Analysis of Surface-Water Data Network in Kansas for Effectiveness in Providing Regional Streamflow Information

By K. D. MEDINA

With a section on

THEORY AND APPLICATION OF GENERALIZED LEAST SQUARES

By GARY D. TASKER

Prepared in cooperation with the
Kansas Water Office

U.S. GEOLOGICAL SURVEY WATER-SUPPLY PAPER 2303

DEPARTMENT OF THE INTERIOR
DONALD PAUL HODEL, Secretary

U.S. GEOLOGICAL SURVEY
Dallas L. Peck, Director

UNITED STATES GOVERNMENT PRINTING OFFICE: 1987

For sale by the Books and Open-File Reports Section, U.S. Geological Survey,
Federal Center, Box 25425, Denver, CO 80225

Library of Congress Cataloging in Publication Data

Medina, K.D.
Analysis of surface-water data network in Kansas for effectiveness in providing regional streamflow information.

(U.S. Geological Survey water-supply paper ; 2303)
Bibliography: p.
Supt. of Docs. no.: I 19.13:2303
1. Stream measurements—Kansas—Analysis. 2. Hydrological stations—Kansas—Analysis. 3. Streamflow—Kansas. 4. Least squares. I. Tasker, Gary D. Theory and application of generalized least squares. 1978. II. Title. III. Series.
GB1225.K2M43 1987 551.48'3'09781 86–600065

CONTENTS

FIGURES

TABLES

Analysis of Surface-Water Data Network in Kansas for Effectiveness in Providing Regional Streamflow Information

By K.D. Medina

Abstract

This report documents the results of an analysis of the surface-water data network in Kansas for its effectiveness in providing regional streamflow information. The network was analyzed using generalized least squares regression. The correlation and time-sampling error of the streamflow characteristic are considered in the generalized least squares method. Unregulated medium-, low-, and high-flow characteristics were selected to be representative of the regional information that can be obtained from streamflow-gaging-station records for use in evaluating the effectiveness of continuing the present network stations, discontinuing some stations, and (or) adding new stations. The analysis used streamflow records for all currently operated stations that were not affected by regulation and for discontinued stations for which unregulated flow characteristics, as well as physical and climatic characteristics, were available. The State was divided into three network areas, western, northeastern, and southeastern Kansas, and analysis was made for the three streamflow characteristics in each area, using three planning horizons.

The analysis showed that the maximum reduction of sampling mean-square error for each cost level could be obtained by adding new stations and discontinuing some current network stations. Large reductions in sampling mean-square error for low-flow information could be achieved in all three network areas, the reduction in western Kansas being the most dramatic. The addition of new stations would be most beneficial for mean-flow information in western Kansas. The reduction of sampling mean-square error for high-flow information would benefit most from the addition of new stations in western Kansas. Southeastern Kansas showed the smallest error reduction in high-flow information. A comparison among all three network areas indicated that funding resources could be most effectively used by discontinuing more stations in northeastern and southeastern Kansas and establishing more new stations in western Kansas.

INTRODUCTION

The long-range goals and objectives of the State of Kansas for management, conservation, and development of water resources are contained in the State Water Plan (K.S.A. 82a–901). As the primary water-resource planning, policy development, and coordination agency, the Kansas Water Office is specifically charged with responsibility for preparing a plan of water-resources development for all areas of the State and for compiling and collecting data and information on the availability and use of water. To assist in meeting these obligations, the U.S. Geological Survey, in cooperation with the Kansas Water Office and other agencies, has established and maintained a network of surface-water data stations throughout the State. The term "network" is not meant to imply a physical interconnection of the streamflow-gaging stations; the data from the stations provide a basis of regional information that can be correlated, and this is the link between stations that makes a network from the set of stations.

Previous studies include an improved program of stream gaging based on the degree of accuracy with which streamflow characteristics can be defined, the amount of data needed, and the most economical method of obtaining the data (Furness, 1957), a reevaluation of the 1957 plan by Jordan and Hedman (1970), and an evaluation for cost effectiveness of schedules of station operation by Medina and Geiger (1984).

Problem

In planning for the development of water resources, certain hydrologic information is desirable. The type of information considered herein is defined as being inversely related to the variance of estimate of a selected streamflow characteristic. However, the optimum level of hydrologic information available from any specific step of the planning process has not been defined. Generalized relations for the benefits derived from and the costs of obtaining regional hydrologic information have not been developed. Also, the reductions in variance derived from additional data have not been compared with the cost of obtaining the additional data.

Objectives

The objectives of the surface-water data network analysis were to provide a quantitative evaluation of the existing data network's ability to obtain optimum regional information concerning selected streamflow characteristics for Kansas, to assess the effects of adding or eliminating streamflow-gaging stations from the network, and to determine how the network can be improved with the least cost for the information gained. The analysis considered pertinent factors that have not been included in the same manner in previous studies, such as interstation correlation and distinction between sampling error and model error for regional estimation methods. The procedures, therefore, were to analyze historical records from streamflow-gaging stations through regional regression methods to determine the sampling mean-square error related to estimates of medium-, low-, and high-flow characteristics based on selected physical and climatic characteristics, and to determine the changes in the sampling mean-square error resulting from adding or eliminating stations from the network.

Description of Surface-Water Data Program

Streamflow data adequate for determining medium-, low-, and high-flow characteristics are available for 235 sites of existing or discontinued streamflow-gaging stations in Kansas. The number of stations used in this study was reduced to 152 for reasons described later in the report. Many stations have record lengths of more than 50 years. Streamflow data have been collected for varied purposes but not usually for the specific purpose of determining regional streamflow characteristics. Some stations have more years of record than were used in the network analysis because streamflow records following regulation upstream were not used. A summary of the periods of record available and the period of unregulated streamflows used in this study are given in table 1 for each station. The minimum period of record used was 4 years.

NETWORK-ANALYSIS PROCEDURE

Description

The basis for this network-analysis procedure is a generalized least squares regression analysis. A regression equation is developed that relates selected physical and climatic characteristics to a streamflow characteristic that is based on recorded observations at all the stations used in a data network. A detailed discussion of the "Theory and Application of Generalized Least Squares" by Gary D. Tasker is given at the end of this report. A feature of the generalized least squares technique that makes it particularly valuable for data-network analysis is that it partitions the prediction mean-square error at a station into a model-error component (the error due to estimating the true streamflow characteristic by the true regression estimate) and a sampling-error component (the error due to estimating the true regression estimate by the sample regression estimate). Only the sampling-error component is affected by increases in record length or by inclusion of new stations, so the network analysis is limited to this component. The generalized least squares concept recognizes the correlation between data at stations that have concurrent periods of record. The individual station variances are adjusted for the effect of interstation correlation in the computation of the sampling mean-square error. The sampling mean-square errors for various network configurations then are compared. A series of computer programs has been developed to make the computations of sampling mean-square error given the appropriate streamflow data and basin characteristics.

Application

A data network was to be evaluated for its ability to estimate all characteristics of streamflow in Kansas. To keep the network-analysis effort within reasonable limits, three specific streamflow characteristics were selected for use in the analysis. Those selected were judged to be representative of the three general categories of medium flow, low flow, and high flow. The three specific streamflow characteristics are the mean flow, the 30-day, 2-year low flow, and the 1-day, 100-year high flow. The values of these streamflow characteristics were determined from streamflow records at each gaging station used in the analysis. Mean-flow values were taken from the annual data report (Geiger and others, 1983) and other previous annual reports. Low-flow values were obtained from Jordan (1983), and high-flow values from Jordan (1984).

An ordinary least squares regression procedure was used as a preliminary screening of physical and climatic characteristics to determine those that were most significant in estimating streamflow characteristics. Those that were shown to be most significant are listed in table 2 and were used in the network analysis.

The selection of a data network depends on the future value of information obtained from a set of data stations. Therefore, it is necessary to define the period of time in the future (called the planning horizon in this analysis) for which the value of the added information would be determined. Three planning horizons were selected for this analysis, the zero-year for the present condition, one at 5 years, to represent short-term information needs, and another at 20 years, to represent long-term information needs.

The computer programs used for the network evaluation have limitations on the number of data stations that can be included in a network. Therefore, it was necessary to divide the network for the State into three areas. The division was based on climatic and hydrologic characteristics and resulted in the western, northeastern, and southeastern

Table 1. Summary of complete-record streamflow-gaging stations and period of record used in network analysis

Station number	Station name	Period of record	Record used in network analysis (water years)
[a] 06814000	Turkey Creek near Seneca	October 1948 to	1949-82
06815600	Wolf River near Hiawatha	March 1961 to June 1970	1962-69
06818200	Doniphan Creek at Doniphan	May 1960 to September 1970	1961-70
[a] 06344700	South Fork Sappa Creek near Brewster	October 1967 to	1968-82
[a] 06844900	South Fork Sappa Creek near Achilles	July 1959 to	1960-82
06845000	Sappa Creek near Oberlin	March 1929 to June 1932 and June 1944 to September 1972	1930-31, 1945-72
06846000	Beaver Creek at Ludell	March 1929 to June 1932 and November 1945 to October 1953	1930-31, 1947-53
06846300	Beaver Creek at Herndon	October 1962 to September 1969	1963-69
[a] 06846500	Beaver Creek at Cedar Bluffs	May 1946 to	1947-82
[a] 06847900	Prairie Dog Creek above Keith Sebelius Lake	June 1962 to	1963-82
06848500	Prairie Dog Creek near Woodruff	October 1928 to September 1932 and October 1944 to	1929-32, 1945-64
[a] 06853800	White Rock Creek near Burr Oak	October 1957 to	1958-82
06854000	White Rock Creek at Lovewell	October 1945 to	1946-56
06855800	Buffalo Creek near Jamestown	July 1959 to	1960-82
06855900	Wolf Creek near Concordia	April 1962 to November 1981	1963-81
06858000	Rose Creek near Wallace	April 1946 to September 1953	1947-53
[a] 06858500	North Fork Smoky Hill River near McAllaster	October 1946 to September 1953 and July 1959 to	1947-53, 1960-82
06859500	Ladder Creek below Chalk Creek near Scott City	April 1951 to September 1979	1952-79
[a] 06860000	Smoky Hill River at Elkader	October 1939 to	1940-82
06860500	Hackberry Creek near Gove	December 1947 to September 1953	1949-53
[a] 06361000	Smoky Hill River near Arnold	February 1950 to	1951-82
06862500	Smoky Hill River near Ellis	December 1941 to September 1952	1943-50
06863300	Big Creek near Ogallah	October 1955 to September 1968	1956-68
[a] 06863500	Big Creek near Hays	April 1946 to	1947-82
06863900	North Fork Big Creek near Victoria	April 1962 to	1963-82
06864500	Smoky Hill River at Ellsworth	April 1895 to October 1905, July 1918 to July 1925, August 1928 to	1919-24, 1929-50
06866900	Saline River near WaKeeney	October 1955 to September 1966 and October 1981 to	1956-66
[a] 06867000	Saline River near Russell	October 1945 to September 1953 and June 1959 to	1946-53, 1960-82
06867500	Paradise Creek near Paradise	April 1945 to September 1953 and October 1962 to September 1974	1947-53, 1963-74
06868000	Saline River near Wilson	May 1929 to September 1963	1930-63
06868400	Wolf Creek near Lucas	June 1959 to September 1971	1960-71
06868500	Wolf Creek near Sylvan Grove	October 1945 to September 1953	1946-53
06868700	North Fork Spillman Creek near Ash Grove	March 1962 to September 1971	1963-71
06869500	Saline River at Tescott	September 1919 to	1920-64
06870300	Gypsum Creek near Gypsum	October 1954 to September 1971	1955-71
[a] 06871000	North Fork Solomon River at Glade	October 1952 to	1953-82
[a] 06871500	Bow Creek near Stockton	November 1950 to	1952-82
06871800	North Fork Solomon River at Kirwin	August 1919 to June 1925, August 1928 to June 1932, and December 1941 to	1920-24, 1929-31, 1943-54
06871900	Deer Creek near Phillipsburg	October 1966 to September 1981	1967-81
06872300	Middle Beaver Creek near Smith Center	April 1961 to September 1970	1962-70
06872500	North Fork Solomon River at Portis	September 1945 to	1946-54
[a] 06873000	South Fork Solomon River above Webster Reservoir	January 1945 to	1946-82
06873500	South Fork Solomon River at Alton	August 1919 to June 1925, August 1928 to June 1932, and June 1942 to September 1957	1920-24, 1929-31, 1943-55
06873700	Kill Creek near Bloomington	March 1963 to September 1981	1964-81
06874000	South Fork Solomon River at Osborne	March 1946 to	1947-55
06875800	Limestone Creek near Glen Elder	October 1965 to June 1971	1966-70
06876000	Solomon River at Beloit	April 1929 to September 1965	1930-54
06876700	Salt Creek near Ada	June 1959 to	1960-82
06877500	Turkey Creek near Abilene	October 1958 to September 1965	1959-65
[a] 06878000	Chapman Creek near Chapman	December 1953 to	1955-82

Table 1. Summary of complete-record streamflow-gaging stations and period of record used in network analysis—Continued

