Mailing Address:
Water Resources Data System
University of Wyoming, Dept 3943
1000 E University Avenue
Laramie, WY 82071

Physical Address:
Wyoming Hall, Room 249
University of Wyoming
Laramie, WY 82071

Phone: (307) 766-6651
Fax: (307) 766-3785

Funding for WRDS and the creation of this electronic document was provided by the Wyoming Water Development Commission (http://wwdc.state.wy.us)

A STUDY OF
SEVERAL MODELS FOR ESTIMATING
MONTHLY STREAMFLOW RUNOFF

Water Resources Series No. 39

Verne E. Smith November 1973

ABSTRACT

Three models for estimating monthly streamflow at ungaged points are developed and tested. The first model is based simply on runoff per unit area. The second model is a least squares regression of drainage area versus runoff. The third is an elevation dependent model that is developed using ridge regression techniques. The models were tested on thirty years of monthly streamflow data from eight drainage basins in the upper Green River Basin of the Wind River Range.

ACKNOWLEDGEMENTS

The author would like to express his appreciation to Frank J. Trelease, Jr. and Paul A. Rechard for their review and useful comments regarding the report. This study was supported by the Office of Water Planning of the Wyoming State Engineer's Office.

KEY WORDS: hydrology, model, ridge regression, runoff, streamflow

Research on which this report is based was funded in part through the Wyoming Water Resources Research Institute by the Office of Water Resources Research, Department of the Interior under the Water Resources Research Act of 1964.

A STUDY OF SEVERAL MODELS FOR ESTIMATING MONTHLY STREAMFLOW RUNOFF

Introduction

For many hydrologic studies monthly streamflow runoff is one of the most useful kinds of data. However, data frequently are not available at a desired location, and some kind of estimation procedure must be used. Ideally, monthly discharge data, real or estimated, should be available sequentially in time for use directly in detailed studies such as reservoir operation studies, or for development of simulated sequences for some studies. In the study discussed herein several models to estimate monthly runoff, year by year, were tested on a study area to compare model estimates with measured discharge in order to determine which model was "best."

The area used in this study comprises a set of eight drainage basins in the Upper Green River Basin of Wyoming. The specific basins are the: Green River at Warren Bridge, 09-1885, New Fork River below New Fork Lake, 09-1930, Pole Creek below Little Half Moon Lake, 09-1985, Fall Creek near Pinedale, 09-1995, Boulder Creek below Boulder Lake, 09-2020, East Fork River near Big Sandy, 09-2030, Silver Creek near Big Sandy, 09-2040, and Big Sandy River at Leckie Ranch, 09-2125. The basins lie on the western slope of the Wind River Range in west-central Wyoming and are part of the headwaters of the Green River, a principal tributary of the Colorado River. The period of record used was 1941 through 1970.

Several of the basins contain natural lakes that have had small dams constructed on them for irrigation regulation. It was assumed that this regulation was minor and would not affect the natural flow significantly.

In the material presented herein, the results are oriented towards the evaluation of the models that are examined. In practical applications, interest would lie in estimating the runoff at some point on a stream where the actual value is unknown. To do this, known values from selected drainage basins located near the basin in question would be applied to the particular model used. This would provide model parameter values similar to those presented in the results. The independent variable value of the unknown drainage basin would then be substituted into the model to obtain the desired estimate.

Runoff Models

Three basic models and modifications of two of them were studied. The first model assumes that the runoff per unit area over a region is total runoff of the region divided by the total area. This can be expressed in equation form as

$$X_i = \frac{12 \sum_{j=1}^{n} Y_{ij}}{\sum_{j=1}^{n} A_j} \tag{1}$$

where X is the runoff in inches, i the particular month of a particular year, Y the runoff in acre feet, j a particular basin, n the number of basins and A the drainage area in acres. Having determined X_i, the streamflow, $\hat{Y}$, in acre-feet, at any ungaged point can be calculated from

$$\hat{Y}_{ij} = \frac{X_i \ A_j}{12} . \tag{2}$$

Model 2 is a linear least square regression for each month, the areas of the drainage basins being the independent variable. In equation form it is

$$Y_{ij} = a_i + b_i A_j \tag{3}$$

where a is a constant, be is the slope of the regression line and the remaining notation is as defined above.

Model 3 is an elevation dependent model similar to the one developed in an earlier study for annual runoff (Smith, Anderson and Scott, 1973. Let $y^i_{T,t}$ denote the total observed streamflow (y_T) of gaging station i in month and year t. Suppose that the drainage area for that station is partitioned into k elevation bands and let p^i_1, p^i_2, ..., p^i_k denote the proportions of area in the k bands. If A^i_1, A^i_2, ..., A^i_k denote the area in each of the k bands, respectively, and A^i the total area of the drainage for the i$\underline{\text{th}}$ station, the $p^i_j = A^i_j/A^i$ or $A^i_j = p^i_j A^i$, where j denotes a particular elevation band.

