THE RELIABILITY OF LOW FLOW REGIONAL EQUATIONS IN BRITISH COLUMBIA

Rory M. Leith

1978

The reliability of low flow regional equations in British Columbia.

GB
1230
.B7
L45
1978

The reliability of low flow regional equations in British Columbia.

Author(s) Name Rory M. Leith

Risk and Reliability Symposium
Waterloo, Canada
June 26 - 28, 1978

THE RELIABILITY OF LOW FLOW REGIONAL EQUATIONS IN BRITISH COLUMBIA

Type Within This Area

Rory M. Leith
Planning and Studies Section
Water Survey of Canada
#502 1001 West Pender Street
Vancouver, British Columbia
V6E 2M9

SYNOPSIS

The reliability of estimates of low flows is examined by least squares fitting and Kalman filtering. These techniques applied to a network of hydrometric stations in British Columbia allow the contribution of individual stations to be assessed. The changes in reliability of estimates are studied as a function of the changes in the size of the network. The subdivision of the study area is examined by a split test sample with inconclusive results. Factors such as common time base of the data and season of occurrence of low flow are found to influence the accuracy of low flow estimates.

INTRODUCTION

The collection of data is not an end unto itself. The final goal must be to provide information for data users. To maximize the information available to data users, with the constraints of budget and manpower, a data gathering network requires analysis of reliability.

As a network can be considered a system, the analysis of reliability can be made with optimal state estimation techniques. The simplest of these techniques is least square fitting or regression in conjunction with linear Kalman filtering. These techniques also provide a method of making estimates at ungauged sites and gaining an indication of the contributions of a particular station in the network.

This paper describes the network analysis for one hydrologic quantity, mean annual seven-day low flow, for the Water Survey of Canada's hydrometric network in British Columbia. The description of the methods of analysis is brief as these are well described in the literature, particularly by Solomon (1975) and Gura (1976).

PROCEDURE

Mean annual seven-day low flow divided by the contributing drainage area (UMLF) was chosen as the response variable. The units of UMLF are cubic feet per second per square mile.

Eighty-five hydrometric stations were selected for the analysis on the basis of the following conditions:

i) recorded natural or near natural flows;
ii) at least five complete years of record;
iii) recorded flows from basins with drainage areas between 80 and 800 square miles (200 to 2000 square kilometers);
iv) recorded flows from basins for which basin-averaged physiographic parameters are available.

With regard to requirement i), near natural flow refers to flow with mild artificial modification.

Five years were considered to be the minimum necessary to define UMLF. The station records are not on a common time base nor are they necessarily continuous.

Because the predictor variables were basin-averaged physiographic parameters extracted from a 10 km x 10 km grid, the lower drainage area limit was imposed so that the average would be taken over at least three squares. The upper limit restricts the station to sampling flow from primarily one region.

Geophysical data, such as basin-averaged physiographic parameters are usually intercorrelated and do not have uniform variance. Therefore, strictly speaking, an F-test cannot be used for determining the significance of terms in a regression equation, and backward elimination and forward selection would be ruled out in development of the best equation. However, an operational assumption is made that F-tests can be used, at least as a guide to significance, but validation tests and examining residuals for bias and normality become important in selecting the best equation.

In batch analysis the information provided by the regression equation can be measured by R^2 where R^2 measures the proportion of the total variation about the mean $\overline{Y}$ explained by the regression. The reliability could be assessed by the standard error of validation or the mean square of residuals when the equation is applied to a set of stations not used in developing the equation. In this paper an effort is made to relate standard error of calibration or standard error of estimate and the standard error of validation to the number of stations used in developing the equation.

In order to assess the contribution of an individual station a linear Kalman Filter was applied to stations added one by one to a base sample. The change in the trace of the P matrix was used as a criterion of the information added by the station. The trace of the P matrix should decrease monotonically with the number of observations.

The advantage of the linear Kalman Filter is computational; by adding data points one at a time, matrix inversions are replaced by arithmetic divisions. The P matrix, given by $(A^*WA)^{-1}$, is a measure of the noise in the process, that is $E[(\hat{x}-x)(\hat{x}-x)^*]$, where $\hat{x}$ is the vector of estimated parameters and x is the vector of model parameters. A is the matrix of observations and W is the weighting matrix.

A consideration is the adequacy of the model for which the parameters are estimated, in this case how good is the linear model. No tests of goodness of fit are used other than tests of residuals.

To examine the changes in standard errors of calibration and validation with size of calibration sample, sets of random samples without replacement of 16, 32, 48 and 64 stations were taken from the 85 stations. For each sample a regression equation was developed for UMLF. No predictor variable transformations were considered for this part of the study.

Equations were developed by selecting from the correlation matrix those physiographic parameters most highly correlated with UMLF. if parameters were highly correlated, the parameter with the highest F-value was selected as a possible regression variable. The equations were developed by backward elimination procedure of TRIP, Triangular Regression Package of the University of British Columbia. For the smaller samples the number of observations limited the maximum number of variables in an equation.

For each equation the standard error of estimate was taken as the standard error of calibration. When the equation was applied to the stations not used in its development, the mean square of the residuals so produced was the standard error of validation. For each equation the significant physiographic parameters were noted so that overall important parameters could be identified.

Equations for all 85 stations were developed by considering all variables as possible and using backward elimination. Scatter plots were examined for indication of transformation.

