OTHER INVESTIGATIONS

EVALUATION PROCEDURE FOR GEOCHEMICAL DATA
URANIUM RECONNAISSANCE PROGRAM
(82F)

By A. J. Sinclair and W. K. Fletcher
Department of Geological Sciences, University of British Columbia

ABSTRACT

Our preliminary studies indicate that a useful procedure for evaluating regional multi-element geochemical data, such as those obtained during the joint Federal-Provincial Uranium Reconnaissance Program, is a systematic application of readily available standard procedures as follows:

(1)   Sorting of data into provenance groups based on predominant rock type of the provenance region of each sample;

(2)   Evaluation of log probability graphs of each element and pH in each provenance group;

(3)   Qualitative evaluation of a correlation matrix for data in each provenance group;

(4)   Multiple regression studies for each provenance group of data using as successive dependent variables those elements of particular interest.

In addition, moving average maps have proved useful in making regional comparisons of element distribution patterns and general geology.

Additional studies will involve more widespread application of these procedures, the use of multivariate statistical methods, and further subdivision of the data base in terms of field parameters.

INTRODUCTION

We are in the process of conducting an examination of regional stream geochemical data obtained during a joint Federal-Provincial Uranium Reconnaissance Program of NTS areas 82 F and K released in 1978. The purposes of our study are to evaluate the usefulness of individual variables coded in the course of these surveys, to develop a statistical procedure for extracting useful information from the data, and to ultilize the data base as an effective means of defining problems of geological interest that warrant further investigation. Our work is developmental in nature and is confined to data for NTS map-area 82F. Initial work has emphasized a routine approach to both subjective and statistical evaluation of the data. A brief outline of initial procedures and typical results obtained form the basis of this report.

Geochemical data from the Uranium Reconnaissance Program were made available to us on magnetic tapes by the British Columbia Ministry of Energy, Mines and Petroleum Resources. Examination of various published compilations of these data (for example, Uranium Reconnaissance Program, open file data) led to
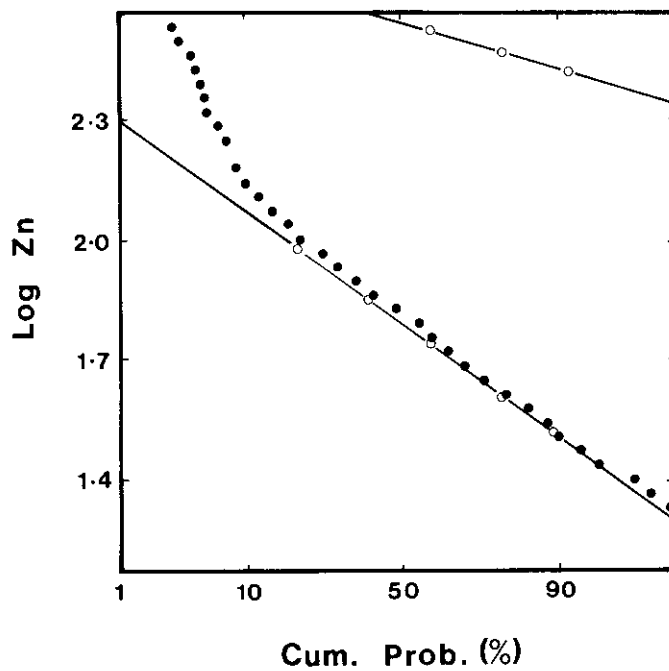
Figure 39. Log probability plot of 515 Zn values for steam sediments from areas underlain by granitic rocks (GRNT-35) in NTS map-area 82F. Black dots are original data, open circles are construction points for partitioning, and sloping straight lines are assumed lognormal populations comprising data set. To a first approximation the lower population can be thought of as background and the upper population can be considered anomalous. The two populations are separated effectively by a threshold of 210 ppm Zn.
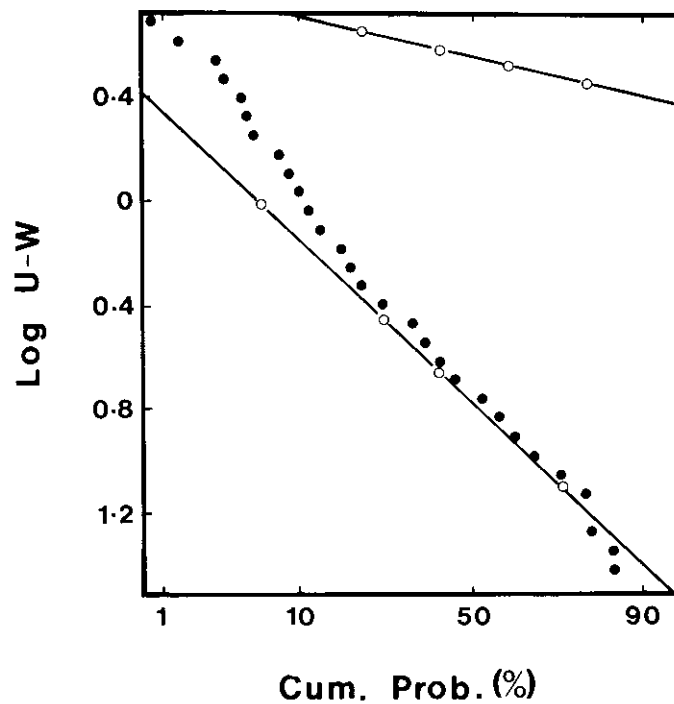