It is reasonable to assume that in any given month and year, t, the contributions to runoff from the different elevation bands are different and that their relative contributions change with t. Therefore, let $y^i_{1,t}$, $y^i_{2,t}$, ..., $y^i_{k,t}$ denote the contribution to total runoff of the k elevation bands. Model 3 in its simplest form is

$$y^i_{T,t} = y^i_{1,t} + y^i_{2,t} + \dots + y^i_{k,t} \tag{4}$$

where T denotes the summation of the elevation bands. If the superscript * indicates runoff per unit area, it follows that

$$y^{i^*}_{T,t} = y^i_{T,t}/A^i = \sum_{j=1}^{k} \frac{y^i_{j,t}}{A^i_j} \frac{A^i_j}{A^i} = \sum_{j=1}^{k} y^{i^*}_{j,t}\, p^i_j \quad . \tag{5}$$

In equation 5 the p^i_1, p^i_2, ..., p^i_k are known constants and $y^{i^*}_{T,t}$ is an observable variable. The quantities $y^{i^*}_{1,t}$, $y^{i^*}_{2,t}$, ..., $y^{i^*}_{k,t}$ are unknown and not observable.

At this point it is necessary to make certain assumptions upon which the model is to be developed. Suppose there are n, $n \geq k$ streamflow gaging stations and that the contribution per unit area in month and year t and elevation band j, j = 1, 2, ..., k, is a constant for all n stations. That is

$$y_{j,t}^{i^*} = \alpha_{j,t}, \quad i = 1, 2, \ldots, n \ , \tag{6}$$

where $\alpha_{j,t}$ is an unknown constant dependent only upon t. This requires a certain uniformity of contribution over the total elevation band. Assuming (6), equation (5) becomes

$$y_{T,t}^{i^*} = \sum_{j=1}^{k} \alpha_{j,t} P_j^i, \quad i = 1, 2, \ldots, n \ , \tag{7}$$

which is the standard form of the general regression equation without intercept. If $\underline{y}_{T,t}^*$ denotes the vector of observed runoff per unit area, $\underline{\alpha}_t$ the vector of unknown parameters, and P the n by k matrix of proportions, then the least square estimate of $\underline{\alpha}_t$ is given by

$$\underline{\hat{\alpha}}_t = (P'P)^{-1} P' \underline{y}_{T,t}^* \ . \tag{8}$$

The nature of the matrix, P, with the rows being proportions summing to one, forces the system to be non-orthogonal. The estimates in a non-orthogonal system are subject to a number of "errors" and tend to be large in absolute value. Ridge regression techniques introducted by Hoerl and Kennard (1970) and evaluted by Marquardt (1970) have been shown to circumvent many of the difficulties by controlling the inflation and general instability (high variance) associated with least square estimates. The ridge regression estimates, while slightly biased, are usually better than least square estimates for purposes of estimation and physical interpretation of the coefficients in that they result in a smaller mean square error of the estimates. Appendix II contains a brief review of the ridge regression procedure.

The ridge regression estimates denoted by $\underline{\tilde{\alpha}}_t' = (\tilde{\alpha}_{1,t}, \tilde{\alpha}_{2,t}, \ldots, \tilde{\alpha}_{k,t})$ represent the estimated partition by elevation band of $y_{T,t}^{i*}$. Multiplying by A^i in equation (5), the estimated partition of total volume is obtained,

$$\tilde{y}_{T,t}^i = \tilde{\alpha}_{1,t}A_1^i + \tilde{\alpha}_{2,t}A_2^i + \ldots + \tilde{\alpha}_{k,t}A_k^i \qquad (9)$$

where the tilde (~) indicates ridge regression estimates. This equation is the working form of model 3. To obtain meaningful results, model 3 requires that there be sufficient ground relief in the drainage basins to provide a fairly large range of values of proportions of areas within elevation bands. This requirement is met in the study area. Plains area drainage basins may not meet this requirement.

In some cases, models 2 and 3 could produce a negative runoff value at a particular site during the base flow period. Therefore, a constraint was placed on the model so that negative or zero values were set equal to the value of one acre-foot. One was used rather than zero because a logarithm transformation could be applied later if desired, and the value is relatively close to zero. The modified models are designated as 2M and 3M.

Logically, models 3 or 3M should provide better results than the other models during the snowmelt runoff period because these models take elevation into consideration, and snowmelt runoff is at least partially a function of elevation. Assuming homogeneity of physical and climatological characteristics within elevation bands, model 3 should provide very good estimates of snowmelt runoff. This assumption appears to hold true as shown later for the month of June.

By examining the monthly values, it appears that September through March can be considered to be the base flow months. Therefore, the models were first applied to these months.

Results

To determine how good each model was at estimating monthly runoff, several statistical parameters were considered. It was decided that residuals would be most useful because they could be used to compare models. The residuals were computed as the observed value minus the value calculated from the particular model. The per cent error was computed to be the residual divided by the observed value times 100.

Tables 1 through 5 present typical results for October 1943 for the models. Note that the small observed runoff values are subject to large per cent errors. This is due to the division by small numbers (the observed runoff). This type of error occurs throughout all of the results of the base flow months and tends to make the results appear worse than they might really be if small flows are not too important.

Computations were run for the base months for each year and each model. This constitutes a vast amount of print-out that is too voluminous to present herein. Therefore, the results are summarized in Tables 6 through 12. Here the average absolute per cent residual errors and their totals and means are presented for each model.

Ridge regression constants from 0 to .6 in increments of .2 were used in model 3. It was found that a constant of .4 gave the best estimates and the results shown for models 3 and 3M are for a constant of .4.