Station number	Station name	Period of record	Record used in network analysis (water years)
06878500	Lyon Creek near Woodbine	December 1953 to September 1974	1955-74
06879200	Clark Creek near Junction City	October 1957 to September 1965	1958-65
a/ 06884200	Mill Creek at Washington	October 1959 to	1960-82
a/ 06884400	Little Blue River near Barnes	April 1958 to	1959-82
06884500	Little Blue River at Waterville	June 1922 to June 1925, August 1928 to April 1958	1923-24, 1929-57
a/ 06885500	Black Vermillion River near Frankfort	October 1953 to	1954-82
06886000	Big Blue River at Randolph	April 1918 to September 1960	1919-59
06886500	Fancy Creek at Winkler	December 1953 to September 1971	1955-71
06888000	Vermillion Creek near Wamego	April 1936 to June 1946, January 1954 to June 1972	1937-45, 1955-71
06888300	Rock Creek near Louisville	October 1958 to September 1965	1959-65
a/ 06888500	Mill Creek near Paxico	December 1953 to	1955-82
06889100	Soldier Creek near Goff	March 1964 to	1965-82
06889120	Soldier Creek near Bancroft	March 1964 to	1965-82
06889140	Soldier Creek near Soldier	March 1964 to	1965-82
06889160	Soldier Creek near Circleville	March 1964 to	1965-82
06889180	Soldier Creek near St. Clere	March 1964 to April 1981	1965-80
06889200	Soldier Creek near Delia	October 1958 to	1959-82
a/ 06889500	Soldier Creek near Topeka	May 1929 to September 1932 and August 1935 to	1930-32, 1936-82
a/ 06890100	Delaware River near Muscotah	July 1969 to	1970-82
06890500	Delaware River at Valley Falls	June 1922 to September 1967	1923-67
06890600	Rock Creek near Meriden	March 1963 to September 1970	1964-70
06891500	Wakarusa River near Lawrence	April 1929 to	1930-77
a/ 06892000	Stranger Creek near Tonganoxie	April 1929 to	1930-82
06893080	Blue River near Stanley	October 1974 to	1975-82
06893300	Indian Creek at Overland Park	March 1963 to	1964-82
06893350	Tomahawk Creek near Overland Park	October 1974 to September 1982	1975-82
a/ 06910800	Marais des Cygnes River near Reading	May 1969 to	1970-82
06911000	Marias des Cygnes River at Melvern	October 1939 to September 1974	1940-72
a/ 06911500	Salt Creek near Lyndon	September 1939 to	1940-82
a/ 06911900	Dragoon Creek near Burlingame	March 1960 to	1961-82
06912000	Switzler Creek at Burlingame	August 1954 to June 1961	1955-60
06912500	Hundred and Ten Mile Creek near Quenemo	September 1939 to	1940-63
06913000	Marais des Cygnes River near Pomona	July 1922 to February 1938 and October 1968 to	1923-37
06913500	Marias des Cygnes River near Ottawa	August 1902 to October 1905 and October 1918 to	1919-63
a/ 06914000	Pottawatomie Creek near Garnett	October 1939 to	1940-82
06915000	Big Bull Creek near Hillsdale	July 1958 to	1958-80
06916000	Marais des Cygnes River at Trading Post	October 1928 to September 1958	1929-58
06916500	Big Sugar Creek at Farlinville	February 1929 to June 1932, November 1948 to September 1958, and July 1959 to September 1970	1930-31, 1950-58, 1960-70
a/ 06917000	Little Osage River at Fulton	November 1948 to	1950-82
a/ 06917380	Marmaton River near Marmaton	May 1971 to	1972-82
07138650	White Woman Creek near Leoti	October 1966 to	1967-82
a/ 07139800	Mulberry Creek near Dodge City	March 1968 to	1969-82
07140700	Guzzlers Gulch near Ness City	April 1961 to October 1980	1962-80
a/ 07141200	Pawnee River near Larned	October 1924 to	1925-82
07141780	Walnut Creek near Rush Center	October 1969 to	1970-82
a/ 07141900	Walnut Creek at Albert	May 1958 to	1959-82
07142300	Rattlesnake Creek near Macksville	October 1959 to	1960-82
07142575	Rattlesnake Creek near Zenith	May 1973 to	1974-82
07142620	Rattlesnake Creek near Raymond	April 1960 to	1961-82
07142860	Cow Creek near Claflin	October 1966 to October 1981	1967-81
07142900	Blood Creek near Boyd	April 1962 to September 1980	1963-80
07143300	Cow Creek near Lyons	April 1938 to September 1951, and October 1961 to	1939-51, 1962-82
07143600	Little Arkansas River near Little River	October 1959 to October 1971	1960-71
07143665	Little Arkansas River at Alta Mills	June 1973 to	1974-82
07144000	East Emma Creek near Halstead	April 1963 to October 1970	1964-70

Table 1. Summary of complete-record streamflow-gaging stations and period of record used in network analysis—Continued

Station number	Station name	Period of record	Record used in network analysis (water years)
a/ 07144200	Little Arkansas River at Valley Center	June 1922 to	1923-82
a/ 07144780	North Fork Ninnescah River above Cheney Reservoir	July 1965 to	1966-82
07144800	North Fork Ninnescah River near Cheney	October 1950 to September 1964	1951-64
07144850	South Fork South Fork Ninnescah River near Pratt	March 1961 to September 1980	1962-80
a/ 07145200	South Fork Ninnescah River near Murdock	August 1950 to September 1959, and June 1964 to	1951-59, 1965-82
07145500	Ninnescah River near Peck	April 1938 to	1939-64
07145700	Slate Creek at Wellington	April 1969 to	1970-82
07146570	Cole Creek near DeGraff	March 1961 to March 1980	1962-79
a/ 07147070	Whitewater River at Towanda	October 1961 to	1962-82
07147100	Whitewater River at Augusta	April 1951 to September 1955	1952-55
07147600	Timber Creek near Wilmot	March 1962 to September 1968	1963-68
07147800	Walnut River at Winfield	December 1921 to	1923-80
a/ 07149000	Medicine Lodge River near Kiowa	May 1895 to October 1896, October 1937 to September 1950, October 1954 to September 1955, and June 1959 to	1938-50, 1955, 1960-82
a/ 07151500	Chikaskia River near Corbin	August 1950 to September 1965 and October 1975 to	1951-65, 1976-82
07155590	Cimarron River near Elkhart	April 1971 to	1972-82
07156010	North Fork Cimarron River at Richfield	April 1971 to	1972-82
07156100	Sand Arroyo Creek near Johnson	April 1971 to	1972-82
07156220	Bear Creek near Johnson	October 1966 to	1967-82
a/ 07157500	Crooked Creek near Nye	August 1942 to	1943-82
07157900	Cavalry Creek near Coldwater	October 1966 to October 1981	1967-81
07165700	Verdigris River near Madison	October 1955 to September 1976	1956-76
07166000	Verdigris River near Coyville	August 1939 to	1940-59
07166500	Verdigris River near Altoona	October 1938 to	1939-59
07167000	Fall River near Eureka	October 1946 to September 1976	1947-76
a/ 07167500	Otter Creek at Climax	August 1946 to	1947-82
07168500	Fall River near Fall River	May 1939 to	1940-48
07169500	Fall River at Fredonia	October 1938 to	1939-48
a/ 07169800	Elk River at Elk Falls	January 1967 to	1968-82
07170000	Elk River near Elk City	October 1938 to September 1969	1939-65
07170500	Verdigris River at Independence	October 1921 to	1922-48
07170700	Big Hill Creek near Cherryvale	October 1957 to	1958-80
a/ 07172000	Caney River near Elgin	October 1938 to	1939-82
07179500	Neosho River at Council Grove	October 1938 to	1939-64
07179600	Four Mile Creek near Council Grove	March 1963 to September 1971	1964-71
07180000	Cottonwood River near Marion	October 1938 to September 1968	1939-67
a/ 07180400	Cottonwood River near Florence	June 1961 to	1962-67
a/ 07180500	Cedar Creek near Cedar Point	October 1938 to	1939-82
07181000	Cottonwood River at Elmdale	October 1922 to September 1932	1923-32
07181500	Middle Creek near Elmdale	October 1938 to September 1950	1939-50
07182000	Cottonwood River at Cottonwood Falls	April 1932 to July 1971	1933-67
07182250	Cottonwood River near Plymouth	March 1963 to	1964-67
07182400	Neosho River at Strawn	October 1948 to June 1963	1949-62
07183000	Neosho River near Iola	February 1898 to December 1903 and October 1917 to	1918-62
07183100	Owl Creek near Piqua	July 1959 to October 1970	1960-70
07183500	Neosho River near Parsons	October 1921 to	1922-62
a/ 07184000	Lightning Creek near McCune	October 1938 to September 1946, and October 1959 to	1939-46, 1960-82
07184500	Labette Creek near Oswego	October 1938 to September 1945	1939-45

a Station that must continue in operation.

Table 2. Hydrologic, climatic, and economic characteristics at streamflow-gaging stations used in network analysis

Station number	Station name	Drainage area (square miles)	Channel slope (feet per mile)	Stream length (miles)	Average annual precipitation (inches)	50-year 24-hour precipitation (inches)	Representative data cost ratio
06814000	Turkey Creek near Seneca	276.0	5.89	55.4	34.0	6.5	0.98
06815600	Wolf River near Hiawatha	41.0	10.0	18.6	36.0	6.5	--
06818200	Doniphan Creek at Doniphan	4.15	44.6	3.94	37.0	6.6	--
06844700	South Fork Sappa Creek near Brewster	74.0	10.8	24.6	19.0	4.9	0.98
06844900	South Fork Sappa Creek near Achilles	446.0	7.0	124.9	19.5	4.9	0.98
06845000	Sappa Creek near Oberlin	1,063.0	7.33	144.1	19.5	4.9	-
06846000	Beaver Creek at Ludell	1,460.0	8.11	140.3	20.5	4.8	--
06846300	Beaver Creek at Herndon	1,535.0	8.0	171.0	21.0	4.8	--
06846500	Beaver Creek at Cedar Bluffs	1,618.0	7.72	181.0	19.0	4.9	0.98
06847900	Prairie Dog Creek above Keith Sebelius Lake	590.0	7.11	139.7	21.5	5.1	1.05
06848500	Prairie Dog Creek near Woodruff	1,007.0	5.61	213.0	21.5	5.1	--
06853800	White Rock Creek near Burr Oak	227.0	6.95	40.3	26.5	5.8	1.05
06854000	White Rock Creek at Lovewell	345.0	6.12	58.0	27.0	5.8	--
06855800	Buffalo Creek near Jamestown	330.0	6.15	50.0	27.5	5.9	--
06855900	Wolf Creek near Concordia	56.0	8.79	19.0	28.5	6.1	--
06858000	Rose Creek near Wallace	28.5	8.0	13.6	18.0	4.9	--
06858500	North Fork Smoky Hill River near McAllaster	670.0	7.84	181.9	17.5	4.7	0.98
06859500	Ladder Creek below Chalk Creek near Scott City	1,460.0	6.87	223.0	18.0	5.0	-
06860000	Smoky Hill River at Elkader	3,555.0	13.2	149.3	18.0	4.9	1.05
06860500	Hackberry Creek near Gove	460.0	6.71	106.2	20.0	5.2	--
06861000	Smoky Hill River near Arnold	5,220.0	11.4	194.0	19.0	5.0	1.05
06862500	Smoky Hill River near Ellis	5,630.0	10.9	224.0	19.5	5.0	--
06863300	Big Creek near Ogallah	297.0	6.27	109.2	22.0	5.5	--
06863500	Big Creek near Hays	594.0	5.8	174.1	22.5	5.6	0.98
06863900	North Fork Big Creek near Victoria	54.0	8.3	27.8	24.0	5.8	0.98
06864500	Smoky Hill River at Ellsworth	7,580.0	8.95	327.0	21.0	5.3	--
06866900	Saline River near WaKeeney	696.0	7.17	162.8	21.0	5.2	0.98
06867000	Saline River near Russell	1,502.0	6.86	277.8	22.5	5.5	0.98
06867500	Paradise Creek near Paradise	212.0	7.29	75.7	24.5	5.8	--
06868000	Saline River near Wilson	1,900.0	6.28	314.2	23.0	5.6	--
06868400	Wolf Creek near Lucas	163.0	16.4	21.8	23.5	5.9	--
06868500	Wolf Creek near Sylvan Grove	261.0	11.5	31.5	23.5	5.9	--
06868700	North Fork Spillman Creek near Ash Grove	26.1	14.0	14.6	26.0	6.0	--
06869500	Saline River at Tescott	2,820.0	5.02	384.8	24.0	5.7	--
06870300	Gypsum Creek near Gypsum	120.0	9.54	27.7	31.0	6.5	--
06871000	North Fork Solomon River at Glade	849.0	7.79	206.9	22.5	5.3	1.05
06871500	Bow Creek near Stockton	341.0	6.73	123.3	22.5	5.4	0.98
06871800	North Fork Solomon River at Kirwin	1,367.0	7.60	222.7	23.0	5.4	--
06871900	Deer Creek near Phillipsburg	65.0	16.5	21.4	24.0	5.5	--
06872300	Middle Beaver Creek near Smith Center	71.0	11.1	27.9	25.0	5.7	--
06872500	North Fork Solomon River at Portis	2,315.0	7.29	263.0	23.5	5.5	--
06873000	South Fork Solomon River above Webster Reservoir	1,040.0	8.29	183.4	22.0	5.4	1.05
06873500	South Fork Solomon River at Alton	1,720.0	8.38	235.9	23.0	5.4	--
06873700	Kill Creek near Bloomington	52.0	10.9	24.7	25.0	5.8	--
06874000	South Fork Solomon River at Osborne	2,012.0	7.93	262.2	23.0	5.5	--
06875800	Limestone Creek near Glen Elder	210.0	6.66	46.0	26.5	5.8	--
06876000	Solomon River at Beloit	5,530.0	6.30	330.0	24.5	5.6	--
06876700	Salt Creek near Ada	384.0	4.65	73.7	27.0	6.1	--
06877500	Turkey Creek near Abilene	143.0	6.67	34.6	32.0	6.5	--
06878000	Chapman Creek near Chapman	300.0	4.25	58.0	30.5	6.3	0.98
06878500	Lyon Creek near Woodbine	230.0	5.45	49.1	32.5	6.6	--
06879200	Clark Creek near Junction City	200.0	6.12	52.9	33.0	6.6	--
06884200	Mill Creek at Washington	344.0	4.58	62.4	30.5	6.1	0.98
06884400	Little Blue River near Barnes	3,324.0	4.33	235.6	29.0	5.8	1.05
06884500	Little Blue River at Waterville	3,509.0	4.26	247.1	29.0	5.8	--
06885500	Black Vermillion River near Frankfort	410.0	5.72	40.3	33.5	6.4	1.05
06886000	Big Blue River at Randolph	9,100.0	2.69	265.0	29.5	5.9	--
06886500	Fancy Creek at Winkler	174.0	8.4	38.0	32.5	6.3	--

Table 2. Hydrologic, climatic, and economic characteristics at streamflow-gaging stations used in network analysis—Continued

