--	--	--	1.251	45.7	.193
Q_3	--	--	--	--	--	--	--	--	--	--	195.4	.618
	-3.308	--	--	3.810	--	--	--	--	--	--	116.1	.430
	-8.237	.875	--	5.447	--	--	--	--	--	--	41.3	.175
	-7.186	.854	--	6.321	--	-1.781	--	--	--	--	36.9	.157
	-7.488	1.125	--	6.457	--	-1.951	--	-0.256	--	--	34.6	.148
	-9.165	1.264	--	7.334	--	-2.110	--	- .341	--	.854	33.0	.141
Q_4	--	--	--	--	--	--	--	--	--	--	252.4	.719
	-4.314	--	--	4.566	--	--	--	--	--	--	136.1	.482
	-9.964	1.004	--	6.444	--	--	--	--	--	--	40.6	.172
	-12.102	1.070	--	7.464	--	--	--	--	--	1.114	37.5	.160
	-13.219	1.388	--	7.846	--	--	--	- .278	--	1.469	35.2	.150
Q_5	--	--	--	--	--	--	--	--	--	--	168.8	.562
	-1.714	--	--	2.883	--	--	--	--	--	--	124.3	.453
	-7.145	.965	--	4.688	--	--	--	--	--	--	31.0	.133
	-5.198	.961	--	4.147	--	--	--	--	-0.890	--	28.4	.122
Q_6	--	--	--	--	--	--	--	--	--	--	160.7	.544
	.588	.650	--	--	--	--	--	--	--	--	120.2	.443
	-6.137	1.062	--	3.897	--	--	--	--	--	--	36.2	.154

Table 3.--Continued.

						Regression Coefficients						
Y	a	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	SE%	SE logs
Q_7	--	--	--	--	--	--	--	--	--	--	141.4	0.498
	0.739	0.560	--	--	--	--	--	--	--	--	112.7	.419
	-5.540	.945	--	3.638	--	--	--	--	--	--	37.1	.158
	-3.452	.941	--	3.059	--	--	--	--	-0.954	--	34.7	.148
Q_8	--	--	--	--	--	--	--	--	--	--	152.1	.525
	.005	.692	--	--	--	--	--	--	--	--	108.6	.402
	-5.994	1.060	--	3.477	--	--	--	--	--	--	36.2	.154
Q_9	--	--	--	--	--	--	--	--	--	--	192.8	.613
	-1.958	--	--	2.859	--	--	--	--	--	--	149.3	.518
	-8.019	1.076	--	4.872	--	--	--	--	--	--	44.1	.186
	-5.611	1.072	--	4.204	--	--	--	--	-1.101	--	41.3	.175
SD_{10}	--	--	--	--	--	--	--	--	--	--	168.8	.562
	-1.640	--	--	2.802	--	--	--	--	--	--	127.3	.461
	-6.873	.930	--	4.541	--	--	--	--	--	--	46.8	.197
SD_{11}	--	--	--	--	--	--	--	--	--	--	290.6	.776
	-5.276	--	--	5.080	--	--	--	--	--	--	144.4	.501
	-11.195	1.051	--	7.047	--	--	--	--	--	--	39.3	.166
SD_{12}	--	--	--	--	--	--	--	--	--	--	268.5	.745
	-5.696	--	--	5.259	--	--	--	--	--	--	111.5	.418
	-10.386	.833	--	6.817	--	--	--	--	--	--	44.5	.187
	-8.106	.829	--	6.184	--	--	--	--	--	-1.042	42.0	.178

Table 3.--Continued.