TABLE 1

TYPICAL RESULTS OF MODEL 1

October 1943

Mean Inches Runoff = 0.2664

Station	Observed Runoff Ac-ft	Estimated Runoff Ac-ft	Residual Ac-ft	Per Cent Error
09-1885	10,020	6,650	-3,370	-33.6
09-1930	818	514	-304	-37.1
09-1985	396	1,243	847	214.0
09-1995	74	529	455	614.3
09-2020	596	1,847	1,251	209.9
09-2030	682	1,125	443	65.0
09-2040	53	645	592	1,117.1
09-2125	1,250	1,336	86	6.8

Average Error = 269.5%

Average Absolute Error = 287.2%

TABLE 2

TYPICAL RESULTS OF MODEL 2

October 1943

$$\hat{Y}_j = -1{,}073 + 0.03592A_j$$

Station	Observed Runoff Ac-ft	Estimated Runoff Ac-ft	Residual Ac-ft	Per Cent Error
09-1885	10,020	9,686	-334	-3.3
09-1930	818	-241	-1,059	-129.4
09-1985	396	939	543	137.0
09-1995	74	-218	-292	-394.3
09-2020	596	1,916	1,320	221.4
09-2030	682	748	66	9.7
09-2040	53	-29	-82	-155.1
09-2125	1,250	1,088	-162	-13.0

Average Error = 40.9%

Average Absolute Error = 132.9%

TABLE 3

TYPICAL RESULTS OF MODEL 3

October 1943

$\tilde{\alpha}_1 = 0.1206$ $\tilde{\alpha}_2 = 0.6106$ $\tilde{\alpha}_3 = -0.2954$ $\tilde{\alpha}_4 = 0.2785$

$k = 0.4$

Station	Observed Runoff Ac-ft	Estimated Runoff Ac-ft	Residual Ac-ft	Per Cent Error
09-1885	10,020	6,688	-3,332	-33.3
09-1930	818	470	-348	-42.5
09-1985	396	753	357	90.1
09-1995	74	382	308	415.9
09-2020	596	934	338	56.7
09-2030	682	898	216	31.6
09-2040	53	-23	-76	-143.2
09-2125	1,250	1,319	69	5.5

Average Error = 47.6%

Average Absolute Error = 102.4%

TABLE 4

TYPICAL RESULTS OF MODEL 2M

October 1943

$$\hat{Y}_j = -1,073 + 0.03592\ A_j$$

Station	Observed Runoff Ac-ft	Estimated Runoff Ac-ft	Residual Ac-ft	Per Cent Error
09-1885	10,020	9,686	- 334	- 3.3
09-1930	818	1	- 817	- 99.9
09-1985	396	939	543	137.0
09-1995	74	1	- 73	- 98.6
09-2020	596	1,916	1,320	221.4
09-2030	682	748	66	9.7
09-2040	53	1	- 52	- 98.1
09-2125	1,250	1,088	- 162	- 13.0

Average Error = 6.9%

Average Absolute Error = 85.1%

TABLE 5

TYPICAL RESULTS OF MODEL 3M

October 1943

$\tilde{\alpha}_1 = .01206$ $\tilde{\alpha}_2 = 0.6106$ $\tilde{\alpha}_3 = -0.2954$ $\tilde{\alpha}_4 = 0.2785$

$k = 0.4$

Station	Observed Runoff Ac-ft	Estimated Runoff Ac-ft	Residual Ac-ft	Per Cent Error
09-1885	10,020	6,688	-3,332	-33.3
09-1930	818	470	- 348	-42.5
09-1985	396	753	357	90.1
09-1995	74	382	308	415.9
09-2020	596	934	338	56.7
09-2030	682	898	216	31.6
09-2040	53	1	- 52	-98.1
09-2125	1,250	1,319	69	5.5

Average Error = 53.2%

Average Absolute Error = 96.7%

TABLE 6

Average Absolute Per Cent Residual Error

September

Year	Model 1	Model 2	Model 3	Model 2M	Model 3M
1941	512.9	115.5	193.5	105.9	193.5
1942	246.3	88.3	113.8	88.3	113.8
1943	608.9	518.6	139.8	518.6	82.2
1944	2494.6	1348.5	170.3	1348.5	170.3
1945	161.8	57.6	72.0	56.5	72.0
1946	983.5	191.2	163.5	63.1	163.5
1947	89.6	48.7	52.4	48.7	52.4
1948	10310.6	5350.4	834.8	96.1	76.2
1949	1422.4	241.6	206.4	65.1	206.4
1950	192.1	70.7	52.9	70.7	52.9
1951	198.0	155.4	43.8	155.4	43.8
1952	1052.8	147.7	214.9	73.5	214.9
1953	645.3	135.1	91.5	88.5	91.5
1954	101.4	68.5	59.7	65.9	59.7
1955	171.6	150.4	63.6	79.9	63.6
1956	221.2	122.6	158.7	122.6	133.3
1957	51.4	33.8	35.1	33.8	35.1
1958	13099.7	296.8	1971.4	55.9	111.3
1959	70.0	61.3	36.2	55.3	36.2
1960	1023.6	352.2	105.7	62.4	105.7
1961	285.5	149.5	130.6	55.8	130.6
1962	239.5	70.9	181.6	70.9	181.6
1963	97.4	54.7	47.4	45.3	47.4
1964	150.8	126.8	124.0	126.8	124.0
1965	55.5	67.1	44.9	67.1	44.9
1966	180.4	75.6	87.6	75.6	87.6
1967	90.9	89.6	79.6	89.6	79.6
1968	68.6	40.1	32.6	40.1	32.6
1969	241.8	313.9	260.6	313.9	257.9
1970	881.0	313.2	517.2	60.8	517.2
Total	35949.4	10856.3	6286.0	4200.3	3581.4
Average	1198.3	361.9	209.5	140.0	119.4

