# MINERAL RESOURCE POTENTIAL: ROUYN-NORANDA REGION, QUEBEC

by

Pervez Akhtar Umar

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Geological Sciences
McGill University
Montreal, Quebec

March 1978

Two roads diverged in a wood, and I -
I took the one less traveled by,
And that has made all the difference.

Robert Frost
1874 - 1963

MINERAL RESOURCE POTENTIAL:

ROUYN-NORANDA REGION, QUEBEC

by

Pervez A. Umar

## ABSTRACT

The prediction of unknown regional mineral potential is an integral part of the exploration process. Quantitative techniques enable the exploration planner to more completely analyze alternative investment opportunities. Such decisions are critical for mining regions like Rouyn-Noranda, Quebec, which are characterized by a highly developed mineral-based economy, the exhaustion of known endowment, and declining discovery rates.

In this study, the multivariate statistical techniques of regression analysis and discriminant analysis are applied in making estimates of undiscovered base metal endowment in the Rouyn-Noranda region. Factor analysis has been applied in data reduction and variable selection.

The study demonstrates the strengths and the weaknesses of the statistical techniques used, and the problems associated with their application to the analysis of geological data. Solutions to these problems are suggested.

The undiscovered base metal endowment in the region, based on geological relationships alone, is estimated at a minimum of 131 million dollars. This is not significant under current economics of mine development. It is therefore suggested that additional inputs in terms of newer concepts will be required to realize a greater measure of the theoretically possible endowment in the region.

The results obtained should provide guidelines for geological research, and for further mineral exploration, not only in the Rouyn-Noranda region, but also in other mining regions requiring resource potential evaluation.

POTENTIEL DES RESSOURCES MINERALES

REGION DE ROUYN-NORANDA, QUEBEC

par

Pervez A. Umar

## RÉSUMÉ

La prédiction du potentiel minéral inconnu d'une région est une partie intégrale du processus d'exploration minière. La conception d'un programme d'exploration devrait se baser sur des techniques quantitatives qui permettent de mieux analyser différentes options d'investissement. De telles décisions sont très importantes pour des régions minières telle Rouyn-Noranda, au Québec, se caractérisant par une économie forte basée sur l'industrie minérale, le quasi-épuisement des ressources minérales connues et un déclin important dans le taux de découvertes de gisements économiques.

Dans cette étude, les techniques de régression et d'analyse discriminante à variables multiples sont utilisées pour prédire les quantités non découvertes de métaux de base dans la région de Rouyn-Noranda. L'analyse des composantes principales est utilisée pour réduire les données de base et choisir les variables les plus importantes.

L'étude démontre les avantages et les désavantages des techniques statistiques utilisées ainsi que les problèmes découlant de leur application à des données géologiques. Certaines polutions pouvant résoudre ces problèmes sont alors

suggérées.

Basé uniquement sur des relations géologiques, le potentiel minéral en métaux de base non découvert à date est estimé à au moins 131 million de dollars. Ceci n'est pas très important par rappórt aux conditions présentes de l'économie minière. Il est donc suggéré d'incorporer plus d'information dans le modèle, en terme de nouveaux concepts métallogéniques et de détails géochimiques et géophysiques pour réalizer une meilleure mesure du potentiel minéral possible de la région.

Les résultats obtenus pourrout guider des programmes futurs de recherche géologique et d'exploration minière, non seulement dans la région de Rouyn-Noranda, mais aussi dans d'autres régions minières nécessitant une évaluation de potentiel minéral.

# ACKNOWLEDGEMENTS

This study has been carried out in the Departments of Geological Sciences, and Mining and Metallurgical Engineering, under the joint supervision of Dr. J.S. Stevenson, Dawson Professor of Geology, and Dr. B.W. Mackenzie, Professor of Mining Engineering and Director, Mining Programme, McGill University, Montreal.

I acknowledge the guidance, interest and support of Professors Stevenson and Mackenzie throughout the study. They have given freely of their time, as often as required, for advice and discussions on all aspects of the study, and have critically supervised both the research and the preparation of the thesis manuscript. I wish to express my deepest gratitude to them.

Professor Mackenzie supported the cost of research and provided ample computer time. His material help enabled the completion of the study and the thesis. I acknowledge this support most gratefully.

Dr. J. Dugas, Director, Mineral Deposits Service, Quebec Department of Natural Resources, Quebec, provided all geological maps used in the study. He also supplied mine production and reserves compilation data and other related geological information for the Rouyn-Noranda region. I benefitted from several discussions with him on the geological problems of the region. I acknowledge this with gratitude.

My thanks to Dr. George P.H. Styan, Associate Professor in

Mathematics, for many discussions on the application problems of multivariate statistical techniques. He also allowed free use of his personal library.

My thanks to Dr. J. Zwartendyk, Head, Resources and Development Section, Mineral Policy Sector, Department of Energy, Mines and Resources, Ottawa, for discussions on resource classification schemes, and for providing me up to date literature on resource classification and modelling.

I am most grateful to my colleague, Mr. Michel Bilodeau, now Lecturer in the Department of Mining and Metallurgical Engineering, for help in solving computer programming problems. Mr. Bilodeau also translated into French the abstract of this thesis.

I had several discussions with Dr. F.P. Agterberg, Head, Geomathematics Section, Geological Survey of Canada, Ottawa, on various aspects of the study. I acknowledge this with gratitude.

The encouragement, support and patience of my wife Kishwer, and the good humour, faith and optimism of our daughter, Aisha, made it all both pleasant and worthwhile. It is not possible to thank them enough.

I gratefully acknowledge a scholarship received from the University of Engineering and Technology, Lahore, Pakistan, a Carl Reinhardt scholarship from the Department of Geological Sciences, a B.J. Harrington fellowship from the Department of Mining and Metallurgical Engineering, and a summer research fellowship from the Faculty of Graduate Studies and Research, McGill University, Montreal, Quebec.

# TABLE OF CONTENTS

## LIST OF TABLES

# LIST OF FIGURES

# CHAPTER I

## INTRODUCTION

### 1.1 General Statement

Minerals are non-renewable resources. The occurrence of economic mineral deposits results from a complex interaction of favourable geological factors. Such deposits are developed to production on the basis of economic and technological consideration to become assets and positive contributors to the national economy. The demand for mineral products is derived from the existing socio-economic and technological environment, creating a need for mineral exploration. However, mineral exploration itself is based fundamentally on geology.

In any new area, existing geological information is at best of a reconnaissance nature. However, once an economic discovery is made, the area is subjected to detailed geological studies and a variety of interpretations on ore genesis. The information base keeps on improving as ore reserves become depleted. When discoveries are no longer readily forthcoming, it becomes increasingly necessary to evaluate the area in terms of its potential for further mineral resources so that the justification for exploration effort and investment can be analyzed. This need will be felt at one time or another in any mining area that is on the decline both in terms of mineral production and economic discoveries. One such mining area is the Rouyn-Noranda region, an area comprised by Duprat, Dufresnoy, Beauchastel and Rouyn townships in north-west Quebec.[1] This area, to be referred to as the Rouyn-Noranda region has been selected in this study for evaluation of copper and zinc resource potential.

### 1.2 Resource Potential Forecasting and Exploration Planning

Mineral resource potential evaluation and exploration planning are closely related; the former is an attempt to forecast a future condition, and the latter is an attempt to control and utilize the forecast condition. Certainly, an organization[2] that can forecast and react to the future in an optimally effective and timely manner has a competitive edge. This is all the more important in the business of mining with a heavy long-term investment commitment and an ever increasing time lag between discovery and production.

Mining company planning as described by Mackenzie (1969) is based on the objectives of profit, survival and growth. Government directed organizations have to pursue similar objectives if the operations are to be efficient

---

[1]See Figure 1.

[2]Whether corporate or government.

and competitive. The profit objective is largely a function of cost-price relationships within a given technological, sociological, political and geographical framework. However, in order to survive and grow, the mining organization must have a well planned exploration programme so that as reserves deplete, they are replenished with newly discovered ore meeting its profit criterion. If the organization does not find new ore, its reserves will decline and deplete, and it will no longer be able to survive in the business of mining; the question of growth, therefore, will not arise. Mackenzie (1973) gives a comprehensive discussion of different exploration strategies in line with the above objectives.

Forecasting is a most essential element of exploration planning for both government and corporate organizations. It can be the subjective judgment or intuition of the planner, or a more rigorous estimate based on the quantified relationships between mineral endowment as known and associated geology. A quantitative technique will enable the organization to better analyze its alternative investment opportunities. However, in the final analysis, it is the subjective judgment based on quantitative analyses that will probably prevail.

Mineral exploration is essentially an investment in information gathering, the objective being the maximization of profits through a planned replacement of depleting reserves. It includes all activities that convert the known and unknown category of resources[1] into the category of "ore reserves". It also includes efforts that narrow the focus of search to areas of more favourable ore potential. The present study is designed towards this particular objective.

Planning as defined by Elliot-Jones (1973) is the process of determining a desirable future condition, and deciding how to proceed from the existing to the desired state. In exploration planning, the existing state is often that of dwindling reserves, as is the situation in the Rouyn-Noranda region. The future desired state is an economic discovery. The decision to proceed from the existing deteriorating state of reserves to the desired state is based on a planned sequencing of the exploration process. And since at each stage of exploration planning, decision making is likely to result in the development of more information, the forecasting model can be continuously improved and better defined as to endowment-geology relationships. For a given geological environment in a specified region, the forecasting model can become an integral part of the planning process. The information flows to and from the forecasting model and the actual exploration activity, with the decision maker between the two. This can continue in a given region until a stage is reached where the marginal level of information for a given outlay does not justify additional investment in exploration.

---

[1] See Section 4.6.

investment. The region can then be considered as approaching exhaustion under the existing state of information, economics and technology. Such a decision can be most optimally based on a forecasting model such as the one developed in this study.

An optimal mineral exploration plan should maximize the present value of an economic discovery for a given expenditure made. This requires that the expenditure on exploration should be as little and as late as possible within the existing policies and priorities of the organization (Herfindahl and Kneese, 1974). Such a strategy is possible through a sequential process in which, at each stage, smaller and smaller sub-areas are selected for development of more detailed information for exploration decision making. This strategy can be best deployed through a quantitative modeling of endowment-geology relationships at various levels of information and detail.

## 1.3 Need for the Study

The cost of exploration per unit of supply has been increasing over the past 25 years. Martin et al (1976) have estimated that to meet the estimated discovery requirements of mineral deposits in Canada, an average of 332 million dollars[1] per year will have to be spent on exploration

---

[1]All monetary values used in this study are in constant 1975 dollars.

activities between the years 1971 and 1995. This estimate is more than three times the historical annual average of 98 million dollars during the period 1951 to 1970.

Inevitably, as exploration proceeds in an area, the returns in the form of economic discoveries diminish after the more obvious larger, higher grade and near surface deposits have been discovered. This is reflected in a deteriorating relationship between costs and returns.

New prospective mining areas in Canada are both remote and climatically inhospitable. Exploration costs will therefore be high, and if an economic discovery is made, mining operations are typically capital intensive. The need for infrastructure in a new remote area will increase the lead time to bring on-stream new production capacity, thus requiring long-term financial arrangements. The current uncertain and unpredictable socio-economic and political environment has made investment capital scarce. Furthermore, government incentives to the mining industry have declined in recent years.

It is imperative therefore, that the mineral wealth in existing mining areas be fully exploited within the limits of economic justification. To do so requires estimates of the resource potential of existing mining areas to form the basis for exploration planning. The role of this study is to support this type of planning in an important area, the Rouyn-Noranda region.

## 1.4 Statement of the Problem

Given a mining area with well known geological information and a history of mineral development and production activity, the problem is to estimate quantitative relationships existing between the known mineral endowment and the associated geological factors, to assess the statistical reliability of the estimated relationships, and to use these relationships in making estimates of undiscovered resource potential in the area.

The area under study is the 400 square-mile Rouyn-Noranda region comprising the townships Duprat, Dufresnoy, Beauchastel, and Rouyn. The resource potential estimates are based on the known copper-zinc deposits in the region and their associated geological characteristics.

# CHAPTER 2

## PREVIOUS WORK RELATED TO STUDY

### 2.1 General Statement

Nearly all studies having relevance to the present work were designed for application over areas much larger than the one selected for the present study. The areas studied by Agterberg, et al. (1972), Harris (1965), Azis, et al. (1972), DeGeoffroy and Wu (1970), and Allais (1957), were respectively 80, 240, 400, 850, and 960 times the size of Rouyn-Noranda region under investigation. Each of these studies had its own objectives, terms of reference, and approach. It should, however, be mentioned here that when relatively small areas comprising a mining region, or center of mineralization of a certain age and type, are considered for resource potential evaluation, there is the advantage of greater uniformity of geological detail, terminology, and reliability. Geological relationship, therefore, can be better studied quantitatively, and relationships better defined. Small-sized areas are, of course, subject to the statistical disadvantages inherent in a small-sample size.

## 2.2 Resource Potential Models

A number of mathematical models have been developed for evaluating regional mineral resource potential at different information levels. All these models in some way attempt to relate a specific known endowment to its quantified geographical or geological environment[1] within an existing, extrapolated, or assumed framework of economics and technology. In broad terms, the models can be categorized as following:

(1) The spatial model;

(2) The geochemical model;

(3) The multivariate analysis model;

(4) The subjective probability model.

These and other less important models not falling in the above categories are reviewed below.[2]

### 2.2.1 The Spatial Model

The spatial analysis model assumes that mineral endowment is a function of area only, being equally distributed in unit areas within a given geographical or geological region.

---

[1]Geochemistry and geophysics are here assumed to be part of the geological environment.

[2]The review of spatial models has been drawn in part after Mackenzie (1971).

The model is therefore appropriate in information deficient situations where, on a broader reconnaissance level, it is desired to have an estimate of the overall regional endowment potential rather than to provide small potentially favourable exploration targets within it.

Spatial analysis is based essentially on the extrapolation of estimated distribution characteristics of known endowment in selected well-explored reference areas to the study area under evaluation. However, for a meaningful analysis, the reference and the study areas should have a similar geological environment, or otherwise be large enough to incorporate a number of geological environments so that there is a high probability of some of them being common in both of them.

Allais (1957) made the first study in which the concept of probability theory was applied in estimating the economic potential of mineral exploration. His focus was the Algerian Sahara, an area of 386,000 square miles. Allais compiled statistics such as the number of mining districts, and their gross production values for the world's well explored mining regions, such as France, North Africa, and the western United States. He found that the number of mining districts per unit area showed a close fit with the Poisson distribution. The probability of exactly X occurrences in the Poisson distribution is

$$f(X) = \frac{\mu^X e^{-\mu}}{X!} \qquad \text{for } X = 0,1,2,\ldots$$

Where $\mu$ is the mean number of occurrences per unit area or cell, and $e = 2.71828$. If the parameter $\mu$ is known, the whole distribution can be written. Thus, by knowing the number of mining districts, mineral deposits, and their value per unit area in the well-explored reference cells, the probability of 0,1,2,...n deposits occurring per unit area in the study area could be estimated.

Allais estimated that for the 386,000 square miles of the Algerian Sahara, about twenty exploitable deposits could be discovered as a result of exploration. He predicted the net gain of the exploration effort at 50 billion francs with a 0.35 probability of realization, and a 0.65 probability of losing twenty billion francs. Perhaps more importantly, he pointed out that the success or failure of the venture depended on whether or not the few large deposits expected in the Sahara are discovered.

Allais assumed conditions of ignorance for the study area and also ignorance on a macro scale in the choice and application of the reference areas. He did not therefore attempt to make any distinction between areas of higher or lower mineral potential. He noted, however, that the range of his estimates could be narrowed by geological and other related information inputs.

DeGeoffroy and Wu (1970) made a similar study over

340,000 square miles of greenstone belts in the Canadian Shield. They used a negative binomial distribution to represent the spatial distribution of deposits in the Shield. They predicted a total of between 1,375 and 1,581 ore deposits worth a total estimated value of between \$155 and \$432 billion.[1] Like Allais, DeGeoffroy and Wu did not make distinctions between areas of high and low potential. However, their results have a greater relevance to the Shield area than had Allais' results to the Algerian Sahara for the reason that a similar geological environment is available both in the reference and the study areas.

Derry (1973) made estimates of the potential endowment in the Canadian North using spatial analysis. However, by using a combination of age and type of sediments and volcanics both in the reference and the study areas, he was able to define his predictions in terms of specific metals.

Spatial analysis is useful in evaluating the benefits of exploration over large unexplored areas. However, it cannot provide guidelines for mineral exploration within such areas. Under conditions of ignorance, spatial analysis does provide a basis for better exploration investment decisions in the light of the information developed.

---

[1] \$87-\$243 billion in 1968.

### 2.2.2 The Geochemical Model

The geochemical model represents an attempt to estimate the endowment of a metal as a function of its average concentration and distribution in the earth's crust. The model does not define target areas for exploration within the studied geochemical environment. However, it has the advantage of being independent of geological data with its inherent problems,[1] and of economic data which are subject to changes with time.

McKelvey (1960) observed that the contained metal content of mineable reserves of many elements in the United States are equal to their estimated crustal abundance[2] multiplied by a factor of between $10^9$ and $10^{10}$. A similar linear relationship was noted by Sekine (1963) for the metal reserves in Japan, the multiplying factor in this case being between $10^8$ and $10^9$. These relationships are too close to be fortuitous. Thus they offer a possibility of making resource estimates on a general level.

Brinck (1967, 1972) is the chief advocate of the geochemical model in resource evaluation. By defining the

---

[1]Incompleteness, lack of uniformity, changing interpretations.

[2]Expressed as a percentage.

relationship between the distribution properties[1] of metal concentration in geochemical sized samples of rock, and the distribution properties of the metal[2] in known deposits, Brinck develops a measure which he calls the "specific mineralizability"[3] of that metal for the given environment. Mathematically,

$$\alpha = \frac{\sigma^2}{3\ln D/d}$$

in which,

$\alpha$ = Specific mineralizability

$\sigma$ = Logarithmic standard deviation of initial sample[4]

d = Linear equivalent[5] of volume of sample

---

[1]Brinck assumes that the tonnages and grades of ore in deposits, and the concentrations of metals in geochemical sized samples are lognormally distributed. He therefore uses their median values and logarithmic standard deviations in his calculations.

[2]In terms of both grade and tonnage.

[3]Specific mineralizability, also called the "absolute dispersion" is the tendency of a metal to occur in the form of an ore deposit. This measure is expressed as a percentage.

[4]The sample here means the reference ore deposits.

[5]The term linear equivalent is a function that depends on both the volume and shape of a body. Matheron (1971) has calculated a set of curves from which the linear equivalent can be directly determined for any given environment given the lengths of its three dimensions. It provides a means of making inferences from one size of environment to another, ie., geochemical sample size to an ore deposit size.

D = Linear equivalent of volume of the environment, assumed as the dry land area of the earth's crust to a depth of 1.5 miles.

Brinck uses the specific mineralizability of a metal, as calculated from its ore reserves, to calculate the metal's resources for various grades and sizes. This he calls the "reference distribution" based on the parameters $\gamma_R$ and $\sigma_r$ which are respectively, the average abundance of the metal in the earth's crust[1] and its logarithmic standard deviation in the ore deposits. The parameters $\overline{X}_s$ and $\sigma_s$, the median concentrations in the results of geochemical surveys and the specific mineralizability from the geochemical surveys respectively, are then used with the reference distribution to estimate the resource potential of the surveyed environment with respect to the world's potential indicated in the reference distribution.

In assuming the whole dry land earth mass as the total environment of average metals abundance, Brinck ignores the role of geological processes, even though it is known that different metals show a well established and consistent association with specific rock types and stratigraphies. The model has not found wide application because of its very general approach. However, it may perhaps be useful conceptually for exploration planning on a reconnaissance level.

---

[1]Brinck uses the average crustal abundance of an element as being equal to its median value.

### 2.2.3 The Multivariate Model

The physical occurrence of mineral endowment is a function of certain geological processes, a modified and incomplete evidence of which is reflected in regional geological variables. If the known mineral endowment and the associated geological variables could be quantitatively related the resulting model could then be used as an endowment predicting tool in a similar geological environment. This is the basis of the multivariate statistical model. However, different techniques may be applied depending upon the objective sought.

The application of multivariate statistical analysis in resource potential evaluation was first demonstrated by Harris (1965). The basic postulates of Harris' multivariate model are as follows:

$$W = f(L,S,F,A)$$

$$P(W) = f(L,S,F,A,W)$$

Where,

W = a measure of mineral wealth

P(W) = probability of occurrence of W

L = age and type of rock

S = structural forms

F = rock fracturing

A = age of igneous activity and contact relationships.

In other words, the occurrence of mineral wealth[1] is a function of the listed geological variables and the probability of occurrence of the mineral wealth is a function of its joint occurrence with the geological variables.

Harris, after dividing his reference area into 243 cells, each 20 miles square, made a series of measurements in terms of areas, lengths, and counts of variables in each cell. He used multiple-discriminant[2] analysis and classification analysis by Bayesian statistics to define the relationship of probabilities, mineral wealth, and associated geology. A total of six discriminant groups were used in classification.

The resulting discriminant function from the reference area was extrapolated by Harris to a total of 144 cells in an area outside the reference area. Of these, 19 cells were classified as favourable for further exploration. This constituted a greatly reduced target area.

Agterberg, _et al_. (1972), applied regression analysis in making estimates of the probability of occurrence of

---

[1]Harris based his study on the base metal deposits in New Mexico and Arizona. The geological variables therefore reflect the processes of the region. He extrapolated the results for predictive purposes to porphyry copper deposits in the state of Utah.

[2]The procedure of discriminant function analysis is described and applied later in this study.

copper and zinc in the Abitibi belt of the Canadian Shield. Since the presence or absence of an ore deposit is used as the value criterion for the response variable, their approach in effect is equivalent to a two-group discriminant analysis. Predicted metal probabilities of the 644-cell[1] area, based on a 26 geological and geophysical variables equation, are contoured using moving averages of 16 adjoining cells at a time. Peaks in the contours other than those occurring over known mining regions indicate potentially favourable areas for future exploration.

Other applications of multivariate analysis include those by Harris (1968) in Alaska, by Kelly and Sheriff (1969) in British Columbia, by Harris and Azis (1970) in Mexico, by Sinclair and Woodworth (1970) in British Columbia, and by DeGeoffroy and Wignall (1970) in southwestern Wisconsin.

Multivariate statistical methods can provide meaningful resource potential estimates on both regional and cell bases. They must, however, be applied with an awareness of the methodological limitations and any shortcomings of the data base. Any multivariate statistical model is best for the system on which it is developed. For this reason, special care must be exercised when extrapolating the model outside the reference area. The need for such a caution would vary

---

[1]Each cell is 10x10 Km., or about 39 square miles in area.

for the type of endowment sought and on the level of information available.

### 2.2.4 The Subjective Probability Model

The subjective probability model is based on the process of extracting and quantifying individual judgment about uncertain quantities. The technique can be considered an extension of the multivariate model in that the endowment-geology relationship is estimated by the opinions of geologists instead of the explicit geologic and known endowment information. In other words, the subjective model assumes that the endowment potential of a region is a function of the experience, knowledge and insight of the geologists familiar with the geology of the region.

Subjective probability analysis was first applied in evaluating endowment potential of the Canadian northwest by Harris, Freyman, and Barry (1970), and by Barry and Freyman (1970). The technique was also applied by Azis, et al. (1972), in making estimates of the undiscovered endowment of the Canadian Shield in Manitoba. Various methods of subjective probability encoding and their applicability are discussed by Spetzler and von Holstein (1975).

For the application of this technique, the geologists selected and willing to participate are provided with a set of geological maps of the study area which are divided into equal area cells and containing all relevant and available information such as existing deposits, their tonnages, grades, and values. They are also provided with a questionnaire containing well defined and unambiguous questions on the commodity sought, its categories to reflect ranges of economic significance at various levels of grades, tonnages, and probabilities.

Since individual opinions are involved, any two geologists can arrive at different probability assignments for endowment in the same cell. It is for this reason that the interview process is so important in this method. Spetzler and von Holstein recommend the following steps in the interview process (p. 352):

(1) Motivation: Rapport with the subject is established and possible motivational biases are explored.

(2) Structuring: The structure of the uncertain quantities is defined.

(3) Conditioning: The subject is conditioned to think fundamentally about his judgment and avoid cognitive biases.

(4) Encoding: The subject's judgment is quantified in probabilistic terms.

(5) Verifying: The responses obtained in the encoding are checked for consistency.

The geologists may be allowed access to the replies obtained from other geologists so that if they may want to revise their original opinions, they may do so. This is called the Delphi process. And finally, a Monte Carlo simulation may be used to average out the replies and to obtain results at the desired confidence level.

The subjective probability method has the advantage that the subjective knowledge accumulated by exploration geologists, and their instinct and insight can be quantitatively included in decision making. The problems of incomplete and non-uniform geological information, and some of the limitations of statistical techniques themselves are avoided by this method. However, application of the method requires a reliable sample size of geological expertise within the study area. This is often difficult because of the reluctance of private company geologists to participate due to the competitive nature of exploration. Government agencies in general are cooperative but their experts may not have the first-hand local experience of the study region so essential to a meaningful application of the methodology.

### 2.2.5 Other Models

The models described in this section are semi-statistical in nature. Their application is only on a reconnaissance level. The following is a brief review.

In a presidential address to the Society of Economic Geologists in 1949, Nolan (1950, p. 604) observed that:

> If mineralisation has occurred fairly uniformly throughout a major geologic province, it is safe to conclude, if large enough areas are involved, that a comparable number of mining districts of various sizes may be expected in that part of the province covered by younger rocks as is found in the exposed areas.

Considering the non-statistical forum of the address, this was an important and innovative suggestion. The conclusions arrived at by Allais (1957) are based on a postulate conforming to the above suggestion.

Bates (1959) used Spearman's coefficients of rank correlation for determining favourable uranium-vanadium areas in the Colorado plateau. The Spearman's rho[1] is similar to the ordinary correlation coefficient except that it requires the use of rankings rather than the absolute values of variables in the computation of the coefficient.

Bates noted that 78 percent of all uranium-vanadium deposits discovered in the Colorado plateau since 1944 fell within the favourable areas outlined in his study. Bates was

---

[1]Spearman's rho, denoted $r_s$ is numerically the following:

$$r_s = 1 - \left[\frac{6(\sum_{i=1}^{N} D_i^2)}{N(N^2-1)}\right]$$

...cont'd

also able to eliminate a particular formation as being unfavourable for this type of deposit thereby reducing the area size for exploration.

Botbol (1971) made a study of thirty base-metal mining regions in the United States using a technique he developed called characteristic[1] analysis. The technique can rank large arrays of geological characteristics in order of their decreasing typicality.[2] Botbol numerically coded his data as one or zero on the basis of the presence or absence of the characteristics in the reference cells. The data matrix[3] is multiplied by its transpose so that the rows of the resulting matrix are logic vectors. The square root of the sum of squared common occurrences is called the typicality of the corresponding characteristic. The most typical characteristics based on their ranking are used as a reference base with which comparisons can be made with characteristics of other areas.

Characteristic analysis has the advantage of defining the relative importance of various geological factors for a

---

Where $D_i$ is the difference associated with the particular individual i, and N is the number of individuals observed.

[1]"Characteristics" is synonymous with the terms factors, or explanatory variables as applied to geological data.

[2]Typicality implies frequency of joint occurrence of a characteristic with another.

[3]The columns of the matrix represent the individual cells, and the rows denote the characteristics to be ranked.

particular type of ore deposit.[1] In this way, the number of variables for subsequent multivariate analyses can be reduced. Further, the ranking of typicalities indirectly helps exploration by inviting greater attention to the more important characteristics at the geological mapping stage. The method does not, however, make numerical estimates of the resource potential in the study area.

## 2.3 The Present Study

The studies that have been described in this chapter have all been applied on a reconnaissance level over large regions. The present study focuses on the application of multivariate techniques over a small area, the size of an average mining region. The advantage here is the availability of uniform and detailed geological information and well accepted concepts on ore genesis. However, there are problems relating to small sample size in statistics. The anticipated benefits in using small areas and detailed geological information include predictions of narrower, better defined target areas for exploration, more precise estimates of endowment potential and an ability to statistically evaluate geological concepts. These are described in the following chapters.

[1] To the extent that the frequency of joint occurrence is a measure of the importance of a characteristic.

# CHAPTER 3

## THE STUDY REGION: ROUYN-NORANDA

### 3.1 General Statement

The Rouyn-Noranda region lies about 400 miles northwest of Montreal, Quebec;[1] it is centred about the twin cities of Rouyn and Noranda, and comprises the townships Duprat, Dufresnoy, Beauchastel, and Rouyn. The region has been the centre of exploration activity since the discovery by Ed Horne in 1920 of the copper-gold deposit that was to become the Horne Mine in 1927. Other discoveries that followed in the region include the Amulet C and Upper A orebodies, and Old Waite in 1925, Aldermac in 1927, Amulet F orebody in 1929, Amulet Lower A orebody in 1938, Quemont in 1945, East Waite in 1949, Vauze in 1957, Norbec in 1961, and Millenbach in 1966 (Dugas, 1966; Simmons, et al., 1973)[2]. Between the years 1927 and 1974, the metals produced and reported in reserves exceeded 2.23 million tons of copper, 1.15 million tons of zinc, 28.76 million ounces of silver, and 13.14 million ounces of gold; this production was all from the massive sulphide deposits. Table 1 summarizes the production and reserves figures for these metals on an individual mine basis.

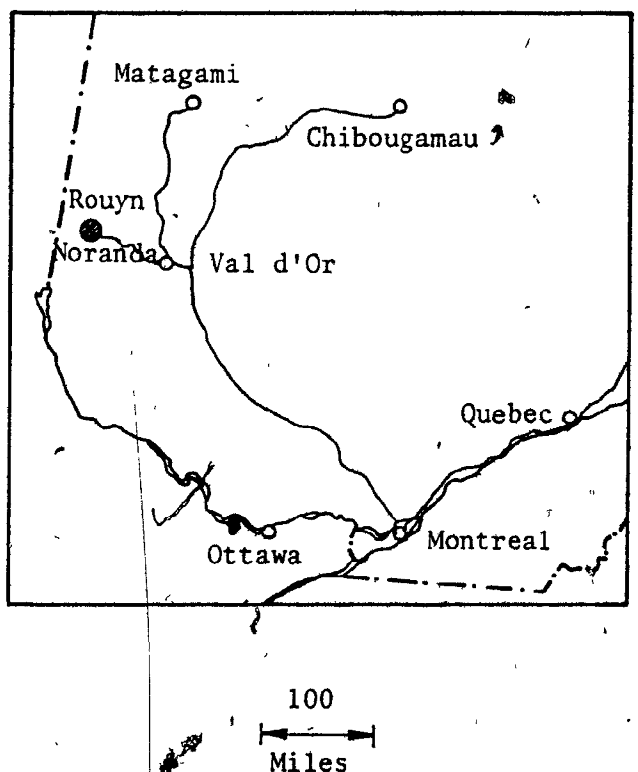
---

[1]See Figure 1.

[2]The Corbet deposit, discovered in 1974, is being developed for production.

FIGURE 1

LOCATION MAP, ROUYN-NORANDA



# TABLE 1

COPPER-ZINC PRODUCTION AND RESERVES
IN THE ROUYN-NORANDA REGION
(AFTER SPENCE AND SPENCE, 1975)

| Mine | Years | Ore (tons) | Copper (tons) | Zinc (tons) |
|---|---|---|---|---|
| *Copper-Zinc* | | | | |
| Aldermac | 1931-43 | 2,057,100 | 30,845 | - |
| Horne | 1927-70 | 56,264,700 | 1,226,018 | - |
| Millenbach # | 1971-* | 2,415,000 | 3.45% | 4.35% |
| Norbec | 1964-70† | 2,800,200 | 93,242 | 134,034 |
| Quemont | 1949-70 | 15,013,000 | 183,801 | 283,991 |
| Vauze | 1961-64 | 385,000 | 11,150 | 3,600 |
| Waite Amulet | 1930-62 | 9,658,000 | 404,009 | 352,921 |
| (A | | 5,872,000 | | |
| B,C,D,E, Bluff | | 596,000 | | |
| F | | 290,000 | | |
| O. Waite | | 1,245,000 | | |
| E. Waite) | | 1,655,000 | | |
| | | 88,593,000 | | |
| *Zinc* | | | | |
| Delbridge | 1969-70 | 400,000 | 2,170 | 34,000 |
| D'Eldona | 1950-52 | 86,500 | 14 | 4,360 |
| West MacDonald | | 1,030,000 | 125 | 30,000 |
| | | 1,516,500 | | |
| Mobrun | -* | 3,000,000 | 0.69% | 2.18% |

\# Current producer

* Reserves, 1972

† Last available data.