Figure 40. Log probability plot of uranium in water (U-W) for 515 sample sites underlain predominantly by granitic rocks (GRNT-35) in NTS map-area 82F. Black dots are original data, open circles are construction points for partitioning, and sloping lines are assumed lognormal populations comprising the data set. To a first approximation the lower population can be considered background and the upper population can be considered anomalous. A threshold of 2 ppb U separates the two populations.

the selection of data from map-area 82F as the basis for most of our studies. Our preliminary evaluation procedure then developed as follows:

    (1)   Production of machine-contoured moving average maps;

    (2)   Sorting of data on the basis of dominant rock type in the provenance region of each water or stream sediment sample;

    (3)   Interpretation of probability graphs obtained separately for each continuous variable for each provenance group of data;

    (4)   Linear correlation studies within each provenance group of data;

    (5)   Multiple regression analysis for specific dependent variables within each provenance group of data.

Additional multivariate statistical studies will be done in the future.

MOVING AVERAGE MAPS

We have examined moving average maps for all variables for which data are available. A moving average map is generally obtained from weighted averages of all values in a window or cell, assigning this average to the cell centre, and repeating the procedure for a network of cells over the data field. Purpose is two-fold: (1) to smooth some of the local variability in the data, and (2) to produce a regular grid of smoothed data that is amenable to straightforward contouring procedures. In constructing such maps three parameters can be varied: averaging procedure (weighting), the size and shape of the window in which data are averaged, and the amount of overlap of adjoining windows. There are no set rules in defining these variables but an important control is the mean spatial sample density and the uniformity of geographic distribution of sample sites. Uranium Reconnaissance Program samples have, in general, been collected to a uniform sample density of close to 1 per 12.5 square kilometres; consequently, 12.5 square kilometres is a minimum window size that would produce little in the way of smoothed data. On the other hand, with too large a window extreme smoothing is obtained with substantial loss of information. After several trials we settled on a 7-kilometre by 7-kilometre cell with 50 per cent overlap. Moving average maps were prepared for all elements. Examination of these maps (Fig. 38) led to the following general conclusions for map-area 82F (west half):

    (1)   Ni and Co highs correlate generally with volcanic terranes such as the Rossland Formation and to a lesser extent fine-grained clastic sequences such as the Slocan Group;

    (2)   U and Mo highs are mostly confined to local areas underlain by plutonic rocks of Mesozoic and Tertiary ages;