TABLE 7

Average Absolute Per Cent Residual Error

October

Year	Model 1	Model 2	Model 3	Model 2M	Model 3M
1941	642.6	310.0	260.2	310.0	260.2
1942	185.0	70.8	140.5	70.8	140.5
1943	287.2	132.9	102.4	85.1	96.7
1944	194.2	71.0	55.1	50.0	55.1
1945	404.9	115.6	149.7	105.7	107.2
1946	139.7	89.9	79.0	89.9	79.0
1947	116.9	41.1	42.5	39.5	42.5
1948	56.3	77.7	37.1	77.7	37.1
1949	994.1	404.2	241.3	90.6	148.0
1950	301.1	83.4	100.5	39.7	100.5
1951	67.3	69.7	50.9	69.7	50.9
1952	123.9	95.9	104.0	95.9	96.7
1953	1087.0	206.5	412.6	123.4	301.3
1954	834.1	157.7	448.7	109.2	212.3
1955	126.4	46.9	61.1	46.9	61.1
1956	241.9	174.7	46.7	81.7	46.7
1957	415.6	103.7	152.9	77.6	130.3
1958	15.9	15.1	9.0	15.1	9.0
1959	1167.2	176.9	406.4	134.6	181.5
1960	61.4	17.2	38.2	17.2	38.2
1961	218.6	31.3	119.7	31.3	119.7
1962	37.4	30.7	14.7	30.7	14.7
1963	531.3	215.1	298.9	70.2	288.3
1964	105.2	70.3	33.7	70.3	33.7
1965	428.4	171.2	368.3	171.2	284.2
1966	43.7	51.4	36.8	51.4	36.8
1967	167.8	71.7	89.8	71.7	85.9
1968	97.9	101.6	40.7	101.6	40.7
1969	51.5	62.1	38.3	62.1	38.3
1970	250.2	78.9	112.3	78.9	112.3
Total	8416.4	3345.2	4091.9	2469.7	3249.6
Average	280.5	111.5	136.4	82.3	108.3

TABLE 8

Average Absolute Per Cent Residual Error

November

Year	Model 1	Model 2	Model 3	Model 2M	Model 3M
1941	65.7	38.3	37.3	38.3	37.3
1942	137.4	74.3	96.3	74.3	96.3
1943	340.7	198.5	144.3	117.6	144.3
1944	89.9	46.2	23.9	44.6	23.9
1945	337.6	254.8	57.5	121.5	57.5
1946	94.4	55.9	54.5	55.9	54.4
1947	64.3	23.2	36.5	23.2	36.5
1948	18.6	34.3	14.8	34.3	14.8
1949	602.9	139.6	130.8	134.5	114.2
1950	182.0	54.5	71.9	35.1	71.9
1951	26.3	25.9	18.2	25.9	18.2
1952	66.8	27.2	28.7	27.2	28.7
1953	499.1	160.4	238.1	99.0	172.7
1954	659.4	503.8	431.7	426.7	399.3
1955	271.1	201.2	123.0	128.9	123.0
1956	188.3	142.1	64.9	114.4	64.9
1957	460.3	437.9	167.3	210.8	167.3
1958	24.4	24.6	12.9	24.6	12.9
1959	467.7	168.4	140.5	66.5	140.5
1960	61.4	38.3	30.8	38.3	30.8
1961	58.8	34.4	34.5	34.4	34.5
1962	39.2	30.3	20.9	30.3	20.9
1963	710.0	250.4	210.1	79.1	144.1
1964	103.4	46.1	33.5	46.1	33.5
1965	448.0	286.6	177.7	131.1	150.4
1966	64.9	40.1	56.5	40.1	56.5
1967	171.8	60.3	90.7	60.3	90.7
1968	163.4	135.7	71.0	135.7	71.0
1969	17.0	4.9	12.1	4.9	12.1
1970	243.5	121.6	222.1	121.6	205.1
Total	6678.3	3659.5	2852.9	2525.1	2628.4
Average	222.6	122.0	95.1	84.2	87.6