As of 1977,[1] the only remaining producer in the region is the Millenbach mine. The depleted condition of ore reserves in the Rouyn-Noranda region therefore underlines the need for an assessment of any additional mineral potential of the region, an assessment that would assist in exploration planning and investment decision making.

## 3.2 Geology of the Rouyn-Noranda Region

### 3.2.1 Introduction

This thesis is predicated on the postulate that ore deposits result from the interaction of specific geological processes that were responsible for:

(1) Extracting by some process the metals contained in the earth's crust;

(2) Transporting the extracted metals to near the surface of the earth in some form or medium, the transportation itself being facilitated by additional geological processes;

(3) Depositing the concentrate resulting from changes in the physico-chemistry of the transporting agent;

---

[1] Since January, 1975, the Norbec mine has acted as a standby to the Millenbach mine, supplying the millfeed from its stockpiled ore. The Horne mine ceased operations in July, 1976.

(4) Preserving the deposited metals by further geological processes.

The characteristic interaction of geological processes that resulted in ore localization was probably random. In fact, some of the more critical factors are still unknown. It was only a decade ago that all base metals deposits in the Rouyn-Noranda region were considered to be of hydrothermal epigenetic origin as originally defined by Lindgren (1933). The same deposits are now believed to be syngenetic, and the result of volcanic processes. While geological thought on ore genesis may change over time, the pattern of ore occurrences in the region does not change. It is for this reason that the geological description in the following sections outlines the regional aspects first before concentrating on the Rouyn-Noranda region itself. This approach is essential from an exploration point of view and for a better comprehension of the results of multivariate analyses applied in this study. As Guild (1976, p. 709) observes, systematic exploration should be based on a genetic model in search of answers to the following questions:

(1) How did the deposit form?

(2) Where were the conditions favourable?

(3) What ancillary features of broader extent might aid in zeroing in on the target?

The answers may only be known conceptually, but they are essential for a meaningful application of quantitative analyses in resource potential assessment.

### 3.2.2 General Statement

Rouyn-Noranda is one of the several Archean regions in the so-called Abitibi greenstone belt of the Canadian Shield, a belt that contains clusters of volcanogenic, base metal, massive sulphide deposits. The following section reviews the geology of the Abitibi belt. This is followed by a description of the geology of the Rouyn-Noranda region itself, and a discussion of the genesis of the massive sulphide ore deposits in the region.

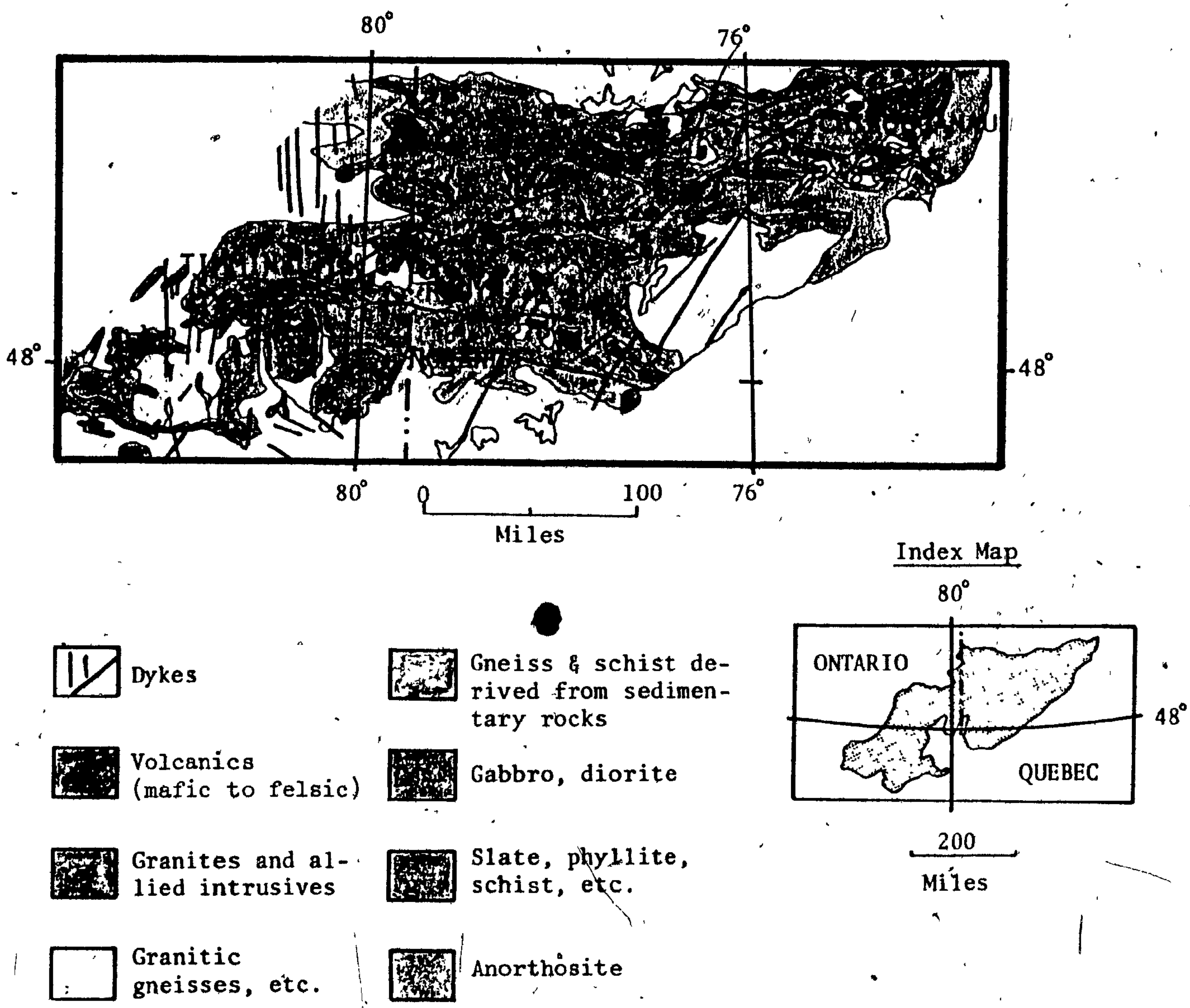
### 3.2.3 The Abitibi Belt

The Abitibi greenstones belt occupies the southern part of the Superior Province, and is the largest single greenstone belt in the Canadian Shield (Fig. 2). In economic terms, it is also the most important.

The rocks in the belt range from mafic to felsic volcanic flow rocks and pyroclastics to sedimentary rocks. These have been intruded by a large number of dykes, sills, and irregular bodies representing a wide spectrum of igneous rocks. The rocks are estimated by Goodwin and Ridler (1970)

FIGURE 2

THE ABITIBI VOLCANIC BELT
(AFTER DOUGLAS, 1970)



to be distributed as follows in the belt:

| | |
|---|---|
| Mafic volcanics | 45.6% |
| Granitic rocks | 32.3% |
| Sediments | 16.0% |
| Felsic volcanics | 3.6% |
| Mafic intrusions | 2.5% |

The volcanic rocks in the belt are of both tholeiitic and calcalkaline affinity. Chemically, they are poor in both potassium and titanium. Descarreaux (1973) observes that except for the oceanic tholeiites, the rocks are poor in calcium relative to magnesium. Baragar (1968) in his study also notes a similar chemistry that is close to that of the circum-oceanic basalts except for the lower potash, lime and iron oxides. Based on the average suite index[1] of the rocks, Descarreaux concludes that the rocks would fall within the range of the basalt-andesite-rhyolite association typical of orogenic belts. Goodwin and Ridler also regard the belt as orogenic and define it as "a remnant of a bilaterally symmetrical intratectonic orogen rather than a conventional asymmetrical continental-oceanic tectonic interface, i.e., an island arc".

---

[1]Suite index is a number equalling

$$(Na_2O+K_2O)^2/(SiO_2-43),$$

and is used to analyze the process of magmatic differentiation. See Barth (1962, p. 168) for detail.

Their concept is explained in Figure 3.

Wilson, <u>et al</u>., 1965, and more recently, Dimroth, <u>et al</u>., 1973, have compared the belt to the present day island arcs, built of highly complex groups of shield volcanoes whose configuration and spacing have changed with time.

Gélinas and Brooks (1974, p. 336) on the basis of detailed chemical and quench-texture studies of the belt have suggested that "the more northerly part of Abitibi volcanic pile studied represents a deeper, more basic section of the volcanic sequence, and could possibly be a more primitive base upon which the island arc was built, the island arc being typified by the rocks south of the DDM[1] break". They also point out the ambiguity that the suggested more primitive base in the northern part has a lower grade of metamorphism, the prehnite-pumpellyite facies, compared to the southern part which is in the greenschist facies. Baragar (1968) has also observed chemical differences between the northern and southern parts of the belt. From a different perspective, Krogh and Davis (1971) in their age dating of rocks across the belt have observed a younging trend towards the south.

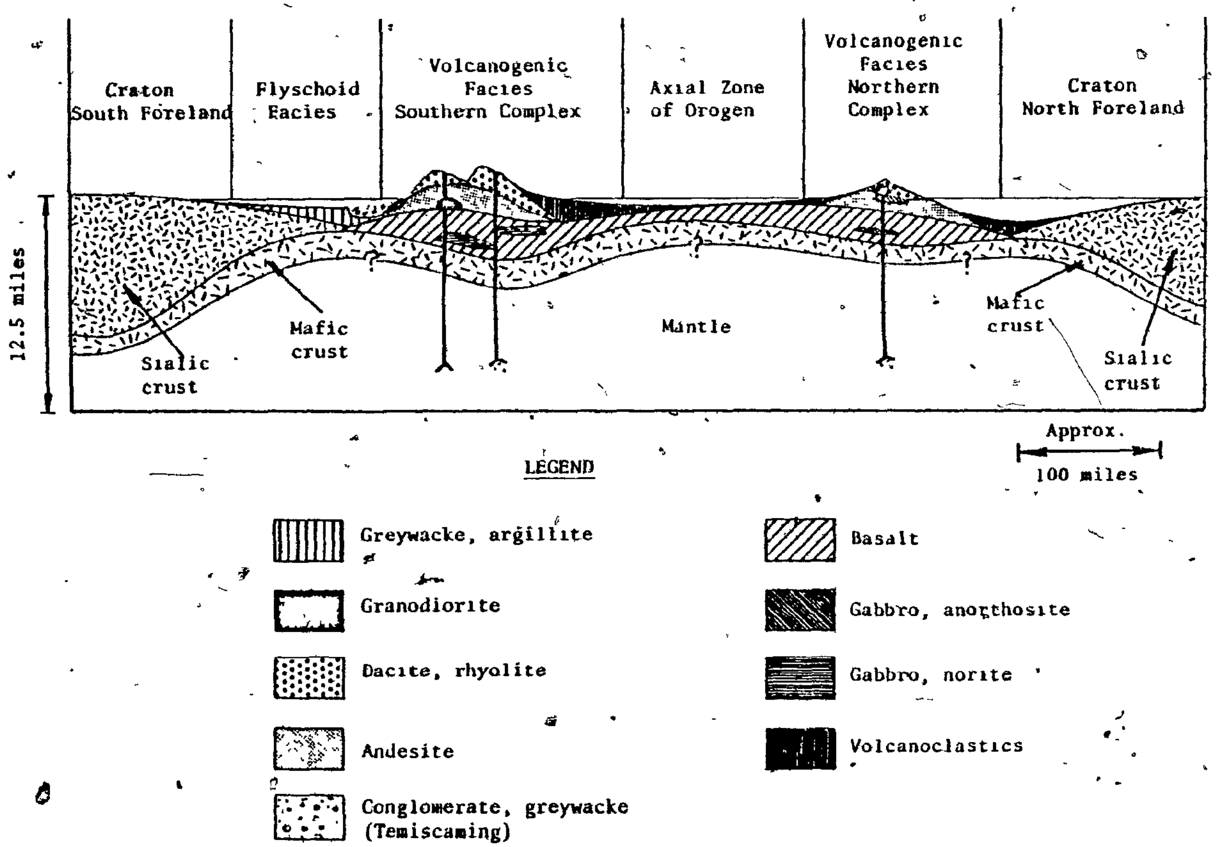
The volcanic rocks are generally weakly metamorphosed except in the vicinity of granitic intrusions. The metasedimentary belt has been metamorphosed to an amphibolite or higher grade.

---

[1]The Duparquet-Destor-Mannville break.

FIGURE 3

HYPOTHETICAL TECTONIC RECONSTRUCTION OF ABITIBI BELT
(AFTER GOODWIN AND RIDLER, 1970)



On the basis of facies, tectonics, structures and geophysics, Goodwin and Ridler have outlined nine volcanic complexes in the Abitibi belt (Figure 4). The complexes have close spatial relationship with the base metal mining regions in the belt. Of these nine, the most important complex is located in the south-central part of the Abitibi greenstone belt and includes the cluster of massive sulphide deposits of the Rouyn-Noranda region. Other complexes contain the base-metal regions of Chibougamau, Matagami, and Timmins. The implication of a relationship between the volcanic processes in the complexes and the massive sulphide deposits is therefore obvious. This aspect is elaborated in latter sections.

### 3.2.4 The Rouyn-Noranda Region

#### General Geology

The Rouyn-Noranda region, because of its economic potential and good outcrop exposure, has been well mapped and studied. The general geology of the region is shown in Figure 5.

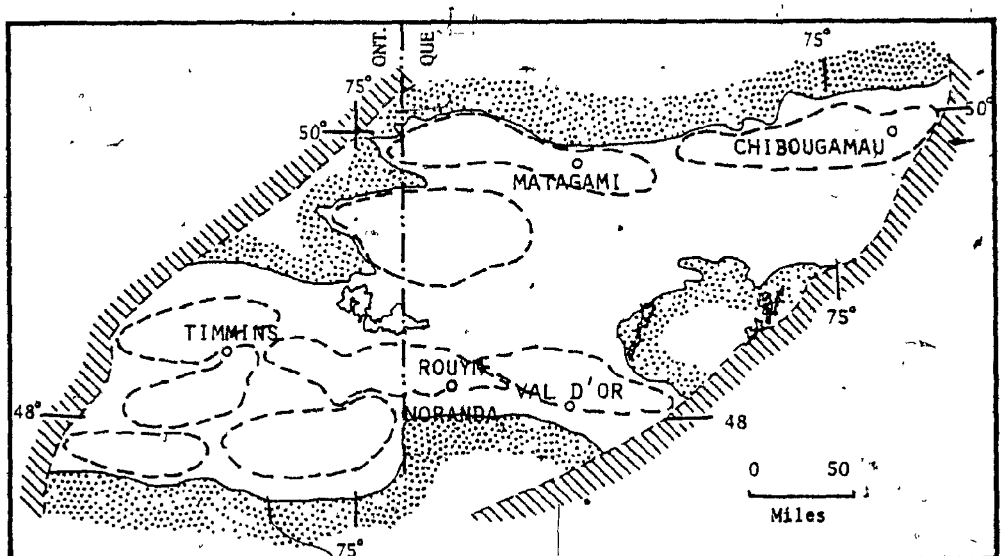
The general stratigraphy of the region[1] is as follows:

---

[1]After Douglas (1970).

FIGURE 4

DISTRIBUTION OF VOLCANIC COMPLEXES IN ABITIBI BELT[1]
(AFTER GOODWIN AND RIDLER, 1970)




[1]Boundaries of volcanic complexes are shown by heavy broken lines.

FIGURE 5

GENERAL GEOLOGY, ROUYN-NORANDA REGION
(AFTER DUGAS ET AL., 1965)


0 — 4 Miles

**LEGEND**

| | |
|---|---|
| Diabase dykes | Dolomite, conglomerate, etc. |
| Granite rocks, including gneisses | Sedimentary rocks |
| Mafic rocks | Rhyolitic rocks and pyroclastics |
| Paragneisses, etc. | Mafic lavas |
| Ultramafic rocks | |

CADILLAC GROUP

Conglomerate
Graywacke
Shale

Unconformity

BLAKE RIVER GROUP

Basic volcanics
Acidic volcanics
Tuffs

PONTIAC GROUP (KEWAGAMA)

Graywacke
Shale
Metasediments

The most important group in terms of ore occurrence is the Blake River group. This group is composed mainly of andesites and rhyolites with relatively lesser amounts of basalt and dacite: There may, however, be more basalt present than originally believed, for as Descarreaux (1973) observed on the basis of chemical analyses, there has been a tendency amongst geologists to give names to volcanic rocks in the region that are too felsic; this is particularly true of basalts which have often been called andesites. On the base map used in this study, basalt and andesite have been combined and treated as one unit.

The Blake River group overlies the graywackes of the

Kewagama group,[1] the contact being characterized by interbedding and interfingering of sediments and volcanics. Wilson (1962) regards this contact as an unconformity. On the south with an apparent unconformity, the more basic volcanics of the Blake River group overlie the metamorphosed graywacke and shale of the Pontiac group.

The stratigraphic units are discontinuous, with the flows pinching out over short distances. The thicknesses also vary, in particular for the more viscous felsic units. Baragar (1968) estimates that the volcanics in the region attained a true thickness of at least 40,000 feet.

Most of the felsic rocks are concentrated in the central and eastern parts of the region and most likely represent centres of eruptive activity. No effusive centres have determined for the basic and intermediate rocks (Spence, 1967).

Roscoe (1965) has observed that the basic volcanic units become more siliceous in stratigraphic ascending order through the sequence. Baragar in his study of the Duparquet section north of Duprat township notes a decline in the colour index of rocks from about 40 at the base to 20 at the top of the main limb, and rising again to about 30 in the south limb. He related the decrease in colour index to the increase in the alumina content. This does not itself suggest increasing

---

[1]Ambrose (1941) correlated Kewagama group with the Pontiac.

acidity with increasing stratigraphic height. Baragar has suggested that the increase in alumina content may be related to prolonged volcanism causing enrichment of the plagioclase component and depletion of the ferromagnesia and titaniferous minerals to mesocratic high alumina lavas. The concept of prolonged volcanism may find support in the study by Krogh and Davis (1971) who found a younging trend towards the southern parts of the Abitibi belt.

In general, the volcanic pile is conformable despite numerous alternations of felsic and intermediate rocks. There is a gradual evolution of rhyolites from andesites, the dacites being commonly present in transistion. However, rhyolites and andesites are also present in sharp contact with one another, both above and below each other. The contacts show no signs of any erosion or sedimentation but they do often contain layers of chert and tuff; these layers constitute good marker horizons over short distances. Discontinuous and irregular belts of breccia are present in volcanic rocks.

All the volcanic products are submarine. As evidence, Spence and Spence (1975, p. 94) cite the following features:

(1) The lack of oxide facies such as banded iron formation;

(2) The poor development of vesicularity in the lavas, reflecting a limited escape of volatiles due to a high hydrostatic pressure;

(3) The great lateral extent of lava flows, especially of fluidal rhyolites;

(4) The paucity of pyroclastic products, implying rare explosive activity due to a high confining pressure and, where found, their restricted distribution;

(5) The lack of erosional products;

(6) The presence of pillows throughout the vertical extent of individual andesitic formations which are as much as 3,000 feet thick.

In view of the absence of areally extensive sheets of aquagene tuffs, Dimroth, et al., (1973), have also suggested a reasonably great depth of eruptive activity, certainly below more than 330 feet.

Lithology

Mafic to intermediate flows:

For this study, all volcanic flows more basic than rhyolite[1] have been included in the mafic to intermediate flows category. These include basalt and andesite with minor dacite and trachyte. These rocks exhibit a broad range of

---

[1] I.e., rocks with $SiO_2$ less than 68% (Spence and Spence, 1975, p. 91).

physical features including, pillows, variolites, amygdules and laminations. Flow breccia is frequently present. The thickness of individual flows normally does not exceed about 100 feet. The contacts between the flows are sharp and may include intercalated breccia. Feeders to these flows have not been recognized with certainty. However, Gilmour (1965), Van de Walle (1972), Dimroth, et al. (1973), and Spence and Spence (1975) have suggested that the older dioritic-gabbroic dykes may have acted as such.

Acidic rocks

Rouyn-Noranda region is exceptional in having a high proportion of rhyolites within its pile of calcalkaline rocks. The rhyolites tend to be concentrated in the centre of the region, decreasing towards north, south, and west. Spence and Spence (1975) suggest the development of the pile from a centre that has migrated eastward along an east-west axis, an axis that is now occupied by the Flavrian and Lake Dufault granites. The rhyolites include both homogeneous and heterogeneous types, the physical state being a function of distance from the source and the viscosity of the lava. Spence (1967) and Spence and Spence (1975) have distinguished five different belts of rhyolites on the basis of stratigraphy and their physical state, and suggest that ore deposits are associated with only three of the five belts. This is evidence of multiple cycles of eruption and ore formation. The problem is that it is difficult to distinguish the different rhyolites or even to correlate between similar rhyolites over more than short distances because of lensing and the effects of intrusions and deformation. Although most of the rhyolites in the region are flow rocks, some may be pyroclastic. This is the view of Sakrison (1966), and Larson and Webber (1977). However, detailed interpretation of Archean volcanics, and the recognition of pyroclastics is a problem. A classification and elucidation of such rocks in the region by Dimroth (1977) is helpful in evaluating the flows of pyroclastic origin and acid volcanics.

## Gabbros, diorites, and quartz diorites

These are the most widespread intrusives forming large irregular bodies, both sill-like and cross-cutting, and confined mostly within the volcanic rocks. They probably cover a wide span of time, but the relationships are not clear. The oldest of these, also called meta-diabase may be penecontemporaneous with the intermediate lavas, and may likely have acted as their feeder dykes.

## Granitic rocks

About one-fourth of the study area is underlain by granitic rocks. These include the Lake Dufault granodiorite, and the Powell and Flavrian granites, all in roughly oval-shaped areas.

The Lake Dufault granodiorite is a composite intrusive intersecting andesites, rhyolites, and diorites. Webber (1962) observes that the western part is massive without lineations and contains inclusions of the brecciated intruded rocks. The eastern part shows the effects of shearing, alteration and weathering more prominently suggesting a hybrid origin caused by the combined effects of assimilation and metasomatism. Sakrison (1966) suggests that the eastern half may include a possible pendent of rhyolite in which the West Macdonald ore-body occurs. Wilson (1941) has suggested that in view of the similarity in mineralogy and chemical composition to the quartz diorite, the granodiorite may have been derived from a dioritic magma, but that the differentiation must have taken place at depth because this rock also intrudes quartz-diorite.

The Flavrian Lake granite and its faulted extension, the Powell granite, are mainly enclosed in rhyolite and lie along the axis of an anticlinorium. Both granites have the same mineralogical composition, and chemically are very similar to the rhyolites (Wilson, 1941). Van de Walle (1972) has suggested that these granite stocks are deeply eroded sites of subvolcanic centres that have been feeding most of the rhyolitic and dacitic flows that are distributed concentrically around the centres.

## Other Intrusives

These include the post-ore potassium rich syenite porphyry bodies of the late Archean (Wilson, 1962) and the late diabase dykes. The dykes are continuous, and almost vertical, and intrude all other rocks in the region. There is a controversy as to whether the dykes are pre- or post-mineralization in age (Price, 1934, 1948; Campbell, 1962). As Ridge (1972) concludes, the relationship between the dykes and base metal mineralization in the region is not well understood.

## Structural Geology

The main structural feature of the region is a complex anticlinorium plunging east on an east-west axis. Dips as measured on the rhyolite-andesite contacts are flat at the centre but increase towards the border areas, where they range from 15 to 60 degrees. Spence and Spence (1975) postulate the control of the fold by the original volcanic centres because of the thickening rhyolites in the axis and nose of the folds.

Faulting is widespread in the volcanics. Wilson (1941) believes that faulting is related to the folding of volcanic rocks. He concludes that movements recurred along the major fault zones at intervals from the early Archean to the Proterozoic. He notes that the intrusions of diorite follow some of the major faults in the region indicating that faulting began before the diorites were intruded. Later movements have sheared the diorites but not as much as the volcanics. Diorites also appear to have a structural relationship to the Flavrian granite in being predominantly outward dipping and forming radial dykes about it.

Summary

The following points summarize the aspects of the Rouyn-Noranda region geology most relevant to the present study:

(1) The region has been a centre of volcanic activity within the larger orogenic Abitibi belt.

(2) Andesite and rhyolite are the predominant rocks in the area. However, basalt may be present more prominently than previously believed. Rhyolites appear to both evolve from the andesites via dacites, and also to occur with sharp contacts with the andesites. These features indicate evolutionary and recurring cycles of volcanic activity.

(3) Granites in the region appear to be genetically related with rhyolites, and diorites with andesites.

(4) The dominant structure in the area is an east-west trending anticlinorium. Volcanic activity appears to have controlled the folding. The widespread faulting in the region appears to be related to the folding process and so do the dykes.

(5) It appears that in the Rouyn-Noranda region, practically all aspects of geological processes in the Archean, lithological, structural and ore-forming, were directly or otherwise a consequence of volcanic activity.

### 3.2.5 Economic Geology

The massive deposits in the Rouyn-Noranda region are either copper-rich with lesser amounts of zinc, zinc-rich with lesser amounts of copper, or mainly pyritic with some copper and zinc. Gold and silver are present with all of them.

Spence and Spence (1975, p. 94) list the following features commonly present in the Rouyn-Noranda massive sulphide deposits:

(1) A normally pipe-like zone of chloritic and sericitic alteration, with disseminated and stringer sulphides extending stratigraphically below the ore. This represents the conduit for rising solutions;

(2) Massive sulphides, rooted in the pipe, form stratabound, usually lensoid bodies on the surface of flows or explosive breccias on or above the upper contact of rhyolitic formations;

(3) A metal zoning showing chalcopyrite and pyrrhotite-rich ore overlain by pyrite and sphalerite, and a lateral and outward increase in pyrite and sphalerite;

(4) Layering in the sulphide conforms to that of the enclosing rocks;

(5) Alteration zones and sulphides are cut by the intrusions.

These features are better appreciated in light of the following review of ore genesis of massive sulphide deposits. It is not the object of this thesis to prove or disprove any particular theory of ore formation. This would not conform with the objectivity contemplated in quantitatively relating ore deposits and associated geology. However, statistical deductions are only valid when corroborated with geological thought and field evidence, and for this reason, theories of ore-genesis, past or current, can be used in identifying and isolating fortuitous relationships so that their effect can be controlled and reduced.

Lindgren (1933) and more recently Ridge (1972), in their discussion of the Rouyn-Noranda base-metal massive sulphides, assign the deposits to the hypothermal[1] category of hydrothermal deposits. In fact until about the mid-sixties, the origin of massive sulphides was well accepted to conform to Lindgren's hydrothermal hypothesis, and exploration for these deposits was carried out accordingly.

That exploration has been so successful in the Rouyn-Noranda region can be attributed to the realization of the role of stratigraphy in ore localization (Dugas, 1966). The role of stratigraphy remains unchanged, perhaps even strengthened by the current volcanogenic concept on massive sulphide formation. In view of the extensive field and laboratory evidence accumulated in recent years, the close spatial and genetic relationship between volcanism and massive sulphides is well accepted. This is also evident in the present world wide trend in exploration for massive base-metals deposits.

The world-wide application of volcanogenic concept has received strong support from the detailed studies of the Miocene Kuroko massive sulphide deposits in Japan (Tatsumi, 1970). These deposits are relatively undeformed and are found intimately associated with calcalkaline volcanic rocks in a currently active island arc. Perhaps, the main speculation remaining, regards the nature of the volcanism related

---

[1] I.e., epigenetic deposits, produced by ascending waters of uncertain origin, but charged with igneous emanations, and concentrated and deposited at great depth and a high temperature, 300-500°C.

ore-bearing fluids, and the factors that controlled their movements and precipitated the contained metals as sulphides.

Oftedahl (1958) has suggested vapour transport during volcanism as an effective method of forming such deposits. The suggestion was not, however, well received in view of the extremely low concentrations of metals in a vapour medium. At the other extreme, Hutchinson (1965) suggests molten sulphide flows, but the suggestion does not conform with experimental evidence on sulphide phase relationships. Most workers,[1] however, believe that some form of hydrothermal activity was involved in the transport and deposition of these deposits. The solutions may have been in the form of sulphide complexes or saline brines, but this is still conjectural.

The presence of hydrothermal activity at the time of ore formation is evidenced in the chemistry and mineralogy of the alteration pipes below the present massive sulphide lenses in the region. These pipes in many cases contain sulphide mineralization called stringer ore. The effects of hydrothermal activity are well documented by Riddell (1952), and Sakrison (1966).

The shift to a volcanogenic concept for the origin of Rouyn-Noranda massive sulphide deposits is not a negation of Lindgren's hydrothermal concept. The difference lies

---

[1] E.g., Sakrison (1966); Barnes and Czemanske (1967); Anderson (1969); Ridge (1972); Sangster (1972); Stanton (1972); Spence and Spence (1975).

essentially in the timing of the geological events. The available evidence strongly suggests that ore deposition took place co-eval with volcanic activity within a short span of geological time and this has resulted in the stratigraphic control of ore deposition. There may be more than one stratigraphic horizon of deposition when volcanism is interruptive in nature. It is for this reason that Dimroth, et al., (1974), have stressed that exploration for these deposits in the Rouyn-Noranda region should be based on recognition and mapping of horizons of pyritic chert and shales; these horizons indicate temporary interruption of volcanic activity at stratigraphic levels where rhyolites are present.

In a recent study, Brooks and Gelinas (1977) have observed that chemical stratigraphy is a fundamental property of most Canadian Archean volcanic belts; and that whereas cal-calkaline volcanics are favourable as potentially mineralized zones, the tholeiitic volcanics are relatively barren. However, as an exception, they point out that mineralization in the Timmins region appears to occur in tholeiitic volcanics. The significance of the exception is evident in that the region contains one of the most important massive sulphide deposits in the world, the Kidd Creek deposit. The implications in terms of ore genesis and exploration, are obvious.

## 3.3 Concluding Statement

Stanton (1972, p. 540) after a detailed discussion of volcanogenic processes in massive sulphide formation sums up the situation as follows:

> All this is, however, no more than a reasonable hypothesis. Clearly there is no aspect of the 'origin' of these deposits that can be said to have been solved. Indeed we have barely shaken ourselves free from the all embracing--and hence highly inhibiting plutonic replacement theory and are hardly past the threshold of a new attack on the problem. Such a stage in the investigation of so important a group of deposits is, however an intriguing and exciting one.

It is against the above background that multivariate statistical analysis is applied in quantitatively relating known endowment and related geology. The strength of this analysis lies in its objectivity.

Basic geological measurements used as data are not subject to change except in detail. The resulting model is therefore pertinent to what is actually observed and measured. However, since no statistical analysis can prove a cause-effect relationship between variables, both the selection of relevant variables and interpretation of results should conform to the current albeit subjective, theories of ore forming processes. This can result in a bias from the individual's perceived understanding of ore genesis and thus requires an objective approach. There can also be a case where the relationship between geology and endowment does not conform to what is accepted as the process of ore formation. This indicates that either the model or the geological theory on ore formation needs to be re-assessed for validity, and if necessary revised.

# CHAPTER 4

## THE DATA BASE

### 4.1 General Statement

Duprat, Dufresnoy, Beauchastel and Rouyn, the four townships comprising the area under present study have been mapped in detail on a scale of a thousand feet to an inch by the Quebec Department of Natural Resources. This Department has also prepared compilation maps of the four townships on scales of two inches and one inch to a mile, and a regional map of 1/4 inch equal to one mile.

The detail on the small scale regional map[1] is too scant for a meaningful quantitative analysis. On the other hand, the large scale quarter township maps[2] are not suitable either because of the complexity of detail. Moreover, the information on them has been mapped by different geologists over different periods of time, and thus, there have arisen problems of uniformity. From the pragmatic point of view of including the optimum detail combined with the ease of making measurements, the compilation map on a scale of two inches to a mile was chosen for the study. This map also has the

---

[1] i.e., 1/4" = 1 mile.

[2] i.e., 1" = 1,000 feet.

advantage of uniformity of detail and nomenclature.

As a first step, the 400 square mile area represented by the four townships mentioned, was divided into subareas; these will be referred to as "cells" throughout the following discussion. The subdivision of an area is based on the following considerations:

- The total size of the area being considered;
- The "grain" of the geological information on base map;
- The objective of the study;
- The statistical approach contemplated;
- Pragmatism.