Station number	Station name	Drainage area (square miles)	Channel slope (feet per mile)	Stream length (miles)	Average annual precipi- tation (inches)	50-year 24-hour precipi- tation (inches)	Represent- ative data cost ratio
06888000	Vermillion Creek near Wamego	243.0	5.5	44.2	34.5	6.5	--
06888300	Rock Creek near Louisville	128.0	10.6	32.2	33.5	6.5	--
06888500	Mill Creek near Paxico	316.0	10.5	40.1	34.5	6.7	0.98
06889100	Soldier Creek near Goff	2.06	25.1	2.94	35.0	6.6	0.98
06889120	Soldier Creek near Bancroft	10.5	18.0	6.24	35.0	6.6	0.98
06889140	Soldier Creek near Soldier	16.9	14.6	9.29	35.0	6.6	0.98
06889160	Soldier Creek near Circleville	49.3	10.8	20.2	35.0	6.6	0.98
06889180	Soldier Creek near St. Clere	80.0	9.2	31.2	35.0	6.6	--
06889200	Soldier Creek near Delia	157.0	6.56	54.7	35.5	6.7	0.98
06889500	Soldier Creek near Topeka	290.0	5.55	71.0	35.5	6.7	1.05
06890100	Delaware River near Muscotah	430.0	5.80	52.4	36.0	6.7	1.15
06890500	Delaware River at Valley Falls	922.0	4.63	68.6	36.0	6.6	--
06890600	Rock Creek near Meriden	22.0	11.9	12.8	36.5	6.7	--
06891500	Wakarusa River near Lawrence	425.0	3.78	70.3	37.0	6.9	--
06892000	Stranger Creek near Tonganoxie	406.0	2.86	73.9	37.0	6.8	1.05
06893080	Blue River near Stanley	46.0	15.0	12.4	40.0	7.0	0.98
06893300	Indian Creek at Overland Park	26.6	12.1	16.5	39.0	6.9	0.98
06893350	Tomahawk Creek near Overland Park	23.9	16.8	11.0	40.0	7.0	--
06910800	Marais des Cygnes River near Reading	177.0	6.21	43.7	37.0	7.0	1.15
06911000	Marias des Cygnes River at Melvern	351.0	4.17	75.2	36.0	7.0	--
06911500	Salt Creek near Lyndon	111.0	5.80	38.0	37.0	6.9	0.98
06911900	Dragoon Creek near Burlingame	114.0	6.63	39.2	36.0	6.8	1.05
06912000	Switzler Creek at Burlingame	26.3	11.4	13.8	37.0	6.8	--
06912500	Hundred and Ten Mile Creek near Quenemo	322.0	6.7	34.0	37.0	6.9	--
06913000	Marais des Cygnes River near Pomona	1,040.0	3.41	104.0	37.0	7.0	--
06913500	Marias des Cygnes River near Ottawa	1,250.0	2.84	124.0	37.0	7.0	--
06914000	Pottawatomie Creek near Garnett	334.0	4.40	50.0	39.5	7.2	0.98
06915000	Big Bull Creek near Hillsdale	147.0	8.12	24.2	39.5	7.0	--
06916000	Marais des Cygnes River at Trading Post	2,880.0	2.08	207.7	39.0	7.1	--
06916500	Big Sugar Creek at Farlinville	198.0	8.03	35.5	40.5	7.2	--
06917000	Little Osage River at Fulton	295.0	4.97	51.5	40.5	7.3	0.98
06917380	Marmaton River near Marmaton	292.0	5.89	46.4	41.0	7.4	1.15
07138650	White Woman Creek near Leoti	750.0	12.6	80.8	17.0	5.1	0.98
07139800	Mulberry Creek near Dodge City	73.8	7.3	25.8	21.0	5.2	0.98
07140700	Guzzlers Gulch near Ness City	58.2	9.64	34.6	21.5	5.7	--
07141200	Pawnee River near Larned	2,148.0	4.18	172.0	21.5	5.7	0.98
07141780	Walnut Creek near Rush Center	1,256.0	5.97	152.0	22.0	5.9	0.98
07141900	Walnut Creek at Albert	1,410.0	5.36	179	22.0	5.7	0.98
07142300	Rattlesnake Creek near Macksville	784.0	4.96	94.8	23.0	6.1	0.98
07142575	Rattlesnake Creek near Zenith	1,052.0	4.10	165	25.5	6.3	0.98
07142620	Rattlesnake Creek near Raymond	1,167.0	4.10	179	26.0	6.2	0.98
07142860	Cow Creek near Claflin	43.0	6.73	17.6	25.0	6.1	--
07142900	Blood Creek near Boyd	61.0	9.82	19.2	24.5	6.0	--
07143300	Cow Creek near Lyons	728.0	3.44	64.1	26.5	6.2	0.98
07143600	Little Arkansas River near Little River	71.0	8.32	17.7	27.5	6.4	--
07143665	Little Arkansas River at Alta Mills	736.0	3.58	73.8	31.0	6.7	0.98
07144000	East Emma Creek near Halstead	58.0	9.00	17.3	32.0	6.7	--
07144200	Little Arkansas River at Valley Center	1,327.0	2.30	124.0	30.5	6.6	0.98
07144780	North Fork Ninnescah River above Cheney Reservoir	787.0	5.85	72.2	26.5	6.5	0.98
07144800	North Fork Ninnescah River near Cheney	930.0	5.36	95.5	26.5	6.5	--
07144850	South Fork South Fork Ninnescah River near Pratt	21.0	10.6	13.4	24.0	6.4	--
07145200	South Fork Ninnescah River near Murdock	650.0	7.13	94.7	26.0	6.6	0.98
07145500	Ninnescah River near Peck	2,129.0	4.80	128.0	27.0	6.6	--
07145700	Slate Creek at Wellington	154.0	6.08	43.0	30.5	7.1	0.98
07146570	Cole Creek near DeGraff	30.0	7.36	17.7	35.0	7.0	--

Table 2. Hydrologic, climatic, and economic characteristics at streamflow-gaging stations used in network analysis—Continued

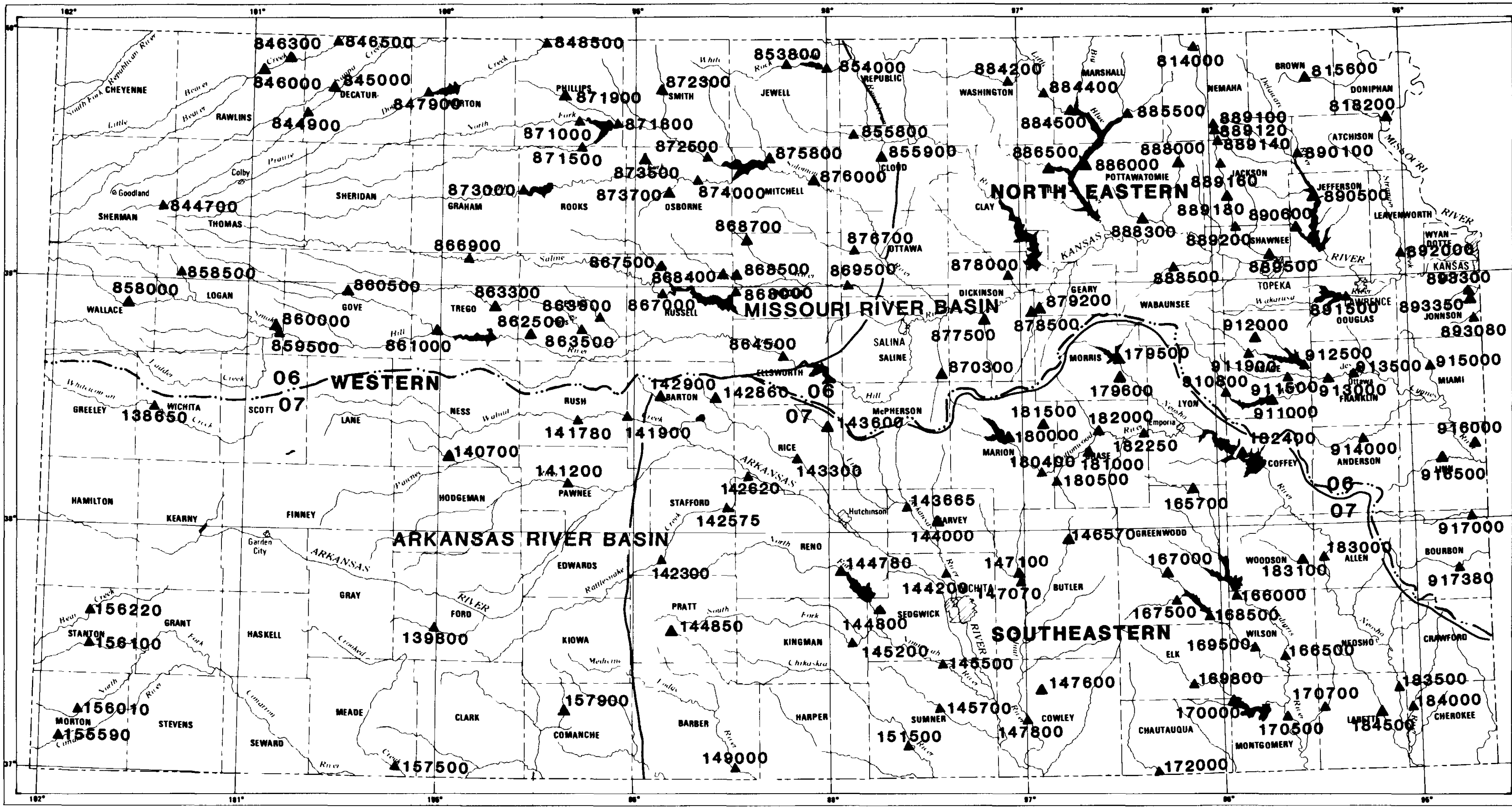
Station number	Station name	Drainage area (square miles)	Channel slope (feet per mile)	Stream length (miles)	Average annual precipitation (inches)	50-year 24-hour precipitation (inches)	Representative data cost ratio
07147070	Whitewater River at Towanda	426.0	4.15	49.2	33.5	6.9	0.98
07147100	Whitewater River at Augusta	456.0	3.20	55.9	33.8	7.1	--
07147600	Timber Creek near Wilmot	63.0	9.90	27.0	35.0	7.2	--
07147800	Walnut River at Winfield	1,880.0	2.50	128.0	34.5	7.1	--
07149000	Medicine Lodge River near Kiowa	903.0	8.27	108.1	25.0	6.5	0.98
07151500	Chikaskia River near Corbin	794.0	7.79	90.9	27.5	6.9	0.98
07155590	Cimmarron River near Elkhart	2,899.0	17.5	177.1	18.0	5.3	0.98
07156010	North Fork Cimmarron River at Richfield	463.0	16.5	85.4	17.6	5.3	0.98
07156100	Sand Arroyo Creek near Johnson	619.0	15.2	125.3	17.0	5.2	0.98
07156220	Bear Creek near Johnson	835.0	13.9	122.0	17.5	5.2	0.98
07157500	Crooked Creek near Nye	1,157.0	4.23	127.0	21.0	5.8	0.98
07157900	Cavalry Creek near Coldwater	39.0	8.61	17.5	23.3	6.2	--
07165700	Verdigris River near Madison	181.0	11.2	39.2	36.5	7.1	--
07166000	Verdigris River near Coyville	747.0	4.98	91.6	37.0	7.2	--
07166500	Verdigris River near Altoona	1,138.0	3.33	134.6	37.5	7.2	--
07167000	Fall River near Eureka	307.0	9.95	38.9	36.0	7.1	--
07167500	Otter Creek at Climax	129.0	13.2	27.8	36.0	7.2	0.98
07168500	Fall River near Fall River	585.0	6.28	61.0	36.5	7.3	--
07169500	Fall River at Fredonia	827.0	5.46	75.8	36.5	7.4	--
07169800	Elk River at Elk Falls	220.0	9.21	41.1	36.2	7.4	0.98
07170000	Elk River near Elk City	575.0	5.25	74.6	36.5	7.3	--
07170500	Verdigris River at Independence	2,892.0	2.68	168.4	37.0	7.3	--
07170700	Big Hill Creek near Cherryvale	37.0	9.10	24.2	38.5	7.6	--
07172000	Caney River near Elgin	445.0	7.39	60.6	35.0	7.4	0.98
07179500	Neosho River at Council Grove	250.0	4.88	29.2	34.0	6.7	--
07179600	Four Mile Creek near Council Grove	55.0	15.5	18.5	34.0	6.8	--
07180000	Cottonwood River near Marion	329.0	5.54	40.9	32.5	6.7	--
07180400	Cottonwood River near Florence	754.0	4.52	62.9	34.0	6.9	1.15
07180500	Cedar Creek near Cedar Point	110.0	9.42	18.2	34.0	6.9	0.98
07181000	Cottonwood River at Elmdale	1,045.0	3.74	83.7	33.0	6.9	--
07181500	Middle Creek near Elmdale	92.0	3.69	45.3	34.0	6.8	--
07182000	Cottonwood River at Cottonwood Falls	1,327.0	3.19	96.5	33.5	6.8	--
07182250	Cottonwood River near Plymouth	1,740.0	2.80	109.0	36.0	6.9	--
07182400	Neosho River at Strawn	2,933.0	2.75	114.7	35.0	6.9	--
07183000	Neosho River near Iola	3,818.0	1.84	190.6	36.5	7.0	--
07183100	Owl Creek near Piqua	177.0	5.87	28.4	38.5	7.3	--
07183500	Neosho River near Parsons	4,905.0	1.85	303.0	38.0	7.1	--
07184000	Lightning Creek near McCune	197.0	3.43	45.8	40.5	7.5	0.98
07184500	Labette Creek near Oswego	211.0	4.74	34.2	39.0	7.6	--

areas shown in figure 1. The geographical distribution of the gaging stations used also is shown in figure 1. A separate network analysis was completed for each of the three areas.

An additional limitation was that a truly realistic set of interstation correlation coefficients could not be determined, and the matrix computations could be performed only when equal coefficients were used. Therefore within each area, constant correlation coefficients that approximated the average values were used: 0.5 for mean flow, 0.6 for low flow, and 0.4 for high flow. Work is ongoing to relate the correlation coefficients to distances between gaging stations, and this approach may prove fruitful in the future for defining the correlation-coefficient matrix.

At gaging stations where reservoir operations control downstream discharge, streamflow cannot be predicted from physical and climatic characteristics; therefore, only the unregulated parts of those records were used for regressions. The amount of regulation on streamflow imposed additional limits on the stations to be included in the analysis. The stations on the larger streams where the drainage area exceeds 10,000 square miles were not included because there are no streams of this size that are not already affected by regulation of flows by reservoirs and diversions.

A representative annual cost was assigned to the information that was obtained from each gaging station. Costs were based on 1984 expenses for operation, maintenance, and data compilation at a gaging station. At stations that have been discontinued and at stations where flow is now regulated, cost was irrelevant because the decision to discontinue the station has already been made or the flow is



EXPLANATION

———	NETWORK BOUNDARY
▲ 887500	COMPLETE-RECORD STREAMFLOW-GAGING STATION AND LAST SIX DIGITS OF NUMBER
—··06/07··—	BASIN BOUNDARY--06, first two digits of station number in Missouri River basin; 07, first two digits of station number in Arkansas River basin

Figure 1. Location of complete-record streamflow-gaging stations and network areas used in surface-water network analysis.

now regulated; therefore, the stations can provide no new data for regional information. Representative costs of obtaining information for the stations used in this study are included in table 2 as ratios to the average cost for all stations.

Subject to the limitations described above, a generalized least squares regression procedure was used to calculate the coefficients for regressions of three streamflow characteristics based on the selected physical and climatic characteristics. The values of the streamflow characteristics computed from the resulting regression equations were compared with values obtained from the actual streamflow records. The sampling mean-square error from this comparison was used to evaluate the ability of the network to estimate regional streamflow characteristics. The network sampling mean-square error was determined for possible future gaging operations ranging from operation of all current stations plus new stations to future operation of just one station.

The generalized least squares analysis provides a measure of the deviation of the combination of physical and climatic characteristics for each station from the mean combination for all stations. The five stations that had the largest deviations were selected as examples of new stations to be used for each streamflow characteristic. Each new station is a fictitious station with physical and climatic characteristics similar to an existing or former station. The new stations were chosen because it is believed that actual new stations with physical and climatic characteristics similar to the fictitious new stations would be the most effective stations (in terms of improving the regression models) to add to the network.

RESULTS OF NETWORK ANALYSIS

The regional regression equation that relates physical and climatic characteristics to selected flow characteristics is the basis for determining the effectiveness of the network. The equations developed for the three separate areas of the network are not considered the best for predictive purposes because the analysis did not consider all feasible combinations and transformations of the independent variables. The purpose of the analysis was to use regression equations for analyzing the sampling mean-square error for indications of effectiveness, assuming that the physical and climatic characteristics chosen are representative. An analysis of the relative impacts of adding or discontinuing streamflow-gaging stations as measured by relative changes in network average sampling mean-square error is described. Where sampling mean-square error is mentioned subsequently in this report, it should be understood to mean the average sampling mean-square error for a data network.