Residuals were tested by probability plots, geographical plots, plots against predicted UMLF and by calculation of skew and kurtosis.

A geographical plot of residuals provided evidence for subdividing the province into three regions: Southeastern British Columbia, Coast and Vancouver Island, and Interior. For the Coast and Island region, scatter plots suggested transformation of the predictor variables.

With the transformed variables, equations were developed and tested. The value of this subdivision or regionalization was tested by examining standard errors of residuals before and after subdivision.

The linear Kalman Filter as described by Gura (1976) can be summarized:

$$P_{k+1} = P_k - P_k A^*_{k+1}(W^{-1}_{k+1} + A_{k+1}P_k A^*_{k+1})^{-1} A_{k+1}P_k \tag{1}$$

$$\hat{x}^{k+1} = \hat{x}^k + P_k A^*_{k+1}(W^{-1}_{k+1} + A_{k+1}P_k A^*_{k+1})^{-1}(y^{k+1} - A_{k+1}\hat{x}^k) \tag{2}$$

The data is added one station at a time, so A_{k+1} is a row vector of the predictor variables, W_{k+1} is a scalar. The quantity in brackets in equations (1) and (2) is also a scalar so that the inversions are reduced to arithmetic divisions.

There are two ways of starting the recursion cycle, first by finding $\hat{x}^1$ and P^1 from a base sample, and second by guessing the parameter vector and its covariance.

In this study the weight of each station was taken as 1. Several combinations of 16 stations were taken to identify a median base in terms of the trace of the P matrix. Once this base had been identified, stations were added to it to examine changes in standard error of estimate, trace of P matrix and standard error of validation as the number of stations in the regions increased.

To assess the effect of a more limited time base a subsample of 69 stations with records between 1950 and 1972 was analyzed. Stations do not necessarily have records over the complete period, for example there may be records from 1962 to 1970.

The effect of time of occurrence of low flow was examined by establishing two categories. Basins in Category 1 had low flows primarily in the winter. Stations have lows equally distributed between fall and winter were excluded from analysis.

RESULTS

Results of the random sampling equation development are summarized in Figure 1. Mean values and mean values plus and minus one standard deviation have been plotted to indicate the spread in standard errors and standard deviations for each size of random sample. In the following discussion the units of UMLF, means, standard deviations and errors are cubic feet per second per square mile.