Cells that are too small make measured data in individual cells approach dichotomy.[1] On the other hand, for a fixed size area, the choice of large sized cells will result in a decreased sample size for an effective statistical analysis. While small cells have the advantage of providing a more specific focus for exploration, the larger cells have a greater variety of geology in them and are thus amenable to developing more effective relationships.

Against the above background, it was decided that dividing the total area into 64 equal sized square cells, each 6.25 square miles in area would be the most practicable

---

[1] I.e., of the "present" or "absent" type.

solution. Therefore, each township of ten by ten miles square area was divided into sixteen cells, each a square of 2.5 by 2.5 miles. The division of each township was done by drawing equidistant lines parallel to the township boundaries.[1]

## 4.2 Measurements Made

Ore deposits result from distinct geological processes following physico-chemical laws. In varying degrees of modification, distortion and completeness, a record of these processes is available in the rocks and structures observed today.

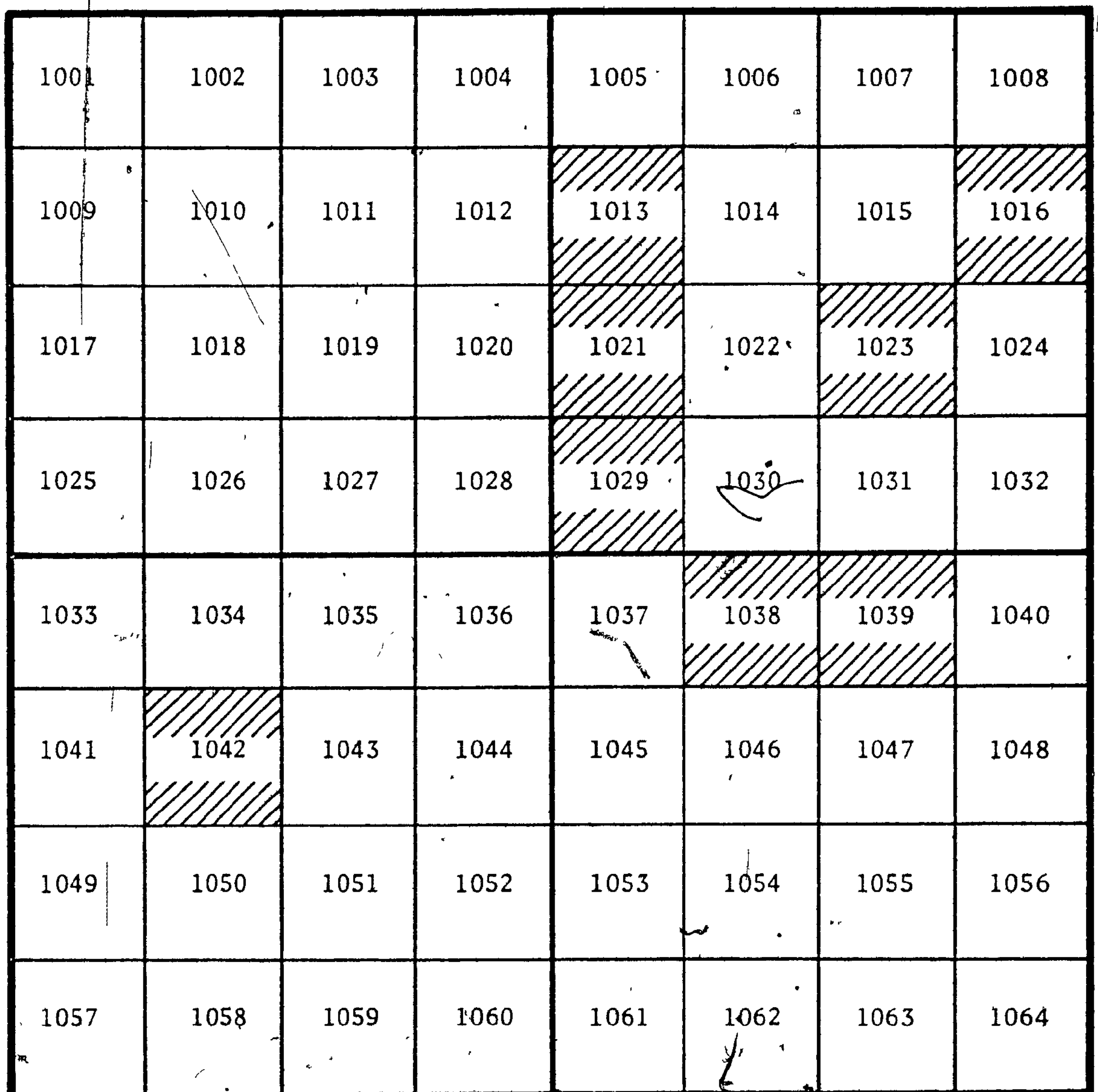
If the volume of a rock type was responsible for ore formation, it can now only be approximated by its surficial area, for the measurements along the depth are least known. If some timing was involved in the ore-forming event, then stratigraphy or contact lengths between formations may be indicative. Contacts with igneous intrusions can be evaluated for evidence of their contribution to the formation of ore deposits. And finally, if any structures were involved in the formation of ore deposits, or were themselves a result of ore forming processes, their measurements can be usefully incorporated along with the areas of rock formations and the contact lengths amongst them. Therefore, in each of the 64

---

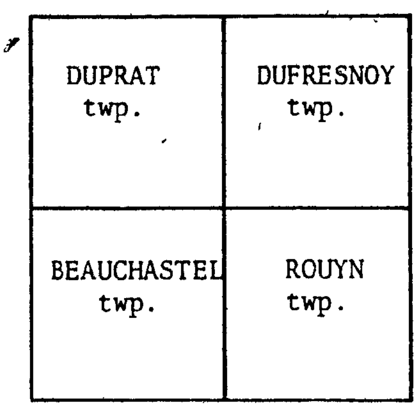
[1]See Figure 6.

# FIGURE 6

## CELL DISTRIBUTION IN THE STUDY REGION[1]




INDEX MAP




[1]Cells containing known endowment are shown hatched.

cells of the study area, measurements were made of the areas of rock formations, contact lengths between every possible pair of these formations, and of synforms, antiforms, dykes and faults.

Areas of formations were measured using a planimeter. The measurements were made four times for each formation in a cell, and the result averaged. The total of all areal measurements in a cell was recalculated to bring the total to 6.25 square miles, the theoretical cell size. The following are the formations measured on the base map:

| Rock type/Formation | Coding |
|---|---|
| Biotite, hornblende paragneises, etc. | AREA 1 |
| Tuff, agglomerate | AREA 2 |
| Rhyolite | AREA 3 |
| Andesite, basalt, dacite, trachyte | AREA 4 |
| Graywacke, arkose (Temiscaming) | AREA 5 |
| Conglomerate (Temiscaming) | AREA 6 |
| Peridotite, pyroxenite | AREA 7 |
| Diorite, gabbro | AREA 8 |
| Rhyolite porphyry | AREA 9 |
| Syenite, monzonite | AREA 10 |
| Granite, granodiorite | AREA 11 |
| Areas of lakes and rivers, geology on which was not extrapoled | AREA 12 |
| Graywacke, conglomerate, etc. (Huronian) | AREA 13 |

Linear measurements were made on all possible contacts between pairs of formations and on structural elements: synforms, antiforms, dykes, and faults. These measurements were made using a pair of dividers with a constant spacing of 0.1 inch. Visual interpolation was made for lengths less than 0.1 inch. Each of these measurements was made twice, averaged, and converted into miles. The following are the contact lengths measured and their coded variable names.

| CODING | Contact between formation: |
|---|---|
| CNTL 1 | Paragneisses & andesite/basalt |
| CNTL 2 | Paragneisses & conglomerate (Temiscaming) |
| CNTL 3 | Paragneisses & peridotite |
| CNTL 4 | Paragneisses & granite/granodiorite |
| CNTL 5 | Paragneisses & graywacke (Huronian) |
| CNTL 6 | Tuff/agglomerate & rhyolite |
| CNTL 7 | Tuff/agglomerate & andesite/basalt |
| CNTL 8 | Tuff/agglomerate & graywacke (Temiscaming) |
| CNTL 9 | Tuff/agglomerate & diorite/gabbro |
| CNTL 10 | Tuff/agglomerate & granite/granodiorite |
| CNTL 11 | Rhyolite & andesite/basalt |
| CNTL 12 | Rhyolite & graywacke (Temiscaming) |
| CNTL 13 | Rhyolite & diorite/gabbro |
| CNTL 14 | Rhyolite & rhyolite porphyry |
| CNTL 15 | Rhyolite & granite/granodiorite |
| CNTL 16 | Andesite/basalt & graywacke (Temiscaming) |
| CNTL 17 | Andesite/basalt & conglomerate (Temiscaming) |
| CNTL 18 | Andesite/basalt & peridotite |
| CNTL 19 | Andesite/basalt & diorite/gabbro |
| CNTL 20 | Andesite/basalt & rhyolite porphyry |
| CNTL 21 | Andesite/basalt & syenite/monzonite |

| | |
|---|---|
| CNTL 22 | Andesite/basalt & granite/granodiorite |
| CNTL 23 | Andesite/basalt & graywacke (Huronian) |
| CNTL 24 | Graywacke (Temiscaming) & conglomerate (Temiscaming) |
| CNTL 25 | Graywacke (Temiscaming) & diorite/gabbro |
| CNTL 26 | Graywacke (Temiscaming) & syenite/monzonite |
| CNTL 27 | Graywacke (Temiscaming) & graywacke (Huronian) |
| CNTL 28 | Conglomerate (Temiscaming) & syenite/monzonite |
| CNTL 29 | Conglomerate (Temiscaming) & granite/granodiorite |
| CNTL 30 | Conglomerate (Temiscaming) & graywacke (Huronian) |
| CNTL 31 | Peridotite & graywacke (Huronian) |
| CNTL 32 | Diorite/gabbro & rhyolite porphyry |
| CNTL 33 | Diorite/gabbro & syenite/monzonite |
| CNTL 34 | Diorite/gabbro & granite/granodiorite |
| CNTL 35 | Diorite/gabbro & graywacke (Huronian) |
| CNTL 36 | Rhyolite porphyry & syenite/monzonite |
| CNTL 37 | Syenite/monzonite & graywacke (Huronian) |
| CNTL 38 | Rhyolite & syenite/monzonite |

Structural parameters are directional features and they are therefore assigned to one of the following groups based on their direction.

| Direction Group | Coding |
|---|---|
| East-west to north-east | 1 |
| North-east to north-south | 2 |
| North-south to north-west | 3 |
| North-west to east-west | 4 |

The structures are therefore coded as shown below to reflect their direction.

| Structural Parameter | Coding |
|---|---|
| Synforms | SNFM 1 to SNFM 4 |
| Antiforms | ANFM 1 to ANFM 4 |
| Dykes | DYKE 1 to DYKE 4 |
| Faults | FOLT 1 to FOLT 4 |

All measurements, areal and linear, are made cumulatively for the particular variable in each cell.

In view of the cell size chosen, and because of the variety of geology present in the area, only some of the above variables are actually present in any particular cell. Variables not present in a cell are given zero value. As shall be explained in later sections, only some of the variables are significant in quantitative modelling.

4.3 Data Compiled

Ore production and reserves figures of copper and zinc[1] for the mines in the region were compiled from the following sources.

- Quebec Dept. of Natural Resources mineral inventory cards;

---

[1]And also, for the associated silver and gold.

- Canadian Mines Handbooks;
- National mineral inventory cards at the Department of Energy, Mines, and Resources, Ottawa;
- Canadian Minerals Yearbooks;
- Company annual reports;
- Unpublished record at the office of the Resident Geologist, Rouyn-Noranda region.

The total production and reserves figures converted into contained copper and zinc tonnages are assigned to the cells on the basis of their known characteristics. These values are referred to as the known endowment. To obtain a common value denominator for copper and zinc, their tonnages were converted into dollar values using the 1975 prices of 63 and 37 cents per pound respectively, and then added.

Of the 64 cells in the region, only eight contain known ore deposits with production history and measured reserves.

## 4.4 The Known Endowment

In developing a multivariate statistical model, it is necessary to relate the known mineral endowment of the region to its associated geological characteristics. Despite their great geological age, early Precambrian, these characteristics can be reasonably mapped and interpreted.

But the associated mineral endowment can never be fully known even after an area has been intensively explored and interpreted. Under these conditions, the most reliable estimate of endowment is the sum of what has been produced and current ore reserves. In this study, this will be referred to as the "known endowment". Unfortunately, the known endowment is not the whole endowment. Of the two contributors to the known endowment, i.e., production and reserves, the former is the more reliable estimate because it was produced and reported in terms of both tonnage and grade. But production itself is dependent upon the technology and economics of the time, and as Harris (1975) points out, these effects cannot be isolated or removed. Production in actual practice is the material mined above a selected cut-off grade. No record is generally available of the marginal or lower grade material left inside the mine that is not reported as part of reserves.

Reserves, unlike production, are subject to great variation, again depending upon the economics and technology at a given point in time. An increase in metal prices will permit a lower grade material to be mined and thus increase the mineable reserves. Advances in technology have a similar effect. And because of the exponential tonnage grade relationships present in some ore deposits, the effect on tonnage of mining lower grade ore can be considerable. Yet the reserves as measured are above an economic cut-off grade and, thus, are not a complete estimate of what is really known to exist.

Further, the reported reserves are dependent on an individual company's policy of disclosing information and may be biased by the existing socio-economic environment. There are also problems related to the terminology[1] used in reporting resources. This too can bias the overall estimate.

It is obvious, therefore, that the known endowment as assessed in this study is incomplete, and therefore represents the "minimum possible" estimate. The forecasts made using this known endowment, will therefore, be conservative estimates.

## 4.5 Problems Related to Geological Data

An objective study of geological phenomena requires that a certain level of objectivity be maintained in the measurement of geological information, in particular field mapping. A significant amount of subjective information accumulates in the mapping process for the following reasons:

- Lack of sufficient rock exposure;
- Lack of a third dimension in viewing rock formations;
- Altered, metamorphosed, and deformed state of the rock;

---

[1]For problems related to resource terminology, see Section 4.6

- Scale of mapping, and the time available for mapping;
- Judgment of the geologist relative to his training, experience and current geological concepts.

Regional data therefore tend to be non-uniform in quality and interpretation, and sometimes continuity. In reviewing problems of applying mathematical techniques in geology, Agterberg and Robinson (1971, pp. 569-570) observe that "the product available to the mathematician conveys the geologist's interpretations of field and laboratory observations and measurements, but not the observations and measurements themselves". Frequently, the interpretations made are related to the particular objective of the geologist's study, and may contain emphasis on factors less relevant to a quantitative study of unknown endowment. This is particularly a problem with older data which may have to be re-interpreted for a meaningful application. Geologists, apart from their possible skepticism of mathematical studies, are also hampered in their work by the absence of a universally accepted classification of rocks. There are a number of rock classification systems based on various criteria such as geochemistry, mineralogy, and textures. However, it was only in 1973 that a sub-commission of the International Union of Geological Sciences[1],

[1] I.e., I.U.G.S.

submitted its recommendations for classifying igneous rocks. In practice, field geologists continue to classify a rock by its locally adopted name when first mapped, or by their own subjective judgment.

It is obvious therefore, that a geomathematical study inherits a certain bias even before it gets started. This bias will be less for a well developed mining region which justifies detailed study over a long period of time and, thus, results in a standardization of geological nomenclature.

## 4.6 Resource Classification Problems

A number of problems relating to geological data are discussed in Section 4.5. Similar problems with more serious possible consequences exist in the case of resource information. Part of the problem with both geological and resource data is that they are dependent upon the particular objective at the time of their measurement or compilation. The ultimate application of this information may be quite different from the initial objectives. However, standardizing the terminology alone can ameliorate the situation and improve the foundation for objective studies.

Practically all definitions of reserves and resources are adaptations of earlier sets of definitions with intent to eliminate ambiguity, increase precision, and to account for changing uses and perspectives. The Department of Energy, Mines, and Resources, Ottawa, recommends[1] the following usage for metalliferous and industrial minerals:

| | |
|---|---|
| Ore: | A naturally occurring, solid mineral-bearing substance from which one or more valuable constituents could be profitably extracted by mining and separation under the conditions prevailing at the time of the appraisal. |
| Ore reserves: | Ore tonnage that can be reasonably assumed to exist. It requires an indication of accuracy of measurement in accordance with the Department classification table.[2] |
| Resources: | These are identified and merely surmized concentrations of naturally occurring solid, liquid, or gaseous materials in or on the earth's crust from which specific commodities are estimated to be obtainable economically with a specified probability and within a specified time span, under explicit assumptions. |

The U.S. Bureau of Mines uses a similar set of definitions. However, its resource definition is not as precise since it does not specify a time span over which the resource could be considered economically feasible.

The first comprehensive attempt to define and clarify the meaning of reserves and resources was made by a committee of the Society of Economic Geologist (Blondel and Lasky, 1956). This was followed by a classification of mineral resources by McKelvey (1972, 1975) shown in Figure 7. The U.S. Bureau of Mines has adopted this classification for

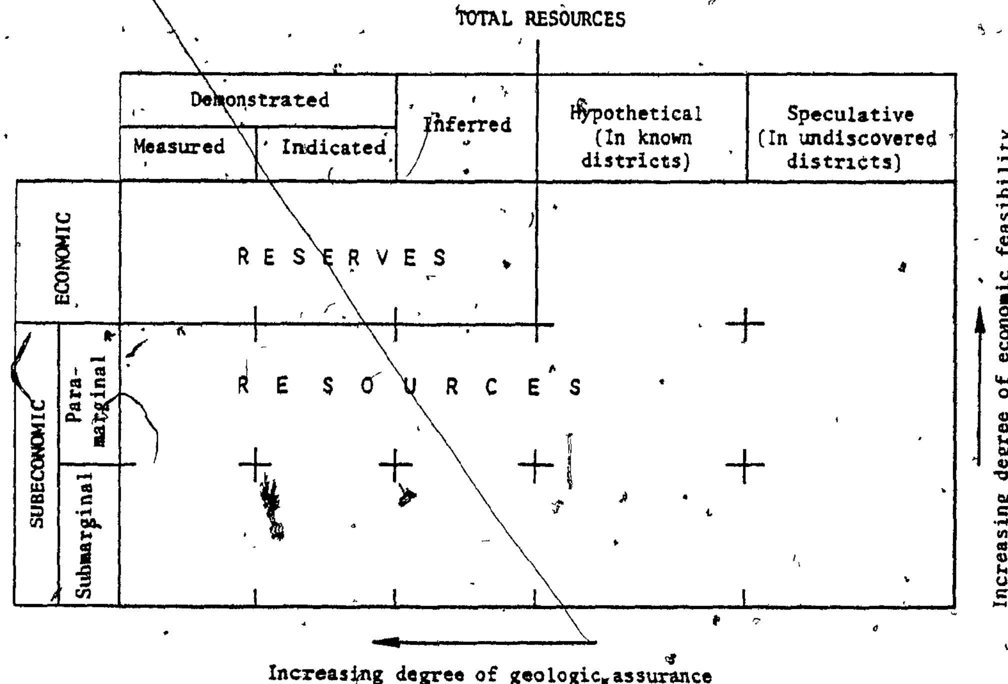
---

[1]See Zwartendyk (1975).

[2]See Figure 8.

FIGURE 7

CLASSIFICATION OF MINERAL RESOURCES
(AFTER McKELVEY, 1975)



its usage. Since McKelvey's classification scheme is rather static, J. Zwartendyk[1] and his associates at the Department of Energy, Mines, and Resources, Ottawa, have made certain modifications, and recommend the modified classification scheme for the Department's use.[2] A simplified form of this scheme has been presented by Azis, et al., (1977).[3] In a continuing work on improving resource terminology, Schanz (1975) has made a series of recommendations in a detailed report for Resources for the Future.

The classification diagrams shown are self-explanatory, and it is not intended to go into details. However, the work being done is still essentially at an academic level, both in Canada and the United States. The need for active participation by the mining industry is essential if the resource classification schemes are to be tested for application and general adoption. Standardized resource terminology will assist in developing a better inventory of what is actually known. More importantly, it will make it easier to relate what is known, economic and uneconomic, to the associated geological environment in helping to make forecasts of unknown resources. From an exploration investment point of

---

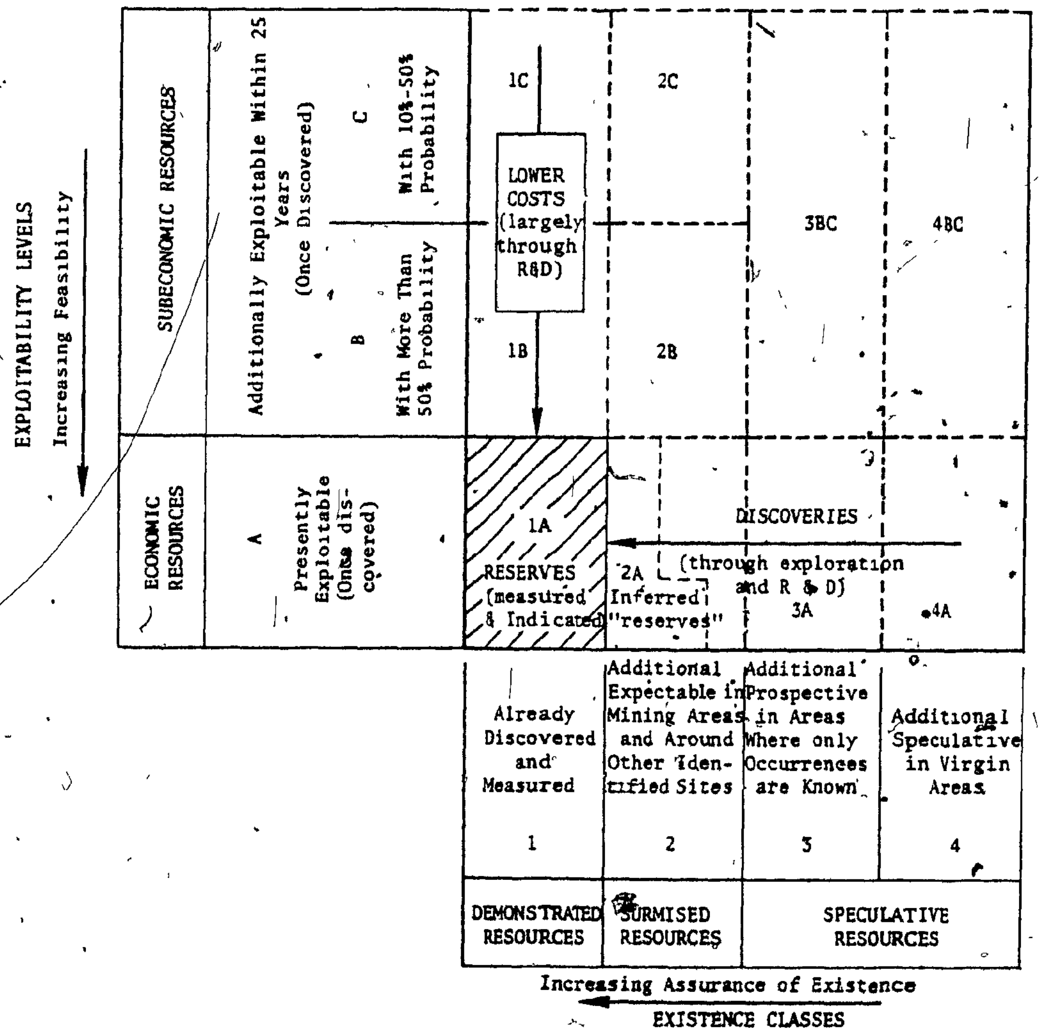
[1]See Zwartendyk (1975).

[2]See Figure 8.

[3]See Figure 9.

FIGURE 8

DEPARTMENTAL RESOURCE CLASSIFICATION SCHEME
ENERGY, MINES AND RESOURCES, OTTAWA
(AFTER ZWARTENDYK, 1975)

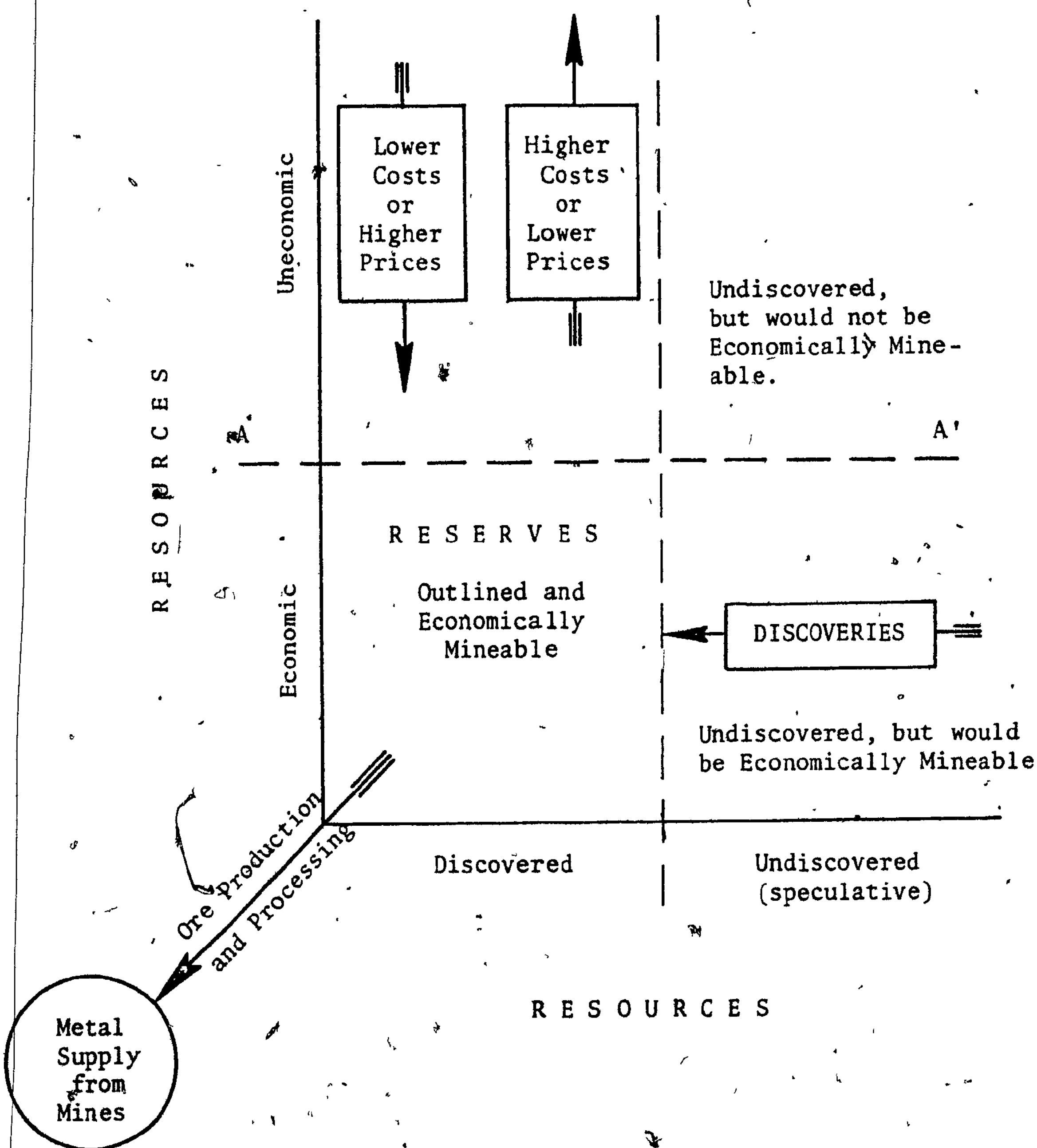



RESERVES (measured & indicated) = 1A (i.e., demonstrated economic resources)
RESOURCES = RESERVES plus all other numbered areas
RESOURCE BASE = RESOURCES plus indefinite area beyond top of diagram

<u>Note:</u> It has been found impossible in practice to make distinctions between 3B and 3C, and between 4B and 4C.

FIGURE 9

A RESOURCE CLASSIFICATION SCHEME
(AFTER AZIS, ET AL., 1977)



view, this will be of direct benefit to the mining companies, particularly within a mining district, and to government and corporate economists in their long range planning. If ore forming processes are similar in both space and time as indicated by geological studies, it should be reasonable to expect that different categories of resource will behave in a similar manner. The characteristics of an ore deposit could be thus predicted as reserves deplete. Mining companies perceive reserves differently, and what they report as reserves is based on practical planning problems, and the company policy on reporting in a competitive business environment.

Under the existing conditions, the most reliable estimates of known resources in Rouyn-Noranda region are the production and reserves[1] figures. These are referred to as the known endowment in this study, and are the basis of all forecasts made using multivariate statistical methods.

## 4.7 Distribution Characteristics of Data

Most of the general studies in statistics have dealt with normal data, and consequently, most techniques and in particular, tests of significance, have been developed for the normal distribution. Natural data, however, are skewed and

[1]The reserves are as reported by the companies and not necessarily in conformity with any of the classifications presented in this section.

this is especially true of geological data. In the present case both the measured and compiled data are positively skewed, the degree of skewness being different for each variable. There is also a tendency towards dichotomy because of a large number of zeros in the data.

Most statistical methods of data analysis require that the observations conform to a normal distribution. Transformations may therefore have to be made to normalize the data. Most commonly the transformation is logarithmic or of the $X^{1/N}$ type, particularly, the square-root member of the family. Jöreskog et al (1976) provide a review of various transformations.

Harris (1965) used a number of transformations to reduce the skewness of his data. The 400 square mile cell size chosen by Harris makes this possible because of a variety of geology present in such large sized cells. Harris also used factor scores instead of the raw variables because being uncorrelated, they are likely to be more normally distributed.

Agterberg, et al. (1972), circumvented the problem of skewness by converting the data base into a dichotomous form, coding a variable equal to one if present in a cell and zero if absent.

In this study, the multivariate techniques used are multiple regression analysis, discriminant analysis and factor analysis. Regression analysis does not require normally distributed data for explanatory variables. However, normality is one of the assumptions in factor analysis and discriminant analysis. The objective of factor analysis in this study is to obtain insight into the structure of the data base so that the most relevant variables could be selected for regression and discriminant analysis. And discriminant analysis is used here to compare the technique with regression analysis.

A series of transformations was therefore attempted on the data base, but because of the presence of a large number of zeros, no significant improvement in normalizing is obtained. The large number of zeros result from from the relatively small cell size used in the study. However, an increase in the cell size while incorporating a greater geological variability will proportionately reduce the number of cells because the total regional area is fixed at 400 square miles. On the other hand, converting the data base into a dichotomous form will fail to give the necessary weightage to individual variables relative to their areas or lengths.

It is felt that while transformation may help make the data base more amenable to tests of significance, particularly in discriminant analysis, an artificial barrier is created between the experimenter and the technique. Geological data are unique in the sense that they are of both evolutionary and interruptive nature. They are evolutionary in that various rock types evolve through the process of magmatic differentiation. They are interruptive in that different cycles may be involved in the evolutionary process. They are also interruptive in that igneous intrusives cut across the existing rock formations. But the data base is measured at a single point in time. Therefore, when using quantitative techniques it is necessary to isolate and remove any spurious contributions to the model from strictly spatially correlated variables. When the information base is as well developed as in the present study of the Rouyn-Noranda region, it is essential to incorporate the role of the relevant variables as fully as possible and to observe their relative contributions. These contributions should conform to accepted geological theories on ore genesis. For these reasons, raw data base is used in the analyses and the possible effects of violating the normality assumption discussed in the appropriate cases.

The models are validated by the "leaving one out" method[1] in which the known endowment cells, one at a time are assumed to have no endowment and their value predicted on the basis of remaining cells. The results obtained indicate that the techniques used have the robustness to accommodate violation of the normality assumption.

A number of statistical problems such as multicollinearity, resulting from the peculiar nature of geological data are discussed under regression and discriminant analyses

---

[1]The leaving one out method is demonstrated in Chapters 8 and 9.

and the approaches described to minimize their effects.

## 4.8 Reduction of Data Dimensionality

The objective of data reduction is to achieve an optimal balance between simplicity for comprehension and interpretation, and the desired level of relevant detail for adequate representation. The need for data reduction is necessary in the present study because there are a total of 67 explanatory variables, some of which do not warrant inclusion because they are shown to have no apparent pertinence to the formation of ore deposits. Their inclusion, in addition to increasing computing costs, can also cloud significant relationships. Some of the variables may be so highly correlated that only one or two of them may give sufficient representation.