Each gaging station contributes a share of the overall information provided by the network. However, the amount of information provided by each station depends on the variability of streamflow, the combination of physical and climatic characteristics, and the length of record at the end of each planning horizon. Therefore, the data at each station will have a unique impact on the sampling mean-square error for the network. Also, the cost of obtaining the information varies among stations. Therefore, the results of the network analysis indicate the relative contribution to reducing the sampling mean-square error versus cost. The results are summarized in graphs similar to figure 2.

Pertinent features of the graphs that summarize the results for the different network areas, streamflow characteristics, and network strategies are explained in figure 2. A graph including new stations has been used as the example in order to show the features. The stations are plotted as points of sampling mean-square error versus the ratio of total annual cost to 1984 cost. Each point represents the sampling mean-square error that would result and the annual cost if the station represented by the plotted point, plus all stations plotted to the left of it, were operated after 1983 for the number of years of the planning horizon. The cost ratio is the sum of the individual station costs expressed as a ratio to the summation of all station costs for the currently operated stations used in the analysis for the particular flow characteristic. The zero-year horizon (status at the end of the 1983 water year) is shown as a plotted point on the y-axis.

The points representing sampling mean-square error are arranged on the graphs so that the station that is most effective in reducing the sampling mean-square error is at the left (after the stations that must continue in operation), and each station toward the right is progressively less effective. Stations that must be operated for purposes other than regional information (such as project operation, hydrologic forecasts, and interstate compact administration) are not plotted individually because they are considered as a group that must be operated, in contrast to stations that could be individually selected for discontinuance. Included in the group of stations that cannot be discontinued are unregulated stations that are used to identify long-term trends and stations that are used to collect data at a site for a specific and current purpose. The point plotted within the space reserved for stations that must be operated represents the total contribution of that group of stations.

The series of symbols plotted on each graph will be referred to as "curves," even though the implied curves have not been drawn. The steep part of each curve represents the stations that are the most effective in reducing the sampling mean-square error; in this example, some of those stations are new. The flat part of the curve represents stations whose future operation would contribute very little to reducing the sampling mean-square error and could be considered for discontinuance, with their costs applied toward new stations that would contribute more toward reducing the sampling mean-square error.

Table 1 identifies the stations that must continue in operation, and a table of results for each area identifies the

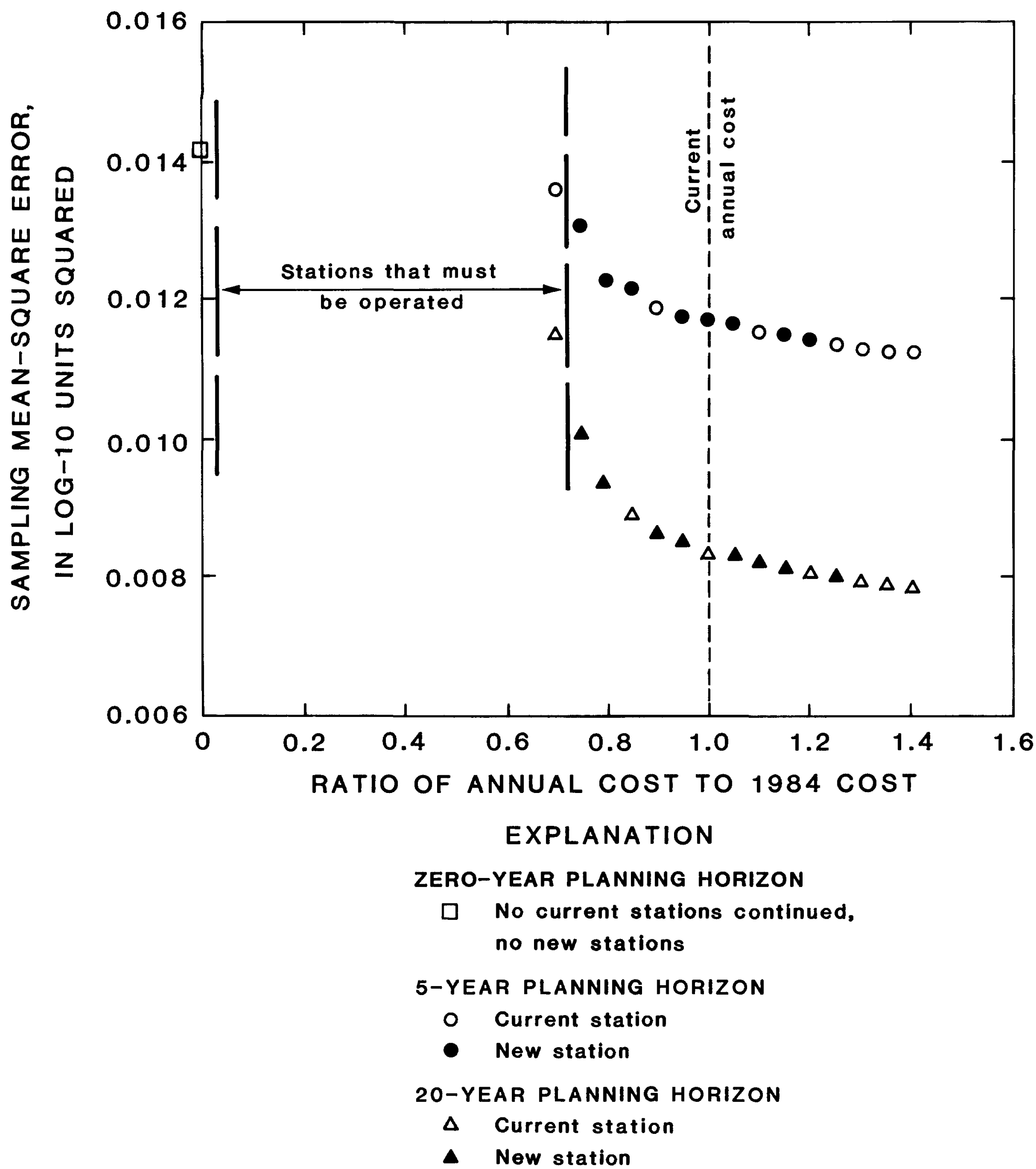


Figure 2. Explanation of pertinent features for graphs of sampling mean-square error versus cost.

order of effectiveness of the other stations (tables 3–5). The composite ranking for the currently operated stations and the new stations of an area is based on the relative ranking of the station for each streamflow characteristic using the 20-year planning horizon. For the case in which a station was not used for low or high flow, the rank for mean flow was used as the best estimate of the rank. Tables 3–5 list the stations according to the composite ranking, in descending order of importance. The composite ranking provides a means of ranking stations in order of priority, assuming that all flow characteristics are of equal importance.

Western Kansas

Results of surface-water network analyses for western Kansas are illustrated in figures 3–5. Figure 3 illustrates the results of three network strategies to provide regional information on mean flow and shows the effectiveness of the stations in reducing the sampling mean-square error. Figure 3*A* shows the currently operated stations in the order of their effectiveness (ignoring requirements to continue certain stations for purposes other than regional information). If the short-term (5-year) planning horizon is considered, most of the annual data-collection cost would provide little reduction in the sampling mean-square error. The long-term (20-year) planning horizon indicates more reduction of error, yet many of the stations would contribute much less regional information than others.

Figure 3*B* shows the relative effectiveness of the current stations that are operated only for regional information and therefore are candidates for discontinuance. The stations that must be operated are not shown but do contribute regional information that reduces the sampling mean-square error and also provide information for other data uses. This graph shows that the annual data-collection cost could be reduced to about 0.8 of the 1984 cost by discontinuing several stations, without sacrificing a significant amount of reduction of the sampling mean-square error.

Figure 3*C* shows that new stations, selected for their particular combinations of physical and climatic characteristics, would have a dramatic effect on the reduction of the sampling mean-square error. Since a new gaging station would provide information for all three streamflow characteristics, the new stations for low and high flow were considered also. For the short-term (5-year) planning horizon, the new stations would provide virtually all the reduction of

Table 3. Station ranking in order of importance in providing regional streamflow information for western Kansas
[Stations that must continue are not included; they are identified in table 1]

Station number	Station name	Station ranking for streamflow characteristic indicated: 5-year planning horizon, Mean flow	5-year, Low flow	5-year, High flow	20-year planning horizon, Mean flow	20-year, Low flow	20-year, High flow	Composite station ranking, 20-year
[a]/ 06862500	New station	1	1	3	1	1	3	1
[a]/ 06861000	New station	2	2	1	2	2	2	2
[a]/ 06855000	New station	3	3	2	3	3	1	3
[a]/ 07138650	New station	4	4	7	4	4	7	4
[a]/ 06855500	New station	5	6	4	5	6	5	5
[a]/ 06875000	New station	8	5	5	8	5	4	6
[a]/ 06858000	New station	6	7	6	7	8	6	7
[a]/ 06858500	New station	7	8	8	6	7	10	8
[a]/ 07141780	New station	10	10	9	9	10	9	9
[a]/ 07141200	New station	9	9	11	10	9	11	10
[a]/ 07157900	New station	11	11	10	11	11	8	11
07138650	White Woman Creek near Leoti	12	12	12	12	12	12	12
07156100	Sand Arroyo Creek near Johnson	13	13	(b)	13	13	(b)	13
07155590	Cimarron River near Elkhart	14	(b)	(b)	14	(b)	(b)	14
07156010	North Fork Cimarron River at Richfield	15	14	(b)	15	14	(b)	15
07141780	Walnut Creek near Rush Center	16	(b)	13	16	(b)	13	16
07156220	Bear Creek near Johnson	17	15	14	17	15	14	17
06866900	Saline River near WaKeeney	18	16	16	18	16	16	18
06863900	North Fork Big Creek near Victoria	19	17	15	19	17	15	19

[a] New station having basin characteristics similar to those of station number indicated.

[b] Station records not used in analysis for indicated streamflow characteristic.

Table 4. Station ranking in order of importance in providing regional streamflow information for northeastern Kansas
[Stations that must continue are not included; they are identified in table 1]

Station number	Station name	Station ranking for streamflow characteristics indicated						Composite station ranking
		5-year planning horizon			20-year planning horizon			
		Mean flow	Low flow	High flow	Mean flow	Low flow	High flow	20-year
a/ 06889100	New station	1	2	2	1	2	1	1
06889100	Soldier Creek near Goff	2	9	4	2	3	3	2
a/ 06870300	New station	3	1	5	3	1	5	3
a/ 06886000	New station	4	6	3	4	8	4	4
a/ 06892000	New station	10	4	1	11	5	2	5
06893080	Blue River near Stanley	9	(b)	(b)	6	(b)	(b)	6
a/ 06890500	New station	13	3	7	9	4	7	7
06889120	Soldier Creek near Bancroft	5	10	8	5	9	6	8
a/ 06884400	New station	6	5	6	10	6	8	9
a/ 06913500	New station	7	8	10	8	10	9	10
06893300	Indian Creek at Overland Park	11	11	12	7	11	12	11
a/ 06884500	New station	12	7	9	13	7	11	12
06889140	Soldier Creek near Soldier	8	12	11	12	12	10	13
06889200	Soldier Creek near Delia	15	13	14	15	13	13	14
06889160	Soldier Creek near Circleville	14	14	13	14	14	14	15

a New station having basin characteristics similar to those of station number indicated.

b Station records not used in analysis for indicated streamflow characteristics.

Table 5. Station ranking in order of importance in providing regional streamflow information for southeastern Kansas
[Stations that must continue are not included; they are identified in table 1]

Station number	Station name	Station ranking for streamflow characteristic indicated						Composite station ranking
		5-year planning horizon			20-year planning horizon			
		Mean flow	Low flow	High flow	Mean flow	Low flow	High flow	20-year
a/ 07147600	New station	1	1	1	1	1	1	1
a/ 07144850	New station	2	2	2	2	2	2	2
a/ 07181500	New station	3	7	5	3	4	3	3
a/ 07183500	New station	4	5	3	4	6	4	4
a/ 07142620	New station	6	3	7	6	3	5	5
a/ 07170700	New station	5	4	4	5	5	6	6
a/ 07180500	New station	7	6	6	7	7	7	7
a/ 07149000	New station	9	8	8	9	8	8	8
07142620	Rattlesnake Creek near Raymond	8	9	9	8	9	9	9
07142575	Rattlesnake Creek near Zenith	10	(b)	(b)	10	(b)	(b)	10
07145700	Slate Creek at Wellington	11	10	10	11	10	10	11
07143300	Cow Creek near Lyons	14	12	12	12	12	11	12
07143665	Little Arkansas River at Alta Mills	12	(b)	(b)	13	(b)	(b)	13
07142300	Rattlesnake Creek near Macksville	13	11	11	14	11	12	14

a New station having basin characteristics similar to those of station number indicated.

b Station records not used in analysis for indicated streamflow characteristic.

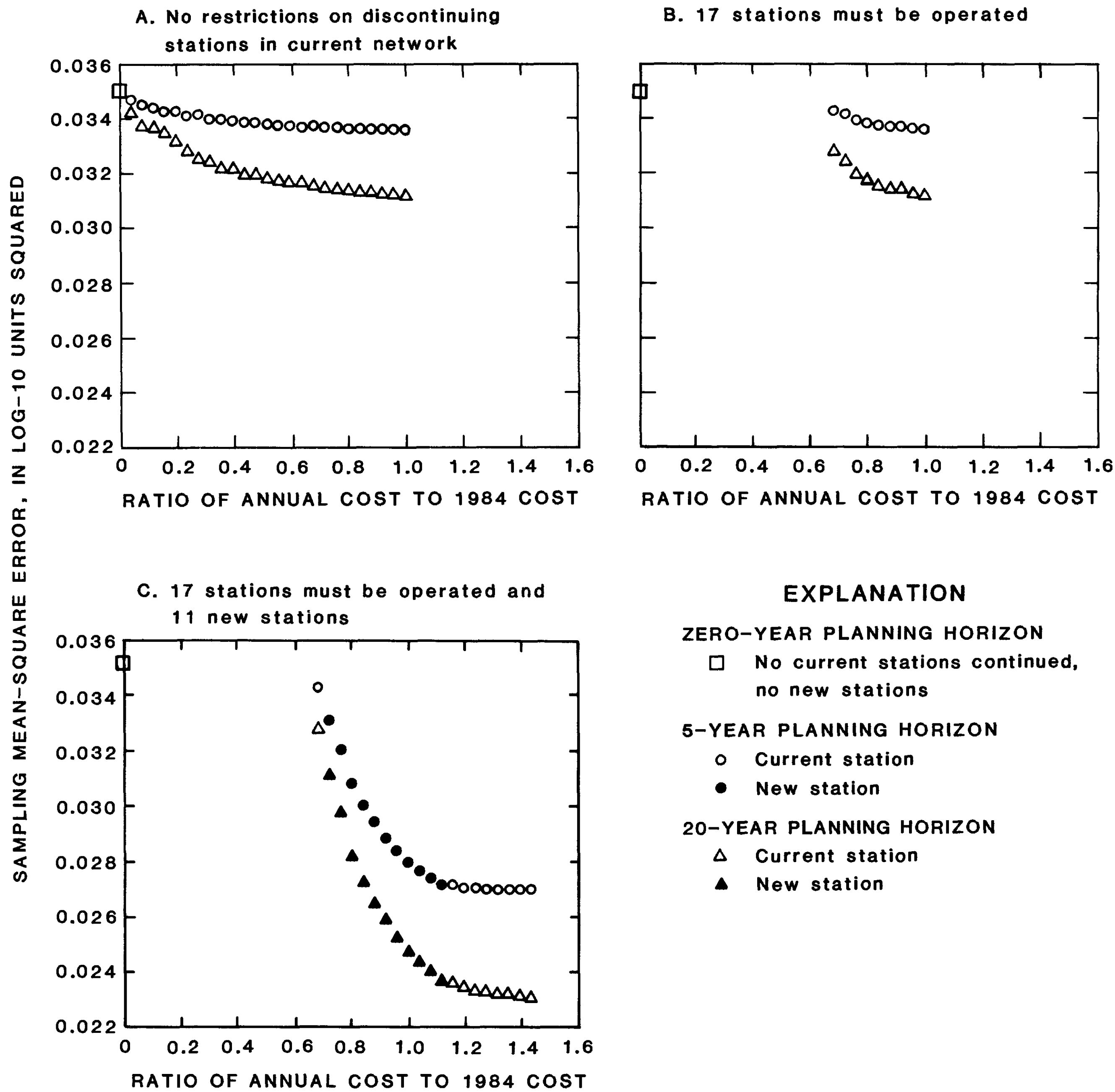


Figure 3. Results of three network strategies to provide regional information on mean flow in western Kansas.

error that could be achieved. For the long-term (20-year) planning horizon, existing stations would provide some of the reduction but the new stations would be by far the most effective. The new stations that were deemed most effective for low and high flow were also effective in the error reduction.