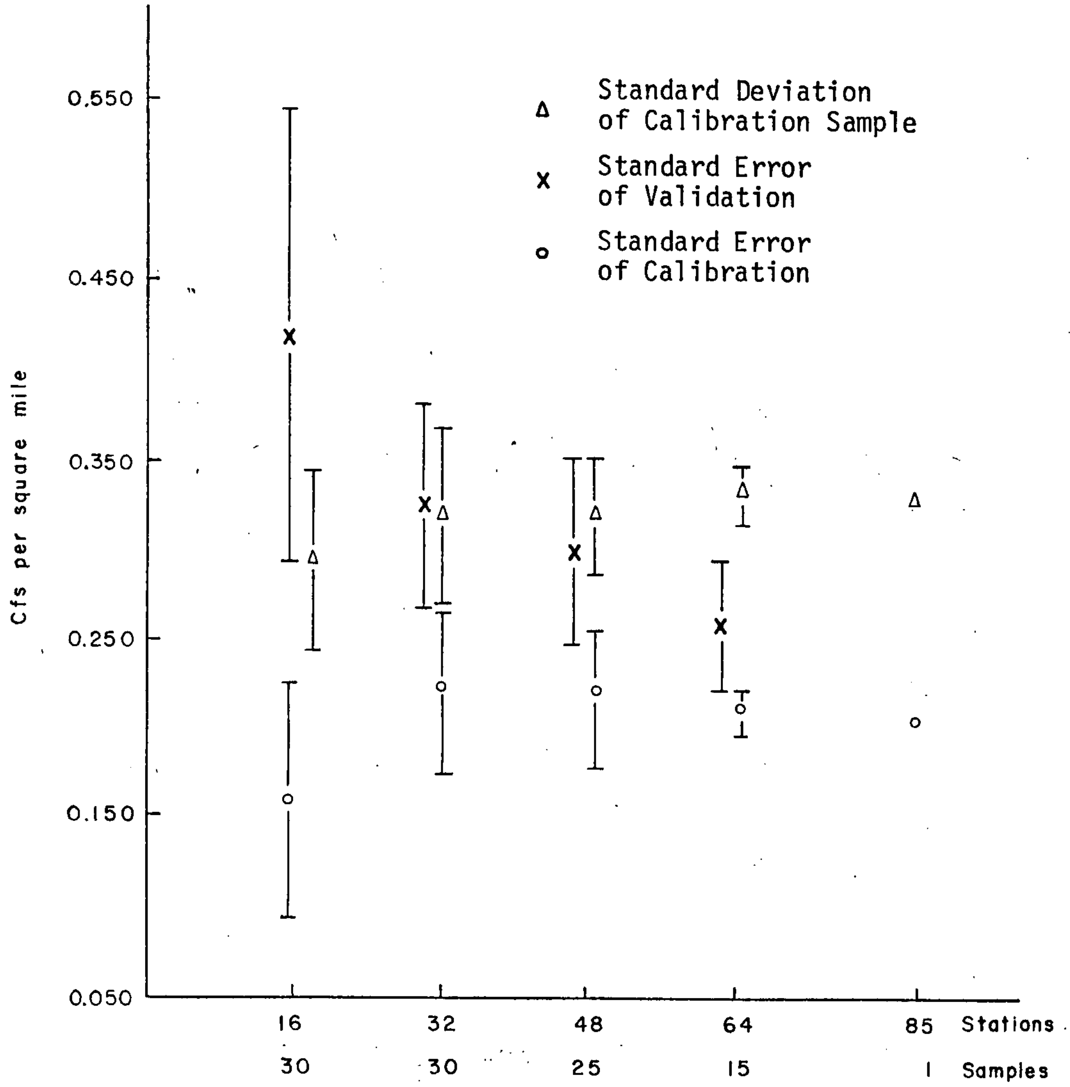

Figure 1. Results of Random Sampling on Standard Deviation, Standard Error of Calibration and Standard Error of Validation

Figure 1 suggests that for the 85-station equation the standard error of validation would fall between 0.200 and 0.260. With the mean of UMLF for the 85-station sample being 0.394, the percent of response mean is between 51% and 66%. The final 85-station equation had a standard error of calibration of 0.204, well below the standard deviation of 0.325. The variability of the 64-station samples is probably underestimated because of the number of common stations appearing in the samples.

Because of the difficulty in determining the best equation by F-tests, two 85-station equations were developed: a 10-variable equation for analysis and a 7-variable equation for quick estimates. The definitions and ranges of the physiographic parameters used in the study are given in Appendix A.

85-Station Equation (10 Variables)

$R^2 = 0.6524$ FPROB = 0.0000 Skew = 0.595
S.E. = 0.2040 52% of response mean Kurtosis = 3.83

$$\begin{aligned} \text{UMLF} = {} & 0.6689 + 0.0130\ \text{NPOSI} - 0.0001170\ \text{ELEV} + 0.005350\ \text{SLP\%} \\ & - 0.0007083\ \text{DSN} - 0.0008782\ \text{DSNW} + 0.02073\ \text{RALKE} \\ & - 0.00009903\ \text{BHW} + 0.000008956\ \text{SENW} + 0.00001215\ \text{SEW} \\ & - 0.000003755\ \text{SESW} \end{aligned} \tag{3}$$

85-Station Equation (7 Variables)

$R^2 = 0.6015$ FPROB = 0.0000 Skew = 0.594
S.E. = 0.2141 54% of response mean Kurtosis = 3.92

$$\begin{aligned} \text{UMLF} = {} & 0.4998 + 0.0111\ \text{NPOSI} - 0.00006876\ \text{ELEV} + 0.006429\ \text{SLP\%} \\ & - 0.0006558\ \text{DSN} - 0.0005961\ \text{DSNW} + 0.0307\ \text{RALKE} \\ & + 0.000004838\ \text{SENW} \end{aligned} \tag{4}$$

A geographical plot of residuals from the 10-variable equation indicated several large residuals in the Coast and Vancouver Island area while in Southeastern British Columbia there is a concentration of positive residuals. Stations in these regions were used in developing separate regional equations. The remaining stations were grouped in the Interior region.

Equation for Southeastern British Columbia with 31 Stations

$R^2 = 0.8680$ FPROB = 0.0000 Skew = 0.264
S.E. = 0.0796 22% of response mean Kurtosis = 1.87

$$\begin{aligned} \text{UMLF} = {} & 4.5774 - 0.0254\ \text{NPOSI} - 0.001135\ \text{DSN} + 0.001533\ \text{DSW} \\ & - 0.0001007\ \text{BHW} + 0.00001231\ \text{SENW} - 0.009276\ \text{SSSE} \end{aligned} \tag{5}$$

Equation for Coast and Vancouver Island with 21 Stations

$R^2 = 0.8466$ FPROB = 0.0001 Skew = 0.209
S.E. = 0.2085 32% of response mean Kurtosis = 1.93

$$\begin{aligned} \text{UMLF} = {} & 4.0151 - 0.00006356\ (\text{NPOSI})^2 - 0.0001874\ (\text{NPOSJ})^2 \\ & - 0.00000005229\ (\text{ELEV})^2 - 0.00000005452\ (\text{BHN})^2 \\ & + 0.0000001737\ (\text{BHW})^2 - 0.0000002577\ (\text{BHSW})^2 \end{aligned} \tag{6}$$

Equation for Interior with 33 Stations

$R^2 = 0.8043$ FPROB = 0.0000 Skew = 0.158
S.E. = 0.1113 44% of response mean Kurtosis = 3.22

$$\begin{aligned} \text{UMLF} = {} & 0.2869 + 0.001322\ \text{NPOSI} + 0.002986\ \text{SLP\%} + 0.0413\ \text{RALKE} \\ & - 0.0001325\ \text{BHW} - 0.000006038\ \text{SESW} - 0.009456\ \text{SSNE} \\ & + 0.004374\ \text{SSE} \end{aligned} \tag{7}$$

The value of regionalization could be assessed by the reduction in standard error of calibration, in particular in the reduction in percent of response mean. However, the number of stations in each region is low, particularly the Coast and Vancouver Island, and by Figure 1 the error produced when the equations are applied to a split sample may be much larger than the standard error of calibration.