Gnanadesikan (1977, p. 6) gives the following conditions that may require reducing dimensionality of multivariate data:

> (1) Exploratory situations in data analysis especially when there is ignorance of what is important in the measurement planning. Here one may want to screen out redundant coordinates or to find more insightful ones as a preliminary step to further analysis or data collection.
>
> (2) Cases in which one hopes to stabilize 'scales' of measurement when a similar property is described by each of several coordinates. Here the aim is to compound the various measurements into a fewer number which may exhibit more stable statistical properties.

(3) The compounding of multiple information as an aid in significance assessment. Specifically, one may hope that small departures from null conditions may be evidenced on each of several jointly observed responses. Then one may try to integrate these noncentralities into a smaller dimensional space wherein their existence might be more sensitively indicated. One such technique that has received some usage is the univariate analysis of variance applied to principle components.

(4) The preliminary specification of a space that is to be used as a basis for eventual discrimination or classification procedures.

(5) Situations in which one is interested in the detection of possible dependencies among observations in high-dimensional space.

When the data are geological, and the geology incompletely resolved, it becomes necessary to check the role of all but those variables whose insignificance is without dispute. Such a situation does exist in the Rouyn-Noranda region even though it has been intensely studied. The same is the case with all mining regions, for geological observations are but indirect evidence of the actual geological processes.

The dimensionality of multivariate data can be reduced by correlation analysis, factor analysis, characteristic analysis, or simply by trial and error based on the perceived significance of individual variables. All these approaches are made use of in this study.

## 4.9 Computational Procedures Used

All computer runs in factor analysis, regression analysis and discriminant function analysis were made using the standard S.P.S.S.[1] (Version 6) programs on the I.B.M. 360 computer at McGill University. A number of regression runs was also made on the C.D.C. 6400 computer at the Department of Energy, Mines and Resources, Ottawa.

---

[1] Nie, N.H., et al. (1975).

# CHAPTER 5

## THEORETICAL METAL ENDOWMENT

Theoretical metal endowment refers to the total endowment, both known and unknown, and whether economic or not. Such an endowment can never be fully known. However, a rough estimate of the order-of-magnitude can be attempted on the basis of the crustal abundance of elements. The implicit assumption here is that the geologic processes that created endowment were highly efficient in extracting metal wealth in accordance with physico-chemical laws. Since crustal abundance is only one of the factors that led to the concentration of endowment, and since other factors relating to transportation and deposition are not known, the estimate will be crude. But in relating what is known to what was theoretically produced, a rough measure can be obtained of how much additional endowment could be expected if the post-mineralization geological processes did not in part, or in full, destroy it.

Goodwin (1965) has noted that the metal content in the volcanic complexes[1] represents integral products of the volcanic cycles and migrated from the parent source to the

---

[1]Of which, Rouyn-Noranda region is one (Goodwin and Ridler, 1970).

volcanic environment in association with the differentiated silicic volcanic rocks; and Krauskopf (1967) has emphasized that sufficient sulphur is present in normal igneous rocks to generate ores from reasonable volumes of rocks. Under the assumptions made therefore, it should not be unreasonable to make estimates of the theoretical endowment in the region.

The average composition of the earth's crust has been estimated by a number of workers in terms of major and trace element contents. As part of the United States Geological Survey program on the data of geochemistry, Parker (1967) compiled this information, sources of which are referred to in his paper. Table 2 summarizes his trace element estimates for copper, zinc, and sulphur in selected igneous rocks and for the crust as a whole.

Shaw, Dostal and Keays (1976) have made estimates of the trace element composition of the Canadian Shield. Their estimates of copper at 14 ppm., and of zinc at 52 ppm. appear to be low when compared to the crustal abundance estimates compiled by Parker. However, since these figures relate to the Canadian Shield[1] which includes the present study area, they are more pertinent for making estimates of copper and zinc endowment in the Rouyn-Noranda region. The proviso is that the trace element estimates adequately represent the

[1] The shield is assumed to be homogeneous for estimates of the trace element abundance.

# TABLE 2

## TRACE ELEMENT ABUNDANCE OF COPPER, ZINC, AND SULPHUR (AFTER PARKER, 1967)

| Rock Types | Copper (ppm)[1] | Zinc (ppm) | Sulphur (ppm) |
|---|---|---|---|
| Ultramafics | 87-100 | 105-130 | 300 |
| Intermediate | 35 | 72 | 200 |
| Felsic granites and granodiorites | 20 | 60 | 400 |
| High calcium granites | 30 | 60 | 300 |
| Low calcium granites | 10 | 39 | 300 |
| Syenites | 5 | 390 | 300 |
| Average for igneous rocks | 55-100 | 40-111 | 520 |
| Average for the earth's crust | 45-55 | 65-83 | 260-520 |

[1]Parts per million.

composition of the original material now solidified as rock.

Using abundance estimates of Shaw, et al., the contained copper metal per cubic mile in the Canadian Shield is $1.73 \times 10^5$ short tons, and of zinc is $6.44 \times 10^5$ short tons. For a maximum feasible mining depth of one mile, and for the 400 square-mile Rouyn-Noranda region, the estimate of contained copper is 69.2 million short tons, and of zinc is 257.6 million short tons.

The known[1] metal endowment of copper and zinc in the region is 2.24 and 1.16 million short tons respectively. In other words, the theoretical estimate of copper exceeds its known endowment by a multiple of 31. Similarly, the theoretical estimate of zinc exceeds its known endowment by a multiple of 222.

As the theoretical endowment figures indicate, the region should contain seven times more zinc than copper. However, since the known endowment of copper is almost twice as much as that of zinc, it should be reasonable to conclude that there is far more potential for zinc in the region than for copper.

The question one may ask is that if indeed more deposits than presently known were concentrated from the

---

[1]This is only the economic endowment, for no estimates are available for the uneconomic endowment.

inherent availability of both sulphur, and, copper and zinc, then how many of them were able to survive the effects of deformation, metamorphism and erosion? It is not possible to isolate and identify individual factors that helped cause or destroy ore concentrations. Most probably the factors were acting jointly. The only approach to understanding the situation would be in the Huttonian concept of surmising causes from the observed effects. This can be done by quantitatively relating known ore deposits to their surrounding geological environment and by applying this relationship in reverse to predict possible locations of any additional prospective deposits. This is the aim of the present study.

# CHAPTER 6

## CHARACTERISTIC ANALYSIS OF DATA

### 6.1 General Statement

The technique of characteristic analysis was described in Section 2.2.5. Botbol (1971), and DeGeoffroy and Wignall (1972) applied characteristic analysis to determine the most commonly present geological features in a large number of ore deposits of a particular type for possible use in mineral exploration.

The focus of the present study is a single mining region, the ore deposits in which are the result of geological processes in a self-contained volcanogenic unit that created massive sulphide mineralization. If all the ore-forming processes were concentrated within this unit, then a crude measure of their relative importance can be obtained by the relative proportion of the characteristics present provided that accepted geological concepts are not violated. For example, it would not be valid to draw conclusions from areal measurements of post-ore sedimentary processes. However, since it is the joint occurrence of geological processes that results in ore formation, characteristic analysis should give a more reliable estimate of the significance of a characteristic than measuring its simple presence in a certain proportion as compared to the other characteristics.

## 6.2 Characteristic Analysis Results

Characteristic analysis was carried out separately for the following sets of data as measured in each of the 64 cells in the study region.

(1) Areas of geological formations;

(2) Contact lengths between formations;

(3) Structural elements: synforms, antiforms, dykes, and faults.

Since similar types of variables are present in each set, a better comprehension of their relative importance can be made. Also, the typicality obtained for each characteristic is converted into a ratio expressed as the percentage of the total typicalities in that set. These percentages are called "relative typicalities". The data used in the analysis are in binary code, i.e., a value of one is assigned if a variable is present in a cell, and zero if not.

The relative typicalities of the areas of geological formations, ranked in a descending order are shown in Table 3. By inspection, the variables may be subdivided into three groups as based on their relative typicalities shown by the broken lines.

The top two groups include essentially igneous rocks

TABLE 3

RELATIVE TYPICALITIES OF AREAS OF GEOLOGICAL FORMATIONS

| Rank | Variable | Relative Typicality |
|---|---|---|
| 1 | AREA 4 | 17.55 |
| 2 | AREA 3 | 16.58 |
| 3 | AREA 8 | 15.73 |
| 4 | AREA 12 | 14.91 |
| 5 | AREA 11 | 9.19 |
| 6 | AREA 2 | 6.90 |
| 7 | AREA 6 | 4.12 |
| 8 | AREA 5 | 3.99 |
| 9 | AREA 10 | 3.05 |
| 10 | AREA 1 | 2.82 |
| 11 | AREA 13 | 2.37 |
| 12 | AREA 9 | 1.79 |
| 13 | AREA 7 | 0.98 |

[1]See page 58 for description of variable names.

the only exception being AREA 12, which represents the areas of lakes.[1] Of these, the first group includes the volcanics andesite, basalt (AREA 4), and rhyolite (AREA 3), and diorite, gabbro (AREA 8). The diorite-gabbro formation is believed to be an intrusive equivalent of the volcanics andesite-basalt. The first group thus represents a typical calcalkaline assemblage.

The second group includes granite, granodiorite (AREA 11), and tuff, agglomerate (AREA 2). Granite, tuff, and agglomerate are essentially chemical equivalents of rhyolite (AREA 3).

The formations in the last group are either metasedimentary rocks, or minor igneous intrusions with apparently no genetic relation to the ore deposits in the region.

Table 4 shows the relative typicalities of contact lengths between geological formations. The breaks in typicalities are shown by broken lines. Since the contact lengths are a direct function of the joint occurrence of formations, the rankings obtained in this case, therefore, correspond with those obtained for areas of formations.[2] The most commonly present contacts are between pairs of the following formations:

---

[1] In geological terms, lakes are a recent phenomenon, and in the Canadian Shield are a result of widespread glaciation. The variable is ranked high because of the common occurrence of lakes with most rock types in the region.

[2] See Table 3.

# TABLE 4

## RELATIVE TYPICALITIES OF CONTACT LENGTHS BETWEEN FORMATIONS

| Rank | Variable[1] | Relative Typicality |
|---|---|---|
| 1 | CNTL 11 | 14.14 |
| 2 | CNTL 19 | 13.62 |
| 3 | CNTL 13 | 13.04 |
| 4 | CNTL 22 | 6.34 |
| 5 | CNTL 7 | 5.69 |
| 6 | CNTL 15 | 5.65 |
| 7 | CNTL 9 | 5.18 |
| 8 | CNTL 6 | 4.47 |
| 9 | CNTL 34 | 3.50 |
| 10 | CNTL 16 | 3.18 |
| 11 | CNTL 24 | 1.81 |
| 12 | CNTL 21 | 1.80 |
| 13 | CNTL 14 | 1.48 |
| 14 | CNTL 17 | 1.46 |
| 15 | CNTL 33 | 1.34 |
| 16 | CNTL 2 | 1.29 |
| 17 | CNTL 38 | 1.18 |
| 18 | CNTL 30 | 1.14 |
| 19 | CNTL 35 | 1.13 |
| 20 | CNTL 23 | 1.12 |
| 21 | CNTL 20 | 1.05 |
| 22 | CNTL 1 | 1.03 |
| 23 | CNTL 28 | 0.91 |
| 24 | CNTL 36 | 0.88 |
| 25 | CNTL 5 | 0.84 |
| 26 | CNTL 10 | 0.84 |
| 27 | CNTL 32 | 0.82 |
| 28 | CNTL 27 | 0.79 |
| 29 | CNTL 12 | 0.77 |
| 30 | CNTL 37 | 0.69 |
| 31 | CNTL 8 | 0.47 |
| 32 | CNTL 29 | 0.45 |
| 33 | CNTL 25 | 0.45 |
| 34 | CNTL 4 | 0.35 |
| 35 | CNTL 18 | 0.32 |
| 36 | CNTL 26 | 0.32 |
| 37 | CNTL 3 | 0.22 |
| 38 | CNTL 31 | 0.22 |

[1]See page 59 for description of variable names.

| | |
|---|---|
| andesite, basalt | (AREA 4) |
| rhyolite | (AREA 3) |
| diorite, gabbro | (AREA 8) |
| granite, granodiorite | (AREA 11) |
| tuff, agglomerate | (AREA 2) |

The remaining contact lengths are sparsely distributed and do not appear to be genetically involved in the ore forming processes.

Table 5 ranks the relative typicalities of structural elements, i.e., synforms, antiforms, dykes, and faults. There do not appear to be any sharp breaks in the rankings. Overall, 36 percent of the structural elements lie in directions ranging from east-west to north-east, 27 percent in directions north-south to north-west, 20 percent in north-east to north-south, and the remaining 17 percent in north-west to east-west. There are no folds present in directions north-east to north-south.

It is stated in Section 3.2.2 that structural elements and volcanism appear to be closely related. However, the role of the structural features in ore formation has not been resolved for the region.

## 6.3 Review of Results

Characteristic analysis has been developed for application to a large number of mineral deposits or mining districts.

# TABLE 5

## RELATIVE TYPICALITIES OF STRUCTURAL ELEMENTS

| Rank | Variable | Relative Typicality |
|---|---|---|
| 1 | FOLT 1 | 19.11 |
| 2 | DYKE 2 | 11.46 |
| 3 | DYKE 1 | 11.36 |
| 4 | FOLT 3 | 11.27 |
| 5 | DYKE 3 | 10.35 |
| 6 | FOLT 2 | 8.20 |
| 7 | FOLT 4 | 5.25 |
| 8 | DYKE 4 | 4.87 |
| 9 | SNFM 4 | 4.09 |
| 10 | SNFM 1 | 3.02 |
| 11 | ANFM 4 | 2.87 |
| 12 | SNFM 3 | 2.80 |
| 13 | ANFM 1 | 2.72 |
| 14 | ANFM 3 | 2.63 |
| 15 | SNFM 2 | 0.00 |
| 16 | SNFM 2 | 0.00 |

[1]See page 61 for description of variable names.

In this study, the technique is used over a single mining region to highlight the most typical geological characteristics as an aid in data reduction. The technique serves the objective, and at the same time emphasizes the calcalkaline volcanogenic nature of the region and its structural attributes.

The volcanogenic nature of the region is well known from its lithology. What characteristic analysis does is to quantitatively express relationships that might otherwise not have been conspicuous. This it does by emphasizing the commonness of a characteristic rather than its actual measurement relative to other characteristics. For example, in the Rouyn-Noranda region, while the cumulated areas of diorite, gabbro (AREA 8), rhyolite (AREA 3), and andesite, basalt (AREA 4) are in the ratio of 1:2:4, their relative typicalities are in the ratio of 1.00:1.05:1.12, i.e., the three are almost equally significant. The analysis thus indicates that the role of rhyolite and diorite, gabbro, relative to that of andesite, basalt is more "typical" than shown by their respective areas. This is achieved by the binary coding of the variables which considers the presence rather than the actual measurement of a variable.

It is of interest to note that the most typical contact length determined by characteristic analysis is CNTL 11, i.e., between rhyolite and andesite, basalt. The contact length is important in that nearly all base metal deposits in the region occur at or near this contact. In this way, the technique can help in selecting the most relevant variables for further statistical study. This is true only if the relevance itself is a function of the joint commonness of characteristics. Such a situation does exist in the Rouyn-Noranda region which was a centre of volcanic activity, and in which all base metal deposits are directly related to that activity.

Characteristic analysis, lacks the probabilistic and predictive resolution needed for estimating the unknown mineral endowment of a region. It is more a mathematical manipulation than a statistical procedure.

# CHAPTER 7

## FACTOR ANALYSIS

### 7.1 General Statement

Factor analysis, as defined by Mather (1976) is the determination of a set of descriptive concepts which summarizes the relationships among the components of a system of interacting variables. The aim of the technique is to explain relationships among correlated variables in terms of a relatively few underlying factor variates, thus reducing the dimensionality of the problem for more incisive interpretation.

Factor analysis is not a predictive tool in resource forecasting. However, when the basic postulate of resource evaluation is the interaction and integration of geological processes in ore formation, the technique becomes most useful in analyzing the apparent relationships existing between geological variables that indirectly are a measure of the processes themselves. The factor analyst must therefore, have some "a priori" knowledge of the system under study so that meaningless or misleading interpretations regarding "cause" and "effect" can be avoided. This is particularly true of geological data in which the roles of different age relations of variables, and of the unknown third dimension[1] are not

---

[1] I.e., depth.

fully resolved. It may seem paradoxical that an "a priori" knowledge is required to understand and interpret the outcome of factor analysis when the technique of factor analysis itself is supposed to identify fundamental and meaningful associations amongst the variables. Actually, the outcome of factor analysis is a re-expression of the information content of the data in a manner that highlights previously unsuspected relationships. The identification of these relationships is only possible against a background of known geological criteria. The greater the level of geological information, the better the insight obtained.

## 7.2 Methodology

Lawley and Maxwell (1971) provide perhaps the best mathematical treatment of factor analysis. Gnanadesikan (1977), Mather (1976), Overall and Klett (1972), and Cooley and Lohnes (1962) discuss the technique in general terms. From a geological point of view, the technique is well described in Jöreskog *et al* (1976), and Davis (1973). The following review has been prepared from these references.

Factor analysis methods always employ principal component analysis as the starting point. In principal component analysis, a set of p variates, generally called $x_1, x_2, \ldots x_p$, is linearly and orthogonally transformed into an equal number of new variates $y_1, y_2, \ldots y_p$, that are all uncorrelated. These are selected such that $y_1$ has maximum variance, $y_2$ has maximum variance at the same time being uncorrelated with $y_1$, and so on. The objective is to find a minimum number of independent components that will account for most of the variance in the original set of variates.

While principal component analysis is variance oriented, factor analysis is covariance oriented, that is, it interprets the structure within the variance-covariance matrix of the data. Principal components are in fact the eigenvectors of this variance-covariance matrix.

In factor analysis, the basic assumption is that

$$x_i = \sum_{r=1}^{k} \lambda_{ir} f_r + e_i \qquad (i = 1, 2, \ldots p)$$

where, $x_i$ are the p original variates, $f_r$ is the rth common factor, k is the specified number of factors, and $e_i$ is a random residual variable affecting $x_i$. The coefficient $\lambda_{ir}$ is the loading of ith variate on the rth factor.[1]

Assuming a multivariate normal distribution, the p x p matrix of variances and covariances will include as its

[1]A factor is a vector weighted in proportion to the amount of the total variance which it represents. The elements in the factor are called its loadings. Factor scores are measurements of a factor, defined as the weighted combination of several original variables.

diagonal elements the following variances:

$$VAR_{ii} = \sum_{i=1}^{k} \lambda_{ir}^{2} + \text{Var } e_i \quad \text{(where } k < p)$$

the off-diagonal elements, i.e., the covariances will be:

$$COV_{iq} = \sum_{i=1}^{k} \lambda_{ir} \times \lambda_{iq}$$

where $\lambda_{ir}$ is the ith measurement of variable r, and $\lambda_{iq}$ is the ith measurement of variable q.

The resulting matrix of variances and covariances [VC] is equal to the product of a p x k matrix of factor loadings [FL] multiplied by its transpose plus a p x p matrix of unique variances [var $e_i$], which accounts for the variance not included when the summation is done from 1 to k instead of 1 to p, and where k, the number of factors is less than p, the number of variates. When k and p are equal, the result is the same as that given by principal component analysis.

Eigenvalues and eigenvectors are then calculated for the standardized variance-covariance matrix.[1] However, the eigenvectors must be normalized so that they define a vector of unit length. This is done by simply dividing each eigenvector by the square root of the summation of the squares of

[1]This becomes the correlation matrix because of standardization.

the eigenvectors. Multiplying each normalized vector by the square root of the associated eigenvalue results in a factor vector. Arranging the elements of a factor vector in matrix form gives the factor matrix, a matrix which contains the coefficients of relationship between the original variables and the derived factor variates.

While the dimensionality of the problem is reduced by factor analysis, more meaningful results can be obtained by factor rotation so that high loadings are obtained for a few variables, and the rest of the loadings in a factor are low. This is the varimax rotation solution. The final solution has the form:

$$[Z] = [T] \times [FL]$$

where [FL] is the original matrix of factor loading, [T] is a non-singular transformation matrix,[1] and [Z] is the matrix resulting from varimax rotation. According to Cooley and Lohnes (1962), the varimax solution has the advantage that the resulting factors tend to be invariant under changes in the composition of the test battery, i.e., small changes in the sample of tests should not affect the basic inferences drawn. Such a procedure is used in this study.

The desirable properties of a good factor solution, after Overall and Klett (1972, p. 90) include:

---

[1]Which causes the orthogonal rotation because $[T]' \times [T] = [I]$, the identity matrix.

(1) Parsimony;

(2) Orthogonality, or at least relative independence;

(3) Conceptual meaningfulness.

That is, a lower number of factors should explain most of the variance and each factor should be independent, representing a unique source of variation. If the above properties are not obtained in the final solution, then it is likely that factor analysis is not a suitable model.

Before factor analyzing a set of data, the following aspects of variables should be evaluated as discussed by Mather (1976, p. 242):

(a) The type of relationship existing among the variables. Factor analysis is concerned with linear relationship and deviation from this assumption can effect results in a manner difficult to predict.

(b) The number of factors to be expected. The implication is that the factor analyst has some insight into the probable nature of the factors, and can predict the number of factors. One way is to extract all possible factors and then decide the number to be retained. A more practical way is to retain all factors having an eigenvalue greater than one, i.e., to retain those factors containing a greater variance than the original standardized data. This approach is adopted in the present study.

(c) <u>The nature of factors to be expected.</u> This means that "a priori" knowledge of the geological processes in the area is necessary. Obviously, something must be wrong if two highly antithetic geological variables load significantly on the same factor.

(d) <u>The variables to be included.</u> Here again, "a priori" knowledge of the geological processes is necessary. If the objective is to analyze all possible geological relationships in the area, then no pre-selection is required. In the present case, the objective is to observe how mineral endowment in Rouyn-Noranda region relates to geological variables. There is thus no reason to include post-ore geological aspects, and therefore, all areas of sedimentary rock formations and their contact lengths can be excluded. This leaves the igneous rocks, the volcanics and the later intrusives for analysis. Since there is some evidence that the later intrusives may have been a part of the original volcanic processes, they have been included as variables. And finally, all structural elements in the region are also retained. There is no question that dykes and faults are later features than the volcanism with which ore formation was associated. But synforms and antiforms are believed to be related to volcanism,[1] and at least some of the faults and dykes may have been a consequence of this folding. Therefore, no preselection is done in case of structural elements. The variables used are listed in Table 6.

(e) The inter-factor relationships. Since an orthogonality of factor is desired in this study, the varimax rotation method is used. The factors, therefore are believed to be free of any correlation among them.

The decision on the number of factors to be retained is an arbitrary one. Most commonly all factors having an eigenvalue greater than one are retained since they contain a greater variance than the original standardized variables. However, Overall and Klett note that factors defined by three or more variables having loadings in excess of 0.35 have been found in their experience to be both stable and replicable. They also state that statistical data reduction is usually considered to be adequate and effective when the number of factors is approximately one fourth the number of original

[1]See Section 3.2.2.

variables, and the variance accounted for is 50 to 75 percent of the total variance. This range is acceptable particularly when the objective of factor analysis is to explain the correlations among variables in terms of a minimum number of factors.

## 7.3 Factor Analysis of Data

### 7.3.1 Variables Analyzed

Factor analysis is performed on three sets of 38 variables each. The first two sets include, respectively, the contained metal tonnage of copper and zinc as a variable. The third set uses the dollar value of cumulated copper and zinc tonnages as a variable. The remaining 37 variables are the same in each of the three sets. Table 6 shows the variables used.

Two procedures for factor analysis are applied:

(i) keeping the diagonal elements[1] of the correlation matrix at one, and

(ii) replacing these variances by the communality estimates of the variables followed by varimax rotation.

[1] I.e., variances.

TABLE 6

VARIABLES USED IN FACTOR ANALYSIS

| No. | Variable Name | Variable Description |
|---|---|---|
| 1 | (Value) | Contained tonnage of copper, or zinc, or their total dollar value |
| 2 | AREA 2 | Area of formation: tuff agglomerate |
| 3 | AREA 3 | Area of formation: rhyolite |
| 4 | AREA 4 | Area of formation: andesite, basalt |
| 5 | AREA 7 | Area of formation: peridotite |
| 6 | AREA 8 | Area of formation: diorite, gabbro |
| 7 | AREA 9 | Area of formation: rhyolite porphyry |
| 8 | AREA 10 | Area of formation: syenite, monzonite |
| 9 | AREA 11 | Area of formation: granite, granodiorite |
| 10 | CNTL 6 | Contact length between: AREA 2 & AREA 3 |
| 11 | CNTL 7 | Contact length between: AREA 2 & AREA 4 |
| 12 | CNTL 9 | Contact length between: AREA 2 & AREA 8 |
| 13 | CNTL 10 | Contact length between: AREA 2 & AREA 11 |
| 14 | CNTL 11 | Contact length between: AREA 3 & AREA 4 |
| 15 | CNTL 13 | Contact length between: AREA 3 & AREA 8 |
| 16 | CNTL 14 | Contact length between: AREA 3 & AREA 9 |
| 17 | CNTL 15 | Contact length between: AREA 3 & AREA 11 |
| 18 | CNTL 18 | Contact length between: AREA 4 & AREA 7 |
| 19 | CNTL 19 | Contact length between: AREA 4 & AREA 8 |
| 20 | CNTL 20 | Contact length between: AREA 4 & AREA 9 |

TABLE 6

(CONTINUED)

| | | |
|---|---|---|
| 21 | CNTL 22 | Contact length between: AREA 4 & AREA 11 |
| 22 | CNTL 34 | Contact length between: AREA 8 & AREA 11 |
| 23 | SNFM 1 | Synform length EW to NE |
| 24 | SNFM 2 | Synform length NE to NS |
| 25 | SNFM 3 | Synform length NS to NW |
| 26 | SNFM 4 | Synform length NW to EW |
| 27 | ANFM 1 | Antiform length EW to NE |
| 28 | ANFM 2 | Antiform length NE to NS |
| 29 | ANFM 3 | Antiform length NS to NW |
| 30 | ANFM 4 | Antiform length NW to EW |
| 31 | DYKE 1 | Dyke length EW to NE |
| 32 | DYKE 2 | Dyke length NE to NS |
| 33 | DYKE 3 | Dyke length NS to NW |
| 34 | DYKE 4 | Dyke length NW to EW |
| 35 | FOLT 1 | Fault length EW to NE |
| 36 | FOLT 2 | Fault length NE to NS |
| 37 | FOLT 3 | Fault length NS to NW |
| 38 | FOLT 4 | Fault length NW to EW |

In the latter case, an iterative procedure is used after replacing the variances by the communality estimates by extracting the same number of factors by re-factor analyzing, and replacing the communality estimate by the newer, improved estimate. The iteration continues until the difference between two successive communality estimates are negligible, or if after a particular iteration, any one or more of the communalities exceeds one. The results obtained by the two procedures give essentially similar insights into the relationships between the variables. The second approach, however, results in a greater parsimony, and therefore, the analysis of results is based on this approach.

Factor loadings for copper and zinc sets are shown graphically in Figures 10 and 11. The variables shown on the diagrams are all positively[1] loaded on their respective factors, each with a value greater than 0.20. This is an arbitrary decision to avoid crowding of the diagrams with non-significant variables. The suggestion by Overall and Klett that 0.35 gives better results when three or more factor loadings of at least this value are present is incorporated in the discussion. The 0.35 threshold is marked on the diagrams with broken horizontal lines. In line with the objective of this study, the factors considered to be important are those

---

[1]The exceptions are copper and zinc. These are shown on the diagrams whether they show positive or negative loadings.

on which either copper or zinc loads most heavily.

7.3.2 The Copper Set

A total of 14 factors are extracted on factor analyzing[1] the copper set. Of these 14, nine have eigenvalues in excess of one, and cumulatively account for 85.3 percent of the variance present. Table 7A shows the eigenvalues associated with each of the 14 factors, and the variances explained by them individually and cumulatively. The following is a review of the associations of variables as they load on individual factors shown in Figure 10.

Factor #1 contains variables predominantly of the tuff, agglomerate type. There is a positive but insignificant amount of copper associated with this set of variables.

Factor #2 combines predominantly rhyolite porphyry variables with a negative loading of copper. This is an evidence of the antithetic relationship existing between the two.

Factor #3 is an important factor because of the high positive value of copper. The variables associated with it which have a value higher than 0.35 are the following:

[1]Using the varimax rotation method.

# TABLE 7A

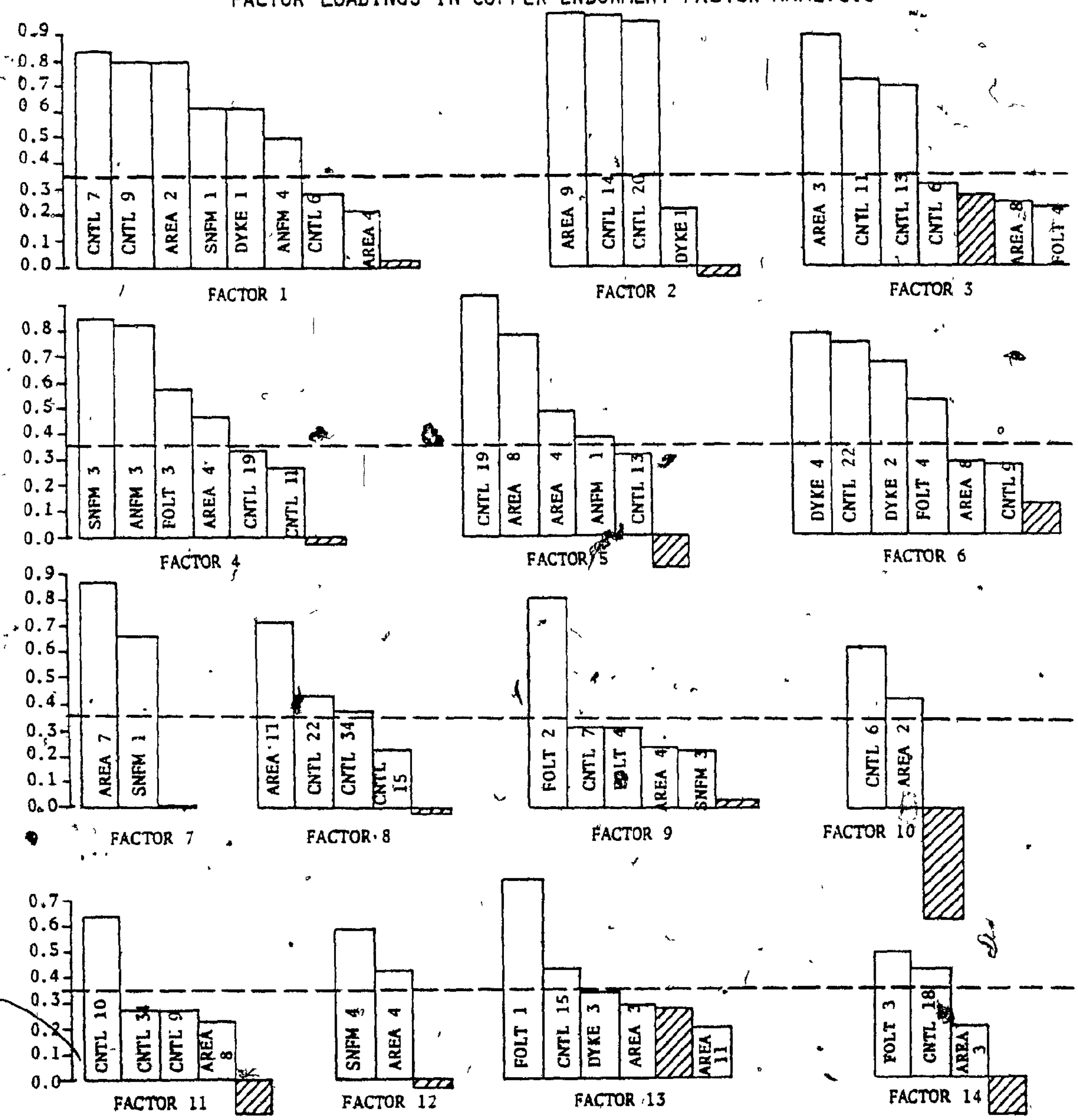
## EIGENVALUES ASSOCIATED WITH COPPER SET FACTORS

| Factor | Eigenvalue | PCT of Var | CUM PCT |
|---|---|---|---|
| 1 | 4.238 | 16.7 | 16.7 |
| 2 | 3.701 | 14.6 | 31.2 |
| 3 | 3.155 | 12.4 | 43.7 |
| 4 | 2.580 | 10.2 | 53.8 |
| 5 | 2.075 | 8.2 | 62.0 |
| 6 | 1.857 | 7.3 | 69.3 |
| 7 | 1.453 | 5.7 | 75.0 |
| 8 | 1.398 | 5.5 | 80.5 |
| 9 | 1.213 | 4.8 | 85.3 |
| 10 | 0.945 | 3.7 | 89.0 |
| 11 | 0.836 | 3.3 | 92.3 |
| 12 | 0.707 | 2.8 | 95.1 |
| 13 | 0.645 | 2.5 | 97.6 |
| 14 | 0.603 | 2.4 | 100.0 |

[1]Using the varimax rotation method.