TABLE 9

Average Absolute Per Cent Residual Error

December

Year	Model 1	Model 2	Model 3	Model 2M	Model 3M
1941	111.0	70.5	59.5	70.5	59.5
1942	291.1	237.4	149.2	237.4	149.2
1943	334.6	312.0	94.1	128.7	94.1
1944	139.5	125.7	31.7	110.8	31.7
1945	321.6	236.5	101.3	121.7	101.3
1946	101.3	62.0	65.6	62.0	65.6
1947	72.8	40.4	35.3	40.4	35.3
1948	18.8	21.2	17.6	21.2	17.6
1949	305.1	181.4	47.6	181.4	47.6
1950	101.0	59.0	39.8	59.0	39.8
1951	37.2	36.9	26.6	36.9	26.6
1952	40.5	26.1	19.4	26.1	19.4
1953	612.2	263.9	198.4	85.9	153.5
1954	600.3	550.6	259.4	482.3	259.4
1955	500.3	513.4	134.9	416.0	134.9
1956	182.5	168.1	80.9	163.3	80.9
1957	351.6	325.5	175.8	234.2	175.8
1958	47.1	43.7	25.9	43.7	25.9
1959	298.2	144.0	125.9	73.4	125.9
1960	239.6	82.6	50.9	76.1	50.9
1961	95.9	62.9	50.2	50.3	50.2
1962	54.1	43.1	35.4	43.1	35.4
1963	2556.8	291.2	385.2	75.4	219.4
1964	110.8	58.6	43.5	55.5	43.5
1965	124.9	95.4	53.8	95.4	53.8
1966	56.3	62.5	24.3	62.5	24.3
1967	82.9	55.8	26.2	55.8	26.2
1968	126.5	120.8	54.1	120.8	54.1
1969	51.7	39.8	30.9	39.8	30.9
1970	250.2	131.5	105.5	105.8	104.1
Total	8126.4	4462.6	2548.8	3375.7	2336.6
Average	270.9	148.8	85.0	112.5	77.9

TABLE 10

Average Absolute Per Cent Residual Error

January

Year	Model 1	Model 2	Model 3	Model 2M	Model 3M
1941	109.8	70.7	58.3	70.7	58.3
1942	302.4	277.3	166.4	277.3	166.4
1943	416.6	447.6	94.2	185.6	92.6
1944	110.6	81.2	26.4	60.5	26.4
1945	325.9	246.1	95.2	140.0	91.0
1946	81.1	62.7	54.1	62.7	54.1
1947	79.7	55.2	41.6	55.2	41.6
1948	13.7	10.9	10.5	10.9	10.5
1949	148.2	91.5	33.8	91.5	33.8
1950	85.9	53.7	38.4	53.7	38.4
1951	38.6	33.8	28.7	33.8	28.7
1952	34.7	26.8	20.9	26.8	20.9
1953	165.1	55.2	52.2	43.6	52.2
1954	503.1	508.4	147.3	475.8	147.3
1955	764.0	773.9	269.0	755.7	269.0
1956	87.8	78.6	48.4	78.6	48.4
1957	261.4	241.3	140.5	211.5	140.5
1958	39.1	35.7	23.6	35.7	23.6
1959	380.8	114.5	174.3	62.0	174.3
1960	805.9	125.0	96.8	103.3	79.8
1961	49.4	29.9	28.5	29.9	28.5
1962	33.2	29.2	21.0	29.2	21.0
1963	6421.1	1201.3	845.4	56.4	307.9
1964	96.9	60.4	57.2	60.4	57.2
1965	76.2	78.1	70.3	78.1	70.2
1966	44.8	48.8	26.8	48.8	26.8
1967	110.0	102.3	45.7	89.4	45.7
1968	120.6	107.3	51.9	107.3	51.9
1969	57.9	29.7	37.6	29.7	37.6
1970	200.9	130.1	43.4	106.1	43.4
Total	11965.5	5207.3	2848.4	3470.1	2288.1
Average	398.9	173.6	94.9	115.7	76.3

TABLE 11

Average Absolute Per Cent Residual Error

February

Year	Model 1	Model 2	Model 3	Model 2M	Model 3M
1941	119.3	93.0	53.5	85.3	53.5
1942	287.8	246.1	147.4	246.1	147.4
1943	382.8	416.6	80.2	200.2	78.2
1944	97.4	69.6	21.1	56.8	21.1
1945	335.6	275.6	93.6	159.4	90.8
1946	74.0	74.7	51.6	74.7	51.6
1947	87.2	56.9	45.0	56.9	45.0
1948	30.7	34.0	12.3	34.0	12.3
1949	56.9	58.6	19.7	58.6	19.7
1950	87.3	51.3	43.3	51.3	43.3
1951	55.0	51.7	42.1	51.7	42.1
1952	30.0	24.2	21.4	24.2	21.4
1953	101.4	51.6	22.4	51.6	22.4
1954	169.8	155.8	54.2	150.2	54.2
1955	210.1	176.5	72.8	167.7	72.8
1956	68.4	60.0	35.9	60.0	35.9
1957	148.9	111.6	54.6	93.1	54.6
1958	31.4	31.9	17.3	31.9	17.3
1959	305.2	97.6	158.4	56.0	158.4
1960	895.3	344.0	91.7	120.3	91.7
1961	65.6	46.4	26.9	44.1	26.9
1962	29.9	22.4	11.1	22.4	11.1
1963	150.3	60.3	33.5	60.3	33.5
1964	119.1	91.7	64.9	81.5	64.9
1965	114.1	94.3	67.5	94.3	67.5
1966	70.3	45.1	27.3	45.1	27.3
1967	96.2	95.0	42.3	84.2	42.3
1968	112.7	94.2	44.3	94.0	44.3
1969	59.6	79.6	52.2	79.6	52.2
1970	127.6	114.8	48.4	109.6	48.4
Total	4519.9	3225.1	1556.9	2545.3	1552.2
Average	150.7	107.5	51.9	84.9	51.7