Regression Coefficients

Y	a	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	SE%	SE logs
SD_1	--	--	--	--	--	--	--	--	--	--	290.2	0.776
	-5.816	--	--	5.336	--	--	--	--	--	--	127.9	.461
	-11.244	0.964	--	7.140	--	--	--	--	--	--	37.3	.159
	-10.663	1.178	-0.385	6.820	--	--	--	--	--	--	35.5	.151
SD_2	--	--	--	--	--	--	--	--	--	--	269.0	.746
	-4.222	--	--	4.366	--	--	--	--	--	--	161.2	.546
	-10.506	1.116	--	6.454	--	--	--	--	--	--	52.3	.218
SD_3	--	--	--	--	--	--	--	--	--	--	197.6	.623
	-2.85	--	--	3.562	--	--	--	--	--	--	129.5	.466
	-7.853	.888	--	5.223	--	--	--	--	--	--	58.8	.243
SD_4	--	--	--	--	--	--	--	--	--	--	330.1	.830
	-5.631	--	--	5.557	--	--	--	--	--	--	148.6	.515
	-11.649	1.069	--	7.557	--	--	--	--	--	--	44.6	.188
	-10.885	1.350	- .506	7.136	--	--	--	--	--	--	41.8	.177
	-8.863	1.317	- .454	6.596	--	--	--	--	-0.960	--	39.6	.168
SD_5	--	--	--	--	--	--	--	--	--	--	161.2	.544
	-1.341	--	--	2.668	--	--	--	--	--	--	123.1	.450
	-6.659	.945	--	4.435	--	--	--	--	--	--	35.2	.150
	-4.04	.939	--	3.708	--	--	--	--	-1.197	--	30.6	.131

Table 3.--Concluded.

Y	a	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	SE%	SE logs
						Regression Coefficients						
SD_6	--	--	--	--	--	--	--	--	--	--	157.0	0.538
	0.722	0.645	--	--	--	--	--	--	--	--	117.9	.436
	-5.547	1.029	--	3.633	--	--	--	--	--	--	48.0	.202
SD_7	--	--	--	--	--	--	--	--	--	--	162.3	.548
	.600	--	--	2.321	--	--	--	--	--	--	134.6	.481
	-5.895	.940	--	4.081	--	--	--	--	--	--	55.7	.231
	-3.080	.935	--	3.299	--	--	--	--	-1.286	--	52.6	.219
SD_8	--	--	--	--	--	--	--	--	--	--	161.2	.545
	.260	.672	--	--	--	--	--	--	--	--	117.8	.436
	- .611	1.063	--	3.692	--	--	--	--	--	--	44.7	.188
	-3.770	1.058	--	3.062	--	--	--	--	-1.070	--	42.1	.178
SD_9	--	--	--	--	--	--	--	--	--	--	179.5	.586
	-1.856	--	--	2.969	--	--	--	--	--	--	132.7	.476
	-7.441	.992	--	4.824	--	--	--	--	--	--	39.3	.167

Using Physical, Climatic, and Mapped-Flow Characteristics as Variables

Ten variables were found significant at the .05 level in at least one of the final relationships when all 25 independent variables were tested in various logical combinations. The results of this set of regressions presented in table 4 are less complete than those in table 3 because fewer dependent variables were used. The option which allowed for recomputing the regression equation with the least significant variable deleted was often not used for these equations. Mapped values of mean annual runoff were found to be significant in all relationships tried. This variable apparently serves the same purpose as the average annual precipitation term in the other relationships. Variability index was a significant variable for all the low-flow characteristics and some of the mean monthly discharges. Ratio of 1 percent duration to mean flow was significant in the 2- and 10-year peak relationships. Main channel length and basin width were used in the place of drainage area in some relationships where this gave lower standard errors of estimate. Drainage area and length might have given nearly equal results.

Table 4.--Regression coefficients and standard errors for relationships developed using physical, climatic, and mapped-flow characteristics as independent variables.

General equation: $\log Y = a + b_1 \log A + b_2 \log L + b_3 \log W + b_4 \log T_1 + b_5 \log Aa + b_6 \log Sn + b_7 \log qm + b_8 \log v + b_9 \log Rlm + b_{10} \log Mr.$

All regression coefficients are statistically significant at the 1-percent level except those underlined which are significant at the 5-percent level.