Table 1 shows the results for a split sample of 18 stations. The lumped standard error of validation for the regional equations is lower than that of the 10-variable overall equation. However, an F-test indicates the difference is not statistically significant.

Table 1. Split Sample Test

Station Number	Residuals for 10-Variable 85-Station Equation	Residuals for Southeast British Columbia	Residuals for Coast and Vancouver Island	Residuals for Interior	Observed UMLF
08GA024	0.181		0.099		0.918
08NH120	-0.188	-0.076			0.099
08DC006	0.147		0.097		0.425
07FC003	-0.597			-0.047	0.003
08LG048	-0.412			-0.144	0.159
08KA001	0.081			-0.070	0.280
08NG004	0.112	0.169			0.121
08ND014	0.232	0.222			0.649
08EB004	-0.067			-0.023	0.236
08HF001	-0.334		-0.964		0.292
08KH008	-0.241			-0.408	0.312
08GA054	1.71		0.908		1.89
08KH019	0.054			0.015	0.060
08HF002	0.251		-0.186		0.964
08ME015	-0.511		-0.170		0.496
08FA001	0.315		-0.576		0.781
08MF009	0.043		0.266		0.843
08DD001	0.261		0.074		0.604
Sum of Squares	2.72	0.084	2.24	0.195	
Number of Variables	10	6	6	7	
Mean Square of Standard Error of Validation	0.623		0.501		

$$\text{FCAL} = \frac{2.72/(18-10-1)}{2.52/(18-7-1)} = 1.54$$

FTAB(7,10) = 3.14 at 5%

The residuals are not normally distributed for either the 10-variable 85-station equation or the regional equations, given in Figure 3.

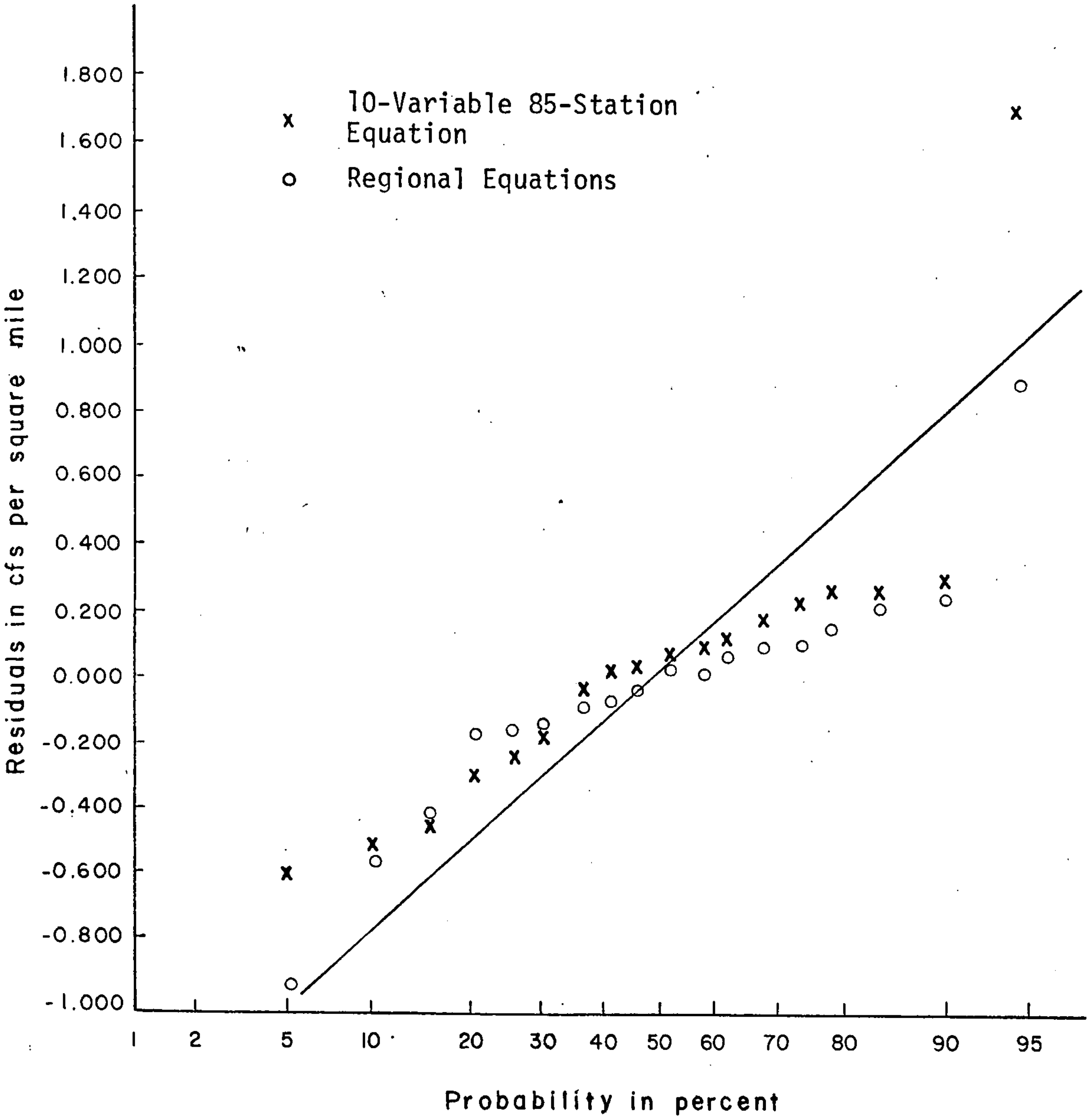

Figure 3. Probability Plot of Residuals for Split Sample

Figure 4 shows the results for Kalman filtering applied to the 85-station sample. The base sample is the first 16 stations in an alphabetic listing of all 85. The eight most frequently significant physiographic parameters, given in Appendix B, are used in the model. The standard error of calibration is an unstable estimate of reliability for small degrees of freedom; not until around 30 degrees of freedom or 40 stations does the estimate settle down. Here reliability is closeness to the minimum standard error of validation. The settling down of the standard error of calibration coincides with the decrease in the rate of change of the P matrix.