TABLE 12

Average Absolute Per Cent Residual Error

March

Year	Model 1	Model 2	Model 3	Model 2M	Model 3M
1941	141.3	54.7	79.1	54.7	79.1
1942	318.0	277.1	156.0	277.1	156.0
1943	161.0	170.2	31.5	121.6	31.5
1944	72.0	50.3	12.1	50.0	12.1
1945	194.7	108.2	38.4	90.8	38.4
1946	74.5	89.7	50.5	89.7	50.5
1947	61.2	42.1	33.1	42.1	33.1
1948	38.2	38.0	13.3	38.0	13.3
1949	37.1	25.9	19.1	25.9	19.1
1950	58.4	37,8	33.4	37.8	33.4
1951	42.2	23.5	24.5	23.5	24.5
1952	33.9	17.0	23.3	17.0	23.3
1953	98.6	51.8	23.3	51.8	23.3
1954	73.3	60.2	28.1	60.2	28.1
1955	119.7	93.8	36.0	83.5	36.0
1956	70.9	74,7	30.2	71.1	30.2
1957	99.2	62.0	23.2	52.6	23.2
1958	23.3	25.7	14.1	25.7	14.1
1959	298.2	77.3	182.5	50.3	182.5
1960	59.3	66.9	25.1	61.4	25.1
1961	52.9	27.7	24.6	27.7	24.6
1962	17.8	14.6	10.3	14.6	10.3
1963	131.3	55.7	35.2	55.3	35.2
1964	81.2	53.4	38.1	52.1	38.1
1965	117.0	90.3	60.0	90.3	60.0
1966	47.7	37.0	19.4	37.0	19.4
1967	72.9	65.0	29.5	61.1	29.5
1968	135.5	129.1	39.2	119.1	39.2
1969	102.3	118.7	76.1	118.7	76.1
1970	172.4	164.4	50.0	139.3	50.0
Total	3006.0	2202.8	1259.2	2040.0	1259.2
Average	100.2	73.43	42.0	68.0	42.0

Tables 6 through 12 indicate that, except for October and November, the ridge regression model provides the best estimates. Since this model should produce the best estimates during the snowmelt runoff period, only model 3M was examined for April through August. Table 13 presents a summary of the results of these computations.

It is apparent that the estimates are improved for the snowmelt runoff months over those of the base flow months, with June runoff being estimated the best. May streamflows frequently are estimated well, but occasional large errors make the average error fairly high. July estimates can be considered to be fair.

Reviewing the assumption of no significant lake regulation, the results were examined to determine if regulation was appreciable. For station 09-1930 it was found that April and May were consistently overestimated and July was underestimated, indicative of irrigation regulation on New Fork Lake. Unfortunately, this error in assumption was not discovered until the voluminous computations had been completed. The results would be improved somewhat for April, May and July if station 09-1930 were deleted for those months.

Conclusions

Of the several models tested, the elevation dependent, ridge regression model appears to produce the best estimates of monthly streamflow for the high runoff and summer periods. For the base flow period, models 2M and 3M estimate about equally good. The estimates from this model should probably be considered to be fair to poor for the base flow months. However, it should be kept in mind that the large residual per cent errors are primarily due to division by a small number (small observed runoff values). The larger streamflows are generally estimated in the range of 5 to 40 per cent error.

Model 3M assumes that runoff is solely a function of elevation. That is, the precipitation pattern, temperatures, vegetation, soils, geologic formations and other factors are homogeneous within elevation bands. This assumption appears to be reasonable for the high flows. For low flows, other factors need to be taken into account.

Since the high flow periods are usually the ones of greatest interest in hydrologic studies, the elevation dependent, ridge regression model is recommended for estimation of streamflow in drainage basins in or near the mountains during the high runoff period. Appendix III presents a procedure for using this model.

TABLE 13

Average Absolute Per Cent Errors

On Model 3M

Year	April	May	June	July	August
1941	125.4	15.4	8.4	40.5	133.1
1942	115.8	59.0	11.7	25.3	51.0
1943	89.7	13.9	16.1	18.3	65.7
1944	23.2	16.5	15.5	20.1	75.5
1945	32.1	63.5	18.2	13.9	55.3
1946	65.6	12.0	17.7	24.3	204.3
1947	39.1	33.7	11.1	13.0	80.7
1948	25.9	23.7	12.8	65.7	85.3
1949	18.4	73.7	10.9	24.3	128.9
1950	55.7	157.0	13.2	11.3	72.5
1951	21.6	19.2	12.9	18.5	86.4
1952	30.5	18.4	16.0	14.5	100.4
1953	26.9	48.9	12.9	36.2	55.7
1954	23.9	32.5	16.4	23.3	239.1
1955	41.3	54.8	11.9	47.3	112.3
1956	38.8	90.3	10.3	49.7	213.7
1957	36.3	243.7	17.0	13.5	91.6
1958	21.1	18.9	15.4	77.7	312.5
1959	94.7	460.0	13.6	54.5	83.5
1960	34.6	29.2	17.9	62.2	290.5
1961	34.5	180.6	15.2	66.5	117.7
1962	48.9	145.3	11.5	19.2	151.3
1963	29.8	55.6	17.2	32.1	101.9
1964	37.2	51.8	10.5	19.5	254.9
1965	57.5	75.7	14.1	12.1	61.6
1966	22.9	31.3	15.5	50.2	154.1
1967	30.7	191.2	15.5	11.5	130.2
1968	18.3	49.8	13.1	29.3	59.8
1969	65.9	19.6	12.3	13.5	60.0
1970	20.4	149.2	15.2	40.8	132.3
Total	1326.6	2434.4	418.1	948.9	3762.0
Average	44.2	81.1	13.9	31.6	125.4

APPENDIX I - REFERENCES

Hoerl, A.E. and Kennard, R.W. "Ridge Regression: Applications to Nonorthogonal Problems." Technometrics (February, 1970), 69-82.