Regression coefficients

Y	a	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	SE%	SE log
$V_{30,2}$	2.707	--	--	--	--	--	--	--	--	--	--	161.8	0.5464
	3.316	--	--	--	--	--	--	0.654	--	--	--	106.0	.4006
	1.051	0.838	--	--	--	--	--	.916	--	--	--	19.2	.0833
D_{10}	.319	1.070	--	--	--	--	--	.979	--	--	--	45.3	.191
D_{50}	-.949	1.033	--	--	--	--	--	.932	-1.782	--	--	60.5	.249
D_{80}	-3.010	1.401	--	--	--	--	--	1.199	-5.182	--	--	96.0	.3707
Q_a	2.156	--	--	--	--	--	--	--	--	--	--	165.6	.5552
	2.746	--	--	--	--	--	--	.633	--	--	--	114	.4236
	.779	--	1.159	--	--	--	--	1.015	--	--	--	55.9	.2323
	.316	--	.856	0.976	--	--	--	.900	--	--	--	17.8	.0770
$M_{7,2}$ +.01	.184	--	--	--	--	--	--	--	--	--	--	533.4	1.1624
	-1.721	--	--	--	--	0.994	--	--	--	--	--	502.2	1.0316
	-1.346	--	--	--	--	1.338	--	1.112	--	--	--	335.5	.8361
	-1.164	--	--	--	--	1.155	--	1.508	-5.938	--	--	190.1	.6076
	2.620	--	--	--	--	1.182	$\underline{-3.179}$	1.097	-5.661	--	--	178.2	.5829
$M_{7,20}$ +.01	-5.176	1.456	--	--	--	--	--	.970	-5.005	--	--	253.8	.7216

Table 4.--Continued.

Regression Coefficients

Y	a	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	SE%	SE logs
P_{50}	3.881	0.356	--	--	--	--	--	0.457	--	--	--	62.1	0.255
P_{20}	3.803	.344	--	--	--	--	--	.497	--	--	--	53.6	.223
P_{10}	3.713	.399	--	--	--	--	--	.518	--	-1.388	--	45.6	.194
P_2	3.235	.453	--	--	--	--	--	.617	--	-1.974	--	39.6	.186
$V_{1,2}$	3.531	--	--	--	--	--	--	--	--	--	--	131.7	.4724
	4.128	--	--	--	--	--	--	.642	--	--	--	74.1	.2976
	3.060	--	--	1.066	--	--	--	.625	--	--	--	35.4	.1508
	2.587	--	0.396	.867	--	--	--	.758	--	--	--	23.4	.1008
$V_{3,50}$	2.696	.697	--	--	--	--	--	.688	--	--	--	47.2	.198
$V_{3,20}$	2.649	.675	--	--	--	--	--	.724	--	--	--	37.3	.159
$V_{3,10}$	2.533	--	.594	.835	--	--	--	.732	--	--	--	31.2	.133
$V_{3,2}$	2.002	--	.597	.926	--	--	--	.832	--	--	--	21.6	.092
$V_{7,50}$	2.193	.795	--	--	--	--	--	.716	--	--	--	43.4	.183
$V_{7,20}$	2.138	.780	--	--	--	--	--	.758	--	--	--	34.8	.148
$V_{7,10}$	2.059	.773	--	--	--	--	--	.793	--	--	--	29.8	.128

Table 4.--Continued.