    (3)   Zn, Ag, and Pb have comparable distribution patterns with highs that correlate closely with known mineralized areas such as Slocan camp, Salmo area, Rossland camp, etc.;

    (4)   Cu highs correlate closely with the Rossland Formation and to a lesser extent with Slocan Group;

    (5)   Mn highs clearly are mostly related to known areas of Zn-Ag-Pb mineralization;

    (6)   High concentrations of U in water and F in water are similar and occur sporadically in acid to intermediate plutonic rocks, especially the western and southwestern parts of the Nelson batholith;

    (7)   Large areas of high Fe values correspond to terrane underlain by volcanic rocks such as the Rossland Formation.

Of course, there are local departures from these generalizations that warrant more detailed study. Compared to the original open file NTS map-area, the rolling mean maps permit rapid assessment of regional

TABLE 1. MEANS AND STANDARDS

Variable	$\bar{x}$	s
Zn	92.33	118.9
Cu	14.34	19.89
Pb	24.51	44.05
Ni	13.75	18.40
Co	6.20	3.41
Ag	0.21	0.501
Mn	439.1	254.6
Fe	1.62	0.538
Mo	1.59	1.32
U	23.8	40.3
U—W	0.47	0.921
F—W	92.6	12.18
pH	7.31	0.409

TABLE 2. MEANS AND STANDARD DEVIATIONS FOR TWO MAIN pH GROUPS LOG TRANSFORMED DATA, GRNT—35, 82F

Variable	Original Units*	Slightly Alkaline (n = 405)		Slightly Acidic (n = 101)	
		$\bar{x}$	s	$\bar{x}$	s
Zn	ppm	1.873	0.272	1.709	0.252
Cu	ppm	1.042	0.312	0.937	0.308
Pb	ppm	1.189	0.396	1.053	0.349
Ni	ppm	0.954	0.419	0.792	0.458
Co	ppm	0.742	0.229	0.686	0.274
Ag	ppm	−0.837	0.312	—0.944	0.168
Mn	ppm	2.605	0.201	2.508	0.245
Fe	per cent	0.194	0.150	0.141	0.175
Mo	ppm	0.116	0.210	0.189	0.235
U	ppm	1.147	1.417	1.094	0.433
U—W	ppb	−0.704	0.567	−0.841	0.609
F—W	ppm	1.791	0.402	1.561	0.456
pH	pH	7.44	0.28	6.72	0.18

*All variables except pH have been $\log_{10}$ transformed.

patterns and interelement comparisons. Loss of detail through smoothing is a disadvantage that can be offset by the use of residual maps that we have not yet produced.

SIMPLE STATISTICS

For all Uranium Reconnaissance Program geochemical samples the predominant rock type in the provenance areas of samples has been coded. Data were first sorted on the basis of these provenance groups and in subsequent statistical evaluation these groups were treated separately. The basic assumption here is that for most samples the abundant rock type in the provenance area will exert a dominant control on the chemical character of derivative stream sediments and that different chemical characteristics will result from different chemical variables for provenance group GRNT-35 for map-area NTS 82F.

Means and standard deviations of continuous variables for data set GRNT-35 are listed in Table 1. For the most part standard deviations are as large or larger than their corresponding means indicating highly skewed distributions. Such distributions commonly arise from a lognormal form or from a combination of populations. Consequently, the forms of density distributions must be examined individually for each element. The methodology we have adopted is a subjective approach involving use of probability plots (Sinclair, 1976).

PROBABILITY PLOTS

Three examples illustrate the types of graphs that are common and some of the apparently straightforward interpretations that result.

The probability graph for Zn (Fig. 39) is typical of those obtained for many elements. A centrally positioned sinuous cumulative curve has been partitioned into two lognormal populations that are distinguished effectively with a threshold of 210 ppm Zn (Sinclair, 1974). The large, lower population can be interpreted as normal background and the smaller, upper population can be considered anomalous.