Hoerl, A.E. and Kennard, R.W. "Ridge Regression: Biased Estimation for Nonorthogonal Problems." Technometrics (February, 1970), 55-66.

Marquardt, D.W. "Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation." Technometrics (August, 1970), 591-595.

Smith, Verne E., Anderson, Donald A. and Scott, Richard G. "Elevation Dependent Model for Estimating Annual Runoff." Accepted for publication, Hydraulics Division, American Society of Civil Engineers, December 1973.

APPENDIX II - SUMMARY OF RIDGE REGRESSION PROCEDURES

It is known that multicolinearity among the independent variables in multiple regression problems produces instability of the individual regression coefficients. In applications where the individual coefficients are to be interpreted and estimates made on future observations, such instability is undesirable. Hoerl and Kennard (1970) have introduced an estimation procedure which controls this instability. The procedures developed in those papers were employed in the development of the elevation dependent model and are summaraized here for the convenience of the reader.

Consider the standard linear model

$$\underline{y} = X\underline{\beta} + \underline{e}, \ \underline{e} : N(\underline{0}, \sigma^2 I_n), \tag{10}$$

where X is an n x p matrix of known constants converted into correlation form (i.e. for each column of X, $\underline{x}_i$ $i = 1,2,...,p$ we have $\underline{x}_i \ \underline{1} = 0$ and $\underline{x}_i \ \underline{x}_i = 1$, where $\underline{1}$ denotes a vector of all ones). Let $\lambda_1, \lambda_2, ..., \lambda_p$ denote the characteristic roots of X'X. The least squares estimates of $\underline{\beta}$ are given by $\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{y}$ and var $\{\hat{\underline{\beta}}\} = \sigma^2(X'X)^{-1}$.

If we consider the square error of the estimates, $L^2 = (\hat{\underline{\beta}} - \underline{\beta})' \ (\hat{\underline{\beta}} - \underline{\beta})$, it is easily shown

$$E\{L^2\} = \sigma^2 tr(X'X)^{-1} = \sigma^2 \sum_{i=1}^{p} (1/\lambda_i) \tag{11}$$

$$Var\{L^2\} = 2\sigma^4 tr(X'X)^{-2} = 2\sigma^4 \sum_{i=1}^{p} (1/\lambda_i)^2. \tag{12}$$

If the independent variables possess a high degree of multicolinearity, some of the characteristic roots will be near zero, giving rise to large mean and variance of the square error.

The ridge regression estimates are given by

$$\tilde{\underline{\beta}}(k) = (X'X + kI)^{-1}X'y \qquad 0 \leq k \leq 1 \tag{13}$$

For the ridge estimate, $\tilde{\underline{\beta}}$, the expected value of the square error is given by

$$E\{L^2(k)\} = \sigma^2 \sum_{i=1}^{p} \lambda_i/(\lambda_i + k)^2 + k^2 \, \underline{\beta}'(X'X + kI)^{-2}\underline{\beta} \qquad (14)$$

where the first term denotes the sum of the variances of the estimates and the second denotes the bias squared of the ridge estimates (note that the ridge regression procedure does not produce unbiased estimates). It is shown that there always exists a $k > 0$ for which the expected square error of the ridge estimates of the coefficients is less than that for the least square estimates. Thus the ridge estimates possess a greater degree of stability than do the least square estimates and are often better for purposes of interpretation and estimation.

APPENDIX III - PROCEDURE FOR USING THE ELEVATION DEPENDENT, RIDGE REGRESSION MODEL, 3M

The procedure for applying the elevation dependent model to obtain streamflow estimates is fairly straightforward and is outlined below. USGS 7½' quadrangle maps are a useful tool in the analysis.

1. Determine drainage basins with known streamflow values for concurrent periods near the basin(s) to be estimated. All the basins should be fairly physically homogeneous within elevation bands.

2. Select elevation bands; example: 7,500-8,500 feet, 8,500-9,500 feet, 9,500-10,500 feet and above 10,500 feet. The selection should take into account physical considerations such as timberline, topographic relief, etc. The number of bands used cannot exceed the number of basins, and it is preferable, statistically, to have several more basins than elevation bands.

3. Determine the area within each elevation band for each basin.

4. Calculate the proportions of areas. For each basin, divide each elevation band area by the total area of that drainage basin. The proportions for each basin will total 1. A large range (ideally at least .5, but .3 is acceptable) of values of proportions within bands is desirable. If a large range does not exist, reselection of elevation bands should be considered.

5. Solve for the α estimates (inches of runoff for each elevation band) using data from the basins with known streamflow values. The model is most easily solved by matrix algebra using a computer. The program used in this study is listed below. It is written in BASIC because matrix operations can be easily

programmed in this language. The data input to this program are:

a. The number of drainage basins and the number of elevation bands (statement 1289).

b. The drainage basin areas in square miles (statement 1290).

c. The proportions of areas of elevation bands for each basin (statements 1300 through 1370). Each statement line is the proportions of one basin, and the statements are in order respective to the order of the total drainage area input (statement 1290).

d. The streamflow values of the basins in acre-feet for the particular month of a particular year (statement 1380) in order respective to the order of the drainage area input.