Y	a	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	SE%	SE logs
						Regression Coefficients							
$V_{7,2}$	3.108	--	--	--	--	--	--	--	--	--	--	154.8	0.5300
	3.709	--	--	--	--	--	--	0.645	--	--	--	99.9	.3822
	2.404	--	--	1.303	--	--	--	.625	--	--	--	52.8	.2196
	1.568	--	0.700	.952	--	--	--	.861	--	--	--	21.1	.0913
$M_{120,2}$ +.01	1.226	--	--	--	--	--	--	--	--	--	--	229.6	.6811
	1.902	--	--	--	--	--	--	.726	--	--	--	160.8	.5442
	.222	--	--	1.678	--	--	--	.699	--	--	--	95.4	.3678
	- .886	--	.928	1.212	--	--	--	1.013	--	--	--	63.0	.2579
	- .812	--	.971	1.030	--	--	--	1.162	-1.745	--	--	46.6	.1956
$M_{120,20}$ +.01													
	- .514	--	--	--	--	--	--	--	--	--	--	572.8	1.3109
	-5.190	1.562	--	--	--	--	--	--	--	--	--	511.4	1.0712
	-4.902	1.312	--	--	--	--	--	--	-5.082	--	--	430.8	.9410
	-5.200	1.745	--	--	--	--	--	1.313	-7.538	--	--	221.7	.6648
Q_{10}	2.003	--	--	--	--	--	--	--	--	--	--	177	.5812
	2.659	--	--	--	--	--	--	.705	--	--	--	113	.4208
	.788	--	1.103	--	--	--	--	1.068	--	--	--	61.6	.2526
	.338	--	.808	.949	--	--	--	.956	--	--	--	32.1	.1369
	.316	--	.796	1.002	--	--	--	.913	.504	--	--	30.2	.1294
Q_2	1.841	--	--	--	--	--	--	--	--	--	--	202.3	.6316
	2.589	--	--	--	--	--	--	.804	--	--	--	117.6	.4346
	.668	--	1.132	--	--	--	--	1.177	--	--	--	64.6	.2639
	.180	--	.813	1.028	--	--	--	1.056	--	--	--	30.2	.1290
	1.192	--	.898	.928	--	--	--	1.092	--	--	-1.508	24.9	.1074
	3.200	--	.842	.944	-0.991	--	--	1.172	--	--	-1.349	22.7	.0976

Table 4.--Concluded.

Regression Coefficients

Y	a	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	SE%	SE logs
Q_6	0.325	0.984	--	--	--	--	--	0.791	--	--	--	25.5	0.1095
SD_{10}	.629	.845	--	--	--	--	--	.792	1.242	--	--	42.1	.1780
SD_2	.199	.982	--	--	--	--	--	1.297	--	--	--	37.4	.1589
SD_6	.476	.957	--	--	--	--	--	.740	--	--	--	40.3	.1709

DISCUSSION OF RESULTS

The effect of some of the most important independent variables is obvious and easily understood. For some of the other included variables, the effect is more complex and sometimes contrary to hydrologic reasoning. This can be because the measured variable is either related to, or an index of, the real cause of the variation. In such a case the variable is useful in the prediction equation but does not add to our understanding of the hydrologic system.

The interpretation of standard errors of estimate in the context of expected design errors is not simple either, because part of the error measured by the standard error of estimate is error in the dependent variable. Nash and Amorocho (1966) give a method of computing the standard error of estimate of a point on a frequency curve based on the standard deviation of the annual data and the number of years of record. This method was used with the average length of record (31 years), the average standard deviations of annual data (about 0.45 log units for peaks, 3- and 7-day high flows), and the assumption of a log-Gumbel distribution. The results show that a frequency curve for a station with these characteristics would have average standard errors of 30, 41, 52, and 68 percent for the 5-, 10-, 20-, and 50-year recurrence interval values. These errors are roughly the same size as the standard errors of estimate of the flood and high-flow frequency equations.

Comparison of Results

The three-dimensional bar graph in figure 8 shows a comparison of the standard errors of estimate for selected regression equations developed with and without the use of mapped-flow variables. These are shown, respectively, by the heavy- and light-shaded bars. The unshaded bar represents the standard deviation of the dependent variable. Since the standard deviation of the dependent variable is also the standard error of estimate of the mean, the difference between the heights of the unshaded bar and the shaded bars is equivalent to the amount of explained error, while the height of the shaded bars shows the amount of unexplained error. The comparisons shown in figure 8 are not explicit because of inherent statistical weaknesses in some of the derived relationships. These weaknesses occur when independent variables are used which are not free of error such as those obtained from maps. This error is not considered in the regression model and will not be reflected in the computed standard error of estimate.