Figure 40 is a somewhat comparable probability plot for uranium in water, also for areas underlain predominantly by rock type GRNT-35, and it shows two populations separated by a threshold of 2 ppb uranium. In both these examples possible explanations of the high or anomalous populations include:

(1) Mineralization in the sample provenance area;
(2) Incorrect designation of principal rock type (that is, some high values for GRNT-35 might be low for another provenance group);
(3) Peculiar local chemical conditions favouring a higher than normal dissolution rate for the metal(s) in question;
(4) Contamination by man in the form of industrial waste, sewage, fertilizers;
(5) Biased sampling of stream sediments (for example, an over-abundances of heavy minerals in the sample selected).

Despite the various conditions that might give rise to high values, the methodology leads to recognition of a significant threshold above which values require an explanation and therefore demand further investigation. A useful procedure is to colour-code the relatively small group (about 25 sample sites in the above two cases) of high values on a map and interpret them individually and collectively in the light of known geology.

135

TABLE 3. CORRELATION MATRIX FOR ORIGINAL DATA, GRNT–35, 82F

	Zn	Cu	Pb	Ni	Co	Ag	Mn	Fe	Mo	U	U–W	F–W
Cu	0.220											
Pb	0.534	0.173										
Ni	0.104	0.182	0.082									
Co	0.198	0.506	0.090	0.675								
Ag	0.366	0.207	0.615	0.020	0.073							
Mn	0.368	0.219	0.264	0.084	0.339	0.321						
Fe	0.274	0.321	0.164	0.331	0.683	0.196	0.510					
Mo	0.185	0.106	0.124	0.044	0.140	0.069	0.206	0.162				
U	0.084	-0.030	0.068	-0.025	-0.134	0.044	0.126	-0.111	0.259			
U–W	-0.011	-0.071	-0.037	-0.083	-0.215	-0.029	-0.048	-0.223	0.033	0.509		
F–W	0.089	0.004	0.002	-0.008	-0.135	0.021	-0.002	-0.134	0.116	0.246	0.453	
pH	0.220	0.135	0.081	0.081	0.101	0.108	0.196	0.221	-0.068	0.076	0.212	0.269

A third example is shown to illustrate a somewhat more complex situation of data interpretation. In the case of pH (Fig. 41) three populations seem to be present. The high pH population represents only 1.4 per cent of the data (about seven values greater than pH 8.1). The other two abundant populations overlap somewhat but can be separated reasonably well at a threshold of 7.0. We can therefore designate these two abundant populations as weakly acid and weakly alkaline. Because acid and alkaline conditions represent fundamentally different controls for some metals we subdivided data for GRNT-35 into pH subgroups using the thresholds determined from our probability graph analysis, and examined metal distributions within each of the two major pH groups (weakly acidic and weakly alkaline). Means and standard deviations of all variables in each group are given in Table 2. A systematic difference is apparent for most means; stream sediments and waters from the weakly alkaline group contain higher metal values in most cases than do those from the weakly acidic group. The most obvious exception is molybdenum which shows the opposite relationship. These results are as might be expected and indicate that different background element abundances are to be expected even within a single provenance group of data, depending on the pH category of individual samples.

A further evaluation of these data involved examination of probability graphs for each metal in each of the two principal pH groups. These plots indicated that only rare anomalous samples occur within the weakly acidic pH population, or, in other words, nearly all the anomalous samples recognized in evaluating the total data set for provenance group GRNT-35 are for samples from a weakly alkaline environment.

These three examples illustrate clearly the practical advantages of systematic utilization of probability graphs for interpretations of geochemical data such as those obtained from the Uranium Reconnaissance Program. Of course, the intercorrelations among variables are obscured, considered only in a cursory or subjective manner, or are ignored by such methods. Furthermore, this type of analysis does not lead readily to the recognition of anomalous values that occur below the selected thresholds. More sophisticated analysis is required to achieve these ends.