FIGURE 10

FACTOR LOADINGS IN COPPER ENDOWMENT FACTOR ANALYSIS[1]



[1]Hatched areas show copper loadings.

- area of rhyolite (AREA 3);
- contact length between rhyolite and andesite, basalt (CNTL 11);
- contact length between rhyolite and diorite, gabbro (CNTL 13).

The remaining variables are:

- contact length between tuff, agglomerate, and rhyolite (CNTL 6);
- area of diorite, gabbro (AREA 8);
- faults lying in directions NW to EW (FOLT 4).

All copper ore deposits in the region occur in rhyolites at or near the rhyolite--andesite, basalt contact. These are the two most important relationships shown in factor #3. However, this factor indicates the rhyolite--diorite, gabbro contact to be an important one also. The significance of this variable is not clear in terms of ore occurrence. The explanation may, however, lie in the belief that diorite, gabbro, and andesite, basalt, may be genetically related.[1] The same comments must apply to the positive association of the areas of diorite, gabbro, with copper. The only structural element associated with this factor is the length of of faults lying NW to EW.

[1]See Section 3.2.2.

Factor #4 includes the structural elements synforms, antiforms and faults, all lying in directions NS to NW. Also present are, the area of andesite, basalt, and the contact lengths of this formation with diorite, gabbro, and with rhyolite. The close association of folding, faulting, and the volcanics andesite and basalt, give credence to the belief[2] that volcanism and structural deformation in the region were related and were probably coeval. The negative association of copper with this factor discounts its economic potential.

Factor #5 essentially includes diorite, gabbro, and andesite, basalt, and their associations. The possibility of a genetic association between these rocks in the region has been mentioned above. Copper is strongly antithetic with this association of variables.

Factor #6 is structurally oriented, and includes dykes and faults lying NW to EW and dykes lying NE to NS. Also included are, the contact length between andesite, basalt, and granite, granodiorite, area of diorite, gabbro, and the contact between tuff, agglomerate, and diorite, gabbro. The factor has a moderate positive loading of copper. However, the associations are not clear because the role of structural elements has not been satisfactorily resolved in the region. But the joint presence of dykes, faults and the intrusives, diorite, gabbro, and granite, granodiorite, and

---

[2] See Section 3.2.2.

a positive loading of copper tempt to invoke a hydrothermal explanation.

Factor #8 has loading of mainly granite, granodiorite, with a negative association of copper. Since all the base metal massive sulphide deposits in the region occur in the volcanics, the negative loading of copper is understandable.

Factor #9 has a heterogeneous set of variables in it. However, only one variable has a value greater than 0.35, and therefore, the factor cannot be considered as stable or significant.

Factor #10 shows a strong antithetic relationship of copper with tuff, agglomerate. Such is the observation in the field also.

Factors #11, 12 and 14, all show associations with which copper has negative affiliations. They are, therefore, not significant from an economic point of view.

Factor #13 includes fault length lying EW to NE, contact length between rhyolite and granite, granodiorite, dyke length lying NS to NW, and areas of rhyolite, and of granite, granodiorite. Despite the rather high positive association of copper with it, the factor is not important in that it has an eigenvalue of only 0.64 compared to 3.15 for factor #3 and 1.86 for factor #6.[1]

---

[1]Factors #3 and 6 are the only other factors that have high positive copper loadings.

In summary, copper shows a strong litho-stratigraphic association with the volcanics in factor #3, and is also associated with structurally dominated associations observed in factors #6 and 13. The selection of variables for resource potential evaluation of copper can be based on these three factors.

### 7.3.3 The Zinc Set

Of the 14 factors extracted in the zinc set, there are nine with eigenvalues in excess of one. These nine factors cumulatively account for 84.8 percent of the total variance. The eigenvalues for the 14 factors and the variances accounted for by them are shown in Table 7B. The association of rock types as represented by their loadings on different factors[1] runs parallel to those in the copper set. However, zinc loads significantly on two factors, numbers, 3 and 7 compared to three for copper. In addition, zinc loads positively, though non-significantly on seven additional factors compared to three more for copper. The indication is that zinc has an apparent relationship with a broader range of geological variables. It may also be possible that at the time of zinc mineralization, geological activity may have become more widespread. But this is conjectural and only warrants support in

[1]See Figure 11.

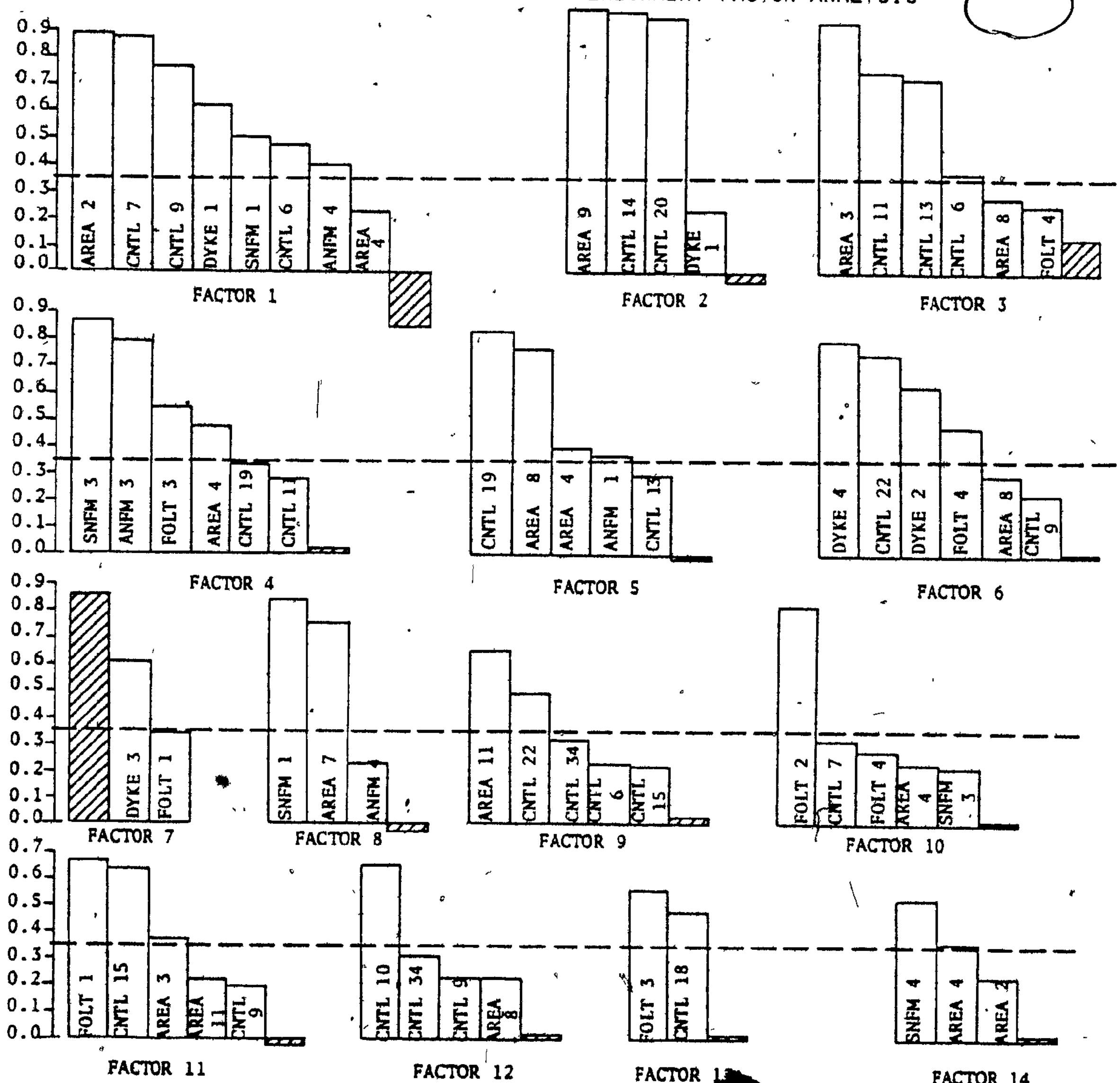
# TABLE 7B

## EIGENVALUES ASSOCIATED WITH ZINC SET FACTORS

| Factor | Eigenvalue | PCT of Var | CUM PCT |
|---|---|---|---|
| 1 | 4.218 | 16.6 | 16.6 |
| 2 | 3.698 | 14.5 | 31.1 |
| 3 | 3.149 | 12.4 | 43.4 |
| 4 | 2.566 | 10.1 | 53.5 |
| 5 | 2.070 | 8.1 | 61.6 |
| 6 | 1.838 | 7.2 | 68.9 |
| 7 | 1.513 | 5.9 | 74.8 |
| 8 | 1.293 | 5.1 | 79.9 |
| 9 | 1.258 | 4.9 | 84.8 |
| 10 | 0.944 | 3.7 | 88.5 |
| 11 | 0.881 | 3.5 | 92.0 |
| 12 | 0.743 | 2.9 | 94.9 |
| 13 | 0.662 | 2.6 | 97.5 |
| 14 | 0.636 | 2.5 | 100.0 |

FIGURE 11

FACTOR LOADINGS IN ZINC ENDOWMENT FACTOR ANALYSIS[1]




[1] Hatched areas show zinc loadings.

that, compared to copper, zinc always occurs at a higher strarigraphical level.

Factor #3 has a positive zinc loading with the following associations:

- area of rhyolite (AREA 3);
- contact length between rhyolite and andesite, basalt (CNTL 11);
- contact length between rhyolite and diorite, gabbro (CNTL 13);
- contact length between rhyolite and tuff, agglomerate (CNTL 6);
- area of diorite, gabbro (AREA 8);
- length of faults lying NW to EW (FOLT 4).

This set of variables is highly important from the point of view of zinc occurrence, and corresponds essentially with those in factor #3 of the copper set. This is an indication of similarity in ore-forming environments of zinc and copper.

Factor #7 shows a high loading of zinc with structural elements, dyke length lying NS to NW (DYKE 3), and with fault length lying EW to NE (FOLT 1). The role of these dykes and faults is not clear beyond their strong spatial correlation with zinc and to a lesser extent with copper. Being post-ore features, these structural elements cannot be the cause of ore formation. However, if they are a consequence of ore-forming processes, they can still be used as effective variables in further statistical analyses subject to the condition that other variables with accepted genetic affiliations are also included with them.

## 7.4 Summary

Factor analysis consists of extracting a parsimonious number of linear relationships from a set of data with the objective of obtaining an understanding of a complex of observed variables in terms of a few underlying factors. The following is a summary of the results obtained:

(1) The three most important stratigraphical associations of variables for copper and zinc are:

  (i) area of rhyolite (AREA 3);
  (ii) contact length between rhyolite and andesite, basalt (CNTL 11);
  (iii) contact length between rhyolite and diorite, gabbro (CNTL 13).

An additional association of zinc is the contact length between rhyolite, and tuff, agglomerate (CNTL 6). This may be related to the observation that zinc occurs at a higher stratigraphic level than copper.

(2) Both zinc and copper are closely associated with NS to NW dyke length (DYKE 3) and EW to NE fault length (FOLT 1). This may be only a spatial association, but cannot be ignored in subsequent statistical analyses.

(3) The associations determined by factor analysis may already be known on a subjective level. What factor analysis does is to provide a more quantitative understanding of the qualitative relationships so that the relative significance of each variable can be determined.

(4) The relationships determined by factor analysis may have a genetic or a spatial base. The distinction can best be made with "a priori" knowledge of the geological processes in the region. The greater this knowledge, the greater the comprehension of inter-relationships present. The fact that the relationships can be isolated and explained has a great significance in variable selection for further statistical analyses.

(5) Factor analysis brings out relationships in accordance with the information base applied. It can indicate previously unsuspected associations. It cannot, however, indicate any relationships not already known or measured. For example, if

either copper or zinc occurs in a different geological environment in the region, and if this occurrence is unknown, factor analysis will not help. But this is also the case with every other multivariate statistical technique.

(6) Factor analysis cannot be used directly in resource forecasting except in a very subjective manner. What it does is to give the factor analyst an additional dimension, an insight that is most essential for further statistical work, so that he can recognize and isolate spurious relationships and minimize their effect. Factor analysis enables the selection of a reduced number of more pertinent variables for an incisive multivariate analysis by other methods.

(7) Factor analysis should prove an excellent guide in mineral exploration, but only in regions where ore deposits are known and the associated geology studied. Instead of concentrating on any individual favourable feature of geology, the explorationist can have a multivariate comprehension based on factor loadings related with the particular endowment sought.

# CHAPTER 8

## REGRESSION ANALYSIS

### 8.1 Introduction

Geological processes represent a complex system in which optimal integration of favourable factors have resulted in ore formation. These processes cannot be duplicated, but the conditions of optimality can be studied by examining relationships between known ore deposits and associated geology. A mathematical function can be formulated to learn more about the underlying relationships and to appreciate the separate and joint effects produced by changes in the geological variables comprising the function. If the function is valid, it can be used as a predictive tool within the system under study, and for extrapolation to similar geological systems. Draper and Smith (1966, p. 2) stress that, "even where no sensible physical relationship exists between variables, we may wish to relate them by some sort of mathematical equation. While the equation might be physically meaningless, it may nevertheless be extremely valuable for predicting the values of some variables from knowledge of other variables, perhaps under certain stated conditions." The most common method of obtaining a mathematical function to summarize a mass of data is by the "least squares" method also called regression analysis.

The objective in the least squares method is to find the values of constants in the chosen function that will minimize the sum of squared deviations of the observed values from those predicted by the equation. The equation consists of two parts: $y$, the dependent or response variable, and $x_1, x_2, \ldots x_p$, the $p$ explanatory variables, also called factors. In this study, the metal content of copper and zinc, and their combined dollar value are used as response variables. The independent variables[1] consist of geological factors, such as areas of formations, contact lengths, and structural elements.

## 8.2 Regression Attributes

Regression analysis is perhaps the most important technique in statistics, and most other multivariate techniques, in some way, are derived from it. A well-fitted regression equation should:

- provide estimates of values of response variables from values of the independent variables. If the correct form of the equation has been chosen, the estimates (predictions) should be both precise and unbiased.

---

[1] "Independent variables" and "explanatory variables" are synonymous terms.

- provide an estimate of the error involved in using the equation as a predictive tool;
- provide a measure of the correlation existing amongst the variables.

Daniel and Wood (1971, p. 5) suggest that a good method of fitting equations to data should:

- use all relevant data in estimating each constant;
- have reasonable economy in the number of constants required;
- provide some estimate of the uncontrolled error in y;
- provide some indication of the random error in each constant estimated;
- make it possible to find regions of systematic deviations from the equation if any such exist;
- show whether the conclusions are unduly sensitive to the result of a small number of runs, perhaps even of one run;
- help to spot sets of data that really are not from separate runs, but actually are from parts of one longer run;
- give some idea of how well the final equation can be expected to predict the responses, both in the overall sense, and at important sets of conditions inside the region covered by the data.

The results obtained from regression analysis, and the problems associated with its application to mineral resource evaluation, are discussed in separate sections[1] in terms of the suggestions listed above.

## 8.3 Assumptions in Regression Analysis

The implicit assumption in regression analysis is that a reasonable linear relationship exists among the unknown parameters of the model. The parameters are estimated by fitting an equation to the available data under the following assumptions:

(1) The values of response variable $y_i$ are normally distributed for a given $x_i$.

(2) The variance of $y_i$ values remains the same for any given $x_i$.

(3) The error terms, $e_i$ are uncorrelated and independent.

The above assumptions can also be expressed in terms of the random error term $e_i$ in that, for a certain $x_i$, $e_i$ has a normal distribution with mean zero and variance $\sigma^2$, the latter being the same for all $x_i$'s. In addition, the third assumption of independence and zero covariance holds for the

---

[1] See Sections 8.12 and 8.13.

error terms. This is because, for a given $x_i$, the variability in $y_i$ is entirely dependent on the random error term $e_i$.

The response variable has to be a quantitative measure but this is not the requirement for explanatory variables which may be qualitative or dichotomous, or a combination of quantitative and dichotomous measures. The response variable is assumed to have a normal distribution, its variance remaining the same for all combinations of explanatory variables. However, it is not assumed that the explanatory variables are normally distributed, or even that they are quantitative measurements (Overall & Klett, 1972).

The effects of violating these assumptions are discussed in Section 8.8.

## 8.4 The Regression Model[1]

The linear regression model describes the linear relationship between a random response vector Y and a set of independent predictor variables $x_i$, $i = 1,2,...p$. The term linearity in the model means that the equation chosen will be linear in the coefficients $\beta_0, \beta_1, \beta_2, ...\beta_p$. The number of variables p in the equation and thus the number of coefficients, cannot be more than n, the number of observations.

[1]Throughout this thesis, the term "regression" will be used to denote linear multiple regression analysis.

Suppose[1] the model under consideration is:

$$Y = X\beta + e$$

where,

$Y$ - is an $(n \times 1)$ vector of observations;

$X$ - is an $(n \times p)$ matrix of known factors;

$\beta$ - is a $(p \times 1)$ vector of coefficients;

$e$ - is an $(n \times 1)$ vector of error terms.

As stated previously, $E(e) = 0$ and $V(e) = I\sigma^2$, the I indicating that the error terms are uncorrelated.

The least squares estimates of $\beta$ are given by:

$$b = (X'X)^{-1} X'Y$$

where b is a vector of estimated $\beta$ values. The fitted values $\hat{Y}$,[2] are obtained by evaluating:

$$\hat{Y} = Xb$$

The elements of the vector b are linear functions of the observations Y which minimize the error sum of squares $e'e$ regardless of the distribution properties of the errors. The vector b provides unbiased estimates of $\beta$ which have a minimum variance

---

[1]The description is based after Draper and Smith (1966).

[2]I.e., the estimates of Y.

of any linear functions of the vector Y elements. And, if the errors are normal, then b is the maximum likelihood estimate of $\beta$.

A quantity $R^2$, called the coefficient of multiple determination is normally used to assess the variation in the data explained by the regression equation. The quantity is actually the square of the multiple correlation coefficient, and calculated as:

$$R^2 = \frac{\Sigma(\hat{Y}_i-\bar{Y})^2}{\Sigma(Y_i-\bar{Y})^2}$$

$R^2$ is often stated as a percentage; the higher it is the better the fitted equation explains the variation in the data, subject to the condition that the improvement obtained in the $R^2$ value by increasing the number of variables is significant under a pre-determined criterion and not because of saturating the regression model.

Another measure that is used in examining a regression term is the standard error of the estimate of Y, a quantity analogous to standard deviation in the sense that it estimates the scatter of the observed values of Y around the computed $\hat{Y}$ values on the regression line. It is defined as:

$$S = \sqrt{\frac{\Sigma(Y-\hat{Y})^2}{n-p-1}} = \sqrt{\text{residual mean square}}$$

If the $\hat{Y}$ values are normally distributed, the standard error of the estimate can be used to set confidence limits of the estimated value at the desired level.

## 8.5 Regression Procedures[1]

Geological information can never be complete and therefore, geological relationships remain a matter of opinion. For this reason, when fitting an equation to data, it is essential for reliability to include as many geological factors as possible. However, this consideration must be balanced against the need for a relatively small number of factors for effective monitoring, and to keep computer costs to a reasonable level. The following is a brief review of the most commonly used regression procedures, and commentary on their application.

### (i) All possible regressions:

This procedure requires developing a set of equations for all possible combinations of explanatory variables, including cases where one or more variables may not be included in the equation. Thus, the number of equations formulated increases exponentially with each additional variable. The equation finally selected is generally the one that explains the

[1]The description in this section is mainly after Draper & Smith (1966).

maximum amount of variability in the data.

This method becomes unwieldy when a large number of factors is being considered. Its utility lies in looking at all possibilities before selecting an equation. In mineral resource evaluation this will not be necessary except under conditions of complete ignorance, in which case the number of variables will be small. Perhaps an entirely different model would be considered under such conditions.

(ii) <u>The backward elimination procedure</u>:

In this procedure, a regression equation containing all variables is first determined followed by a calculation of the partial F-test value for each variable on the assumption that each variable is the best one to enter the equation. Any variable with an F-value below a pre-selected cut-off F-test value is deleted. Then the regression equation is reformulated with the remaining variables. A partial F-value is again calculated for each variable as before, and a new regression equation computed with variables exceeding the F significance level. The procedure continues until no further variables can be deleted.

The procedure is quicker and less costly than calculating all possible regressions. The number of variables must not exceed the number of observations or else they will not all be included at one time. When geological factors are being evaluated for resource assessment, the method proves of use only if all the input factors are believed to have been related in ore formation. Otherwise, the presence of redundant variables in the first equation may lead to unexpected results, particularly if they happen to have a higher level of F-test value than those more directly related in resource formation.

(iii) The forward selection procedure:

This procedure formulates a regression equation by including one variable at a time, performing regression, then including one more variable, and so on. The order of insertion is based on the value of the partial correlation coefficients of the variables not yet included in the equation. The procedure therefore is the reverse of the backward elimination method. It has the advantage of including only those variables that have a significance above a pre-selected F-test value, instead of first computing a large regression equation and then eliminating unnecessary variables. The method does not, however, consider the effect that the insertion of a new variable may have on the contribution of those variables already in the equation. However, this can be rectified by a judicious selection of variables or by using the stepwise regression method. The forward selection method has the advantage of economy in computing time.

(iv) The stepwise regression method:

The stepwise regression method is the same as the forward selection procedure except that after each addition of a variable this method examines the partial F-values of those already in the equation. If the partial F-value of any variable in the equation decreases below a pre-selected F-test value on the addition of a new variable, it is deleted from the equation. The procedure continues until the equation is satisfied. Thus, this method is an improvement on the forward selection procedure.

## 8.6 Procedure Used

The forward selection procedure available in the S.P.S.S. library at McGill University is used in all regression analyses. A series of runs was also made using the stepwise procedure available on the C.D.C. 6400 computer at the Department of Energy, Mines & Resources, Ottawa, and the results were compared with those obtained using the forward selection procedure. The results in the two cases are similar. In an oral communication, Agterberg (1976) stated that in his studies, he too had observed that the results obtained by the two methods are essentially the same.

## 8.7 Selection of Variables

Perhaps the most difficult problem in applying regression analysis to geological data is the selection of the "best" set of explanatory variables for the model. The importance of selecting the right variables cannot be over emphasized. The peculiarity of geological information is that observations and measurements of data are the deformed, modified and incomplete representations of geological processes and not the processes themselves. Geological processes are both evolutionary and interruptive, spread out over long spans of geological time. Yet geological observations are measured in one point in time. If the role of this time dimension is not fully understood, the various stages of geological processes cannot be identified; they can only be approximated from existing evidence. This is the case with the Rouyn-Noranda region, necessitating therefore, a greater number of trials to obtain the appropriate equation, and greater caution in interpretation.

The easiest approach in variable selection is to let the computer do it automatically on the basis of partial correlation coefficients. This approach is also suggested by Agterberg, et al. (1972, p. 27), where they state that "By working with many variables . . . , we admit considerable redundancy in the data base. However, during the multiple regression, the redundancy is automatically eliminated. For this type of multivariate statistical analysis, it may be best to start off with as many variables as possible and to let the elimination of redundancies be done by the computer."

In the present study, it has been observed that extremely misleading results can be obtained by allowing an automatic selection of variables by the computer. The partial correlation coefficients which are the basis for computer selection only reflect the spatial relationships among the variables. They may or may not have any genetic link with the response variable. The computer cannot make any distinction between variables that have spurious or genetic correlations. The following two regression equations[1] illustrate this point.

$$\begin{aligned}\log_{10} \text{Copper} = {} & 4.64 - 1.31 \text{ CNTL } 7 + 0.34 \text{ FOLT } 1 \\ & - 0.02 \text{ CNTL } 13 - 0.18 \text{ AREA } 4 \\ & + 0.33 \text{ CNTL } 34 - 0.40 \text{ AREA } 8\end{aligned}$$

$$\begin{aligned}\log_{10} \text{Zinc} = {} & 5.04 - 25.12 \text{ CNTL } 14 - 0.14 \text{ CNTL } 13 \\ & + 0.20 \text{ FOLT } 1 + 0.37 \text{ AREA } 2 \\ & - 0.17 \text{ FOLT } 3 - 0.02 \text{ AREA } 12\end{aligned}$$

These equations have been obtained for the eight cells that contain known ore deposits. There were no constraints placed on the computer and the selection of variables was done

---

[1] See Section 4.2 for description of variable names.

automatically using the stepwise regression procedure. Each equation has an $R^2$,[1] value of one, indicating that the equation fully accounts for the variability in the data for the eight cells.

Observe that neither of the equations contains the variables, area of rhyolite (AREA 3), or the contact length between rhyolite and andesite, basalt (CNTL 11). These are the two variables with which nearly all massive sulphide deposits are associated in the region. Yet other, less pertinent but more spatially correlated variables are automatically selected by the computer.

Although these equations can be fitted to the data in the eight cells with known ore deposits in them, what is the predictive worth of equations that do not include variables known to be associated with the ore deposits? There are other examples in which the mere presence or absence of a certain variable significantly changes the regression equation. Some of these examples are discussed in the evaluation of results. At this stage, it should be stressed that in regression or any other statistical analyses, the ability to monitor the role of variables can help avoid incorrect conclusions. This requires "a priori" subjective knowledge of

---

[1] $R^2$ is the coefficient of determination. It is a measure of the proportion of variance in Y which is accounted for by the estimated linear regression of Y on the $p$ explanatory variables.

the behaviour of geological factors. Where this knowledge is incomplete, information from a similar geological system can be useful.

Sometimes it may be necessary to include variables that have a spatial, rather than a genetic link with the process of ore formation. This is the case if such variables instead of being a cause of ore forming processes are a result of such processes. Structural features can result from folding associated with volcanism which itself may have been the cause of ore formation. These indirect relationships can be useful.

It is perhaps best to input variables in a pre-determined order based on their genetic affiliation for the response variable rather than on the basis of partial correlation coefficients. However, where there is uncertainty concerning the relative importance of variables, they can be grouped together in sets, the insertion level among the group being pre-established, but the selection from within the groups being left to the computer. This practical procedure is used in this study. The final equation should reflect the geological system under evaluation as logically as the present knowledge of the local geology permits.

The decision as to the initial set of input variables is an extremely important one, for any further selection is based on the resulting contribution to the response variable. The method of letting the computer do the work by inputting the maximum number of variables has already been commented upon. Another method is to use subjective judgment in making the initial selection, based on knowledge and experience of the type of deposits and the area under study. It has been shown in many statistical analyses, and also observed in this study that quantitative relationships extend to variables not earlier believed to have a significance in subjective observations.

Characteristic analysis is a simple way of choosing the initial variables on the basis of their relative typicalities. The assumption is that the frequency of observation of a variable is an indication of the importance of its role in geological processes. This may or may not be the case depending upon the type and complexity of geology in the region. To be acceptable, the selected characteristics should conform to the various concepts on ore formation in the region. The order of input can then be based on decreasing values of relative typicalities.[1]

When the area under evaluation is well known with sufficient information available on both the explanatory and response variables, it helps greatly to perform appropriate factor analyses and then select the variables loaded on the factor accounting for the greatest variance in the response variable. It is believed that this is the best method of

---

[1]See Section 2.2.5, and Chapter 6.

variable selection. An additional advantage of factor analysis is that other geological relationships, previously unsuspected, also are highlighted for a better comprehension of the geology of the region. It should be pointed out, however, that the "cause" and "effect" relationship should not be automatically accepted based on factor analysis. This is a problem of statistically treating geological data. Results have to be monitored with sound judgment at each step.

Since factor analysis after orthogonal rotation results in uncorrelated factors, some workers have used factor scores rather than the actual variables in subsequent statistical analyses. The drawback of this approach is that it builds an artificial barrier between the statistical model formulated and the geological system as described by the variables. It no longer is possible then to observe, understand and keep track of the roles of individual variables.

After the initial variables have been selected, it will be necessary to monitor the role of those most significantly contributing to the response variables. Because of a high correlation, some of these variables can become very sensitive to changes in their values. If such a significant variable happens to be a structural element such as a dyke or a fault, it should be input at a later stage in regression analysis, if at all. For example, the variable DYKE 3 is so highly correlated with zinc that its mere inclusion in the early stage of regression results in a positive value prediction in cell 1027. This cell contains only granite and granodiorite rocks and massive sulphides cannot be expected there; DYKE 3 gives misleading positive results in that cell. The actual contribution of such linear structural elements has to be watched; they are often an interpretation or an interpolation, their recorded length being subject to error by the mapping geologist, and later by the draftsman. It is felt that the most stable measurements on a geological map are the areas of formations and the contact lengths between them. The lengths of dykes, faults, etc. are probably less accurate measurements. However, they can be valid contributors to the response variable and therefore should be included, but their effect closely observed.

The initially selected variables can be reduced in number after regression analysis if any of them do not significantly contribute to the coefficient of multiple determination, the $R^2$ value. It may also be necessary to experiment with additional variables to achieve an optimal combination for a regression model that adequately describes the behaviour of the geological system in the Rouyn-Noranda region.

## 8.8 Treatment of Response Variable Data

Ideally, in any type of regression analysis, raw data for the response variable should be normally distributed. This, however, is not the case for most data relating to natural phenomenon. Suitable transformations[1] may then be necessary to reduce non-normality in the response variables, using the transformed state in formulating the least squares model. It should be emphasized that minimizing the sum of squared deviations for the transformed model is not necessarily the same as minimizing the sum of original, untransformed data; that is, the least squares estimators of the transformed variables will not be the same as those of the original model. For example, in the Rouyn-Noranda region, the regression models formulated on the raw data and on their logarithmic transformations are different in terms of both the coefficients and the explanatory variables included. A decision has therefore to be made between using the raw data thus violating the normality assumption or to make suitable transformations taking the risk of obtaining a least squares fit significantly different from that obtainable from the raw data.

The response variable data of the Rouyn-Noranda region consists of the following:

- known endowment in eight of the 64 cells, in terms of contained metal tonnages of copper and zinc, and their dollar value;
- unknown endowment in the remaining 56 cells.

It is likely that some additional unknown endowment

---

[1]See also Section 4.7.

may be present in the eight cells with known endowment. For this reason, the known endowment defined as the sum of production and reserves of economic deposits in these cells represents the minimum possible endowment. All predictions of endowment in the region should therefore be considered as minimum possible values.

The 56 cells with no known endowment are of the following two types:

- cells that have mineral endowment which is at present undiscovered;
- cells that have no endowment.

At the time of data compilation, there is no way of knowing whether a cell with no known endowment belongs to the first, favourable group, or to the second, barren group. It may be subjectively possible to predict that a cell has no endowment if it does not have any attributes directly or indirectly associated with known ore deposits in the region. On the other hand, however, it will not be possible to subjectively assign a value to the unknown endowment in a cell on the basis of the mere presence of attributes related to known ore deposits.[1] What is important is not so much the

[1]Resource estimates based on quantified subjective probabilities represent a different model, the effectiveness of which is not contradicted by the above statement. See Section 2.2.4.

presence of a favourable set of attributes as the presence of a relationship among these attributes which corresponds to the relationship which exists in the reference cells that have a known endowment. This relationship can only be known after statistical analyses. Such analyses require that all response variables have a quantitative value. Since no endowment is presently known in 56 of the 64 cells in the region, there is no alternative but to initially assign a zero value to the response variables in such cells. Should the mathematical relationships formulated from the known endowment cells hold, then it will be possible to distinguish between cells with no endowment and cells with undiscovered endowment. Until this is done, the data on response variables consists of eight positive values for the cells with known endowment and 56 zeros for cells with zero known endowment. The large number of zeros causes a strong positive skewness in the response variable distribution. The presence of zeros also makes it very difficult to apply any transformation to normalize the data, for whatever the transformation, its effect is uniform on all the zeros, and therefore the overall distribution remains skewed.