As the figure shows, regression equations with mapped-flow variables gave lower standard errors for all low flows, mean flows, and 2-year high flows. For floods and high flows with return periods greater than 2 years, the standard errors were about the same for both sets of regression equations, with the set using physical and climatic characteristics usually slightly better. The addition of mapped-flow variables in the relationship for the 2-year 7-day low flow reduced the standard error from 533.4 percent to 306 percent. For mean flow, the reduction in standard error by the use of mapped-flow variable was from 28.2 percent

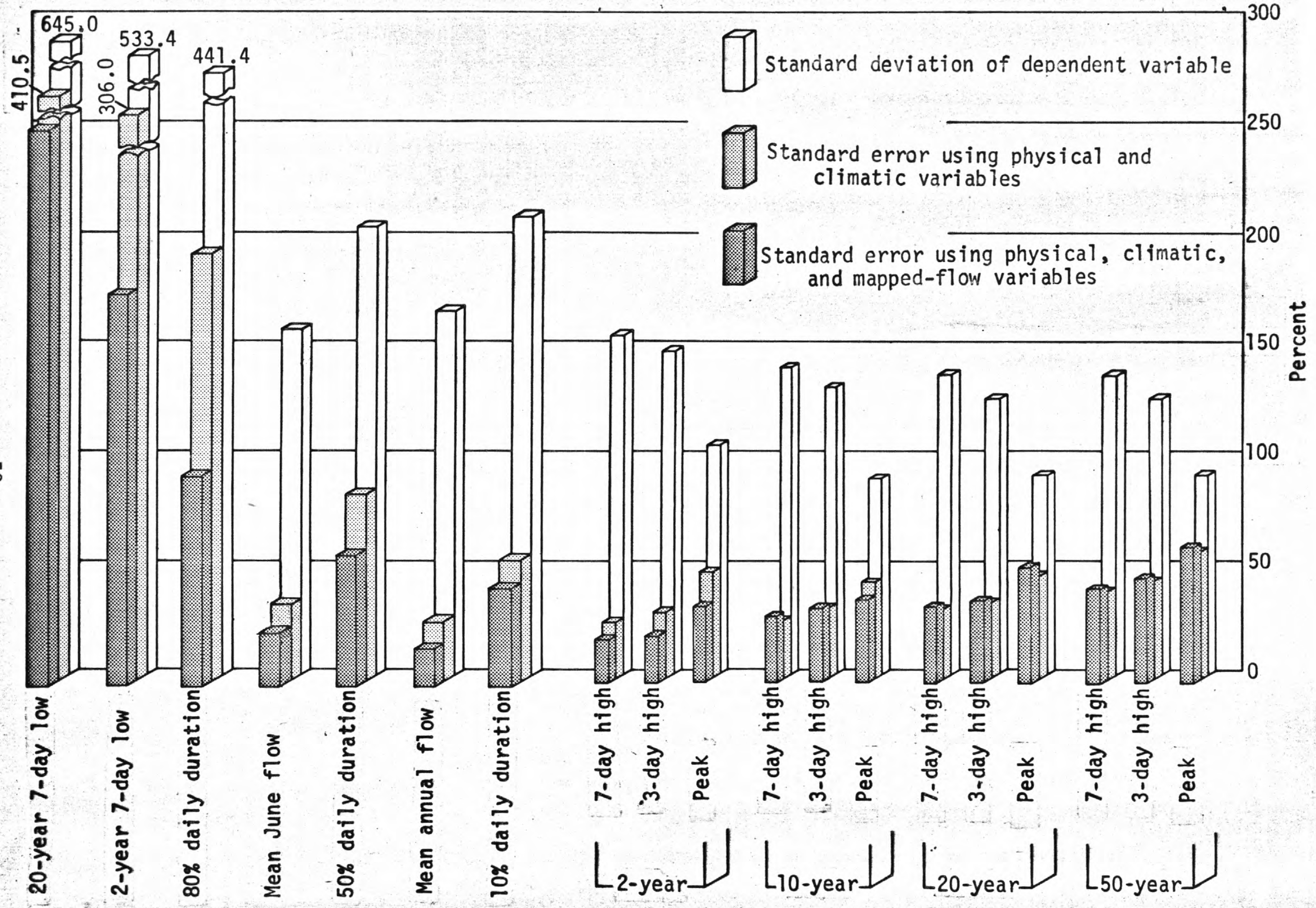

Figure 8.--Standard errors of estimate in the Central United States study-area for selected relationships.

to 17.8 percent. The best regression equation for the 20-year peak using mapped-flow variables had a standard error of 53.6 percent compared to 47.1 percent for the equation using only basin and climatic characteristics. In general, the use of mapped-flow variables appears to be useful for the low-flow and mean-discharge ranges and of little use in high-flow and flood-peak definition. The results must be considered, knowing that most of the stations used in this report were also used in the derivation of the flow-variable maps.

Effects of Interstation Correlation

The regression model is based upon certain assumptions which are not entirely met by hydrologic data. The results must be interpreted on the basis of how well these assumptions are met. The assumptions are that: (1) The independent variables are fixed values and hence do not have probability distributions. (2) The residuals from the regression surface are normally distributed. (3) The variance about the regression surface is constant. (4) The values of the dependent variables are mutually independent. An attempt not to violate (1) was made when selecting streamflow records for analysis. Assumptions (2) and (3) are essentially met by the selection of transforms for the data. Assumption (4) cannot be met because most of the hydrologic data collected within a region are influenced by the same weather systems. Matalas and Benson (1961) have shown that the result of interstation correlation is to increase the variance of the constant and reduce the variance of the b coefficients. This means that the significance of the coefficient may be actually higher than computed, but that increased variance of the constant gives less confidence for the general magnitude of the event than the standard error of estimate indicates.

Use of Results in Improving Stream-Gaging Program

High Flow