Our probability plots differ from those on maps published by the Geological Survey of Canada because we divided data into provenance groups whereas they did not. It is also important to realize that probability graph analysis is not possible for some provenance groups because of too few data. At least 50 data values, and preferably 100 or more, are desirable for probability graph analysis.

CORRELATION

A correlation matrix was obtained for chemical data for sediments and water and is reproduced in Table 3. All significant correlations in this case are positive.

Molybdenum and pH are either uncorrelated with, or weakly correlated with other elements. Strong correlations, defined arbitrarily as those with a correlation coefficient of 0.5 or greater can be used to define four groups of inter-related elements as follows:

(1)	Pb-Zn-Ag
(2)	Co-Cu-Ni
(3)	Co-Fe-Mn
(4)	$U-U_w$

These are expected element groupings from the point of view of common geochemical associations.

Stream width and stream depth have been omitted from the correlation table because they have such low correlation coefficients with the elements represented. Both of these variables have very weak negative
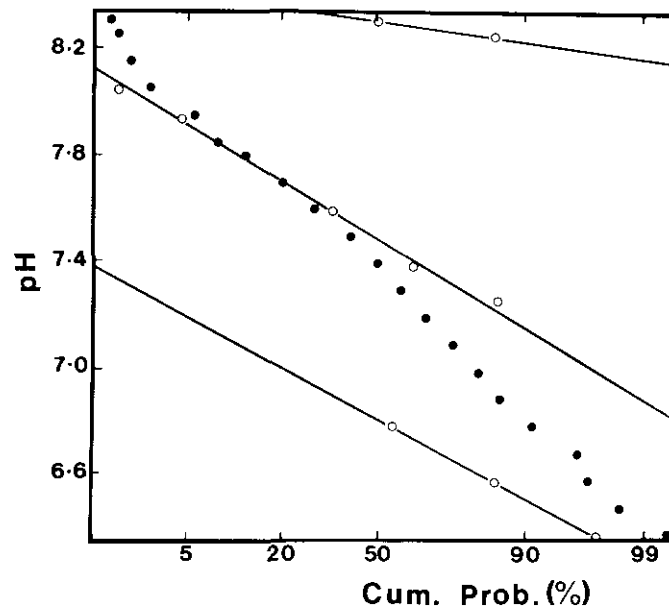
Figure 41. Probability plot of pH values for 515 stream sediment sites from drainage basins underlain by GRNT-35 (Mesozoic granitic rocks) in NTS map-area 82F. Filled circles are original data, open circles are construction points for partitioning, and the three sloping straight lines represent three interpreted normal populations of pH values.
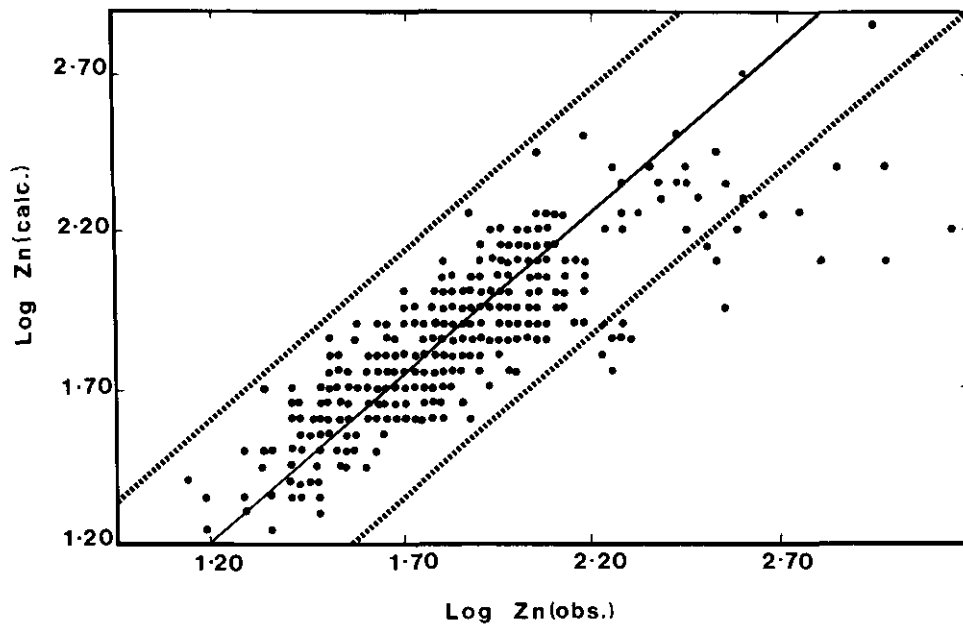