Referring to the lack of normality in the distribution of response[1] variable data, Draper and Smith (1966, p.

---

[1]It is not assumed that explanatory variables are normally distributed or even that they are quantitative measures. (Overall and Klett, 1972, p. 425).

59) note that "An assumption that the errors $\epsilon$,[1] are normally distributed is not required in order to obtain the estimates $b$,[2] but it is required later in order to make tests which depend on the assumption of normality such as t- or F-test, or for obtaining confidence intervals based on the t- and F-distributions."

Regression techniques are robust enough to adapt to moderate deviations from the normality assumption, particularly when the aim of the study is only to find the best fitting least squares function. However, normality in the response variable is essential when interval estimates are to be made or significance tests applied.

## 8.9 The Ore Cells Regression Model

The problems caused by a large number of zeros in the response variable data can be circumvented to a certain extent by computing the regression model on the basis of observations on the known endowment cells alone. The assumption here is that ore forming processes are represented by the geology in these cells, and that the size of the cells is large enough to have accommodated ore forming activity. The relationship

---

[1] $\epsilon$ is the vector of error terms $e_i$.

[2] $b$ is the vector of regression coefficients.

formulated from the known endowment cells can then be applied for predictive purposes over the cells with no known endowment. It is also assumed that sufficient number of known endowment cells are present for the formulated model to adequately represent the geological system under observation. On a reconnaissance level, it may be acceptable to augment the data base by including known endowment cells from other typical mining regions. The increased number of observations will improve the reliability and comprehensiveness of the computed model. It should be obvious, however, that an equation which represents the typical behaviour of a regional system cannot be derived from non-typical data.

When a network of cells is randomly superimposed over a given region, it is likely that the ore deposits will not be centred in their respective cells. There can be instances in which an ore deposit lies so close to the boundary with the adjacent cell that geological relations present in the adjacent cell may also have contributed to its occurrence. Thus, when the objective is to build a model based only on the known endowment cells, it may be more desirable to first centre the ore deposits in their respective cells and then make the necessary measurements assuming that the centre of the cell is close to the centroid of geological activity that led to ore formation. The situation will also depend on the presence of geological trends and intrusions within the region as well as the size and shape of the individual cells. However, to avoid a bias, particularly where geological relations are complex, it may be best to confine the measurements within the randomly placed network of cells. The regression models and forecasts made using the eight known endowment cells in the region are discussed in Section 8.19.1.

## 8.10 The Total Area Regression Model

The Rouyn-Noranda region has been described as an eruptive centre of ore producing volcanogenic activity.[1] The region can be considered as a self-contained geologic unit since there is little evidence of similar eruptive activity or associated ore formation outside it. The quantitative relationships between geology and ore occurrence are therefore believed to be local to the region or its immediate vicinity.[2] Therefore, if a functional relationship exists between the explanatory variables and the known endowment, i.e., with the response variable in one or more cells, then a similar functional relationship should also extend to the cells with no known endowment but which have the necessary integrated presence of explanatory variables statistically similar to those

---

[1]See Section 3.2.2.

[2]On a more generalized level, however, geological relationships can be similar for typical ore deposits in other regions, and of different geological ages.

in the known endowment cells so that the endowment not known can be predicted in them. Such a situation should be true in relatively closed geological systems in which all ore forming activity was confined within limited areas. If such is not the case, then either a state of disequilibrium exists in the response variable data,[1] or otherwise, the functional relationship between the response and explanatory variables is not valid.

The data for the response variables in the region consists of the known endowment as defined for this study, and the unknown endowment to be predicted but initially assigned a zero value. The effect of performing regression analysis with this kind of response variable data is as follows:

- cells with no known endowment which have been assigned a zero endowment value, and which do not have any level of geological similarity with known endowment cells should obtain a predicted value of zero or close to zero. In extreme cases, a negative value may be predicted which while indicating a high degree of dissimilarity will in effect be the same as zero endowment.

---

[1]The situation arises when part of the response variable data has to be arbitrarily assigned a zero value for lack of information on its unknown endowment. Some of the cells may be correctly assigned a zero value, but other cells that do in fact have an endowment which is not known, are wrongly assigned a zero value. This thus creates a disequilibrium in the system.

- cells with no known endowment which have been assigned a zero endowment value and which have a geological similarity with known endowment cells should obtain a positive but a rather low predicted value irrespective of the $R^2$ value, the reason being as indicated earlier, that instead of zero, a higher, but unknown value should have been assigned to them.
- cells which have a known endowment in them should obtain a predicted value lower than their known endowment because similar integration of geology in cells of unknown endowment have their response variables arbitrarily assigned zero values.

In such a situation, it is possible to simply apply regression analysis over the total observations in the region and then to examine the results knowing that the greater the number of unknown endowment cells with incorrectly assigned zero values with respect to the known endowment cells, the lower will be the general level of predictions obtained. These predictions can be examined from a relative point of view, i.e., cells with a higher potential will in general show higher predicted estimates than those with less. Cells with little or no potential can be eliminated by this method, and at the same time other cells can be ranked relative to one another.[1] These rankings will not have any economic implication in strictly objective terms. However, on a purely reconnaissance level, the method should prove useful.

Agterberg et al. (1972),[2] used a modification of this method to obtain an estimate of the relative potential of copper and zinc in the Abitibi area in Ontario and Quebec. On dichotomous data, they used regression analysis as an estimate of the discriminant function separating two populations of cells with response variable set at one if an ore deposit was present, and zero if it was not. The results obtained were contoured to obtain a copper and zinc potential contour map of the area. The response variables in all cells that had an ore deposit in them, were each assigned a value of one regardless of the size of the deposit. Their results, therefore, imply the presence of an ore deposit as a "geological event", the probability of which is predicted by the regression estimate of the response variable. Although no economic considerations are involved in their study, the study is still useful for reconnaissance level exploration.

---

[1]This ranking can vary with different sets of explanatory variables, depending upon how pertinent and how highly correlated they are. The right selection of variables is most important if reliable results are to be obtained. See Section 8.7.

[2]See Section 2.2.3.

## 8.11 Iterative Regression Analysis

The problem of what values to assign to cells of unknown endowment can be resolved using an iterative process of regression analysis which will in effect provide the predicted estimates of endowment in these cells. The implicit assumption is that in terms of information and concepts, the region under evaluation is a closed geological system at the present point in time, and that a state of statistical equilibrium exists between the response and explanatory variables such that if ore endowment is a function of a particular integration of geological variables, then a similar integration of these variables should also reflect a similar level of ore endowment. This assumption is necessary because with additional information or newer concepts, the model will require an updating or a re-evaluation. If this is not done, there will be an increased deviation of the predicted values from the actual observations.

A regression analysis is first performed using the pre-selected set of explanatory variables, and the set of response variables including the subset of known endowment values for cells that have known ore deposits in them, and the subset of zero values assigned to the cells with no known endowment. The regression estimates for cells with no known endowment in them will be higher than zero in cases where any level of geological similarity exists with respect to that in the known ore-bearing cells. These estimates will vary in each case depending on the degree of similarity. Therefore, they are relatively more precise estimates of the endowment in the unknown cases than the arbitrary zeros assigned to each of them.

After the first regression run, a new set of response variables results consisting of the subset of known endowment values which remains unchanged, and the subset of predicted values replacing the zeros in cases where there is no known endowment. The explanatory variables remain the same as in the initial run. The regression is performed again on this new set of data and the predicted values obtained for each cell.

At this stage, the predicted values earlier assigned to the unknown endowment cells are replaced by the newly predicted values. However, the values of response variables in cells of known endowment are kept at their same original values. Regression analysis is repeated on this most recent data, and the predicted values observed for trends with respect to the earlier predictions. The objective here is to see if the predicted values of the cells that do have a known endowment converge towards that value. It shall also be necessary to check on the improvement in the $R^2$ values. The rate of improvement determines the number of iterations required to make the regression results converge on the observed values in the known endowment cells. It may be necessary after a number of runs to increase the number of explanatory variables particularly if the endowment has more than one mode of association with other variables as determined by such methods as factor analysis. Obviously, the first set of independent variables regressed on are the ones that have a greater correlation with the resource constituting the endowment.

The regression procedure is repeated iteratively until a stage is reached where the $R^2$ values approach one and where the predicted values of endowment for the known endowment cells nearly equals the known values. At this point the system can be considered to have stabilized; that is, each cell has now been assigned an endowment value based on the level of geological similarity with the known endowment cells. Any further regression iteration will not significantly change the values in any of the cells. The predicted value of the response variable in each cell is an estimate of unknown endowment in it with a standard error of estimate as determined for the particular iteration. The standard error can be used to set confidence limits at any desired level.

## 8.12 Regression Problems

The problems in regression analysis will vary from project to project, but will in general be related to the completeness, correctness and detail of the data base and to the set of explanatory variables selected. It may not be possible to fully eliminate or even reduce the effect of some of the problems, but the regression analyst must nevertheless be aware of their presence and possible effects when evaluating results.

The factors considered in variable selection have already been discussed in Section 8.7. It should be noted here that the omission of a relevant variable, particularly if it is correlated with those already in the model, may result in biased and inconsistent least squares estimators. The inclusion of an irrelevant variable on the other hand will only make the regression coefficients less efficient.

There are also problems of autocorrelation, heteroscedasticity and multicollinearity, each of which results in inefficiency, bias or error.

Autocorrelation is the grouping together of positive or negative residuals. When autocorrelation is present as a result of an incorrect form of regression model, or because some pertinent variable has not been included in the model, it is possible to rectify the situation by suitable revisions and modifications. However, the problem becomes more complex and difficult when the disturbance terms have an autoregressive structure.

Autocorrelation results in inefficient least squares estimators which, however, may still be unbiased. However, the variances of the parameters and the error terms may be biased. It should be noted that in studies of resource evaluation, autocorrelated positive residuals may be indicators of more favourable areas in terms of mineral potential. This, in fact is the first observation made when regression results are studied for cells with an unknown endowment which have been initially assigned a zero value. In the iterative regression analysis used in this study, it is not until statistical equilibrium is attained in the response variables that the autocorrelation is minimized. Autocorrelation may also suggest phenomena of geologic interest particularly when some relevant variable is not included in the regression model. This may have a bearing on the genesis of ore deposits in the area being studied. Other geological implications are beyond the scope of this study.

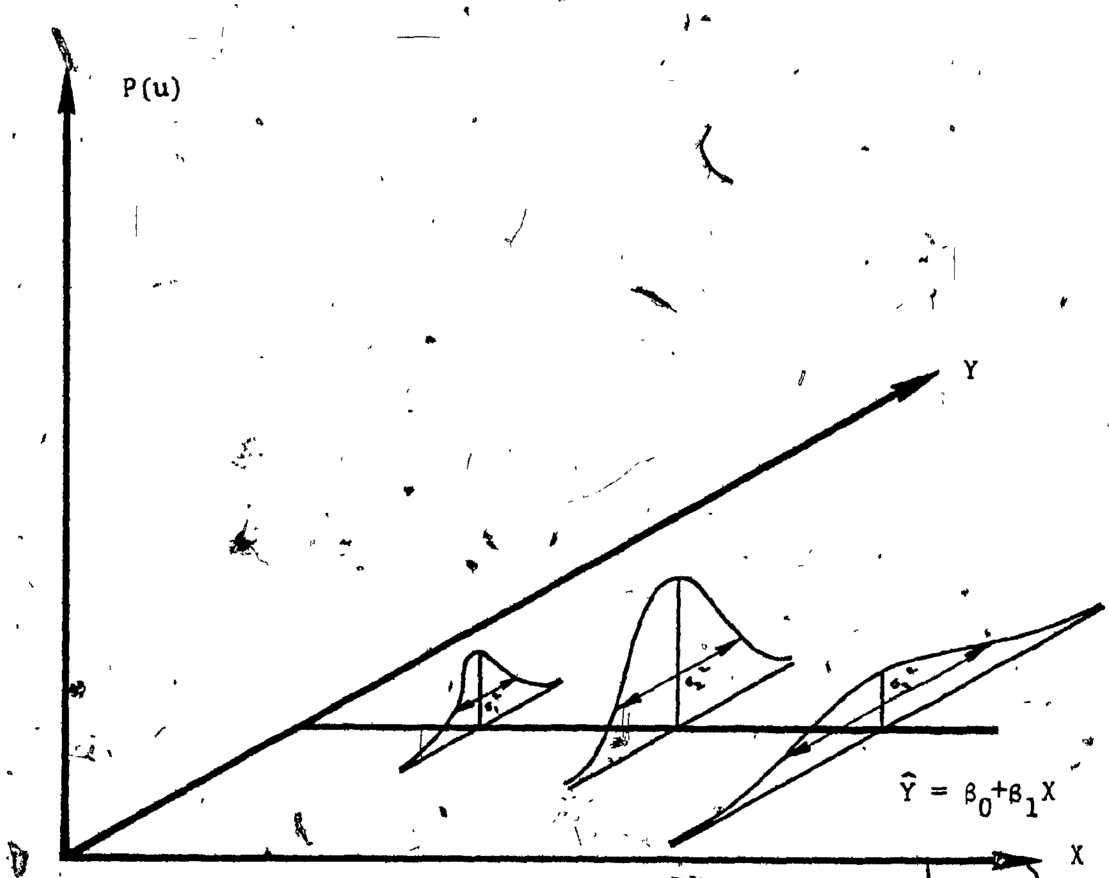
Heteroscedasticity[1] results when the assumption that the variance is constant about the regression line is violated. When this is the case, the least squares estimators do not possess the property of minimum variance of all linear estimators. Although, they will still be unbiased, the calculated confidence intervals and acceptance regions will likely be wrong.

Multicollinearity results when any of the independent variables are correlated. The question is not so much the

[1]See Figure 12.

FIGURE 12

HETEROSCEDASTIC VARIANCES
(AFTER MATHER, 1976)



presence of multicollinearity but rather, the degree of multicollinearity which exists. The regression coefficients will be imprecise when a high degree of multicollinearity exists. This is because the least squares estimators have a high variance.

The degree of multicollinearity will increase as the scale of observations increases, or conversely, as the cell size decreases. This is a problem that arises in all studies where measurements are made on geological observations after superimposing a grid of cells over the area. Geological processes particularly of an evolutionary type occur over large areas. Structural features like dykes and faults may extend from one corner of a map area to the other. The result is that individual cells cannot contain a set of response variables that do not influence or which are not themselves influenced by, the geological relationships present in the adjoining cells. One solution for this is to use cells of large size but this will not be possible over size-restricted areas such as a mining region. To reduce possible harmful effect of multicollinearity in this study, the superimposed set of 64 cells is divided into two subsets by converting the region into a "checkerboard" pattern of cells so that alternating cells are included in each subset. There is no cell that has a common boundary with any other cell in each subset except for the point connections at the corners.[1] Looking at the map of the area with the "checkerboard" cells on it,[2] it is evident that the continuity of geology from one cell to the next is abruptly halted or decreased as a result of including only alternating cells in each subset. Only those geological features that have a distinct direction, either north-east or north-west, appear to continue diagonally into the next cell in the subset, but such cases are rare. The undesirable effect of using the "checkerboard" scheme is that the degrees of freedom are reduced in accordance with the reduction in number of observations. However, when the number of observations is large compared with the number of explanatory variables being used in the analysis, this effect should not be significant. Another difficulty that is observed in this study is that, of the eight known endowment cells, three fall in one subset and five in the other. Thus the number of reference cells, which is always small in a mining region, is further reduced. The predicted endowment therefore in one subset of the checkerboard type data will be based on the geological relations present in the three cells with known endowment, and in the other subset will depend on five such cells. The regression models will also be different in the

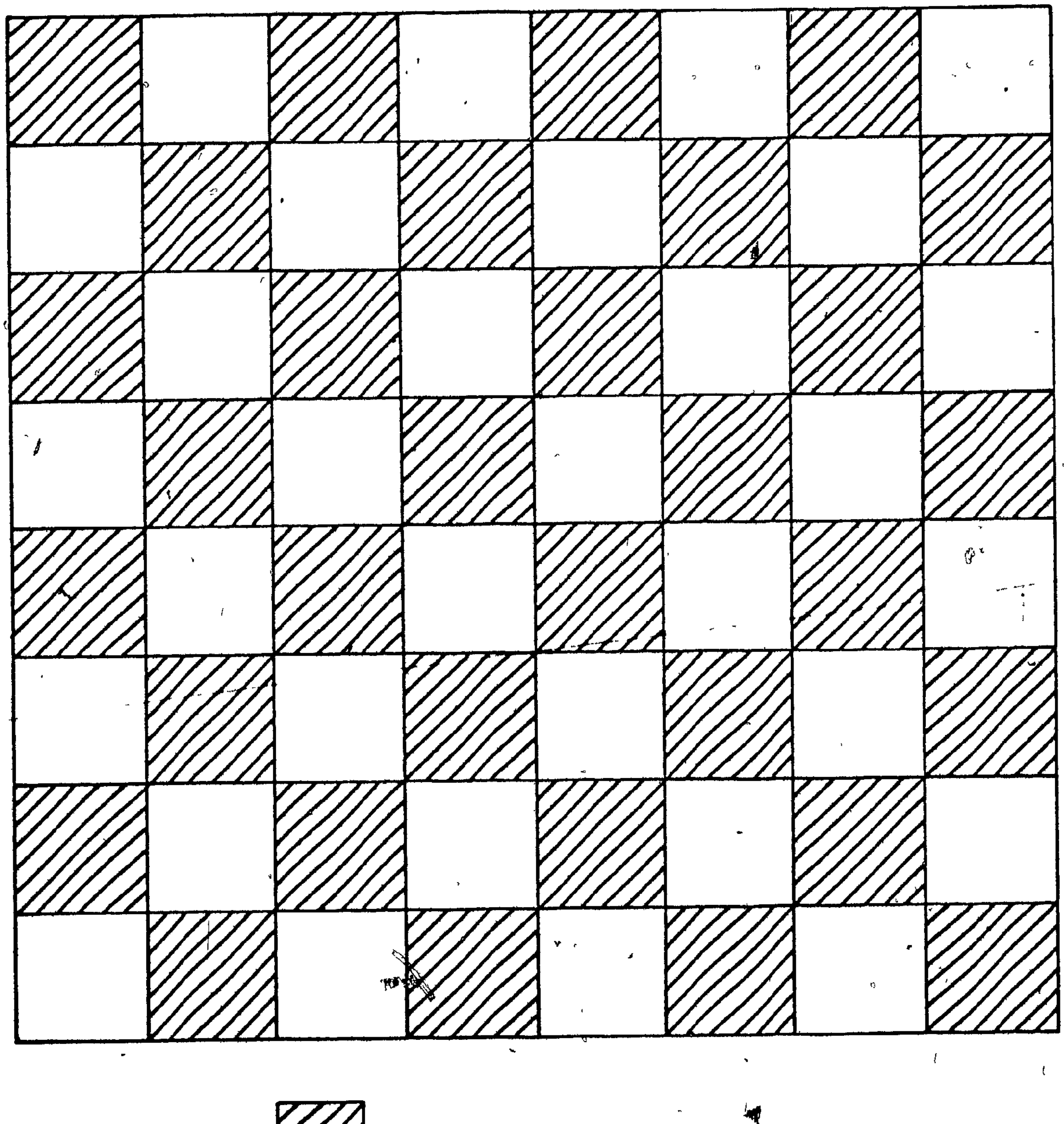
---

[1]See Figure 13.

[2]See Figure 14.

FIGURE 13

CHECKERBOARD-TYPE DIVISION OF CELLS

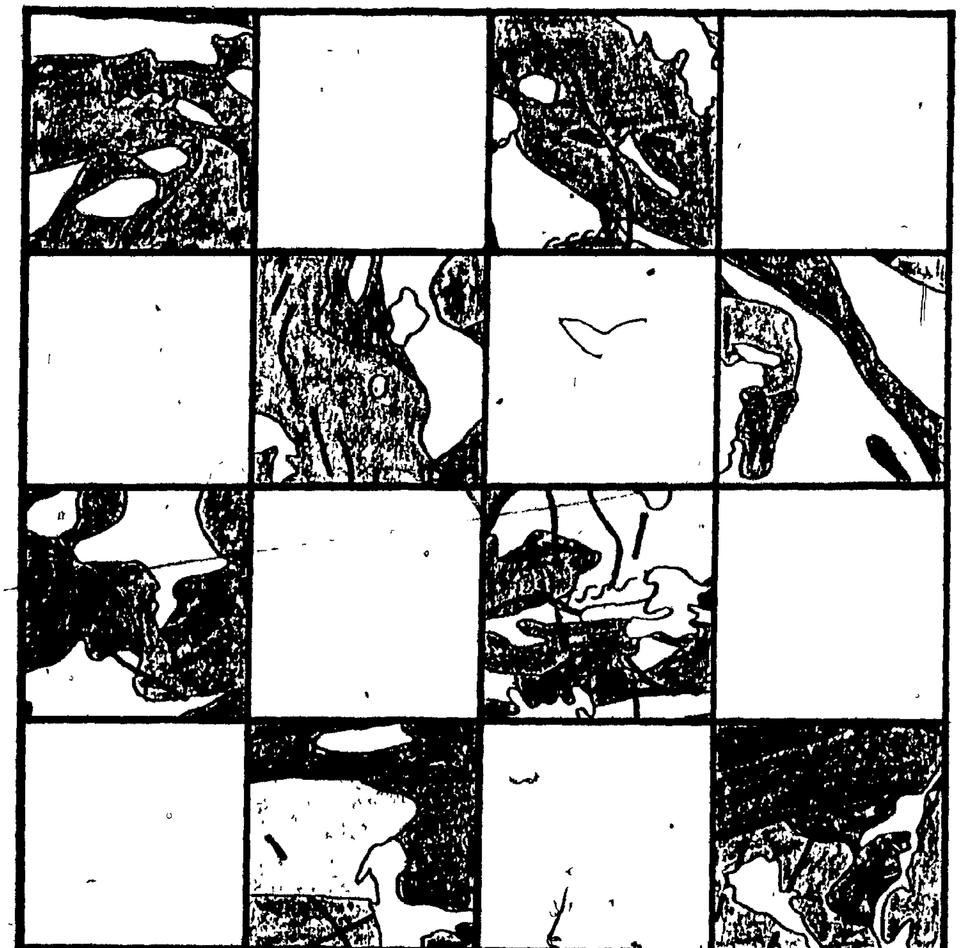


Subset A cells

Subset B cells

## FIGURE 14

THE CHECKERBOARD TYPE DIVISION[1] OF GEOLOGICAL DATA,[2] SUBSET A



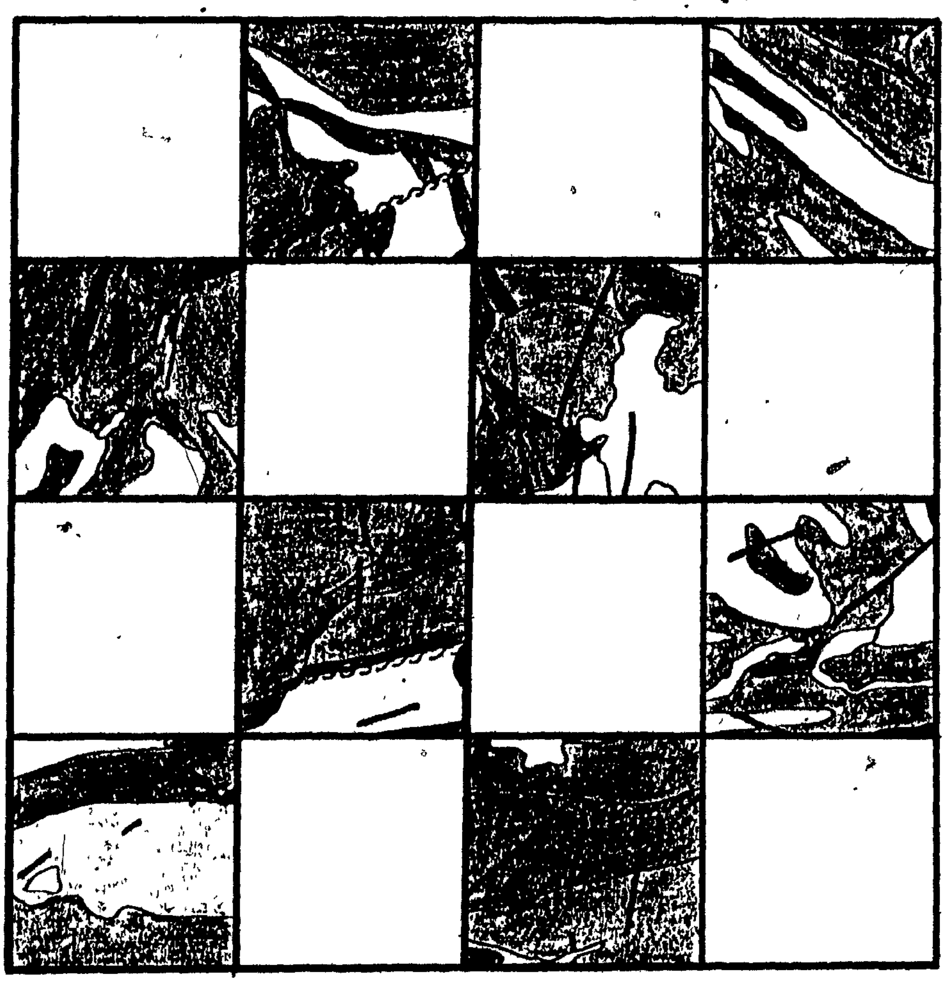
[1]For conceptual reasons, the above map area has been divided into 16 cells instead of the 64 as actually used in the study.

[2]The scale of the map and the colour index used for rock formations are the same as in Figure 5.

............ Con't.

FIGURE 14[1]
(CONTINUED)

THE CHECKERBOARD TYPE DIVISION OF GEOLOGICAL DATA, SUBSET B



[1]See footnotes on previous page.

two cases. Finally, it should be realized that all predictions will be based on what is present in the known endowment cells in each subset. It is not possible to make a prediction of additional endowment in such cells since they are the ones against which the whole system has been calibrated. However, if at any stage new information is available in terms of endowment or geology, the regression model should be changed to accommodate the new data. If the additional information relates to the discovery of new endowment in any cell in excess of that known, then re-evaluation of the regression model by the iterative process might bring out information on additional endowment in the cells that already have a known endowment. Another way of making an estimate of additional mineral endowment in known ore-bearing cells is to further reduce the size of the cells or better still, subdivide each cell into smaller parts and then perform the iterative regression analysis. Needless to say there will always be a certain minimum size of the cell below which it will not be possible to obtain a model based on geological relationships, for regardless of how small or local a certain ore deposit is, the geological processes that produced it are more extensive.

The problems related to geological data are discussed in Section 4.5. It would be pertinent here, to repeat that all geological relationships that are observed and measured are in two dimensions only. Any information as to changing relationships at depth is either unavailable or incomplete.

The most efficient regression model would accommodate information on this third dimension. Since this is not possible, the implicit assumption made is that the observations observed in two dimensions reflect the geological relationships of all three dimensions. The error associated with this assumption is directly related to the complexity of geology and on how the observations at the surface continue at depth.

## 8.13 Regression Results

Iterative regression analysis[1] is separately performed on the 64-cell data set,[2] and on its two 32-cell checkerboard type data components.[3] After eight iterations following the first regression run, no further improvement was noted in the coefficient of determination and, therefore, the iterative procedure was terminated. The results of these runs are shown in Table 8.

Table 9 compares the results of the 64-cell analysis and the checkerboard analysis after the completion of eight

---

[1]See Section 8.11 for description.

[2]The 64-cell data set includes all the 64 cells in the region. This set is referred to as the 64-cell set.

[3]The two 32-cell checkerboard type data components, subset A and B are shown in Figure 13. The two components are jointly referred to as the checkerboard set to distinguish it from the 64-cell set.

## TABLE 8

**RESULTS OF ITERATIVE REGRESSION ANALYSIS ON THE CHECKERBOARD SETS OF DATA**

| Cell No. | Known Endowment | Pre-Iterative Run | Second Iteration | Fourth Iteration | Sixth Iteration | Eighth Iteration |
|---|---|---|---|---|---|---|
| | $x10^6 | $ | $ | $ | $ | $ |
| 1 | - | 471 | 10,723 | 86,548 | 27,179 | 26,757 |
| 2 | - | 32,552 | 3,002,989 | 31,176,618 | 13,086,183 | 13,776,345 |
| 3 | - | 284 | 213 | 111 | 879 | 1,027 |
| 4 | - | 1,316 | 5,989 | 17,431 | 7,158 | 6,968 |
| 5 | - | 1,446 | 17,623 | 63,811 | 52,665 | 56,499 |
| 6 | - | 213 | 832 | 2,40[illegible] | 1,181 | 1,160 |
| 7 | - | 19,013 | 1,038,449 | 4,830,40[illegible] | 9,728,263 | 12,404,578 |
| 8 | - | 174 | 246 | 310 | 150 | 146 |
| 9 | - | 3,552 | 324,177 | 3,637,521 | 1,882,032 | 2,010,861 |
| 10 | - | 2,692 | 706,804 | 29,108,243 | 9,922,870 | 9,806,795 |
| 11 | - | 123 | 870 | 2,483 | 1,552 | 1,591 |
| 12 | - | 3,544 | 408,393 | 5,002,523 | 13,043,616 | 16,016,662 |
| 13 | 18.81 | 2,153 | 201,293 | 1,311,796 | 13,678,044 | 17,501,368 |
| 14 | - | 582 | 10,464 | 68,784 | 36,745 | 36,758 |
| 15 | - | 18 | 41 | 154 | 73 | 69 |
| 16 | 74.48 | 31,469 | 3,207,217 | 14,084,486 | 45,345,271 | 66,723,528 |
| 17 | - | 620 | 13,617 | 104,592 | 59,161 | 59,780 |
| 18 | - | 641 | 12,409 | 126,408 | 74,103 | 74,996 |
| 19 | - | 127 | 246 | 365 | 385 | 380 |
| 20 | - | 570 | 2,408 | 4,077 | 2,689 | 2,750 |
| 21 | 244.66 | 853 | 53,685 | 942,019 | 214,884,950 | 251,214,672 |
| 22 | - | 389 | 20,254 | 540,352 | 57,338 | 53,547 |
| 23 | 27.80 | 27,799,885 | 27,799,821 | 27,799,757 | 27,799,821 | 27,799,821 |
| 24 | - | 15 | 2 | 1 | 2 | 2 |
| 25 | - | 397 | 9,882 | 78,918 | 44,916 | 46,472 |

TABLE 8

(CONTINUED)

| | | | | | | |
|---|---|---|---|---|---|---|
| 26 | - | 120 | 347 | 702 | 604 | 609 |
| 27 | - | 118 | 152 | 171 | 563 | 599 |
| 28 | - | 85 | 54 | 36 | 1,194 | 1,358 |
| 29 | 1,123.84 | 1,894 | 70,543 | 1,209,771 | 726,065,832 | 1,056,956,223 |
| 30 | - | 127 | 420 | 947 | 926 | 951 |
| 31 | - | 100 | 100 | 100 | 100 | 100 |
| 32 | - | 14 | 1 | 0 | 0 | 0 |
| 33 | - | 11,008 | 3,227,822 | 91,305,933 | 57,637,880 | 62,322,376 |
| 34 | - | 95 | 104 | 97 | 57 | 55 |
| 35 | - | 66 | 84 | 110 | 95 | 90 |
| 36 | - | 397 | 247 | 115 | 144 | 146 |
| 37 | - | 25 | 10 | 7 | 1 | 1 |
| 38 | 2,109.28 | 112,440,560 | 2,460,775,533 | 2,809,156,922 | 1,916,266,909 | 2,065,142,384 |
| 39 | 34.88 | 1,806,550 | 20,020,642 | 26,276,287 | 33,388,046 | 34,366,950 |
| 40 | - | 312 | 294 | 242 | 219 | 218 |
| 41 | - | 8,548 | 182,255 | 824,763 | 12,278,859 | 15,076,065 |
| 42 | 43.48 | 1,500 | 107,178 | 1,659,680 | 32,298,399 | 37,577,770 |
| 43 | - | 707 | 4,030 | 11,109 | 22,344 | 24,069 |
| 44 | - | 1,858 | 19,936 | 66,128 | 31,676 | 32,535 |
| 45 | - | 194 | 456 | 825 | 955 | 975 |
| 46 | - | 688 | 6,703 | 26,761 | 26,816 | 27,594 |
| 47 | - | 18 | 6 | 3 | 1 | 1 |
| 48 | - | 772 | 1,783 | 2,882 | 1,450 | 1,385 |
| 49 | - | 1,899 | 6,886 | 10,077 | 9,414 | 9,792 |
| 50 | - | 458 | 4,391 | 23,215 | 542,001 | 661,161 |
| 51 | - | 1,233 | 50,423 | 548,274 | 588,531 | 614,312 |
| 52 | - | 400 | 10,854 | 124,649 | 69,050 | 70,676 |
| 53 | - | 100 | 100 | 100 | 100 | 100 |
| 54 | - | 184 | 348 | 461 | 3,178 | 3,573 |
| 55 | - | 151 | 335 | 561 | 415 | 421 |

## TABLE 8

(CONTINUED)

| | | | | | | |
|---|---|---|---|---|---|---|
| 56 | - | 45 | 7 | 2 | 1 | 1 |
| 57 | - | 118 | 152 | 171 | 77 | 74 |
| 58 | - | 66 | 84 | 110 | 55 | 54 |
| 59 | - | 38 | 15 | 8 | 25 | 26 |
| 60 | - | 66 | 84 | 110 | 55 | 54 |
| 61 | - | 38 | 16 | 9 | 3 | 3 |
| 62 | - | 66 | 84 | 110 | 155 | 158 |
| 63 | - | 118 | 152 | 171 | 292 | 302 |
| 64 | - | 66 | 84 | 110 | 55 | 54 |
| Average $R^2$ | | 0.390 | 0.800 | 0.935 | 0.999 | 0.999 |

Notes:

[1] The analyses were carried out separately on two 32-cell checkerboard-type sets of data. However, the results are shown for all the 64 cells in one table.