Flood volumes were the characteristics best defined by regression analysis. Better definition may be obtainable by collecting longer records and searching for basin characteristics which will affect shorter duration flood volumes, such as basin translation and storage attenuation characteristics. These same characteristics probably would improve flood-peak definition also. For design, longer exceedance intervals are usually used for flood peaks than for other flow characteristics; therefore, records of peak discharge to define station values of these longer exceedance intervals should be longer than records of other flow characteristics.

Mean Flow

The monthly mean discharges with the poorest definition by regression analysis were the winter months December, January, and February. One reason for this probably is regional variation in the lag between winter precipitation and runoff which is not adequately explained by mean annual precipitation and average annual snowfall. This normally is a time of little stress on the water supply of unregulated streams, and the errors may not be of a large practical consequence. Considering both the standard error of the monthly mean discharges and the standard error of the standard deviation of monthly mean discharges, the error of a frequency curve of monthly mean discharges may be quite large in the extreme discharge ranges. Annual mean discharge was defined with the lowest standard error of any of the relationships.

Low Flow

The low-flow relationships developed in this study had large standard errors. Two alternatives appear to be available for improving knowledge concerning low-flow characteristics. One alternative is to devote more research to quantifying the effect of underlying geology on low flow for use with multiple regression. The other alternative is to undertake a much larger program of partial record low-flow data collection to collect as much site data as needed in the foreseeable future. These low-flow data can usually be correlated with data from nearby complete-record gaging stations with longer records to develop low-flow characteristics at the partial-record sites. Either approach should be undertaken only to an extent commensurate with the need for such data.

SUMMARY AND CONCLUSIONS

This study of a 41-gage sample indicates that multiple-regression methods produce useful relations in determining discharge characteristics for most ranges of natural flow. The low-flow relations were poorly defined, and flood-peak relations need improvement.

Regression equations that included mapped-flow terms as independent variables generally had lower standard errors than equations using only basin and climatic characteristics. These lower standard errors may be because mapped-flow terms are more effective in explaining variance of the middle and low-flow ranges and partly because the computed standard errors do not include errors inherent in the mapped-flow variables.