Figure 4 shows the results of three network strategies to provide regional information on low flow in western Kansas. Although the sampling mean-square error (0.152) is substantially higher than for mean flow, there is little difference between the 5-year and 20-year results. The curves in figures 4*A* and 4*B* are relatively flat, indicating that little additional information on low flow can be gained by continuing the current stations. Figure 4*B*, when compared with figure 4*A*, shows that the stations currently being operated for regional information as the principal purpose are more effective for that purpose than the stations operated for other purposes. Figure 4*C* shows the dramatic effect of

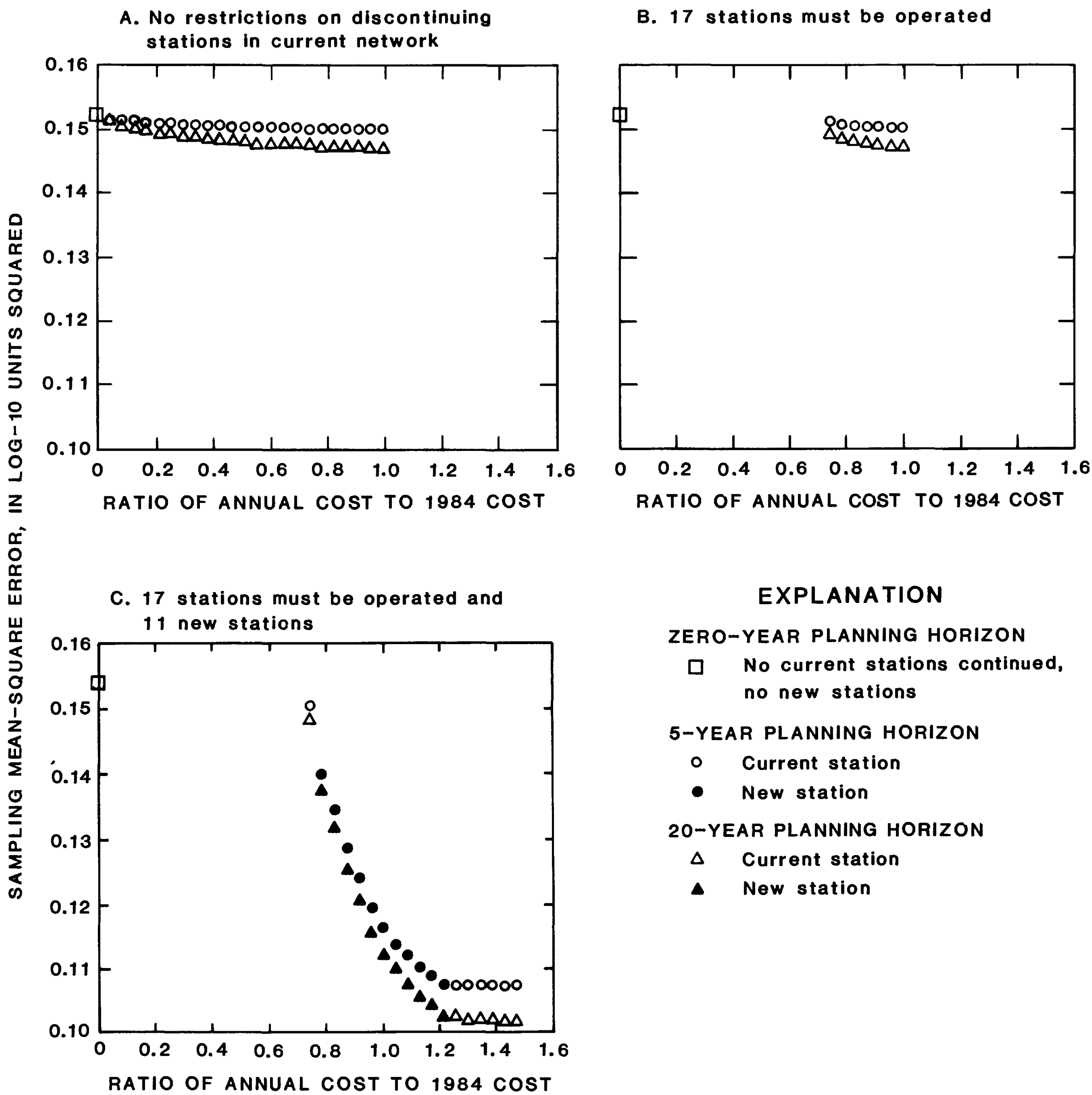


Figure 4. Results of three network strategies to provide regional information on low flow in western Kansas.

adding new stations for low-flow information. The stations for mean and high flow were also effective in reducing the sampling mean-square error. As table 3 shows, the new stations for low-flow information do not coincide with the new stations for mean-flow information.

Figure 5 shows the three network strategies for providing regional information on high flow in western Kansas. The three graphs indicate that, if high flow is the regional information desired, continuation of all stations plus new stations would be beneficial in reducing the sampling mean-square error, particularly in the long-term (20-year) planning horizon. The new stations for mean and low flow also were effective in reducing the sampling mean-square error.

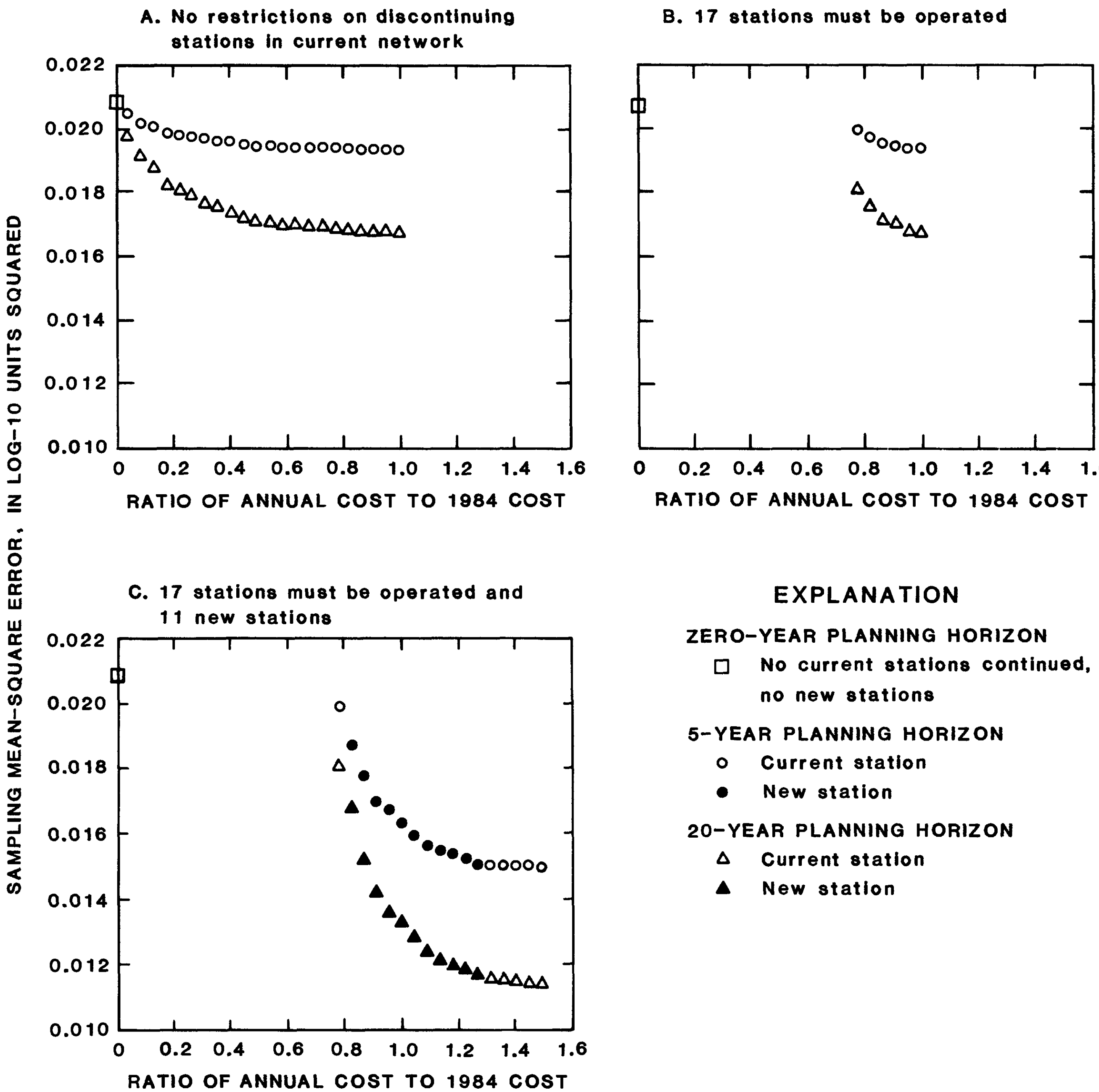


Figure 5. Results of three network strategies to provide regional information on high flow in western Kansas.

Northeastern Kansas

The results of three network strategies for providing regional information in northeastern Kansas are illustrated in figures 6–8. Figure 6 illustrates the results for mean flow. The three graphs indicate that the long-term (20-year) planning horizon would provide the most reduction in the sampling mean-square error. One or two of the existing stations (see table 4 for the station names) would provide nearly all the reduction of error that could be achieved. The addition of new stations would not provide as large a reduction as might be expected. As table 4 shows, the new stations may not always be as effective as some of the existing stations in the reduction of the error. The new stations for low and high flow did not materially affect error reduction.

In figure 7*A*, for low flow, although the sampling mean-square error (0.087) is substantially higher than for mean flow, there is little difference between the 5-year and 20-year results. Figure 7*B*, when compared with figure 7*A*, shows that the stations being analyzed for low-flow regional

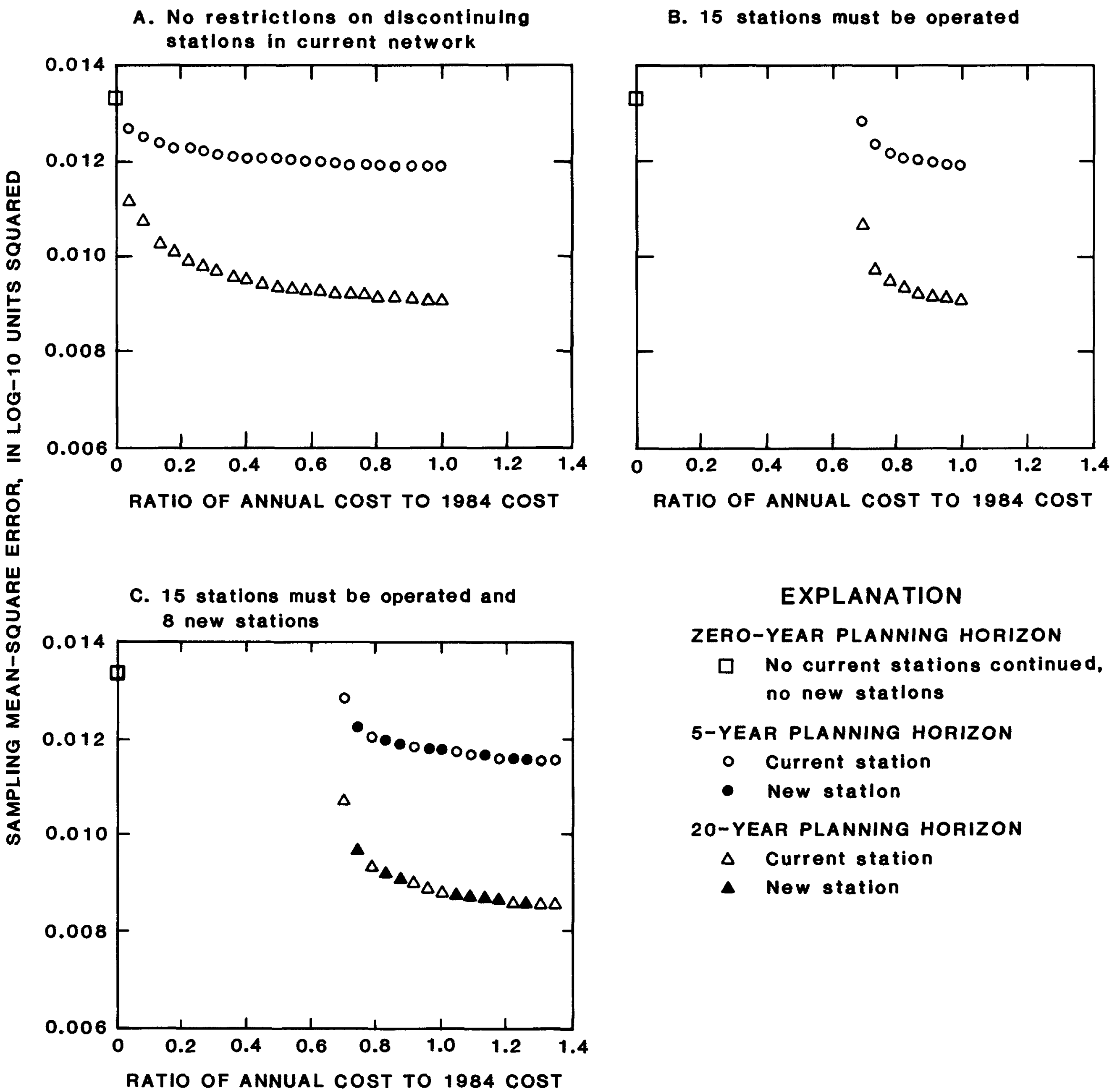


Figure 6. Results of three network strategies to provide regional information on mean flow in northeastern Kansas.

information as the principal purpose are more effective for that purpose than the stations operated for other purposes. Figure 7C shows that adding new stations can have a substantial effect on reducing the sampling mean-square error, and that the reduction could be achieved at slightly lower than the 1984 data-collection cost.