When applied to the Southeast region, Kalman filtering indicates which stations provide the largest changes in the P matrix, given in Table 2.

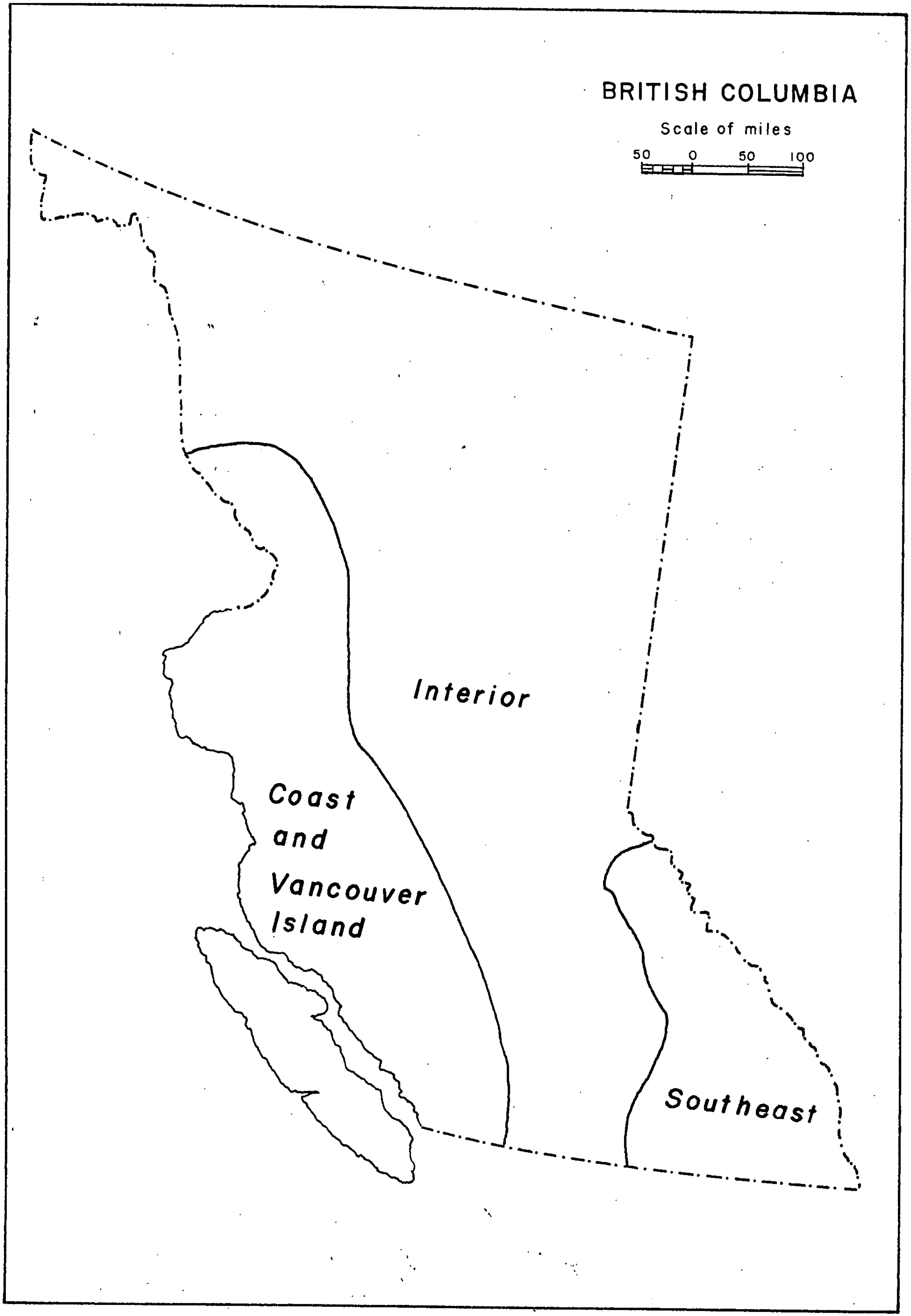

Figure 2. Map of British Columbia Showing the Three Regions for which Equations have been Developed