6. Calculate the streamflow for the desired basin(s). The output of the program is the α estimates. Multiply each α by its respective elevation band area (in acres) of the desired basin(s), sum the products for the basin(s), and divide by 12 to obtain the estimate(s) in acre-feet.

Note that the program provides α estimates for only one month of one year. If estimates for various months and years are desired, the program can be modified easily to perform the same computations repeatedly on successive input streamflow data.

PROGRAM LISTING

```
10 DIM D(11),Z(11),L(4),G(4),F(4),M(4),C(248,12)
20 DIM P(11,4),S(11,4),X(11,4),W(4,11),E(4,11)
30 DIM Y(11,1),V(4,1),J(4,4),Q(4,4),U(4,4),T(4,4),L(4,4)
40 DIM N(44),A(4),B(4),H(4)
50 REM A1 = NØ. OF STATIONS.
60 REM B1 = NØ. OF ELEVATION BANDS.
70 READ A1,B1
80 C1=A1*B1
100 MAT SIZE D(A1),Z(A1),K(B1),G(B1),F(B1),M(B1)
110 MAT SIZE P(A1,B1),S(A1,B1),X(A1,B1),W(B1,A1),E(B1,A1)
120 MAT SIZE Y(A1,1),V(B1,1),J(B1,B1), Q(B1,B1),U(B1,B1)
130 MAT SIZE T(B1,B1),L(B1,B1)
140 MAT SIZE N(C1),A(B1),B(B1)
150 MAT READ D
160 MAD READ P
170 MAT READ Z
200 MAT H= ZER
310 REM AVERAGE ØF P
320 FØR J=1 TØ B1
330 M(J)=0
340 FØR I=1 TØ A1
350 M(J)=P(I,J)+M(J)
360 NEXT I
370 M(J)= M(J)/A1
380 NEXT J
400 REM DETERMINE Y
410 FØR I=1 TØ A1
420 Y(I,1)=(Z(I)/D(I))/640*12
430 NEXT I
440 FØR J=2 TØ B1
450 K(J)=0
460 FØR I =1 TØ A1
470 S(I,J)=(P(I,J)-M(J))**2
480 K(J)=S(I,J)+K(J)
490 X(I,1)=1
500 NEXT I
510 K(J)=SQR(K(J))
520 NEXT J
530 FØR J=2 TO B1
540 FØR I =1 TØ A1
550 X(I,J)=(P(I,J)-M(J))/K(J)
560 NEXT I
570 NEXT J
580 MAT W=TRN(X)
590 MAT J=W*X
600 MAT Q=IDN(B1,B1)
610 Q(1,1)=0
620 A3=A3+1
```

```
650 D5=0
660 FOR K=0 TO .6 STEP .2
670 D5=D5+1
680 ;
690 PRINT" K =  ",K
700 MAT U=(K)*Q
710 MAT T=U+J
720 MAT L=INV(T)
730 MAT E=L*W
740 MAT V=E*Y
750 FOR J=2 TO B1
760 F(J)=V(J,1)/K(J)
770 NEXT J
780 S1=0
790 FOR J=2 TO B1
800 O(J)=F(J)*M(J)
810 S1=O(J)+S1
820 NEXT J
830 G(1)=V(1,1)-S1
840 ;
850 PRINT"                 ALPHA ESTIMATES"
860 PRINT G(1),
870 FOR J=2 TO B1
880 G(J)=F(J)+G(1)
890 PRINT G(J),
900 NEXT J
910 PRINT
920 PRINT"    PRED      OBS        DIFF                  % ERROR"
930 L=0
940 FOR I =1 TO A1
950 FOR J=1 TO B1
960 L=L+1
970 N(L)=P(I,J)*D(I)*640
980 NEXT J
990 NEXT I
1000 L=0
1010 C4=0
1020 C5=0
1030 FOR I = 1 TO A1
1040 C6=0
1050 FOR J=1 TO B1
1060 L=L+1
1070 C6=C6+(G(J)*N(L))/12
1080 NEXT J
1085 IF C6>0 THEN 1090
1086 C6=1
1090 C2=C6-Z(I)
1100 C3=C2/Z(I)*100
```

```
1110 PRINT C6,Z(I),C2,C3
1120 C4=C4+C3
1130 C5=C5+ABS(C3)
1140 NEXT I
1150 C4=C4/A1
1160 C5=C5/A1
1170 ;
1180 PRINT "AVE %ERROR"C4,"AVE ABS % ERROR",C5
1190 H(D5)=H(D5)+C5
1200 IF I4<31 THEN 1260
1210 D6=H(D5)/31
1220 PRINT"SUM OF AVE ABS % ERROR = ",H(D5)
1230 PRINT"MEAN OF AVE ABS % ERROR = ",D6
1260 NEXT K
1280 STØP
1289 DATA 8,4
1290 DATA 468,36.2,87.5,37.2,130,79.2,45.4,94
1300 DATA .268,.341,.142,.249
1310 DATA .219,.218,.127,.436
1320 DATA .174, .167,.253,.406
1330 DATA .295,.237,.206,.262
1340 DATA .153,.108,.271,.468
1350 DATA .07,.315,.278,.337
1360 DATA .059,.139,.566,.236
1370 DATA .143,.374,.204,.279
1380 DATA 10020,818,396,74,596,682,53,1250
```