Figure 8 shows the three network strategies for high flow. The three graphs indicate that if high flow is the regional information desired, continuation of existing stations provides nearly as much reduction of the sampling mean-square error as the addition of new stations. The addition of new stations for mean and low flow contributed very little to the sampling mean-square error reduction.

Southeastern Kansas

The results of three network strategies for providing regional streamflow information in southeastern Kansas are summarized in figures 9–11. Figure 9 illustrates the results for mean flow and shows the effectiveness of the stations in

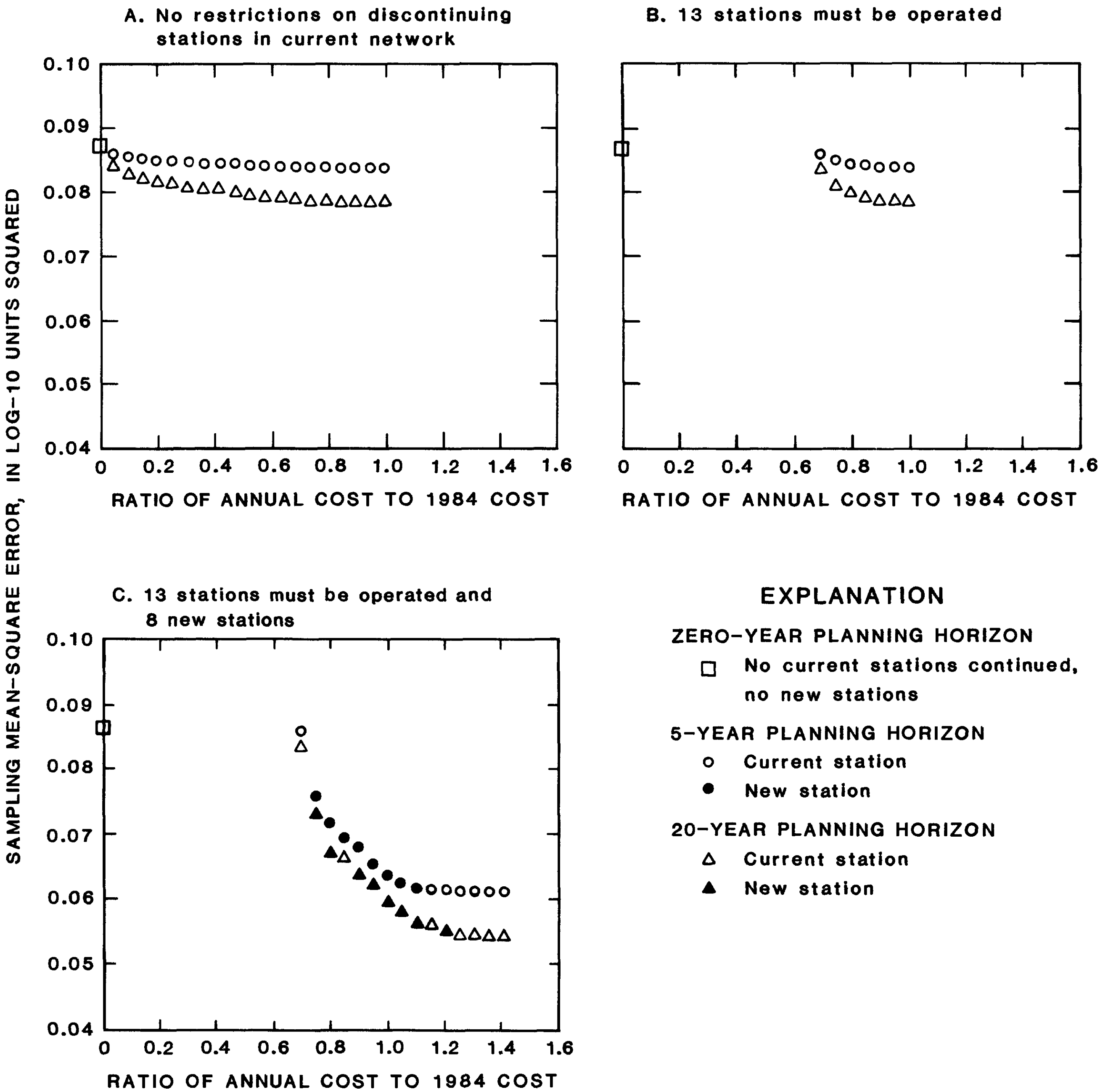


Figure 7. Results of three network strategies to provide regional information on low flow in northeastern Kansas.

reducing the sampling mean-square error. If the short-term (5-year) planning horizon is considered, most of the annual cost would provide little reduction in the sampling mean-square error. The long-term (20-year) planning horizon indicates more reduction of error, yet more than one-half of the annual cost does not provide a significant reduction. Table 5 shows the order of importance for stations considered in the southeastern area.

Figure 9*B* shows current stations that are operated only for regional information and are candidates for discontinuance. The long-term view shows the most reduction in error, and the annual cost of data collection could be reduced to about 0.8 of the 1984 cost without sacrificing a significant reduction of the sampling mean-square error.

Figure 9*C* shows that some new stations, selected for their particular combinations of basin characteristics, would be effective in reducing the sampling mean-square error. Not as many new stations would be needed to achieve the reduction of error as for other flow characteristics in this area. The new stations for low and high flow contributed very little to the sampling mean-square error reduction.

Figure 10 shows the results for low flow in southeastern Kansas. The current sampling mean-square error (0.073) is higher than for mean flow, and there is little

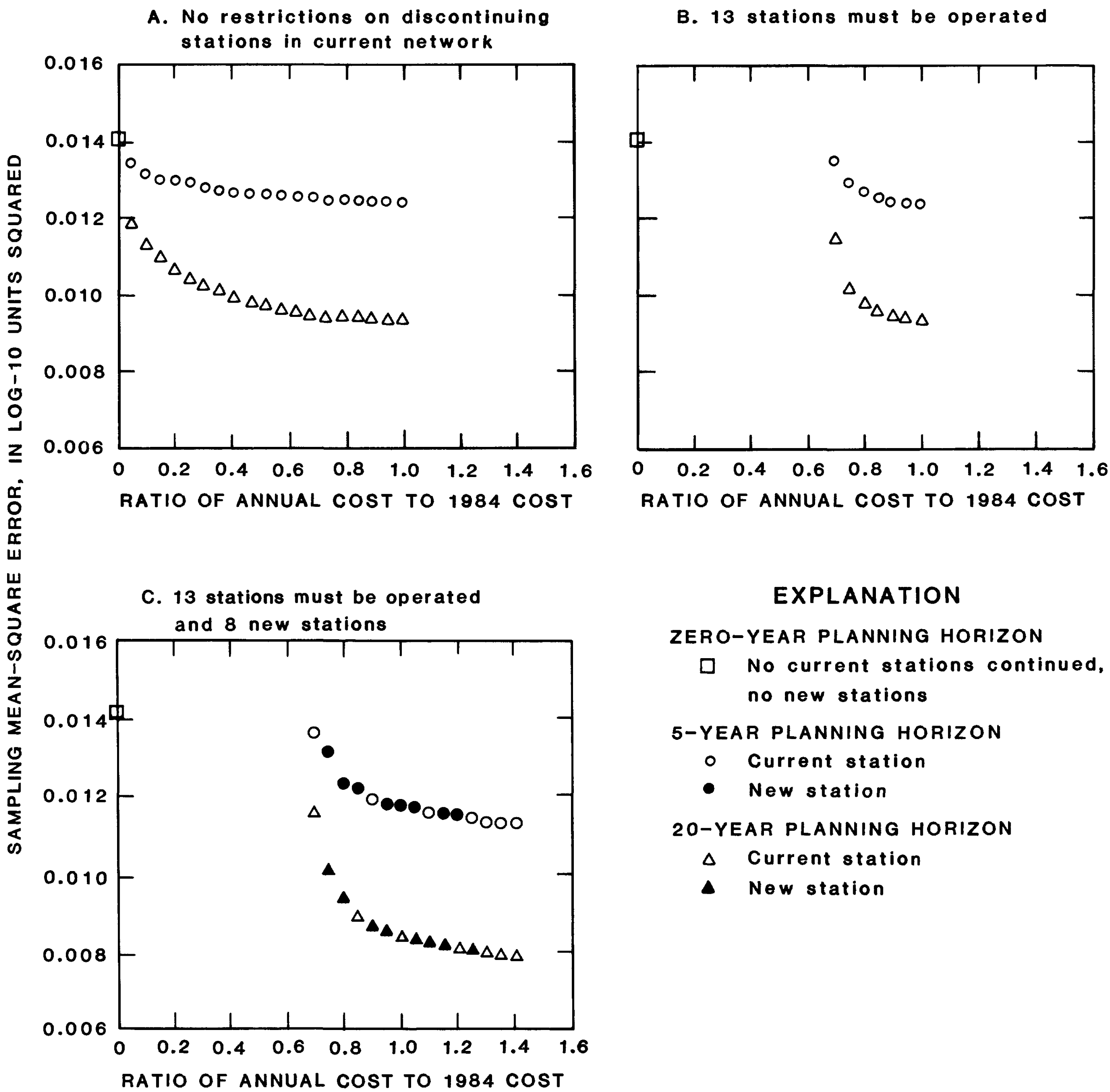


Figure 8. Results of three network strategies to provide regional information on high flow in northeastern Kansas.

difference between the 5-year and 20-year results. The curves in figures 10*A* and 10*B* are relatively flat, indicating that little additional information on low flow can be gained by continuing the currently operated stations. Figures 10*B* and 10*C* show that the stations currently being operated for regional information only could be discontinued if low flow were the only characteristic considered and that adding new stations could reduce the sampling mean-square error. The new stations for mean and high flow could also contribute to error reduction.

Figure 11 shows the results of three network strategies for providing regional information for high flow in southeastern Kansas. The three graphs indicate that if high flow is the regional information desired, continuation of all stations plus new stations would be beneficial in reducing the sampling mean-square error, particularly in the long-term (20-year) planning horizon. The addition of new stations for mean and low flow could help in reducing the error.

COMPARISON OF NETWORK AREAS

Western Kansas had the highest values of sampling mean-square error for all three types of streamflow charac-

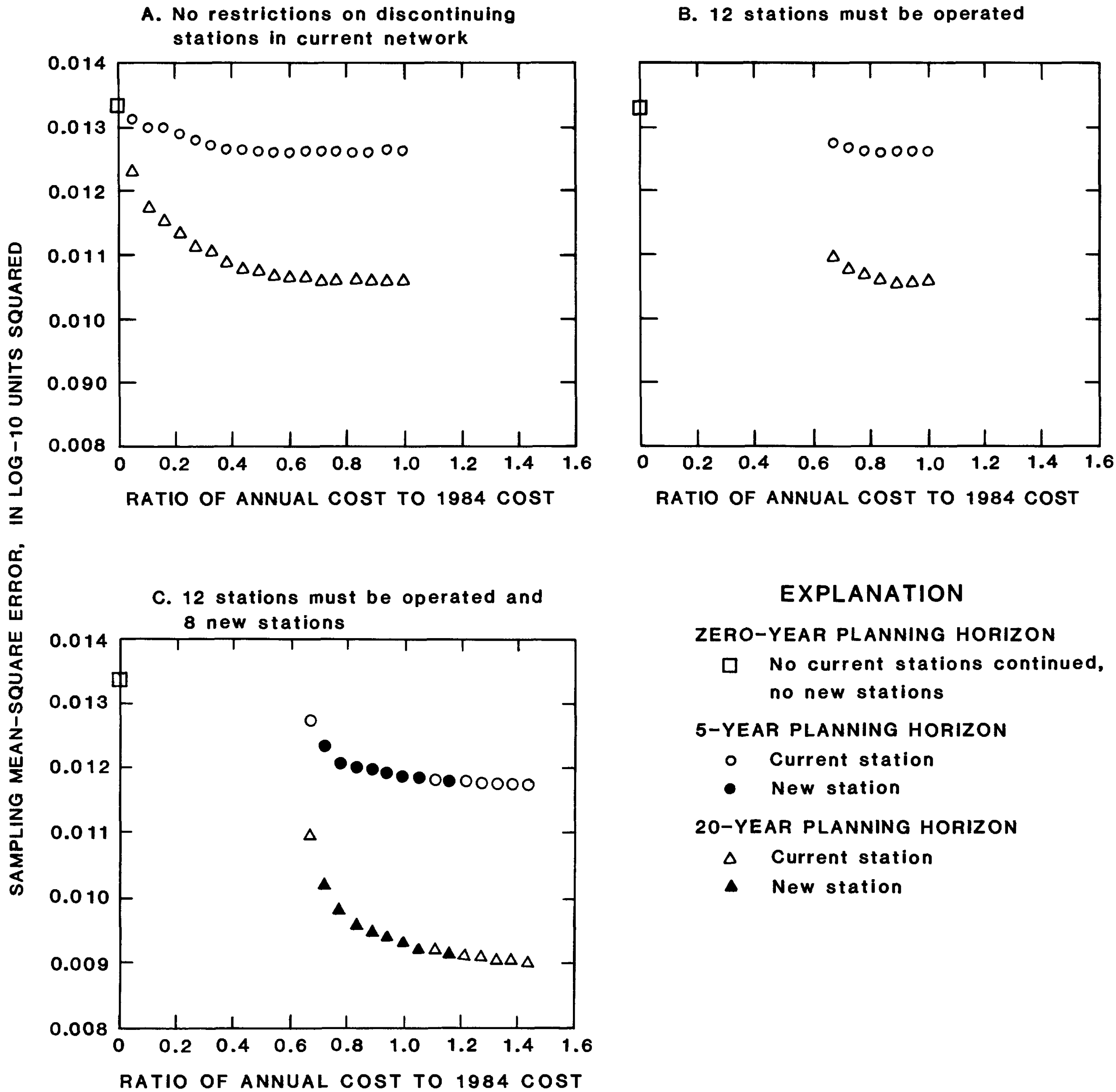


Figure 9. Results of three network strategies to provide regional information on mean flow in southeastern Kansas.

teristics compared with northeastern and southeastern Kansas. Low-flow characteristics had larger sampling mean-square error values than the other characteristics in all three network areas. The magnitude of the sampling mean-square error values for northeastern and southeastern Kansas are very similar for all three streamflow characteristics.

To aid in evaluating the three network areas and the three streamflow characteristics, common points on the graphs were used: the zero-year planning-horizon point; the point that represents the current budget (cost ratio of 1.0) for the 20-year planning horizon with no new stations added; and the point that represents the current budget (cost ratio of 1.0) for the 20-year planning horizon with new stations added in place of some existing stations. The comparison of the amount of reduction of the sampling mean-square error between these points should indicate the relative magnitude of effectively improving the error for the designated area and streamflow characteristic.