Figure 42. Plot of calculated log (Zn), using multiple regression equation of Table 4 versus observed value of log (Zn). Traced from computer print-out, thus abscissa and ordinate are slightly different scales. Plotted values scatter about the central line representing the 'expected relationship.' Fringing lines contain 95 per cent of the values. Points outside the 95-per-cent confidence zone are arbitrarily taken to be anomalous.

correlations with U-W and F-W and pH (maximum r = 0.196) but correlate strongly with each other (r = 0.461).

False correlation is a common problem in treating geostatistical data, commonly arising from a few outlying values that are not representative of the bulk of data (for example, Chapman, 1976). To guard against such a likelihood scatter diagrams are useful.

MULTIPLE REGRESSION

Multiple regression is used to find a systematic relationship between a single (dependent) variable and a combination of other (independent) variables. The method has been applied with apparent success to a wide variety of types of multivariate geochemical data (for example, Spilsbury and Fletcher, 1974). A procedure that we have found particularly useful is backwards stepwise regression (Le and Tenisci, 1978) in which a multivariate relationship is established as above, and individual variables are dropped successively if their coefficients are not significantly different from zero as some preset level of significance (we used 0.05). The method does not include redundant variables, that is, those that are linear combinations of other variables. All variables except pH have been log transformed for this analysis. Results for several dependent variables are given in Table 4. By far the best statistical relationship of all those listed is that for Co as the dependent variable. If data for other elements are used to calculate the expected Co content and a log (Co) value of 0.779 (that is, about 6 ppm Co) is obtained, the true value has a 68-per-cent chance of lying in the range 0.681 to 0.877 (that is, from 4.8 to 7.5 ppm Co).

The relationship for log (Zn) as a dependent variable will serve to illustrate the practical applications of the multiple regression approach to data analysis. For each sample the raw data can be used in the log (Zn) equation to calculate a value of the independent variable (Y_c) which can be compared with the observed value (Y_o). The difference between a calculated and observed values for a single sample represents a residual (d):

$$d = Y_o - Y_c$$

If one can assume that the calculated values represent an average background relationship, then high positive residuals are anomalous in Zn. Similarly, high negative residuals indicate abnormally high amounts of independent variables with positive coefficients and/or abnormally low amounts of independent variables with negative coefficients. Consequently, all samples that depart markedly from the general multivariate background model are unusual and warrant detailed study. Figure 42 is a plot of calculated versus observed log (Zn) values based on the relationship of Table 4. Ideally the observed and calculated values would plot on the central line; in fact, they scatter such that about 95 per cent of samples should plot in the field between the two bounding curves. Points outside this 95-per-cent confidence field represent anomalous samples based on an arbitrary definition. Although we have not done so here, it is convenient in practice to show on diagrams such as Figure 42 the thresholds determined for univariate data from probability graphs for example.

CONCLUSIONS

Our evaluation of geochemical data from the Uranium Reconnaissance Program in British Columbia is in its early stages but already a useful systematic procedure for analysing the data has emerged. Future work will involve applications of multivariate statistical procedures to the data, an evaluation of the usefulness of individual discrete variables, and an evaluation of problems arising from the study and warranting further investigation.