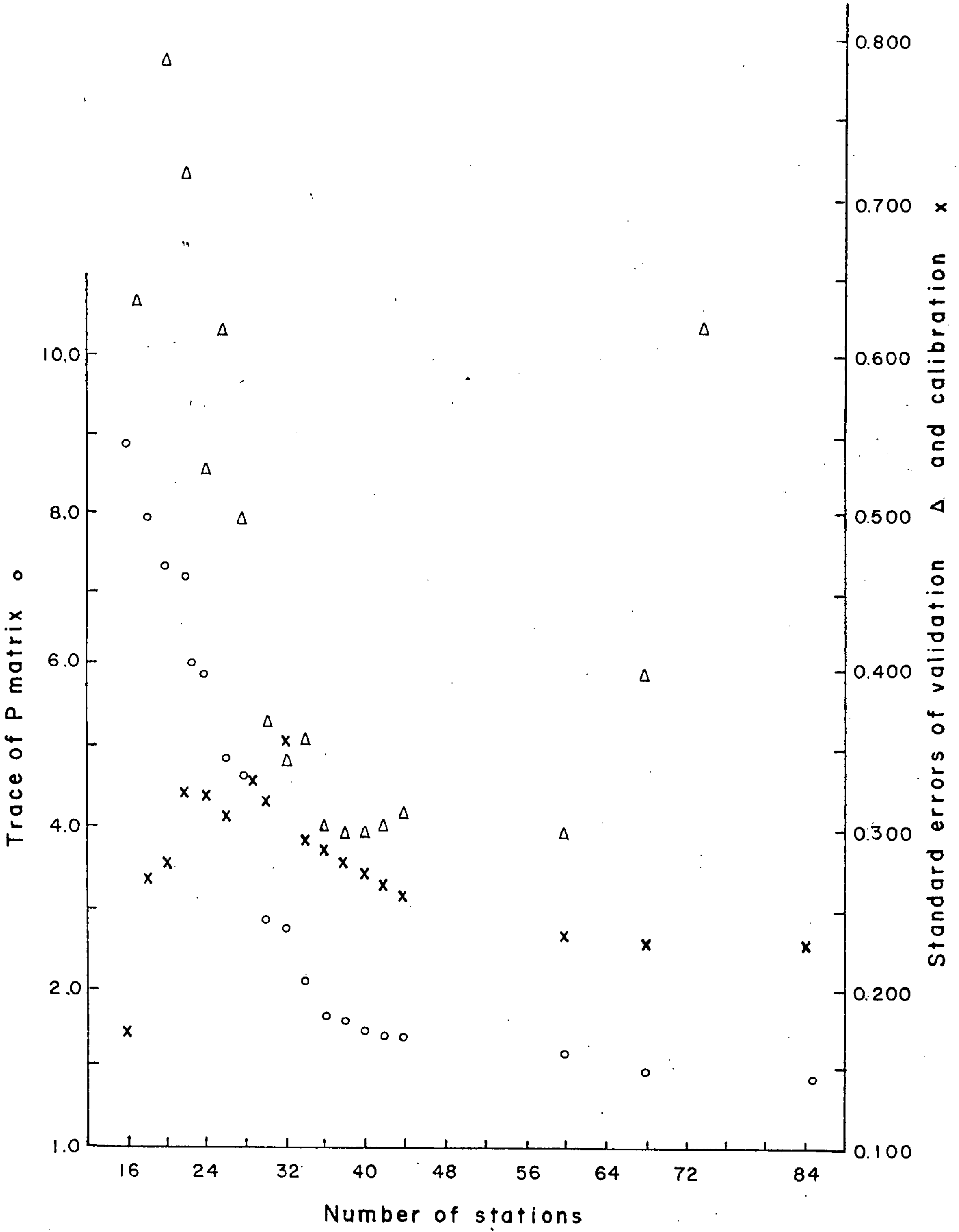

Figure 4. Behaviour of Trace of P Matrix, Standard Error of Calibration, Standard Error of Validation, with Number of Stations Analyzed

Table 2. Changes in Trace of P Matrix

16-Station Base Sample Trace P = 435.4

Station Number	Trace P	Change	Standard Error of Calibration
08NF001	330.9	104.5	0.106
08NE006	434.7	0.7	0.108
08NH066	409.9	25.5	0.104
08NH007	403.3	32.1	0.113
08NH006	190.2	245.2	0.106
08NH034	405.6	29.8	0.107
08NG046	406.9	28.5	0.094
08NE074	184.2	251.2	0.105
08NE044	180.8	254.6	0.106
08LE027	277.3	158.1	0.089
08NG051	412.4	23.0	0.088
08NJ014	431.2	4.2	0.104
08NA011	432.1	3.3	0.111
08NA012	426.7	8.7	0.107
08NF004	428.2	7.2	0.123

For the 69 stations with records between 1950 and 1972 the following equation was developed:

$R^2 = 0.7131$ FPROB = 0.0000 Skew = 0.174
S.E. = 0.174 49% of response mean Kurtosis = 3.57

$$\begin{aligned} UMLF = {} & 0.8306 + 0.0110\ NPOSI - 0.00008874\ ELEV + 0.007203\ SLP\% \\ & - 0.0007726\ DSN - 0.0005977\ DSNW - 0.00005684\ BHW \\ & + 0.000005645\ SENW \end{aligned} \tag{8}$$

The standard error of calibration is considerably lower than the mean standard error of calibration minus one standard deviation for 69 stations in Figure 1. This indicated that bringing the data to a common time base would increase accuracy. The equation would then be applicable to the chosen time period only, so a long period would be desirable. The equation is similar to the short 85-station equation except for BHW instead of RALKE. The standard errors of the coefficients listed in Table 3 indicate a better station equation.