[2] The pre-iterative run means the first regression run, the predicted values from which are input in the iterative runs.

[3] The results are shown for alternate iterations only.

[4] Average $R^2$ is the mean of the $R^2$ values of subsets A and B for the particular iteration made.

[5] The results for the known endowment cells have been underlined.

iterations. The evaluation of predicted values is made after comparing predicted and known endowment levels in the appropriate cells. The closer the convergence[1] the better should be the predicted estimates in cells with no known endowment. Stated another way, the closer the convergence of predicted and known endowment, the better is the statistical equilibrium achieved in the observed geological system.

In the 64-cell analysis, it is not generally possible to obtain a reasonable convergence on known endowment values in spite of a coefficient of determination of 0.991. The exceptions are cell 1038 with a 74 percent convergence, and cell 1023 with a 90 percent convergence. A highly anomalous value exceeding 10 times that known is predicted in cell 1029.[2] And, for the remaining five known endowment cells,[3] an average convergence of only 65 percent is obtained.

Given these results, the predicted values in cells with no known endowment is doubtful. It is evident that observed geological relationships are in disequilibrium. As long as the convergence on known endowment values is

---

[1]Convergence is the ratio of predicted endowment to the known endowment in a cell. Convergence applies only to the cells that have a known endowment in them, i.e., cells #1013, 1016, 1021, 1023, 1029, 1038, 1039, and 1042. See Figure 6 for the location of these cells.

[2]See following paragraph.

[3]I.e., cells #1013, 1016, 1021, 1039, and 1042.

TABLE 9[1]

COMPARISON OF PREDICTED ENDOWMENT IN THE 64-CELL SET ANALYSIS AND THE CHECKERBOARD ANALYSIS USING ITERATIVE REGRESSIONS.[2]

| Cell # | Known Endowment | PREDICTED ENDOWMENT 64-cell analysis | PREDICTED ENDOWMENT Checkerboard analysis |
|---|---|---|---|
| | | ($x10$^3$) | ($x10$^3$) |
| 1001 | - | 596 | 27 |
| 1002 | - | 133,052 | 13,766 |
| 1003 | - | 42 | 1 |
| 1004 | - | 67 | 7 |
| 1005 | - | 1 | 56 |
| 1006 | - | 2 | 1 |
| 1007 | - | 30 | 12,405 |
| 1008 | - | 0 | 0 |
| 1009 | - | 17,980 | 2,011 |
| 1010 | - | 3,238 | 9,807 |
| 1011 | - | 2 | 2 |
| 1012 | - | 1,002,691 | 16,017 |
| 1013 | 18,810 | 8,032 | 17,501 |
| 1014 | - | 9 | 37 |
| 1015 | - | 83 | 0 |
| 1016 | 74,480 | 32,538 | 66,724 |
| 1017 | - | 8 | 60 |
| 1018 | - | 510 | 75 |
| 1019 | - | 0 | 0 |
| 1020 | - | 23 | 3 |
| 1021 | 244,660 | 25,411 | 251,215 |
| 1022 | - | 2,759 | 54 |
| 1023 | 27,800 | 25,033 | 27,800 |
| 1024 | - | 0 | 0 |
| 1025 | - | 516 | 46 |
| 1026 | - | 5 | 1 |
| 1027 | - | 33 | 1 |
| 1028 | - | 0 | 1 |
| 1029 | 1,123,840 | 11,412,138 | 1,056,956 |
| 1030 | - | 2 | 1 |
| 1031 | - | 214,194 | 0 |
| 1032 | - | 0 | 0 |

[1]Results of known endowment cells are underlined.

[2]Results shown are for the 8th iterative run.

...cont'd

TABLE 9[1]

(CONTINUED)

| | | | |
|---|---|---|---|
| 1033 | - | 1,610 | 62,322 |
| 1034 | - | 6 | 0 |
| 1035 | - | 0 | 0 |
| 1036 | - | 8 | 0 |
| 1037 | - | 0 | 0 |
| 1038 | 2,109,280 | 1,551,347 | 2,065,142 |
| 1039 | 34,880 | 34,912 | 34,367 |
| 1040 | - | 0 | 0 |
| 1041 | - | 89,214 | 15,076 |
| 1042 | 43,480 | 3,855 | 37,578 |
| 1043 | - | 104 | 24 |
| 1044 | - | 3 | 33 |
| 1045 | - | 6 | 1 |
| 1046 | - | 65 | 28 |
| 1047 | - | 0 | 0 |
| 1048 | - | 0 | 1 |
| 1049 | - | 58,690 | 10 |
| 1050 | - | 851 | 661 |
| 1051 | - | 16 | 614 |
| 1052 | - | 169 | 71 |
| 1053 | - | 30 | 0 |
| 1054 | - | 5 | 4 |
| 1055 | - | 0 | 0 |
| 1056 | - | 0 | 0 |
| 1057 | - | 0 | 0 |
| 1058 | - | 0 | 0 |
| 1059 | - | 0 | 0 |
| 1060 | - | 0 | 0 |
| 1061 | - | 0 | 0 |
| 1062 | - | 0 | 0 |
| 1063 | - | 0 | 0 |
| 1064 | - | 0 | 0 |
| | $R^2$ = | 0.991 | 0.999 |
| | S.E. = | $1.846x10^6 | $1.050x10^6 |

[1]See footnotes on previous page.

incomplete, the predicted values in the remaining cells should be similarly incomplete or otherwise incorrect. The high predicted value for cell 1029 is the result of multicollinearity which affects the predicted values in varying degrees depending upon the correlation structure among the explanatory variables. A lack of convergence can also result if some relevant explanatory variable has been omitted from the model, but this should be indicated by the lack of improvement in the coefficient of determination with further iterations.

As compared to the 64-cell analysis, a near complete convergence is obtained with regression on the checkerboard set of data, the convergence averaging over 95 percent for the known endowment cells, and ranging from 86 to 103 percent. Also, there is no anomalously high value estimated for any cell as is the case with cell 1029 in the 64-cell analysis.

The high $R^2$ value obtained in the checkerboard analysis is partly because the subdivision of the 64-cell data into two 32-cell components reduces the degrees of freedom. However, the near complete convergence on known endowment and the consistency of predicted values is possible as a result of the control on multicollinearity which the checkerboard approach achieves.

Generally, the results obtained by regressing on the 64-cell, and the checkerboard ~~components,~~ appear to be comparable in the sense that high and low predicted values correspond in each case. However, there is a wide divergence in the estimates of endowment for cells 1007, 1031 and 1049. The checkerboard analysis gives a prediction of 12 million dollars in cell 1007, but the 64-cell analysis predicts only 30,000 dollars. Further, the predicted values in cells 1031 and 1049 are 214 and 59 million dollars respectively using the 64-cell set as compared to checkerboard predictions of zero and 10 thousand dollars.

By superimposing the grid of cells on the geological map of the region, it is evident that not only does cell 1007 lie diagonally north-west of known endowment cell 1016, but also that the geological trends in the two cells are north-west, continuing from one cell into the other. To observe the influence of cell 1016 on cell 1007, a new regression run was made assigning a zero value to the known endowment in cell 1016. The result shows a prediction of six million dollars in cell 1016 compared with its known 74 million dollar endowment. The predicted value in cell 1007 is three million dollars. The resolution of multicollinearity in such a case becomes difficult particularly when there is no other known endowment cell nearby. Obviously, cells 1007, and 1016 have a certain measure of geological similarity between them. Cells like 1007 therefore, should not be ignored in any exploration effort. The potential of this cell will be further evaluated using multiple discriminant analysis.[1]

[1] See Chapter 9.

The 64-cell analysis gives a prediction of 214 million dollars for cell 1031. However, the checkerboard analysis predicts a virtually zero endowment for this cell. The reason for this difference becomes clear when the location of the cell is observed on the geological map. Cell 1031 lies directly south of the known endowment cell 1023. The known endowment in cell 1023 occurs in tuff, agglomerate,[1] a formation that continues into cell 1031. There is no other massive sulphide deposit in the region that occurs in this rock formation. Obviously therefore, the geological relationships present in cell 1031 bear little if any similarity with any known endowment cell except 1023. The application of the checkerboard approach removes the influence of known endowment cell 1023 resulting in the low prediction of value in cell 1031. The original 64-cell analysis estimate for this cell appears to be a result of multicollinearity. However, because of the similarity of this cell with cell 1023, it should mot be dropped from consideration because there is no reason why the unique associations that produced endowment in cell 1023 could not have done so in the similar geological environment continuing into cell 1031.

The 64-cell analysis also gives a high value of 59 million dollars in cell 1049. The checkerboard analysis predicts an endowment of only 10 thousand dollars in this cell

---

[1]I.e., AREA 2.

despite the fact that the regressed checkerboard component also includes the known endowment cell 1042 lying diagonally beside this cell. The only logical explanation for the high predicted endowment in cell 1049 appears to be its joint border with cell 1041, which by the iterative procedure gives a high predicted value even though because of its unknown endowment, it had originally been assigned a zero value. The high value of cell 1049 also appears to be the result of multicollinearity, and this value is therefore spurious.

Because of the changing nature of geological data, it will not be possible to completely eliminate multicollinearity. However, by use of the checkerboard approach, the continuity in the data is broken up, so that the explanatory variables within each set are no longer as strongly related as in the continuous 64-cell analysis.[1] Also, when it is possible to converge on known endowment values, multicollinearity can be considered to have been suppressed. Otherwise, spuriously high values may be obtained with little if any convergence on known endowment.

It has been earlier stated that the efficiency of iterative regression analysis can be judged by the statistical stabilization of geological relationships combined with a convergence on the known endowment in each of the reference cells. It may be possible to statistically stabilize geological

---

[1]E.g., see Figure 14.

relationships; but this may be a result of multicollinearity or inappropriate and insufficient explanatory variables. In such a case, the stabilization of $R^2$, the coefficient of determination, may be misleading. It is the convergence on the known values in the reference cells combined with a stabilized $R^2$ value which is important.

In the Rouyn-Noranda region, the eight reference cells have a total known endowment of 3,677 million dollars. The 64-cell analysis predicts an endowment of 13,093 million dollars in these eight known endowment cells, an increase of 3½ times. Actually, this increase is mainly because of the anomalously high predicted value of $11,412 million in cell 1029. If this anomalous value is reduced to that actually known in the cell, then the total endowment predicted by the 64-cell analysis for the eight reference cells is 2,805 million dollars, or only 76 percent of that known. No further convergence can be obtained with additional iterations.

The checkerboard analysis predicts a total endowment of 3,557 million dollars compared with the 3,677 million dollars known in the eight known endowment cells. This is a close convergence, being 95 percent of that known. Cell 1016 and cell 1042 are the only ones with a relatively low convergence at 90 percent and 86 percent respectively. And unlike the 64-cell analysis, there are no anomalousy high values associated with the reference cells.

The total endowment predicted by the checkerboard set for the whole region is 3,690 million dollars. Since this set predicts a total endowment of 3,557 million dollars for the eight reference cells that have an already known endowment, the predicted value in the remaining 56 cells is 133 million dollars. About 99 percent of this amount is contained in the following cells.

| Cell | Predicted Endowment (Millions of Dollars) |
|---|---|
| 1002 | 13.78 |
| 1007 | 12.40 |
| 1009 | 2.01 |
| 1010 | 9.81 |
| 1012 | 16.02 |
| 1033 | 62.32 |
| 1041 | 15.08 |
| Total | 131.41 |

The rest of the cells can be considered as barren until such time as additional favourable information may become available.

With the exception of cell 1033, the predicted endowment values in the other cells appear to be of a low order.[1]

[1] Sangster (1976) in an oral communication observed that there is no likelihood of a large "Horne type" deposit being discovered in the Rouyn-Noranda region. He stated that smaller deposits could be found in the region.

This situation should be looked at from the fact that of the eight known endowment cells with reference to which the forecasts have been made, four have a known endowment averaging 31 million dollars, the range being from 19 to 43 million dollars. And only two of the remaining four cells have a known endowment in the billion dollars category. The other reason for a low predicted value is that as a result of iterative regression analysis, a part of the total variance is lost in each iteration so that the result tends to develop around the most likely value. This is also evident in that after eight iterations, the standard error of the estimate narrows down to 1.05 million dollars. The predicted results must therefore be simultaneously considered on their absolute values as well as their relative values. It should be clear from Table 9 that following a prediction of 2.01 million dollars for cell 1009, there is a sharp downward break in the continuity of predicted values in the remaining cells. In exploration planning therefore, these figures should mean much more than their absolute estimated values. And since the forecasts made are based strictly on geological data alone, the addition of new dimensions such as geophysical information can further define the quality of estimates for exploration decision making.

## 8.14 Variables Used

The variables used in the first regression and the succeeding four iterations are as follows:

| Variable Name | Variable Description |
|---|---|
| AREA 3 | Area of rhyolite |
| AREA 8 | Area of diorite, gabbro |
| CNTL 6 | Contact length between rhyolite, and tuff, agglomerate |
| CNTL 10 | Contact length between tuff, agglomerate, and granite, granodiorite |
| CNTL 11 | Contact length between rhyolite, and andesite, basalt |
| CNTL 13 | Contact length between rhyolite, and diorite, gabbro |
| CNTL 15 | Contact length between rhyolite, and granite, granodiorite |
| DYKE 4 | Dyke length in directions north-west to east-west |
| FOLT 4 | Fault length in directions north-west to east-west |

The selection of the variables is made on the basis of factor analysis and other considerations described in Section 8.7. The above variables are closely associated with copper and zinc as determined by factor analysis.[1] They are also consistent with the geological concepts as related to the volcanogenic nature of base metal deposits in the region.

After four iterations, to improve convergence and

[1] Factor #3 for both copper and zinc.

the coefficient of determination, $R^2$, the following variables are added to those already in the equation:

| Variable Name | Variable Description |
|---|---|
| AREA 11 | Area of granite, granodiorite |
| DYKE 3 | Dyke length in directions north-south to north-west |
| FOLT 1 | Fault length in directions east-west to north-west |

The selection of these variables is also based essentially on factor analysis. These variables were not included in the initial stage of iteration for the following reasons:

(i) They are less relevant than the variables included earlier. This is evident from both geological concepts on ore genesis, and also from the fact that while the earlier included variables are associated with factor #3, for both copper and zinc, these variables are associated with the less important factors, #13 and #7 for copper and zinc respectively.[1]

(ii) DYKE 3 and FOLT 1 are very highly correlated with known endowment in the region. Their relationship with the known endowment appears

[1]Factor #13 for copper and factor #7 for zinc.

strictly spatial because they are post-ore features. When these variables are included in the early stages of iteration, they completely dominate the regression equation so that even those cells that have post-ore rock formations in them are predicted as favourable if DYKE 3 and FOLT 1 are present in them. It is essential, therefore, to build the base of regression model on variables believed to be genetically related with ore formation. But since DYKE 3 and FOLT 1 may have been a consequence of ore forming processes, their influence must also be considered. For this reason, they are added to the already input variables at a stage where an increase in the $R^2$ value and/or the convergence on known endowment starts to taper off with further iterations. In the present study, this stage is reached after four iterations.

Regression equations are obtained for each of the two components[1] in the checkerboard analysis. Subset A includes the known endowment cells 1016, 1021, 1023, 1039, and 1042 having a total value of 425.30 million dollars. Subset

---

[1]The two components in the checkerboard analysis are referred to as subset A and subset B.

B contains the known endowment cells 1013, 1029, and 1038 having a total value of 3251.93 million dollars. The regression equations for the two subsets after the eighth iteration are as follows:

<u>Subset A</u>

$$
\begin{aligned}
\log_{10} \text{DOLLAR} = & -4.265603 + 0.2859512 \times \text{CNTL } 11 \\
& + 0.7733751 \times \text{AREA } 3 \\
& + 3.526068 \times \text{AREA } 8 - 0.6875897 \times \text{CNTL } 13 \\
& + 0.6636013 \times \text{CNTL } 6 - 0.07064882 \times \text{FOLT } 4 \\
& + 0.1054084 \times \text{FOLT } 1 - 42.99437 \times \text{DYKE } 4 \\
& + 1.081944 \times \text{DYKE } 3 + 0.04084261 \times \text{AREA } 11 \\
& - 1.120562 \times \text{CNTL } 15 + 2.736221 \times \text{CNTL } 10 \\
& (R^2 = 0.999)
\end{aligned}
$$

<u>Subset B</u>

$$
\begin{aligned}
\log_{10} \text{DOLLAR} = & -4.067191 + 1.900224 \times \text{AREA } 3 \\
& - 0.2760116 \times \text{CNTL } 11 + 2.679324 \times \text{AREA } 8 \\
& - 0.2741231 \times \text{CNTL } 6 - 0.2130330 \times \text{CNTL } 13 \\
& + 0.9608059 \times \text{DYKE } 3 - 3.327689 \times \text{DYKE } 4 \\
& + 2.119953 \times \text{FOLT } 4 - 0.3091724 \times \text{AREA } 11 \\
& - 0.02074412 \times \text{FOLT } 1 - 2.264679 \times \text{CNTL } 15 \\
& - 9.070831 \times \text{CNTL } 10 \\
& (R^2 = 0.999)
\end{aligned}
$$

As shown, the logarithm to the base 10 of dollar value of endowment is used as the response variable instead of the untransformed dollar value. The conversion used is:

$$DOLLAR = \log_{10} (\text{Dollar Value} + 0.0001)$$

The constant term 0.0001 is added so that zero values become amenable to logarithmic conversion. The objective of the conversion is to have a moderating effect on the extreme values of the untransformed response variable. The logarithmic conversion also reduces skewness but it cannot eliminate it.[1]

The relative contributions of individual explanatory variables used are shown graphically in Figures 17 and 18. These are also discussed in Section 8.18.

## 8.15 Statistical Stability of the Geological System

In iterative regression analysis, the geological relationships are assumed to be in a state of equilibrium when all the response variables become consistent and do not change with additional iterations. The consistency in the value of the 56 unknown response variables is attained with respect to relationships between the explanatory variables and the unknown endowment in each of the eight reference cells. To test the equilibrium of the system, any one of the non-zero response variables can be assigned a zero value and the regression performed as before to see if convergence does take place. For

[1] This point has been discussed in Sections 4.7 and 8.8.

this reason, each response variable with an originally known endowment has been, one at a time assigned a zero value, and regression performed iteratively on the checkerboard type data.[1] The results are shown in Table 10 and Figure 15. The results shown are for the eight known endowment cells only since they are the ones that can indicate the validity of the relationships developed.

A close convergence of predicted and known endowment values is obtained for all reference cells except cells 1023, 1038, and 1039.

The endowment in cell 1023, as has been pointed out, occurs in a different geological environment, being associated with tuff and agglomerate. No other known deposit in the region has this kind of association. Therefore, when this cell is assigned a zero value, it is not possible for the relationships in the remaining cells to predict any endowment in it. It is also observed from Table 10 that when iterative regression analysis is performed, while convergence in each of the remaining seven known endowment cells takes place gradually, it is cell 1023 that immediately receives its full known endowment in the first regression run. This occurs because it is associated with an environment not present in the

---

[1]The response variable $\log_{10}$ (Dollar + 0.0001) was assigned a zero value. This in effect means a starting point of one million dollar in an untransformed state of response variable.

## TABLE 10[1]

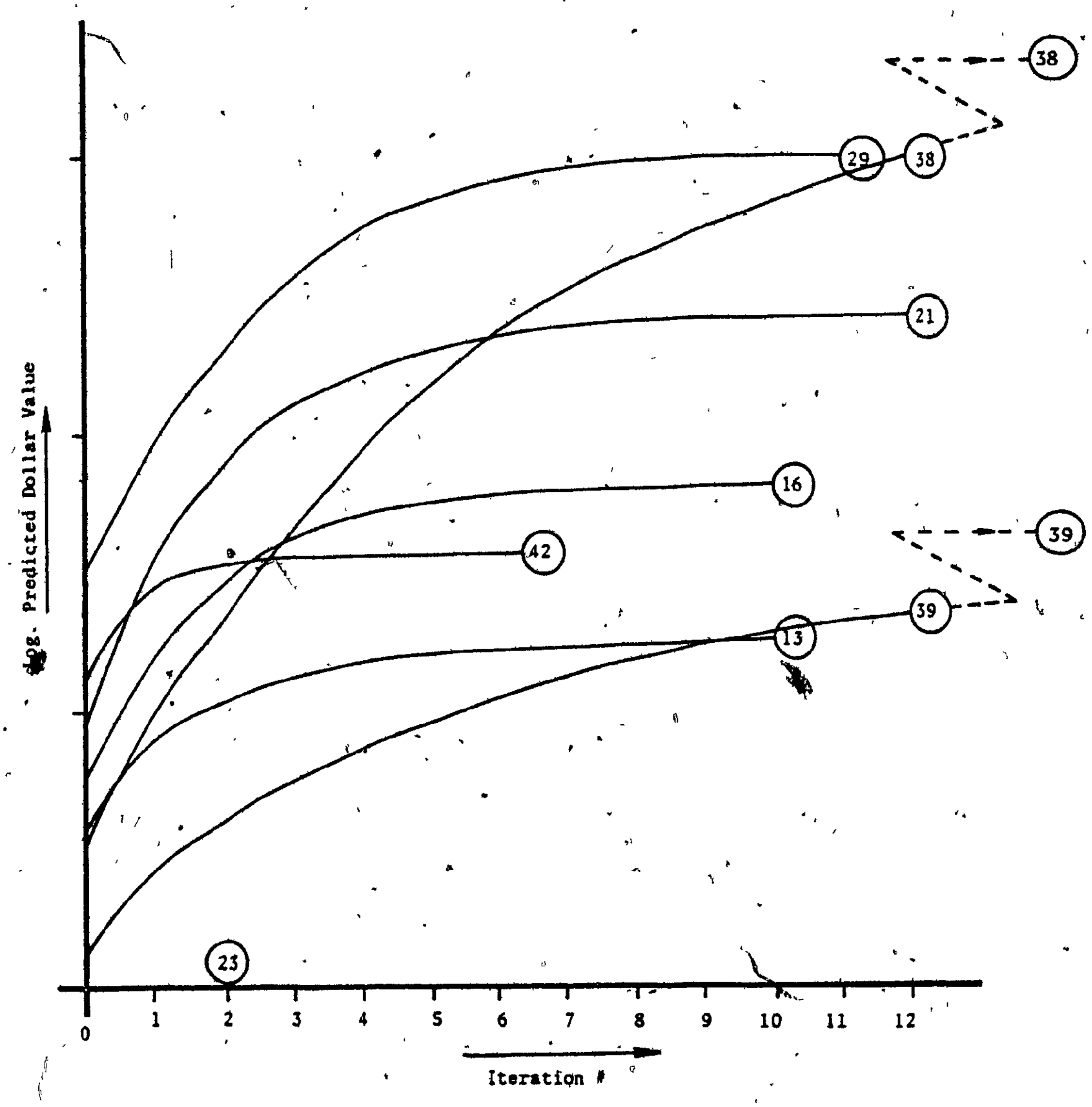
ITERATIVE REGRESSION ESTIMATE CONVERGENCE ON KNOWN ENDOWMENT VALUES ASSUMING ZERO LOG DOLLAR VALUE IN KNOWN ENDOWMENT CELLS, ONE AT A TIME

| | | | INPUT VALUE FOR RESPONSE VARIABLE | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CELL # | KNOWN $ x $10^6$ | ORIGINAL FORECAST $ (after 8 Iters.) | ZERO LOG $ VALUE | 1st FORECAST INPUT | 2nd FORECAST INPUT | 3rd FORECAST INPUT | 4th FORECAST INPUT | 5th FORECAST INPUT | 6th FORECAST INPUT | 7th FORECAST INPUT | 8th FORECAST INPUT | 9th FORECAST INPUT | 10th FORECAST INPUT | 11th FORECAST INPUT | 12th FORECAST INPUT |
| 1013 | 18.81 | 17.50 | 3.76 | 7.66 | 11.22 | 13.77 | 15.37 | 16.30 | 16.83 | 17.12 | 17.27 | 17.36 | - | - | - |
| 1016 | 74.48 | 66.72 | 6.01 | 16.78 | 30.20 | 42.28 | 51.26 | 57.24 | 60.96 | 63.21 | 64.53 | 65.30 | 65.74 | - | - |
| 1021 | 244.66 | 251.21 | 9.50 | 36.55 | 81.82 | 132.51 | 176.81 | 210.11 | 232.96 | 247.80 | 257.12 | 262.87 | 266.37 | 268.49 | 269.76 |
| 1023 | 27.80 | 27.80 | 0.00 | 0.00 | - | - | - | - | - | - | - | - | - | - | - |
| 1029 | 1123.84 | 1056.96 | 34.60 | 73.72 | 203.72 | 382.00 | 563.55 | 716.78 | 831.74 | 911.90 | 965.29 | 999.87 | 1021.88 | 1035.72 | 1044.39 |
| 1038 | 2109.28 | 2065.14 | 3.47 | 9.83 | 23.47 | 48.58 | 89.26 | 148.42 | 227.06 | 323.98 | 436.11 | 559.13 | 688.24 | 818.79 | 946.76 |
| 1039 | 34.88 | 34.37 | 1.73 | 2.76 | 4.08 | 5.66 | 7.48 | 9.44 | 11.49 | 13.57 | 15.60 | 17.55 | 19.37 | 21.06 | 22.59 |
| 1042 | 43.48 | 37.58 | 13.82 | 28.28 | 34.37 | 36.25 | 36.78 | 36.92 | 36.96 | - | - | - | - | - | - |

[1] See also Figure 15.

FIGURE 15

CONVERGENCE ON KNOWN ENDOWMENT VALUES USING ITERATIVE REGRESSION ANALYSIS[1,2]




[1]See text for explanation.

[2]Figures in circles are the last two digits of known endowment cell numbers.

other cells.

Cells 1038 and 1039 both show a convergence similar to, but slower than the other known endowment reference cells. That complete convergence does take place, is apparent from the trends of their curves in Figure 15. These curves have a shape similar to that in every other case, but because of a flatter slope, need an additional number of iterations to fully converge on their known endowment.

## 8.16 The Magusi River Cell

The Magusi River deposit and the New Insco deposit are recent discoveries.[1] They occur close together in the south central part of Hébécourt township adjoining the Rouyn-Noranda region.[2] The two deposits have an estimated total value of more than 75 million dollars.

A grid of cells similar to that used over the Rouyn-Noranda region was superimposed over Hébécourt township, and geological parameters measured in the cell that contains the Magusi River and the New Insco deposits. The measurements are made on a 1" = 1000 ft. quarter township map prepared by the Quebec Department of Natural Resources. This cell is referred to as the Magusi River cell.

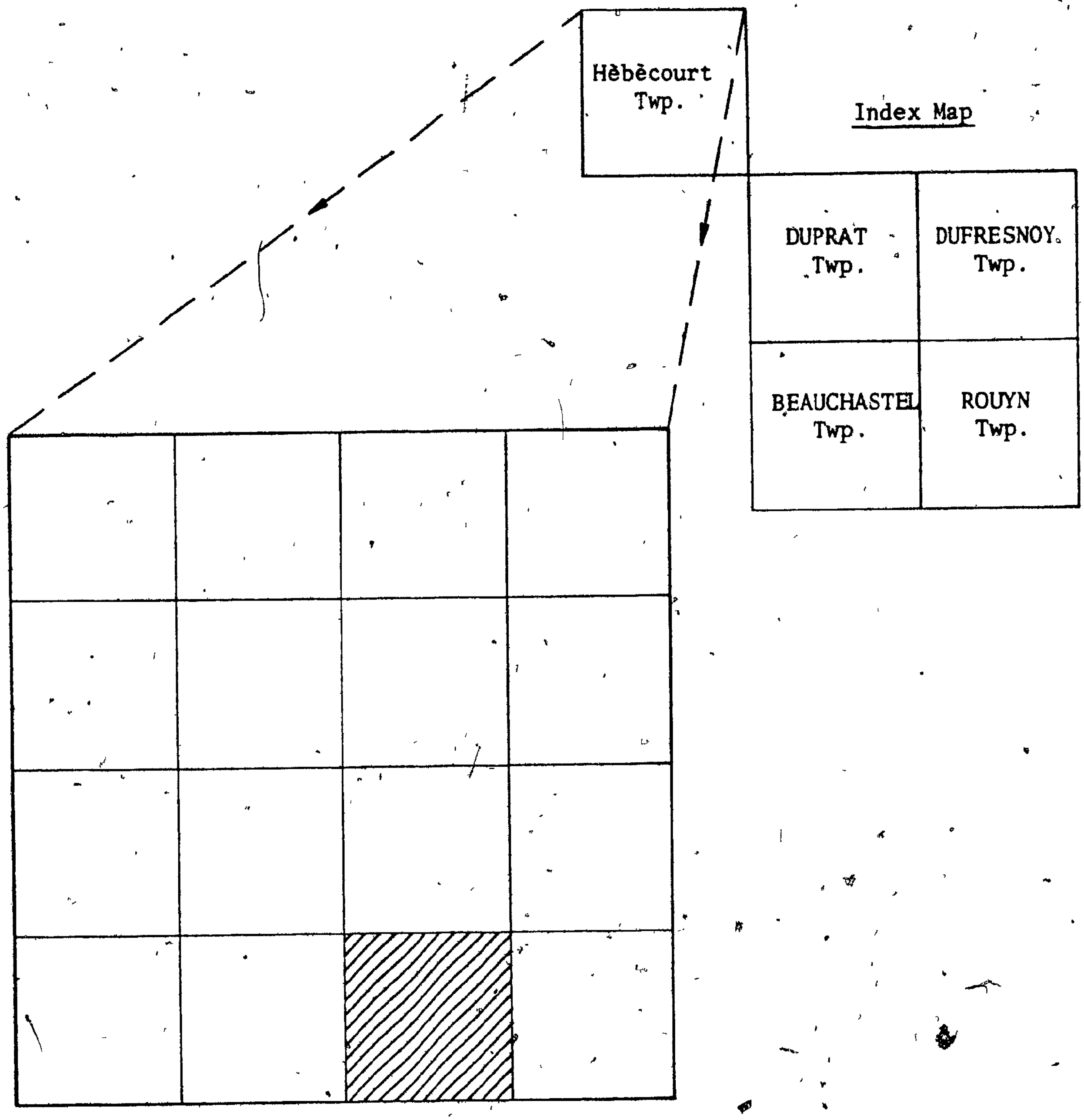
---

[1]These discoveries were made in 1972/73.

[2]See Figure 16.

FIGURE 16

THE MAGUSI RIVER CELL[1]




---

[1]Shown hatched. Rouyn-Noranda, the study region, is comprised of the four townships, Duprat, Dufresnoy, Beauchastel and Rouyn.