The stream-gaging network is operated to obtain both point information and transferable information for the orderly development and utilization of the water-resource system. The network must be flexible enough to provide data for both present and future needs which may utilize entirely different concepts in data reduction, analysis, and transfer. If present needs for data in some ranges of discharge have been met, then the gaging network should be adapted to collect data where the present network is weak and to include the types of data which may be needed in the future. In reference to the Central United States study area, this would mean collection of longer records of peak flow and collection either of more low-flow site data or more research into the causes of variations in low flow. It is also important to remember that if a hydrologic gaging network is to be designed for eventual analysis by multiple regression, stations are needed to satisfy as wide a range as possible of those independent variables which were found to explain streamflow variations.

This study could be extended profitably to cover the entire state of Kansas for both the state-wide generalization of streamflow and to provide the base for a re-analysis of cooperative gaging needs.

CITED REFERENCES

Flora, S. D., 1948, Climate of Kansas: Kansas Board Agriculture Rept. v. 67, no. 285, 320 p.

Furness, L. W., 1959, Kansas streamflow characteristics, pt. 1, Flow duration: Kansas Water Resources Board, Tech. Rept. no. 1, 213 p.

Furness, L. W., Burns, C. V., and Busby, M. W., 1966, Kansas streamflow characteristics, pt. 6B, Base-flow distribution: Kansas Water Resources Board, Tech. Rept. no. 6B.

Lane, E. W., and Lei, Kai, 1950, Streamflow variability: Am. Soc. Civil Eng. Trans., v. 115, p. 1084-1134.

Matalas, N. C., and Benson, M. A., 1961, Effect of interstation correlation on regression analysis: Jour. Geophy. Research, v. 66, no. 10, p. 3285-3293.

Nash, J. E., and Amorocho, J., 1966, The accuracy of the prediction of floods of a high return period: Water Resources Research, v. 2, no. 2, p. 191-198.

U.S. Department of Commerce, Weather Bureau, 1959a, Climates of the States, Colorado: Climatography of the U.S., no. 60-5, 16 p.

__________, 1959b, Climates of the States, Iowa: Climatography of the U.S., no. 60-13, 16 p.

__________, 1959c, Climates of the States, Kansas: Climatography of the U.S., no. 60-14, 16 p.

__________, 1959d, Climates of the States, Missouri: Climatography of the U.S., no. 60-23, 16 p.

__________, 1959e, Climates of the States, Nebraska: Climatography of the U.S., no. 60-25, 16 p.

__________, 1961, Rainfall frequency atlas of the U.S.: Tech. paper no. 40, 115 p.

APPENDIX I

Description of Independent Variables

Alluvial area:--The total area of valley alluvium, as planimetered from 1:500,000-scale State geologic maps.

Average annual evaporation:--Average annual lake and reservoir evaporation in inches, from written communication with the Kansas Water Resources Board.

Average annual snowfall:--Average annual snowfall depth in inches, as determined from a map, Flora (1948).

Average maximum July temperature:--The mean of the maximum daily temperatures for the month of July, from maps published in the USWB series (Climates of the States).

Average minimum January temperature:--The mean of the minimum daily temperatures for the month of January, from a map published in the USWB series (Climates of the States).

Basin elevation:--An index of basin elevation in feet above mean sea level, determined by averaging the elevations of 10 percent and of 85 percent of the distance along the main channel.

Basin width:--Drainage area divided by length.

Bifurcation ratio:--Ratio of the number of streams of a given order in a basin to the number of streams of the next higher order in that basin. For this study the ratio of the first-order to second-order streams was used for a sample of about one-fourth of the basin where the first-order stream was defined as a stream with no tributaries on a 1/2 inch = 1 mile State Highway Commission county highway map, and a second-order stream was defined as a stream with two or more first-order tributaries on the same map.