TABLE 4. EXAMPLES OF MULTIPLE REGRESSION
GRNT—35, 82F

Dependent Variable: log (Zn)
R^2 = 0.684
Standard Error = 0.1598
log (Zn) = 0.3040 + 0.0414 pH + 0.1007 log (F—W) + 0.2334 log (Fe)
              + 0.2629 log (Mn) + 0.1428 log (Ag) + 0.3122 log (Pb)
              + 0.0939 log (Cu)
Rejected variables are: Ni, Mo, U, U—W, Co

Dependent Variable: log (Cu)
R^2 = 0.593
Standard Error = 0.2020
log (Cu) = 0.0020 + 0.0651 log (F—W) — 0.0807 log (U) + 0.0927 log (Mo)
              — 0.3816 log (Fe) + 0.0770 log (Ag) + 0.8098 log (Co)
              + 0.1252 log (Ni) + 0.2245 log (Zn)
Rejected variables are: Pb, Mn, U—W, pH

Dependent Variable: log (Co)
R^2 = 0.834
Standard Error = 0.0983
log (Co) = 0.2078 — 0.0708 log (F—W) — 0.0274 log (U) + 0.6707 log (Fe)
              + 0.0983 log (Mn) + 0.2080 log (Ni) — 0.0718 log (Pb)
              + 0.1886 log (Cu)
Rejected variables are: Zn, Ag, Mo, U—W, pH

Dependent Variable: log (Mn)
R^2 = 0.577
Standard Error = 0.1411
515 observations
log (Mn) = 1.3557 + 0.0541 pH — 0.0483 log (U—W) + 0.1229 log (U)
              + 0.4004 log (Fe) + 0.2688 log (Co) — 0.1125 log (Ni)
              + 0.0649 log (Pb) + 0.2280 log (Zn)
Rejected variables are: Ag, Mo, F—W

Dependent Variable: log (U)
R^2 = 0.556
Standard Error = 0.2840
515 observations
log (U) = 1.1140 — 0.1277 pH + 0.0767 log (F—W) + 0.4331 log (U—W)
              + 0.3733 log (Mo) + 0.5089 log (Mn) — 0.3404 log (Co)
              + 0.1171 log (Ni) + 0.1169 log (Pb) — 0.2232 log (Cu)
Rejected variables are: Ag, Zn, Fe

REFERENCES

Chapman, R. P. (1976): Limitations of Correlation and Regression Analysis in Geochemical Exploration, *Institute Min. Metal.,* Vol. 85, pp. B279-B283.

Department of Energy, Mines and Resources, Canada and *B.C. Ministry of Energy, Mines & Pet. Res.,* Regional Stream Sediment and Water Geochemical Reconnaissance Data, Southeastern British Columbia, NTS 82F, Open File 514 (NGR 25-1977).

Le, C. and Tenisci, T. (1978): UBC TRP Triangular Regression Package, Computing Centre, *University of British Columbia,* 197 pp.

Sinclair, A. J. (1974): Selection of Threshold Values in Geochemical Data using Probability Graphs, *Jour. Geochem. Expl.,* Vol. 3, pp. 129-149.

.... (1976): Application of Probability Graphs in Mineral Exploration, *Assoc. Expl. Geochemists,* Spec. Vol. 4, 95 pp.