Table 3. Comparison of Regression Coefficients and their Standard Errors of Estimate

	69-Station Equation		Short 85-Station Equation	
	Coefficient	Standard Error	Coefficient	Standard Error
Constant	0.8306	0.1987	0.4998	0.2304
NPOSI	0.0110	0.001799	0.0111	0.002076
ELEV	-0.00008874	0.00002726	-0.00006876	0.00002498
SLP%	0.007203	0.001011	0.006429	0.001127
DSN	-0.0007726	0.0001389	-0.0006558	0.0001603
DSNW	-0.0005977	0.0001381	-0.0005961	0.0001548
SENW	0.000005645	0.000001763	0.000004838	0.000002061

After the stations had been separated according to season of low flow occurrence, only 16 had lows in summer and fall. No reliable equation with less than six variables could be developed, but for the 59 stations with low flows in winter,

$$R^2 = 0.7108 \quad FPROB = 0.0000$$
$$S.E. = 0.1529 \quad 37\% \text{ of response mean}$$

$$UMLF = 1.3125 + 0.006537\ NPOSI - 0.0001755\ ELEV - 0.0009850\ DSSW - 0.00008697\ BHW - 0.0001731\ BHSW \qquad (9)$$

The standard error of calibration is lower than the mean minus one standard deviation for 59 stations, given in Figure 1, indicating that different equations could pertain to the seasons of occurrence of low flows.

DISCUSSION

Single values of standard error of calibration, standard error of validation or trace of the P matrix should not be used as estimates of reliability of regional equations. A sequence of values is required. In this study of low flows using least squares fitting of physiographic parameters apparently 30 degrees of freedom are required to obtain reliable estimates of accuracy.

Apparently, although not proved conclusively, regionalization or subdivision of the study area, and bringing the low flows to a common time base improve the reliability of estimates.

If regionalization is required, then the province is undergauged, especially for the Coast and Island Region

Potential additions to the network can be screened by Kalman filtering to give an indication of their contribution to the network by the change in P matrix.

APPENDIX A

DEFINITION OF AND STATISTICS FOR PHYSIOGRAPHIC PARAMETERS USED IN STUDY

Parameters	Symbols	Units	Explanation
Drainage area	AREA	Square miles	Total drainage area for basin
Grid coordinates	NPOSI NPOSJ	Dimensionless	Coordinates for centre of gravity of basin
Elevation	ELEV	Feet	Average elevation of basin
Percent slope x 10	SLP%		Basin slope average over squares included in basin
Azimuth	SLPAZ	Degrees	Angle between west-east direction and horizontal projection of line of steepest descent of local slope plane

Parameters	Symbols	Units	Explanation
Distance to sea			
North	DSN	Kilometers	Distance from centre of gravity of basin to sea in north, northwest, west, and southwest directions
Northwest	DSNW		
West	DSW		
Southwest	DSSW		
Relative area			
Lake	RALKE	Dimensionless	Percentage of area of basin occupied by lakes, forests, swamp, glaciers, and built-up areas Note: ΣRA does not always equal 100.
Forest	RAFOR		
Swamp	RASWP		
Glacier	RAGLC		
Urban	RAURB		
Barrier height			
North	BHN	Feet	Difference between average elevation of basin and highest elevation encountered in north, northwest, west and southwest directions until ocean is reached
Northwest	BHNW		
West	BHW		
Southwest	BHSW		
Shield effect			
North	SEN	Feet	Sum of elevation differential of all ascending stretches of terrain encountered when travelling from ocean shore in north, northwest, west, southwest directions to corresponding point
Northwest	SENW		
West	SEW		
Southwest	SESW		
Signed slope			
Northeast	SSNE	Feet/kilometers	Takes into account general configuration of terrain
East	SSE		
Southeast	SSSE		

Note: Further information and references on these parameters may be found in Hydrometric Network Planning Study for Western and Northern Canada Report 5019-70, November 1970 by the Shawinigan Engineering Company Limited, Section 4.2.1, page 33.

Parameter	Mean	Standard Deviation	Maximum	Minimum
AREA	349	198	800	62
NPOSI	157	42	212	55
NPOSJ	47	36	132	11
ELEV	4603	1328	7017	1136
SLP%	61	23	115	5
SLPAZ	183	98	350	14
DSN	1926	345	2395	958
DSNW	2278	1075	3393	226
DSW	581	250	990	64
DSSW	489	252	989	52
RALKE	1.78	2.82	9	0
RAFOR	70.3	22.7	99	11
RASWP	0.29	1.11	9	0
RAGLC	2.98	6.66	48	0

Parameter	Mean	Standard Deviation	Maximum	Minimum
RAURB	0.0	0.0	0	0
BHN	3135	1235	6733	1160
BHNW	2457	868	5886	690
BHW	2195	841	3920	47
BHSW	1715	939	4490	400
SEN	42800	14500	88490	18300
SENW	132000	63100	200000	7910
SEW	35000	18600	71300	2070
SESW	48300	41800	128200	1850
SSNE	1.14	13.2	32	-26
SSE	1.06	24.4	65	-64
SSSE	0.376	12.5	38	-38