When considering regional mean-flow information, western Kansas showed the most improvement. The sampling mean-square error at the zero-year planning horizon (0.035) would decrease to 0.031 using the 20-year planning

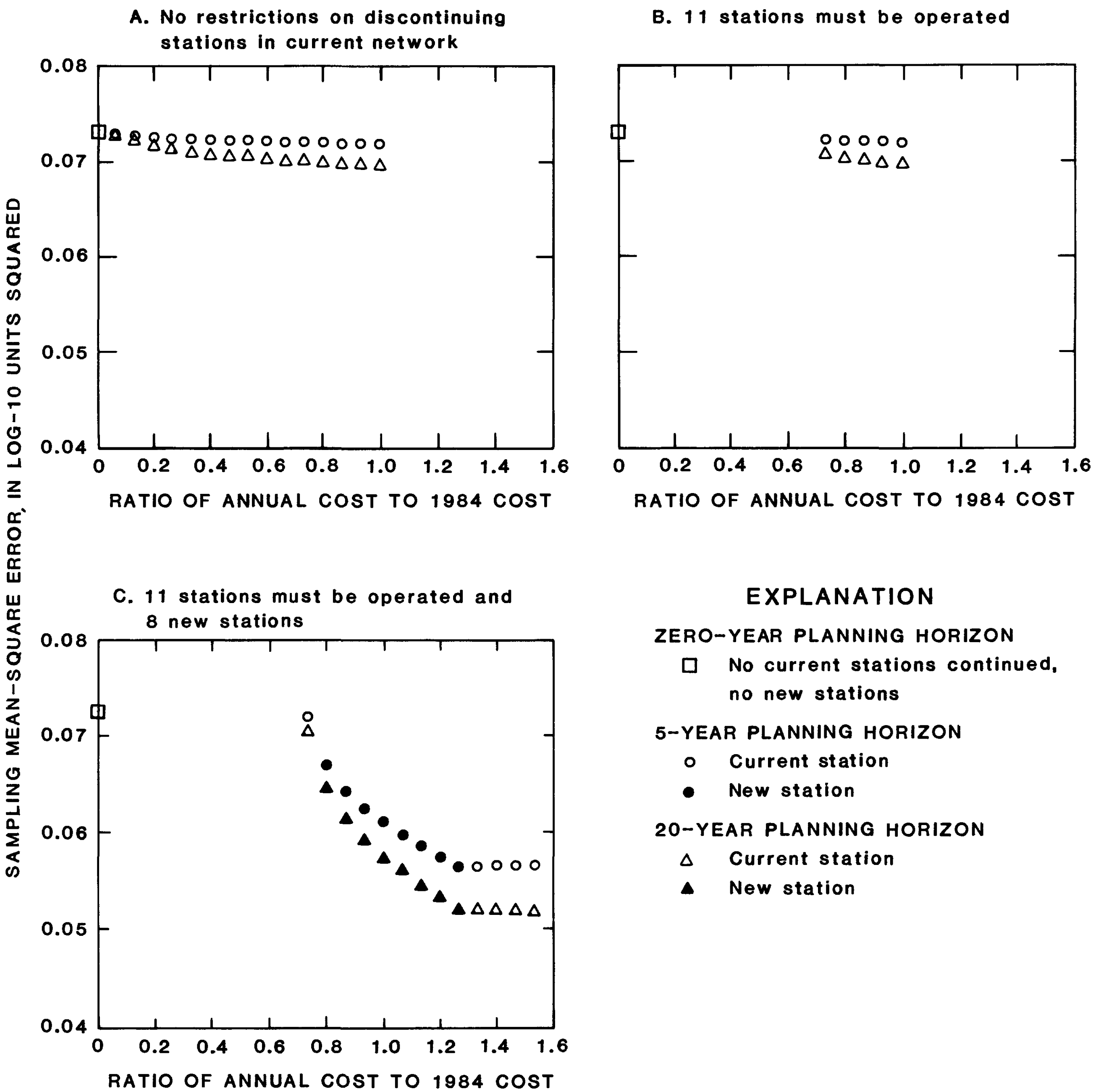


Figure 10. Results of three network strategies to provide regional information on low flow in southeastern Kansas.

horizon and existing stations, and would further decrease to a value of 0.025 by adding new stations, indicating substantial improvement (fig. 3). Northeastern Kansas results showed minimal reduction of the sampling mean-square error by adding new stations (fig. 6). Southeastern Kansas results showed a decrease from the zero-year planning horizon (about 0.013) to the 20-year planning horizon (about 0.011) using existing stations, and a further reduction to about 0.009 by adding new stations (fig. 9).

The addition of new stations for regional low-flow information showed the most reduction of sampling mean-square error for all network areas. Western Kansas could benefit most from the addition of new stations; the sampling mean-square error at the zero-year planning horizon (0.152) would decrease to 0.147 for the 20-year planning horizon using existing stations at a cost ratio of 1.0, and could be reduced further to 0.112 by adding new stations (fig. 4). Northeastern Kansas has a sampling mean-square error of

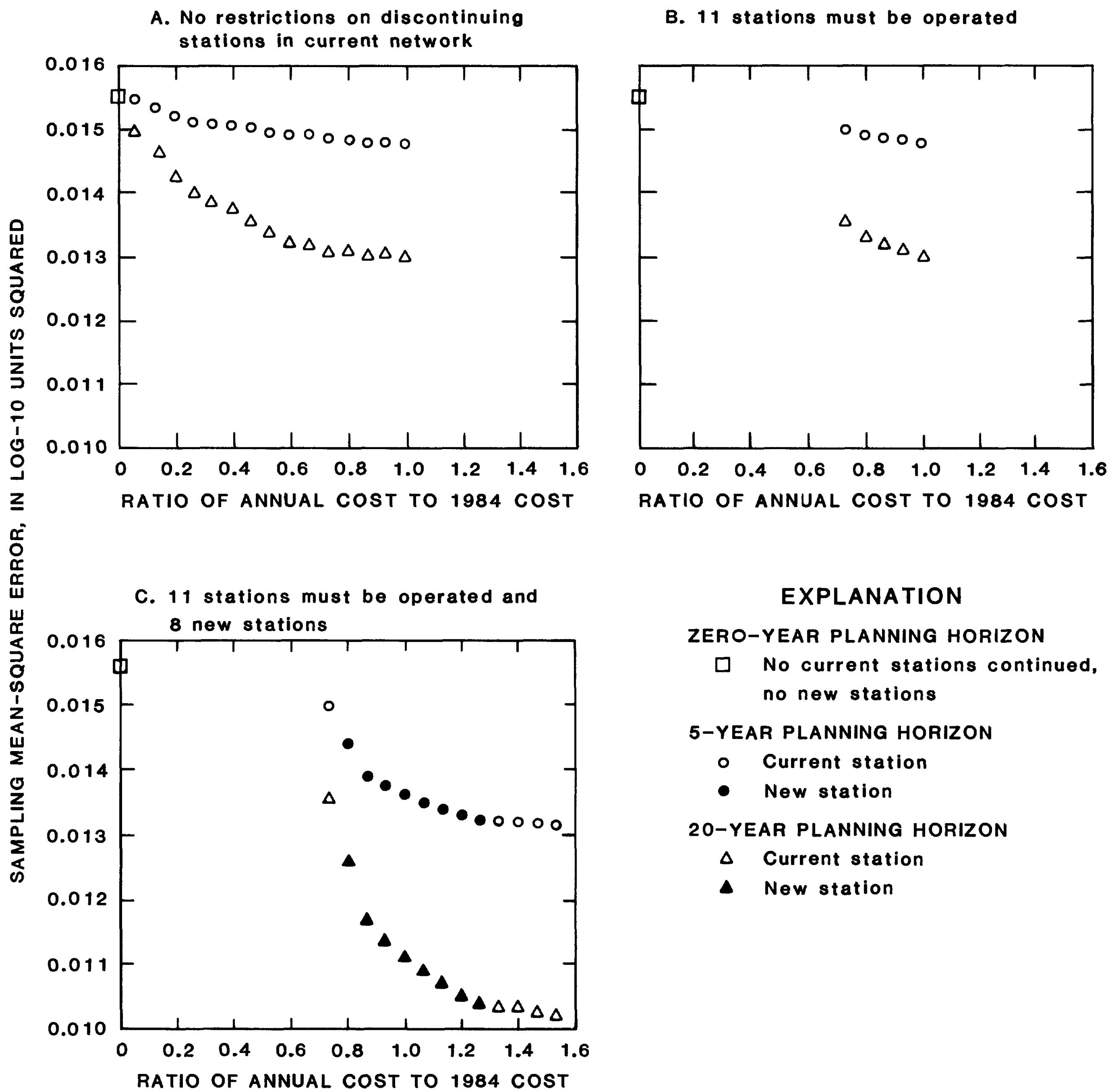


Figure 11. Results of three network strategies to provide regional information on high flow in southeastern Kansas.

0.087 for the zero-year planning horizon; this could be decreased to 0.079 for the 20-year planning horizon using existing stations, and further reduced to 0.060 by adding new stations (fig. 7). Southeastern Kansas showed the smallest decrease of the sampling mean-square error from the zero-year planning horizon (about 0.071) to continuing the same network stations for 20 years (0.070). This value could be improved to 0.058 by adding new stations for regional low-flow information at a data-collection cost ratio of 1.0 (fig. 10).

For regional high-flow information, western Kansas sampling mean-square error could be reduced from about 0.021 for the zero-year planning horizon to about 0.017 by continuing the same network stations, and further reduced to about 0.013 by adding new stations at the 20-year planning horizon (fig. 5). Northeastern Kansas would have a small reduction of the sampling mean-square error (from 0.014 to about 0.009) by continuing the same network stations, and new stations would reduce the sampling mean-square error to only about 0.008 (fig. 8). Southeastern Kansas showed a small sampling mean-square error reduction from the zero-year planning horizon (0.016) to the 20-year planning horizon (0.013) using existing stations, and the smallest sampling mean-square error reduction for all network areas and streamflow characteristics by adding new stations (about 0.011) (fig. 11).

The network analysis indicated that for each streamflow characteristic in each network area some stations could be discontinued without significantly affecting the sampling mean-square error. Funds from the discontinued stations then could be used to add new stations that would contribute more effectively to reducing the sampling mean-square error. The analysis also indicated the amount of additional funding necessary to attain the maximum reduction of sampling mean-square error by continuing the present stations and adding some new stations. If future funding levels must be reduced by discontinuing more stations than are added, the analysis showed that most of the discontinued stations should be in northeastern and southeastern Kansas and the largest number of new stations should be added in western Kansas.

APPLICATIONS AND LIMITATIONS

Example of Application

An example is given to demonstrate the use of the technique as a management tool in decisions regarding regional information. Suppose that additional regional information on low flow is needed in the southeastern area, and that the budget cannot be increased to accommodate operation of any additional stations. A planning horizon of 20 years is used to allow evaluation of possible results. Figure 10 shows that the addition of three new stations would provide an appreciable reduction in the sampling mean-square error. Table 5 shows that the three proposed new stations for low flow are stations that have physical and climatic characteristics similar to the following stations: 07147600 Timber Creek near Wilmot, 07144850 South Fork South Fork Ninnescah River near Pratt, and 07142620 Rattlesnake Creek near Raymond. Since the budget must remain the same, three currently operated gaging stations must be discontinued. Because it is desirable to take maximum advantage of low-flow data in the southeastern area, stations must be discontinued from the other two areas.

The next step is to select the stations to be discontinued. This is done by choosing the stations having low composite rankings within either the western or the northeastern area. Tables 3 and 4 indicate that Saline River near WaKeeney, North Fork Big Creek near Victoria, Soldier Creek near Circleville, and Soldier Creek near Delia are among the stations where additional record collection is least cost-effective for regional information. Therefore, one may select from these four stations the three stations to discontinue.

The next step is to review the data uses for the gaging-station records and the continuing needs of other agencies for those stations. One may refer to Medina and Geiger (1984, table 2) for data uses and funding information for each station. If other agencies do not need continuing data from these stations, the process can go forward. If a data user wants a particular type of data, perhaps an alternative type of gaging can be used to furnish the desired information—for example, a crest-stage gage to record peak flow or a miscellaneous low-flow measuring site that can be correlated with other stations to provide low-flow information, thus satisfying the data uses. If none of the stations can be discontinued, another set of stations are selected and the procedure starts over again.

Uses and Limitations

This report describes a technique that can be used to evaluate and compare alternative gaging plans in terms of effectively collecting regional streamflow information. The technique can be used to gain insight into how much regional information may be lost or gained by decisions to reduce or increase the operating budget. It can be used to evaluate the effects, in terms of regional information, of adding specific new stations to the network. It can also be used to determine which stations are least cost effective for collecting additional regional information. In tables 3–5 and figures 3–11 hypothetical new stations having specific physical and climatic characteristics were assumed to exist for purposes of illustration. In fact, such new stations having exactly the assumed physical and climatic characteristics may be impossible to find. In locating new stations, one must consider factors other than physical and climatic characteristics—for example, stability of gage control, accessibility of the site, and other uses of the data to be collected.

SUMMARY

The surface-water data network in Kansas was analyzed for its effectiveness in providing regional streamflow information using the generalized least squares regression method. Unregulated medium-, low-, and high-flow characteristics were selected to be representative of the regional information that can be obtained from streamflow-gaging-station records for use in evaluating the effectiveness of continuing the current network stations, discontinuing some stations, and (or) adding new stations. The analysis used streamflow records for all currently operated stations that were not affected by regulation and discontinued stations for which unregulated flow characteristics, as well as physical and climatic characteristics, were available.

Three planning horizons were used for comparison—the zero-year planning horizon for the present condition, a 5-year horizon for short-term planning, and a 20-year horizon for long-term planning. The State was divided into three network areas, western, northeastern, and southeastern Kansas, and analysis was made for the three streamflow characteristics in each area, using the three planning horizons. The analysis did not suggest specific sites for new stations but indicated the set of physical and climatic characteristics of a new station that would be influential in reducing the error.

The stations were ranked by the amount of their reduction of sampling mean-square error per unit cost of data collection. Only new stations and stations that are operated primarily for regional streamflow information were shown in this ranking procedure. A composite ranking for the 20-year planning horizon ranks all the stations in order of priority, assuming that all streamflow characteristics are of equal importance. Stations that must be operated for data uses other than regional information were not ranked according to their contributions to reducing the sampling mean-square error because other needs preclude their being considered for discontinuance.

The analysis provided a procedure for reviewing the regional streamflow information currently available and, if additional information is desired, the types of data needed, the possible reduction of the sampling mean-square error that could be expected, the areas of the State where the additional data would be most beneficial, and the relative cost of the information that might be obtained.