Spilsbury, W. and Fletcher, W. K. (1974): Application of Regression Analysis to Interpretation of Geochemical Data from Lake Sediments in Central British Columbia, *Can. Jour. Earth Sci.,* Vol. 11, pp. 345-348.
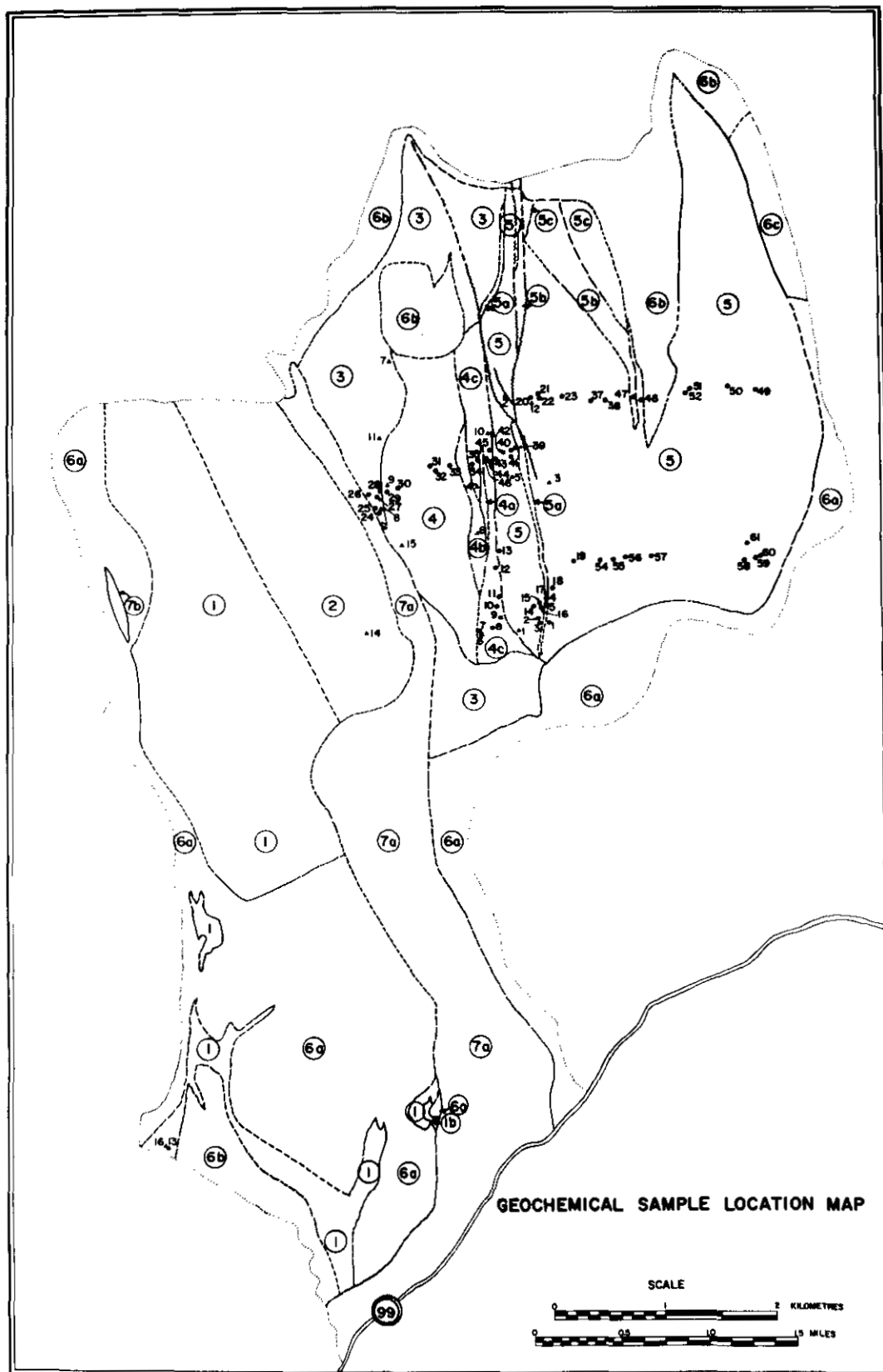
GEOCHEMICAL SAMPLE LOCATION MAP

SCALE

Figure 43.   General location of geochemical samples. Circled numbers represent geological units of Miller and Sinclair (1978); filled triangles are 'whole rock' sample sites; and filled circles are 'minor metal' or lithogeochemistry sample sites.

142