Channel entrenchment depth:--The depth of entrenchment, in feet, of the main channel in a straight reach near the gage. Determined by leveling the difference between the top of the bank in the flood plain and the bottom of the bank where it meets the streambed.

Drainage area:--Total drainage area in square miles shown in the latest Geological Survey water-resources data report, determined by planimeter from the best available maps.

Forested area:--Percentage of drainage area covered by forest and Federal woodland, increased by 1.00 percent. Determined from a compilation of land use on file in the Salina, Kans. State SCS office.

Main channel length:--Thalweg length in miles between gage and basin divide, following the channel which drains the largest basin, as determined from the U.S. Series of Topographic Maps, scale 1:250,000, published by the U.S. Geological Survey.

Mean annual precipitation:--Mean annual precipitation for the period 1921-56, as determined from Theissen polygons for all USWB stations in the study area which had complete records for the 1921-56 period and shown by Furness (1959, table 3).

Meander ratio:--Ratio of the length of the main channel to the length of the valley between the gage and the divide.

Mean discharge:--Map value of 1921-56 mean discharge, from Furness (1966).

Mean-flow length:--The average length in miles of surface and channel flow, determined by imposing a grid of 30 to 40 points over a 1:250,000 map of the basin and determining the average flow lengths from the points to the gage.

100-year 24-hour precipitation:--Determined from USWB Tech. paper no. 40.

Ratio of 1 percent duration to mean flow:--Mapped value of ratio of the 1 percent duration of flow discharge to the mean flow, as determined by Furness (1966).

Slope:--Slope in feet per mile between points 10 percent and 85 percent of the length of the main channel from the gaging station to the divide, as determined from 1:250,000 maps.

Soils infiltration index:--The potential maximum infiltration, in inches, under the average soil moisture conditions preceding an annual flood. Values obtained from information available in State SCS offices.

Streambed grain size:--Median grain size of the sample of streambed material taken near the stream gage, as determined by plotting grain-size analysis on log-probability paper.

Streambed width:--Width of the streambed in feet in a straight reach near the gage, as determined by measuring the distance between the bottoms of the banks.

2-year 24-hour precipitation:--Determined from USWB Tech. paper no. 40.

Valley width:--Width in miles of the river valley at the highway crossing nearest to the gage, as determined from odometer readings.

Variability index:--Map value of variability index, as defined by Lane and Lei (1950) and published in Furness (1966).

APPENDIX II

Glossary of Terms as Used in this Report

Annual value:--The value of a discharge characteristic at a particular gaging station for a particular year such as 1936 peak discharge at station 6-8880.

Dependent variables:--The flow characteristics for which attempts were made to develop regression equations.

Independent variables:--Those physical, climatic, and mapped-flow characteristics which were assumed to influence streamflow.

Mean:--Either the average of annual values used to give a station value or the average of the station values for the 41 stations used in the report.

Pearson type III distribution:--A three-parameter distribution defined by mean, standard deviation, and skew, which may be asymmetrical. The normal distribution is a special degenerate case of the Pearson type III distribution when the skew is zero.

Physical characteristic:--A topographic or geomorphic characteristic measurable either from maps or field inspection.

Skew:--The degree of non-symmetry exhibited by a frequency distribution. Mathematically defined as:

$$g = \frac{n \sum (Xi - \overline{X})^3}{(n-1)(n-2)(SD)^3}$$

where:

- Xi is an item,
- $\overline{X}$ is the mean of the items,
- n is the number of items, and
- SD is standard deviation.

Standard error of estimate:--The standard deviation about the regression surface. Approximately two-thirds of the events lie within one standard error of the regression value.

Standard deviation:--Either the standard deviation of the annual values at the station or the standard deviation of station values for the 41 stations used in this report. Mathematically defined as:

$$SD = \sqrt{\frac{(Xi - \overline{X})^2}{n - 1}}$$

Station value:--A statistical characteristic based on all the available annual records at a particular gaging station such as the 10-year peak discharge at station 6-8880.