APPENDIX B

FREQUENCY OF OCCURRENCE OF PHYSIOGRAPHIC PARAMETERS IN RANDOM SAMPLE EQUATIONS

Parameter	16-Station Equation	32-Station Equation	48-Station Equation	64-Station Equation
NPOSI	2			8
NPOSJ	1	3	17	4
ELEV	4	9	14	13
SLP%	15	24	24	15
SLPAZ	1		1	
DSN	5	3		11
DSNW	7	11	17	15
DSW	6	5	5	8
DSSW	4	4	1	
RALKE	3	7	9	10
RAFOR	1			2
RASWP	1			
RAGLC	3	1		
RAURB				
BHN	4	4	3	
BHNW	2		3	
BHW	8	5	3	4
BHSW	7	6	5	4
SEN	6	4	7	
SENW	10	9	5	7
SEW	2		4	5
SESW	2		1	5
SSNE	2	1	1	
SSE	5	3	2	1
SSSE	5	1		

REFERENCES

Gura, I.A., An Algebraic Approach to Optimal State Estimation, Aerospace Report No. ATR-74(9990)-1, The Aerospace Corporation, El Segundo, California, 1976.

Solomon, S.I., "Parameter Regionalization and Network Design", A Stochastic Approach to Water Resources, H.W. Shen, Fort Collins, Vol. 1, Chapter 12, 1976.

Water Survey of Canada, Low Flows in British Columbia Annual 7-Day Averages, Vols. 1 and 2, Vancouver, 1974.

As the cost of establishing and maintaining hydrometric stations was increasing rapidly and the value of the data, measured by the number of requests, did not seem to increase at a commensurate rate, a systematic examination of the data collection effort was undertaken. Starting in 1970, the Water Survey of Canada in conjunction with the Consultant, Shawinigan Engineering Company, Limited (1970) carried out a deliberate and systematic study of data collection. The objective of this study was to provide guidelines and to give meaningful direction for initiating and expanding a program for the long range development of the hydrometric network.

(Figure 1)

Figure 1 shows the dramatic turn in the size of the network which took place in 1973 after the classification process began.

The study recommendations of particular interest to British Columbia were station classification and the development of a regional network. As stations had previously been located for the design of specific projects some stations had become redundant. With station classification, the purposes of

Revelstoke Region - 16 stations

$$\text{MAF} = 4516 + 16.94 \times \text{AREA} + 139.4 \times \text{RAFOR} - 0.5191 \times 10^5 \times \text{RASWP} - 0.3522 \times \text{SEN} \quad (5) \qquad R^2 = 0.999, \ \text{S.E.E.} = 2120 \text{ cfs}$$

Vancouver Region - 25 stations

$$\text{MAF} = 572.3 + 13.67 \times \text{AREA} + 36.19 \times \text{SLP\%} \quad (6) \qquad R^2 = 0.940, \ \text{S.E.E.} = 2396 \text{ cfs}$$

Windermere Region - 19 stations

$$\text{MAF} = 0.5868 \times 10^5 + 8.486 \times \text{AREA} - 28.16 \times \text{DSN} - 1919 \times \text{RASWP} \quad (7) \qquad R^2 = 0.960, \ \text{S.E.E.} = 3620 \text{ cfs}$$

The regions were named for cities and all are located in the southern portion of the Province. The reader is referred to Leith (1975) for a description of the procedure used and the location of these seven regions. The conclusions of the study were that regionalization by regression appears to be effective as different regions had significantly different equations and in each region the standard error was lower than the standard deviation of the mean annual floods. As regression is a statistical technique, regional equations would have been more satisfying if there had been a physical theory to guide the development of the equations. This need for physical theory or background would probably have been more acute if the hydrologic variable being modeled had been less general than mean annual flood.

However, regionalization by regression does provide through standard error, a means of evaluating the effectiveness of transferring information gathered by the existing network. This then provides an estimate of the effectiveness of the network and a strong indication that the network requires more stations sampling natural flow from basins of under 1500 square kilometers.

Table 4. Definition of Physiographic and Precipitation Parameters

1.	AREA is the total drainage area for the basin.
2-3.	NPOSI and NPOSJ are the coordinates for the centre of gravity of the basin.
4.	ELEV is the average elevation of the basin.
5.	SLP% is the basin slope averaged over the basin.
6-9.	DSN, DSNW, DSW, DSSW are the distances from the centre of gravity of the basin to the sea in the north, northwest, west and southwest directions.
10-13.	RALKE, RAFOR, RASWP, RAGLC are the percentages or relative areas of the basin which are occupied by lakes, forests, swamps and glaciers.
14-17.	BHN, BHNW, BHW, BHSW are barrier heights or differences between average elevation of basin and highest elevation encountered in the north, northwest, west and southwest directions until the ocean is reached.
18-21.	SEN, SENW, SEW, SESW are shield effects defined as the sum of the elevation differential of all ascending stretches of terrain encountered when travelling from the ocean shore at the north, northwest, west and southwest to the corresponding point.
22-24.	SSE, SSNE, SSSE are approximations of regional slope which take into account the general configuration of the terrain in the east, northeast, and southeast directions.
25-26.	MAPRE is the mean annual precipitation averaged over the basin; a second precipitation value is TBPRE, total basin precipitation which equals mean annual basin precipitation multiplied by the area of the drainage basin.

In the second report in a series of regionalization studies in British Columbia, Leith (1976) studied regression of mean annual flow against physiographic parameters for 62 basins throughout British Columbia. Mean annual flow is the mean for calendar year flows over the period of record for a particular station where only complete years of record are considered.