The analysis showed that the sampling mean-square error can be reduced by adding new stations and possibly discontinuing some current stations. The addition of new stations for determining mean flow would be most beneficial in western Kansas and to lesser degrees in the other two areas. The reduction of the sampling mean-square error due to the addition of new stations would have the smallest effect in northeastern Kansas, where continuation of current stations would be more effective. Additional low-flow information would be very important in all three areas, with western Kansas having the most dramatic reduction in the sampling mean-square error and southeastern Kansas benefiting least from the addition of stations for low-flow information. The reduction of the sampling mean-square error from the addition of new stations for high flow would be beneficial in western Kansas, and the effect would be less in the other two areas. Southeastern Kansas had the smallest sampling mean-square error reduction in high-flow information.

A comparison of the results for all areas indicated where funding resources could be most effectively spent and not adversely affect the sampling mean-square error reduction for some of the streamflow characteristics. If future funding levels must be reduced by discontinuing more stations than are added, the analysis showed that most of the discontinued stations should be in northeastern and southeastern Kansas, and the largest number of stations should be added in western Kansas.

THEORY AND APPLICATION OF GENERALIZED LEAST SQUARES

By Gary D. Tasker

Theory of Generalized Least Squares

The problem addressed in this section of the report is how to estimate the parameters of a linear model that specifies the value of a flow characteristic (such as the 100-year peak) as a function of various physical and climatic characteristics (such as the drainage area, channel slope, and precipitation). The data available at each of N stations in a region are summarized by the vector of physical and climatic characteristics, $\mathbf{x}_i$, and a streamflow record of n_i values

$$\left\{y_{i,1},\ y_{i,2},\dots,\ y_{i,n_i}\right\}$$

available for each station i. These may be the annual maximum 1-day streamflow or a transformation thereof, so that the y's are normally distributed. The normality assumption is adopted to facilitate this explanation. Other distributions, such as the log-Pearson Type III, could be employed with appropriate adjustments in the explanation that follows. The normal assumption worked well in Monte Carlo experiments reported by Stedinger and Tasker (in press) even when the actual distribution was log-Pearson Type III.

Let

$$\hat{\mathbf{Y}} = \left(\bar{y}_1 + z_p s_1, \dots,\ \bar{y}_N + z_p s_N\right)^T$$

be the vector of at-station estimates of the $(1/p)$-year streamflow, where $\bar{y}_i$ and s_i are the usual sample estimates of mean and standard deviation of the streamflows at station i and z_p is the standard normal deviate with exceedance probability p. Also, let

$$\mathbf{Y} = \left(\mu_1 + z_p \sigma_1, \dots, \mu_n + z_p \sigma_n\right)^T$$

be the vector of true $(1/p)$-year streamflows, where μ_i and σ_i are the true mean and standard deviation of the $y_{i,t}$ at station i.

The ordinary least squares (OLS) regional regression model can be written

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}\ , \tag{1}$$

where $\mathbf{X}$ is an $(N \times k)$ matrix of physical and climatic characteristics augmented by a column of ones, $\boldsymbol{\beta}$ is a $(k \times 1)$ vector of regression coefficients, and $\mathbf{e}$ is an $(N \times 1)$ vector of random errors with expected values of zero and covariance matrix that is assumed to be of the form $\sigma^2 I_N$. Here,

I_N is an N-dimensional identity matrix and σ^2 is a constant. The OLS estimate of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}}_{OLS}=(\mathbf{X}^T\mathbf{X})^{-1}\,\mathbf{X}^T\,\widehat{\mathbf{Y}}\ . \tag{2}$$

The estimator has a sampling covariance matrix, given the stated assumptions, equal to

$$\mathrm{Var}\,(\widehat{\boldsymbol{\beta}}_{OLS})=\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\,. \tag{3}$$

The OLS model is appropriate when all estimates of $(1/p)$-year streamflows have approximately equal variance and when concurrent streamflows at different stations are uncorrelated or independently distributed. In general, these two conditions are not met because gaged records are of different lengths, concurrent streamflows are often cross correlated, and the natural variability of streamflows between stations is not the same.

A more appropriate model for use in hydrologic regression is the generalized least squares (GLS) model. In the GLS model, the assumptions of equal variance of the $(1/p)$-year streamflows and zero cross-correlation for concurrent streamflows are relaxed. The covariance matrix for the errors in the GLS model is the matrix $\boldsymbol{\lambda}$, whose elements are

$$\lambda_{ij}=\begin{cases}\gamma^2+\sigma_i^2\left(1+0.5z_p^2\right)\Big/n_i & \text{for } i=j\\[2ex] \rho_{ij}\sigma_i^2\sigma_j^2m_{ij}\left(1+0.5\rho_{ij}z_p^2\right)\Big/n_in_j & \text{for } i\neq j\end{cases}\ , \tag{4}$$

where ρ_{ij} is the estimated cross correlation between annual floodflows (or any other streamflow being considered) at stations i and j, m_{ij} is the number of concurrent years of record, and γ^2 is the model-error variance. The model-error variance is a measure of the precision of the true regression model and is defined by

$$\gamma^2=E\left[(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^2\right]. \tag{5}$$

The value of γ^2 is assumed to be independent of $\mathbf{x}_i$, where $\mathbf{x}_i$ is a vector of physical and climatic characteristics at station i. The GLS estimate of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}}_{GLS}=\left(\mathbf{X}^T\boldsymbol{\lambda}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\boldsymbol{\lambda}^{-1}\,\widehat{\mathbf{Y}}\ . \tag{6}$$

This estimator has a sampling covariance matrix, given the assumptions, equal to

$$\mathrm{Var}\left(\widehat{\boldsymbol{\beta}}_{GLS}\right)=\left(\mathbf{X}^T\boldsymbol{\lambda}^{-1}\mathbf{X}\right)^{-1}. \tag{7}$$

The greatest barrier to using GLS methods has been that the value of the covariance matrix $\boldsymbol{\lambda}$ is unknown because it depends on the population values of γ^2, σ_i, and ρ_{ij}. Stedinger and Tasker (1985) proposed use of a reasonable estimator of $\boldsymbol{\lambda}$. Their Monte Carlo simulations showed that use of the estimator of $\boldsymbol{\lambda}$ in equation 6 led to an improved estimator of $\boldsymbol{\beta}$ (in terms of mean-square error) and more accurate estimates of the model variance, γ^2, than did use of competing ordinary and weighted least squares procedures.

The method used to estimate $\boldsymbol{\lambda}$ will be described briefly; Stedinger and Tasker (1985, in press) provide a more detailed description. If $\boldsymbol{\lambda}$ were known, then

$$E\left\{\left(\widehat{\mathbf{Y}}-\mathbf{X}\,\widehat{\boldsymbol{\beta}}_{GLS}\right)^T\left[\boldsymbol{\lambda}\left(\gamma^2,\,\rho_{ij},\,\sigma_i\right)\right]^{-1}\left(\widehat{\mathbf{Y}}-\mathbf{X}\,\widehat{\boldsymbol{\beta}}_{GLS}\right)\right\}=N-k\ , \tag{8}$$

where $\boldsymbol{\lambda}$ is written as $\boldsymbol{\lambda}(\gamma^2, \rho_{ij}, \sigma_i)$ to emphasize its dependence on γ^2, ρ_{ij}, and σ_i. The problem is to obtain estimates of γ^2, ρ_{ij}, and σ_i to enter into equation 4 to obtain an estimate of $\boldsymbol{\lambda}$.

Reliably estimating the cross correlation between flows at each pair of stations, ρ_{ij}, where only shorter concurrent records are available, is a difficult task. It is best done subjectively using good hydrologic judgment. The goal of estimating the ρ_{ij}'s is to capture the essential underlying cross-correlation structure among the streamflows within a region. One suggestion for estimating the ρ's is to classify each pair (i, j) of stations as likely to have a high, medium, or low cross correlation on the basis of the sample cross correlation for the pair and hydrologic judgment. In classifying pairs of stations, one could consider the proximity of the stations, their drainage areas, any tributary relation between them, the type of storms that cause peak flows (or other types of streamflows being considered) in each basin, and other factors.

At first one may consider estimating the σ_i's in equation 4 by their usual sample estimate s_i. However, this leads to a poor estimate of $\boldsymbol{\beta}$ (Stedinger and Tasker, 1985, in press). A better approach is to first regress the s_i against physical and climatic characteristics and then use the regression estimate of $E\left(\sigma_i|x_i\right)$. Finally, γ^2 can be estimated by iteratively searching for solutions to equation 8 to obtain an estimate, $\widehat{\gamma}^2$, of γ^2 for the specified values of ρ_{ij} and $E\left(\sigma_i|x_i\right)$. Some refinements in the procedure actually result in an algorithm that is slightly more involved than the procedure outlined here.

Application to Data-Network Analysis

The goal of a surface-water network analysis is identification of an efficient plan for future stream-gaging activities for various levels of operating budgets. A gaging plan is defined as the identification of a set of stations (currently operated stations or new stations) that will be operated at least until the end of a planning horizon (5–20 years). An efficient gaging plan is a feasible (within the specified operating budget) plan that gives the best value of the objective function. In this network analysis, the objective function is the average sampling mean-square error of a regional regression. This objective will be discussed in greater detail in the next section.

The results of a network analysis can be used to (a) identify an efficient gaging plan for a specific operating budget, (b) provide insight into how much regional information is lost or gained by reducing or increasing the operating budget, and (c) evaluate a proposed gaging plan by comparing it with an efficient gaging plan of the same budget level.

Average Sampling Mean-Square Error

A valuable feature of the GLS technique is that it provides a reliable estimate of the prediction mean-square error, *MSEp*, at a station. The *MSEp* can be partitioned into a model-error component (the error due to estimating the true streamflow characteristic, y, by the true regression estimate, y_R) and a sampling-error component (the error due to estimating y_R by the sample regression estimate, $\tilde{y}_R$). In the GLS technique, only the sampling-error component is affected by increases in record length or by inclusion of new stations, so that attention can be focused on this component as an objective function.

Let $\tilde{\mathbf{x}}_j^T=(1, x_1, x_2,..., x_m)$ be a vector of physical and climatic characteristics at station j. The sampling mean-square error at station j is

$$V_j=\mathbf{x}_j^T\left(\mathbf{X}_w^T\mathbf{\Lambda}_x^{-1}\mathbf{X}_w\right)^{-1}\tilde{\mathbf{x}}_j \ . \qquad (9)$$

Here, w is an index referring to different station-record-length combinations that may be used to obtain estimates of $\boldsymbol{\beta}$. In other words, the w is an index of different gaging plans. Note that $\tilde{\mathbf{x}}_j^T$ need not be a row of $\mathbf{X}_w$. To get a measure of regional streamflow information, $\hat{V}_j$ can be averaged over a representative set of stations.

In this network analysis the average sampling mean-square error is taken over the stations used in the regression analysis so that the average sampling mean-square error for gaging plan w is

$$\bar{V}_w=\frac{1}{N}\Sigma\left[\mathbf{x}_j^T\left(\mathbf{X}_w^T\hat{\mathbf{\Lambda}}_w^{-1}\mathbf{X}_w\right)^{-1}\tilde{\mathbf{x}}_j\right] , \qquad (10)$$

where the summation is over all N stations used in the regression. The matrix $\mathbf{X}_w$ is the set of physical and climatic characteristics (augmented by a column of 1's) corresponding to all the stations used in the regression analysis plus any new stations or reactivated old stations under consideration. The $\hat{\mathbf{\Lambda}}_w$ matrix is the GLS weighting matrix computed by updating the record lengths for each station that is assumed to be operated in plan w during the planning horizon.

The regional information associated with two gaging plans can be compared by calculating $\bar{V}_w$ for each plan. For example, if for plan 1 the average sampling mean-square error is $\bar{V}_1$ and for plan 2 it is $\bar{V}_2$, then plan 2 can be considered better than plan 1 (in terms of regional streamflow information) if $\bar{V}_2<\bar{V}_1$. Furthermore, if the cost of operating under plan 2 is less than or equal to that of plan 1, then plan 2 would be preferred to plan 1.

Generating Efficient Gaging Plans

The problem considered here is how to identify a feasible gaging plan, indexed by w, so as to minimize the average sampling mean-square error. In identifying a feasible plan one must recognize that some stations will continue to be operated because of site-specific data needs. In addition, it may be impossible to continue to operate stations at some sites, perhaps because the site has been flooded by dam construction or the streamflow has become largely regulated and unsuitable for inclusion in a regional analysis.

Let F represent the set of stations that one can choose to operate or not to operate. These may be existing stations, old stations, or new stations. Let S_w be the set of stations operated in gaging plan w. S_w is a subset of F. The problem then is to find the subset S_w^* that minimizes $\bar{V}_w$ and is within a budget limit, B, so that

$$\Sigma c_j \leq B \ , \qquad (11)$$

where c_j is the marginal cost of operating station j during the planning horizon and the summation is over the stations contained in subset S_w.

Fundamentally, the optimization of $\bar{V}_w$ over S_w contained in set F poses a very large nonlinear integer programming problem. Because the number of subsets of F that satisfy equation 11 can be very large, a direct attack on the problem is generally not attractive. Instead, a step-backward type algorithm is used to identify a reasonable estimate of the best gaging plan for a specified budget.

The step-backward algorithm starts by considering all possible stations as being operated during the planning horizon. It then incrementally drops operation of those stations that are eligible to be dropped. The station selected to be dropped at each step is the one that contributes the least reduction of the objective function per unit cost.

SELECTED REFERENCES

Furness, L.W., 1957, Development of a balanced stream-gaging program for Kansas: Kansas Water Resources Board Bulletin 4, 50 p.

Geiger, C.O., Lacock, D.L., Shelton, L.R., Penny, M.L., and Merry, C.E., 1983, Water resources data, Kansas, water year 1982: U.S. Geological Survey Water-Data Report KS–82–1, 485 p.

Jordan, P.R., 1983, Magnitude and frequency of low flows of unregulated streams in Kansas, and estimation of flow-duration curves for ungaged sites: Kansas Water Office Technical Report 17, 55 p.

——— 1984, Magnitude and frequency of high flows of unregulated streams in Kansas: U.S. Geological Survey Open-File Report 84–453, 45 p.

Jordan, P.R., and Hedman, E.R., 1970, Evaluation of the surface-water data program in Kansas: Kansas Water Resources Board Bulletin 12, 49 p.

Medina, K.D., and Geiger, C.O., 1984, Evaluation of the cost effectiveness of the 1983 stream-gaging program in Kansas: U.S. Geological Survey Water-Resources Investigations Report 84–4107, 57 p.

Stedinger, J.R., and Tasker, G.D., 1985, Regional hydrologic regression—1. Regression analysis with ordinary, weighted, and generalized least squares compared: Water Resources Research, v. 21, no. 9, p. 1421–1432.

——— in press, Regional hydrologic regression—2. Model error estimators, estimates of sigma, and log-Pearson Type III distribution: Water Resources Research.

METRIC CONVERSION FACTORS

Factors for converting the inch-pound units given in this report to the International System (SI) of Units are given below:

Multiply inch-pound unit	By	To obtain SI unit
	Length	
inch	25.40	millimeter
mile	1.609	kilometer
foot per mile	0.1894	meter per kilometer
	Area	
square mile	2.590	square kilometer

☆ U S GOVERNMENT PRINTING OFFICE 1987 - 181-407 - 226/40115