The explanatory variables data of the Magusi River cell are added to the checkerboard subsets A and B. The response variable is assigned a zero value in the Magusi River cell while the response variables in all other cases retain their values as calculated by iterative regression analysis on the checkerboard data. An iterative regression analysis[1] is performed on the two subsets, and convergence obtained after four runs. The results are shown in Table 11.

A study of Table 11 shows that different estimates of endowment are obtained for the Magusi River cell using the two checkerboard analyses. The question, therefore, arises as to which of the two forecasts should be considered. The results are different in the two cases because different regression equations are computed for the different relationships present. The two subsets have different known endowments and thus different endowment--geology relationships. Obviously, the choice should depend on the level of similarity between the geological relationships computed in each subset and those existing in the Magusi River cell.

Each of the two checkerboard subsets has attained a state of statistical equilibrium following iterative regression analysis. In other words, each response variable has been assigned a certain value such that further iterations

---

[1]The explanatory variables used are the same as in the earlier checkerboard analysis. See Section 8.14.

## TABLE 11

PREDICTED VALUES IN THE MAGUSI RIVER CELL USING ITERATIVE REGRESSION ANALYSIS

| Run # | SUBSET A Predicted Value (Dollars) | SUBSET A $R^2$ | SUBSET B Predicted Value (Dollars) | SUBSET B $R^2$ |
|---|---|---|---|---|
| 1 | 68,134 | 0.948 | 13,128 | 0.973 |
| 2 | 320,389 | 0.997 | 22,774 | 0.999 |
| 3 | 462,601 | 0.999 | 31,163 | 1.000 |
| 4 | 504,731 | 1.000 | 31,719 | 1.000 |

do not significantly affect the stability of the system being regressed. When a cell with an unknown endowment is added to the cells in the system under equilibrium, an arbitrary zero value has to be assigned to the response variable of such a cell if its endowment is not known.[1] Performing regression analysis on the new set of cells[2] will tend to disturb the equilibrium of the system. The greater the similarity of geology in the cell added to that in the system, the greater will be the disturbance shown in the predicted values of response variables. This is because the response variable in the added cell is assigned a zero value for lack of information on its unknown endowment. The disturbed state of equilibrium will become evident in the change brought about in the coefficient of determination. This coefficient in a system under statistical equilibrium is normally close to one.

It is observed in Table 11 that following the first regression run, the coefficient of determination drops to 0.948 in subset A and 0.973 in subset B, from its original value of one. It is also observed that the geological system achieves equilibrium after four runs in subset A and after only two in subset B. It is therefore apparent that the

---

[1] In this study, the Magusi River cell is assigned an arbitrary zero value to test if it is possible to predict its known endowment from the existing relationships.

[2] Or, observations.

Magusi River cell has more geological similarity with subset A than with subset B.

The endowment predicted in the Magusi River cell after three iterations with subset A is 504,731. Compared to the known endowment in this cell, the predicted estimate is very low. It is, however, observed that while the total computed endowment in subset A[1] is 519 million dollars, it increases to over 527 million dollars when the Magusi River cell is added to the analysis. In another set of runs, the data on Magusi River cell is combined with the 32 original un-regressed observations of subset A, and regression performed iteratively. After eight iterations, the system reaches equilibrium, and the total value of endowment predicted in the subset is now estimated at 667 million dollars. This means an increase of 148 million dollars. Not all this increase can be attributed to the Magusi River cell. In fact, the predicted value for this cell by this method is only $346,276. But the increase itself is a measure of the improved favourable relationships resulting from the addition of this cell to the subset.[2]

---

[1] After eight iteration. See Section 8.13.

[2] A similar set of runs was made with the subset B. The original estimate of endowment over the 32-cell subset is 3,171 million dollars. This increases to 3,234 million dollars, an increase of 63 million dollars following the addition of the Magusi River cell to the original observations and iteratively regressing to convergence.

To observe further the effect of an additional observation on the existing equilibrium of a geological system being regressed, two more runs were made. In the first, the data of cell 1021 was repeated, so that the repeat becomes an additional observation. In the second run, the data of cell 1032 was similarly repeated. The objective was to observe the effect that a high and a low valued cell would have on the equilibrium that a geological system has achieved. Cell 1021 is a known endowment reference cell valued at 244.66 million dollars with a predicted value of 251.21 million dollars. Cell 1032 has no known endowment, and a predicted zero endowment.

Regression, following the addition of cell 1021 has an insignificant but positive effect on the originally estimated endowment values.[1] But this should be expected. When a high valued cell has its data repeated and used as a new observation, it re-emphasizes the existing response-explanatory variables relationships. The system is not disturbed and remains in equilibrium as before. Similarly, the repeat as an additional observation of the barren cell 1032 data has no effect on the equilibrium of the system, and changes in the estimates of endowment are negligible.

In the above runs, there is no effect on the equilibrium of the system because the relationship between response

---

[1]That is, the estimates of response variables in the checkerboard analysis after eight iterations.

and explanatory variables of the repeated cells has already been accounted for in the iterative analyses leading to convergence and equilibrium. On the other hand, when the Magusi River cell; which has some level of similarity with the known endowment cells in the existing system is included, it causes disequilibrium until its response variable has been assigned a value commensurate with the interaction of geology represented by its explanatory variables. But this does not explain the low value predicted in the Magusi River cell even though it otherwise shows a high degree of favourability as explained earlier.

Since the predicted value of the Magusi River cell is anomalously low compared to its known endowment, the cell probably represents a different geological environment locally. Or, the predictive efficiency of the system itself may be poor. The predictive efficiency of the system has already been validated by performing iterative regression analysis with a dummy zero endowment value assigned to known endowment cells, one at a time in each series of runs.[1] However, in the present set of runs, the explanatory variable data of known endowment cell 1021 is repeated as an additional variable. The response variable for this cell is assigned a zero value. The repeated cell data is added to the checkerboard subset A and regression performed iteratively. Convergence

[1]See Figure 15.

is obtained after five iterations. The predicted values of the response variable are shown below for each iterative run.

| Run # | Predicted Value ($\$ \times 10^6$) |
|---|---|
| 1 | 1.02 |
| 2 | 32.30 |
| 3 | 117.71 |
| 4 | 191.01 |
| 5 | 228.95 |
| 6 | 245.02 |

The known endowment of the cell is 244.66 million dollars, and therefore a near perfect convergence is obtained after six runs. This is a measure of the efficiency of the system in equilibrium. But such a convergence is only possible when the added observation lies within the confines of the geological system and not outside it.

The Magusi River cell is rich in pyroclastics; almost one-third of the cell area has been mapped as tuff, agglomerate (AREA 2). Should some of the material presently classified as tuff, agglomerate prove to be rhyolite (AREA 3), the predicted value of endowment will increase significantly. This is possible because the Magusi River area has not yet been studied in as much detail as the Rouyn-Noranda region. As an extreme example, if all the tuff, agglomerate is assumed to be rhyolite, then without any changes[1] in the other explanatory variables, the predicted value of endowment in the Magusi River cell is 17.56 million dollars with checkerboard subset A and 131.55 million dollars with subset B. These estimates are respectively 34 and 603 times the predictions made when the formation, tuff, agglomerate itself is used in the regression models. This is the reason why it is emphasized that the most effective predictions are made within the limits of a system within which the classification of rocks is consistent.

Conversely, if some of the material classified as rhyolite in the Rouyn-Noranda region be regarded as tuff, agglomerate, the regression models developed will have a similar favourable effect on the estimate of endowment in the Magusi River cell.[2]

The reason for the low predicted value of endowment in the Magusi River cell lies essentially in its lying outside the Rouyn-Noranda region which is considered as a closed geological system for statistical analysis. On a local level, the Magusi River cell may lie in a different geological

---

[1] E.g., contact lengths.

[2] Sakrison (1966) stated that most of the rock classified as rhyolite in the Rouyn-Noranda region is of pyroclastic origin. Similarly, Larson and Webber (1977) have indicated that the proportion of pyroclastics in the region is considerably more than reported. See detail in Section 3.2.2.

environment in a litho-stratigraphic sense. However, some of the differences in lithology may be related to differing rock classifications.

## 8.17 Certain Aspects of Known Endowment Cells

In evaluating performance of a statistical analysis, it is useful to be aware of the relationships existing among the cases being used as calibrators for predictive purposes. Such cases in this study are the known endowment reference cells.

Regression analysis[1] has therefore been performed over the 64-cell data but endowment is assumed to be present in only one reference cell. The response variable in this cell is assigned a value of one while the response variables in the remaining seven reference cells and the 56 unknown endowment cells are all assigned a value of zero. After analysis, the procedure is repeated, one at a time for each of the remaining reference cells. A total of eight such runs is thus made. The results obtained are essentially measures of similarity of the particular reference cell of assumed present endowment with the other cells with an

[1]The procedure is similar to that used by Agerberg et al. (1972), but has been modified for the objective described above. See also Section 2.2.3.

assumed zero endowment. Since the response variables have been assigned dichotomous values, i.e., zero and one, the analysis is much like a two-group discriminant analysis except that instead of the cells being assigned to either one of the two groups, the estimated values of the response variable are continuous. These values have been converted into percentages,[1] and are shown for the reference cells only, in Table 12.

One of the striking observations in Table 12 is that individual reference cells[2] are unable to predict any endowment in cell 1042. It is predicted with a 100 percent probability when cell 1042 itself is the one with an assumed present endowment; but then, it does not predict endowment in any other cell either. This uniqueness may be attributed to a different local environment in the cell, it being away from the main cluster of ore deposits in the centre of the region. Cell 1042 is also unique in containing post-ore stocks of syenite-monzonite.

Cell 1021 on the other hand gives positive predictions in all other cells except 1042 and 1016. This cell

---

[1]The results should theoretically lie between zero and one, and are equivalent to probabilities of occurrence of endowment without regard to economics, tonnage or grade. These probabilities have been multiplied by 100 to obtain percentages.

[2]In the description related to Table 12, the term cell means the known endowment reference cell.

## TABLE 12

### PROBABILITY OF OCCURRENCE OF ENDOWMENT IN KNOWN ENDOWMENT CELLS USING ONE REFERENCE CELL AT A TIME

| Probability[1] of Endowment Occurrence in Cell # | Reference Cell Numbers[2] with Assumed Present Endowment | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1013 | 1016 | 1021 | 1023 | 1029 | 1038 | 1039 | 1042 |
| 1013 | 90.70 | -2.53 | 19.09 | -0.30 | -4.94 | -2.78 | 4.33 | 0.00 |
| 1016 | -5.69 | 88.49 | -2.13 | 2.31 | 8.48 | 0.71 | 3.22 | 0.00 |
| 1021 | 18.83 | -3.77 | 43.00 | 2.39 | 25.48 | 2.39 | -3.91 | 0.00 |
| 1023 | 0.06 | 2.76 | 2.32 | 99.17 | -1.98 | -1.19 | 0.69 | 0.00 |
| 1029 | -6.34 | 8.22 | 24.09 | -1.78 | 83.10 | 1.93 | -0.43 | 0.00 |
| 1038 | 1.19 | 0.28 | 4.64 | -0.47 | -0.34 | 95.15 | -1.37 | 0.00 |
| 1039 | 4.19 | -0.66 | 0.19 | 0.64 | -0.84 | -0.78 | 94.49 | 0.00 |
| 1042 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 100.00 |

[1] Probabilities have been expressed in percentages.

[2] Prediction made by a reference cell for itself is underlined.

therefore is more similar to other reference cells. Spatially, it lies within the main cluster of ore deposits in the region.

Both cells 1042 and 1016 are situated away from the cluster of ore deposits lying in cells 1013, 1021, 1029, and 1038. Dugas (1977) in an oral communication points out that the Mobrun deposit lying in cell 1016 is believed to be at a higher stratigraphic level than the other deposits in the region. He further points out that there are no dykes associated with this deposit or occurring in its immediate surroundings. But cell 1016 does have some similarity with other reference cells as evidenced by its positive predicted values in cells 1023, 1029, and 1038.

It is obvious that when only one response variable out of 64 has a value of one assigned to it, and all others are assigned zero values, the regression equations will be different in each regression run. The fact that some of the regression equations cannot predict the presence of endowment in other cells is an evidence that with changing or evolving geological environments, the predicted values may be unexpected, unless and until the regression runs are either confined to the limits of the system or a broad enough response variable information is available to generate a model that reflects the variabilities of explanatory variables.

Another set of runs has therefore been made using a similar approach, but this time assuming endowment to be present in seven of the eight reference cells. From them the probability of occurrence of endowment in the eighth cell is estimated. Eight separate runs are made each with a zero value assigned to the response variable of a different reference cell. A value of one is assigned to the response variables of the seven other reference cells. As before, the unknown endowment cells are assigned a zero value for their response variables. The results of probabilities generated for the reference cells are shown in Table 13.

It is observed from the table[1] that negative probabilities are obtained in case of cells 1016 and 1039, and a zero probability is obtained in cell 1042. These runs confirm the conclusion described earlier that cells 1016 and 1042 represent a locally different geological environment of ore occurrence. Both cells 1016 and 1042 are located away from the main concentration of ore deposits. Further, cell 1016 is very rich in zinc while cell 1042 has no reported zinc in it. Similarly, cell 1039 is zinc-rich even though, the adjoining cell 1038 is copper-rich. It therefore may be related to another cycle of activity within the larger volcanogenic environment in the region. The relationships between intermediate cases are various measures of their similarities as shown in the table.

The purpose of the above exercise has been to

---

[1]See Table 13.

## TABLE 13

PROBABILITY OF OCCURRENCE OF ENDOWMENT USING SEVEN REFERENCE CELLS AT A TIME AND ASSUMING ZERO ENDOWMENT IN THE EIGHTH

| Ref. Cell # | Known Probability % | Predicted by 8 Cells | Predicted Probability by Excluding Cell # (%) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 1013 | 1016 | 1021 | 1023 | 1029 | 1038 | 1039 | 1042 |
| 1013 | 100 | 98.12 | 8.70 | 99.53 | 80.05 | 98.60 | 101.31 | 95.57 | 97.30 | 94.15 |
| 1016 | 100 | 90.52 | 96.83 | -1.13 | 90.61 | 87.92 | 80.54 | 85.32 | 88.05 | 85.89 |
| 1021 | 100 | 84.17 | 62.74 | 99.88 | 44.68 | 78.75 | 68.58 | 94.25 | 88.21 | 95.88 |
| 1023 | 100 | 103.85 | 103.29 | 100.34 | 100.47 | 4.32 | 104.08 | 103.26 | 103.47 | 102.93 |
| 1029 | 100 | 101.27 | 108.80 | 99.68 | 77.52 | 103.80 | 24.72 | 105.41 | 104.18 | 105.17 |
| 1038 | 100 | 98.01 | 98.48 | 100.43 | 95.93 | 98.18 | 98.51 | 0.54 | 99.60 | 100.18 |
| 1039 | 100 | 92.85 | 90.94 | 99.95 | 96.61 | 93.16 | 100.47 | 100.00 | -3.51 | 100.37 |
| 1042 | 100 | 100.02 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 0.00 |

[1] Probabilities have been converted into percentages.

[2] Assumed unknown (zero) endowment cells in a particular run are underlined.

observe the relationships among the known endowment cells. The conclusion is that in spite of the broader similarities among reference cells, there are significant differences in their localized geological environments as determined statistically. All forecasts have been made in reference to these cells, and it is inevitable that some resolution is lost because of the local differences that are present. Once an equilibrium has been attained in the system with iterative regression analysis the model will reflect the joint environment observed in the reference cells. It will not be able to forecast any endowment the environment of which has not been considered in the modelling process. The case of the Magusi River cell appears to belong to this category.

The results shown in Tables 12 and 13 should have a bearing on the genetic history of the region. It is well accepted now that the Rouyn-Noranda region was the centre of volcanic activity that led to the formation of massive base metal deposits. But the different environments associated with ore deposits in their respective cells indicate a pulsatory nature of the volcanism. This thought should agree with the southward "younging" trend in metavolcanics observed across the region by Krogh and Davis (1971), and with the postulate of Spence and Spence (1975) that rhyolites of several different ages are present in the region. And Roscoe (1965) has described the metallic zoning in massive sulphide deposits in the region in which zinc lies at a stratigraphically higher level than copper. These aspects may also be observed regionally in that the southern-most reference cell 1042 has only copper endowment and no reported zinc with it, while the northern-most and outward-most cells 1016 and 1039 respectively are very zinc-rich, and relatively poor in copper.

If volcanism was indeed pulsatory, as appears to be the case, then the predictive effectiveness of the model will be affected when extrapolated outside the immediate environment of the system under study. Alternatively, the model should be modified to accommodate the variabilities in explanatory variables of newer areas. This aspect is mentioned to indicate the usefulness of regression analysis in understanding, confirming or appraising for modification, the existing theories of ore formation in the region. It should be pointed out here that while most comparative studies on the geology of massive sulphide deposits have emphasized features of similarity, their differences have not been as well reported upon. Regression analysis can be a useful tool in this regard. Needless to say, information on differences among massive sulphide deposits can help exploration as much as a knowledge of the similarities in them.

## 8.18 Role of Explanatory Variables in Iteration

The explanatory variables have been selected on the basis of factor analysis. Their input in the regression model has been controlled so that the first two input variables are, area of rhyolite (AREA 3), and the contact length between rhyolite and andesite, basalt (CNTL 11). Rhyolite is the most common host rock, and its contact with the andesite-basalt group of rocks the most persistent stratigraphic feature in ore localization. These are followed by other explanatory variables the inputs of which are determined by a combination of pre-selected inclusion levels and the partial correlation coefficients. The regression models obtained in the checkerboard analysis are given in Section 8.13. A better comprehension of the role of variables can be obtained by studying their standardized coefficients.[1] For comparison, the standardized coefficients have been cumulated and then converted into percentage of the total. The results of the checkerboard subsets A and B are shown in Figures 17 and 18.

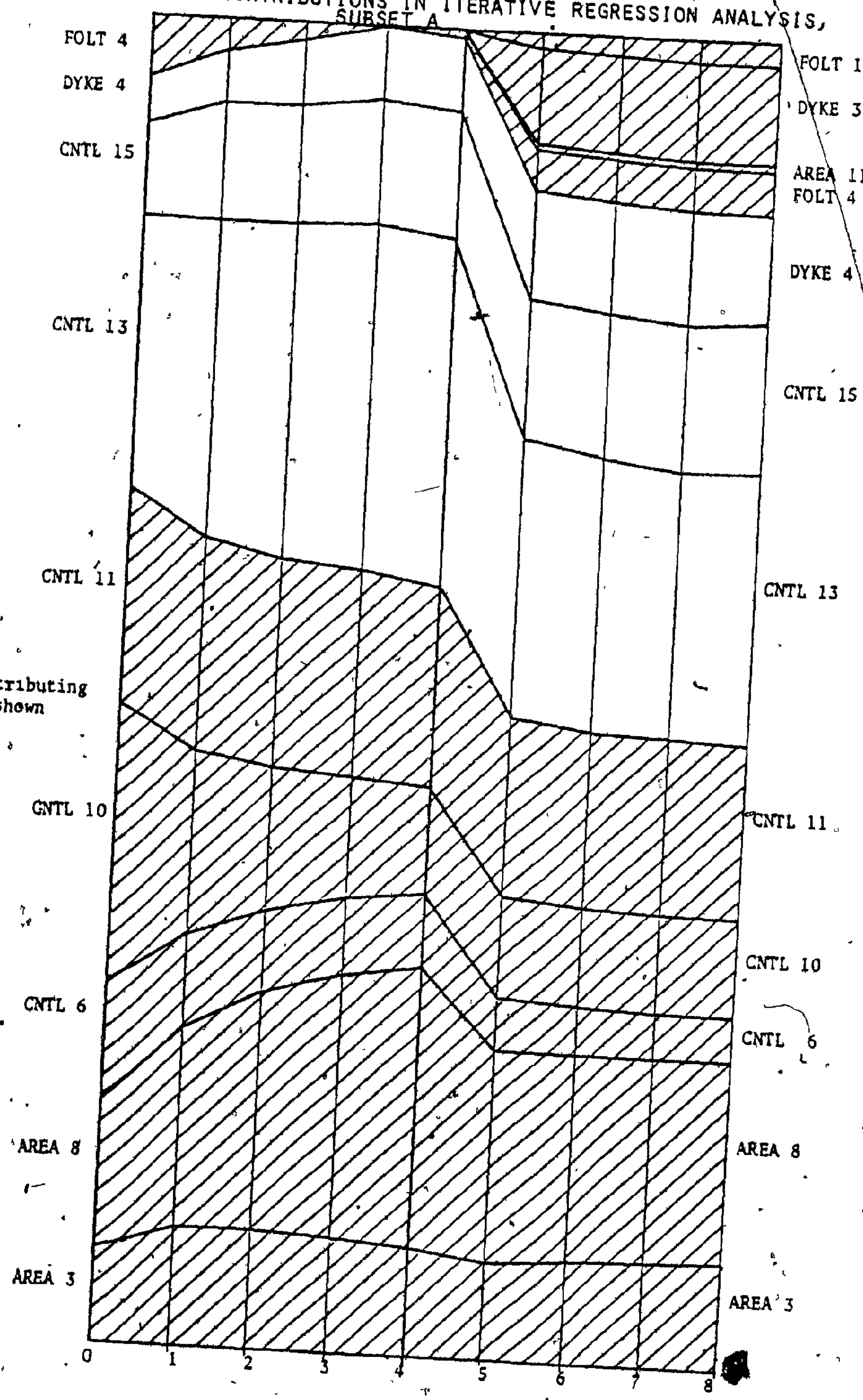
It is observed from Figures 17 and 18 that AREA 3,[2] and AREA 8 both contribute positively to the models in each of the subsets. In addition, the structural elements FOLT 4, FOLT 1 and DYKE 3, all contribute positively to the regression models in each of the subsets.

---

[1]When explanatory variables are measured in different units such as areas and lengths, their relative contributions to the regression model can be studied by standardizing them in unitless form. This is done by computing the model on the standardized variable values rather than the original values.

[2]See Section 4.2 for description of variable names.

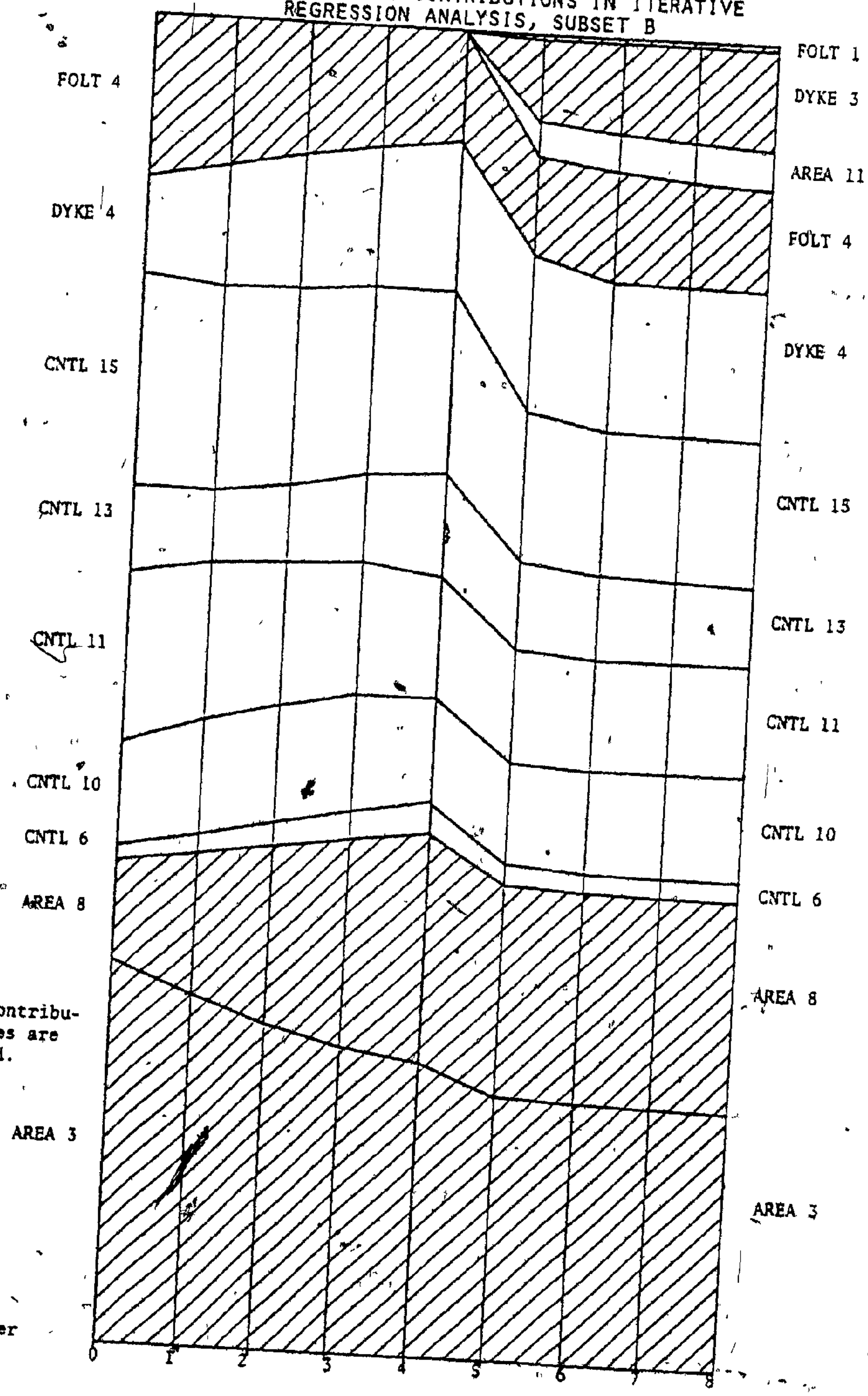
FIGURE 17
EXPLANATORY VARIABLES CONTRIBUTIONS IN ITERATIVE REGRESSION ANALYSIS, SUBSET A




NOTE:

Positively contributing variables are shown hatched.

FIGURE 18
EXPLANATORY VARIABLES CONTRIBUTIONS IN ITERATIVE REGRESSION ANALYSIS, SUBSET B




NOTE:

Positively contributing variables are shown hatched.

On the other hand, CNTL 13, CNTL 15, and DYKE 4, all contribute negatively to both regression models.

The role of structural elements in the region has not been fully resolved, either with respect to one another or with respect to volcanism and related ore deposits. The positively contributing structural elements, however, show some relationship: FOLT 4 and FOLT 1 are both essentially east-west trending features and may be related to the east-west anticlinorium structure of the region. Further, DYKE 3 cuts across the directions of FOLT 1 and FOLT 4, and may be related to them. Any other comments beyond this would be speculative.

The positive contribution of AREA 3 is understandable, but not that of AREA 8. The positive contribution of AREA 8 may be a result of purely spatial correlation or to a deeper underlying cause as reflected by the belief of some regional geologists[1] that diorite and gabbro indicate channelways of the extruded andesites and basalts.

The negative contributions of CNTL 13 and CNTL 15 to each of the two regression models are clear inasmuch as the ore deposits are considered of volcanogenic origin. CNTL 13 and CNTL 15 represent the contact lengths of rhyolite with diorite, gabbro, and with granite, granodiorite

---

[1]E.g., Wilson (1962), Van de Walle (1972), and Spence and Spence (1975).

respectively. Since the ore deposits are believed to be volcanogenic and occurring at a certain evolutionary stage of volcanism, it is more likely that ore localization will take place between two volcanic formations rather than between a volcanic formation and an intrusive. In addition, the presence of an intrusive can only have an assimilative or dispersive effect on any existing massive sulphides. These comments are valid only if the volcanogenic origin of ore deposits is assumed, and may not hold if a hydrothermal epigenetic concept is invoked.

In respect of the remaining variables, their contributions, whether positive or negative, depends upon the particular response-explanatory variables relationships present in the known endowment reference cells in each of the two subsets A and B. For example, subset A has the West Macdonald ore deposit occurring in tuff, agglomerate in cell 1023. CNTL 10, the contact between tuff, agglomerate, and granite, granodiorite, therefore contributes positively to the model. Since no such deposit occurs in the subset B cells, the presence of CNTL 10 in its model contributes negatively.

In subset B, CNTL 11 makes a negative contribution. This may appear unexpected but as observed on the geological map of the region, cell 1038 which has the highest known endowment and thus is a major contributor to the equation, is largely composed of AREA 3 and contains relatively minor

CNTL 11.[1] AREA 3 is correlated with CNTL 11 being one of its component members; because of its predominant presence in the cell, it makes up for the contact length's negative role. The role of multicollinearity should not be forgotten when simultaneously, the areas of formations and the contact lengths between them are considered as explanatory variables. The objective is to balance the role of variables so that in their final form, they best describe the system.

Figures 17 and 18 show changes in the performance of explanatory variables with each iteration, and after the fourth iteration, the overall effect when three more variables are added. As the figures show, the system rapidly converges to equilibrium after the fifth iteration because the relative contributions of variables become consistent. This also means that the system has stabilized with respect to the values estimated for the response variables. The figures also indicate that the predictions based on a single regression run will not be efficient and will be likely to change both with further iterations or addition of variables.

## 8.19 Other Regression Tests

In this section, a review is given of a number of

---

[1]CNTL 11 is the contact length between AREA 3 and AREA 4.

regression tests that preceded the checkerboard models. The objective of the earlier runs was to observe the response and efficiency of the variables under regression, and to make various manipulations leading to the procedure finally adopted. All these runs are non-iterative, and therefore, the predicted results are of a low order of magnitude. As has been described earlier, the low predicted estimates result when the response variables in cells with no known endowment are arbitrarily assigned zero values, thus in effect attenuating the response explanatory variable relationships that are developed from the known endowment cells.

### 8.19.1 The 8-Cell Model

In resource evaluation studies using regression analysis, if the values of the response variable[1] are not known in most of the cells, and if all cells are believed to be part of the same geological system,[2] a regression model can be formulated based only on cells in which the response variable is known. Such a model can then be extrapolated over the remaining cells to make estimates of the unknown response variable. The advantage of this approach is its

---

[1] I.e., mineral endowment.

[2] I.e., they should have been drawn from the same population.

simplicity; also, no initial assumption has to be made regarding the values of response variable in the unknown endowment cells. With only eight cells out of 64 that have a known endowment, the sample size is too small for an effective model, particularly because dissimilarities exist among the reference cells themselves. And with only eight reference cells, the number of explanatory variables that can enter the equation cannot exceed six in addition to the constant term. Such an equation can only be effective if the system is such that each cell contains at least several of the explanatory variables present in the equation. Most of the 67 explanatory variables measured in the region are not present in the known endowment cells, and are therefore not a part of the regression model. Many of these do not even have any relevance with ore formation, but are present in varing amounts in about one-fourth of the region. The application of the regression model should give a zero prediction of endowment in cells that do not have any variables present in the model. But with a positive constant term in the model, the predicted value in the potentially barren cells becomes equal to the constant term. For example, observe the following regression equation.

$$\begin{aligned}\text{Log}_{10}\text{ Copper} = \;& 4.645 - 1.308 \text{ CNTL } 7 + 0.335 \text{ FOLT } 1 \\ & - 0.018 \text{ CNTL } 13 - 0.180 \text{ AREA } 4 \\ & + 0.329 \text{ CNTL } 34 - 0.398 \text{ AREA } 8\end{aligned}$$

All cells that do not have any of the above six variables will still obtain a predicted value of 44,157[1] tons of copper worth over 55 million dollars. In fact, if a cell had only post-ore sedimentary formations, the mere presence of FOLT 1 in it will give a predicted value equal to the constant term plus the positive contribution of FOLT 1 relative to its length. The resulting forecast can then exceed 55 million dollars. Most of the other equations calculated for copper and zinc using different combinations of variables are beset with similar problems.

The above condition is the same as trying to apply a model based on a certain geological system to a different system. This can to an extent be controlled by using larger sized cells so that a greater variety of explanatory variables is present in them, but for a region of fixed size, the number of observations is thereby reduced, and the objective of using small-sized cells cannot be fulfilled. Further, since ore deposits tend to occur as clusters, the number of known endowment cells becomes further reduced.

At a purely reconnaissance level, a broad based regression model can be formulated by incorporating data from known endowment cells outside the study region, taking care that there is a similarity in the characteristics of the geological environment between the study region and those

---

[1]44,157 is the antilogarithm of 4.645, the constant term.

outside it.

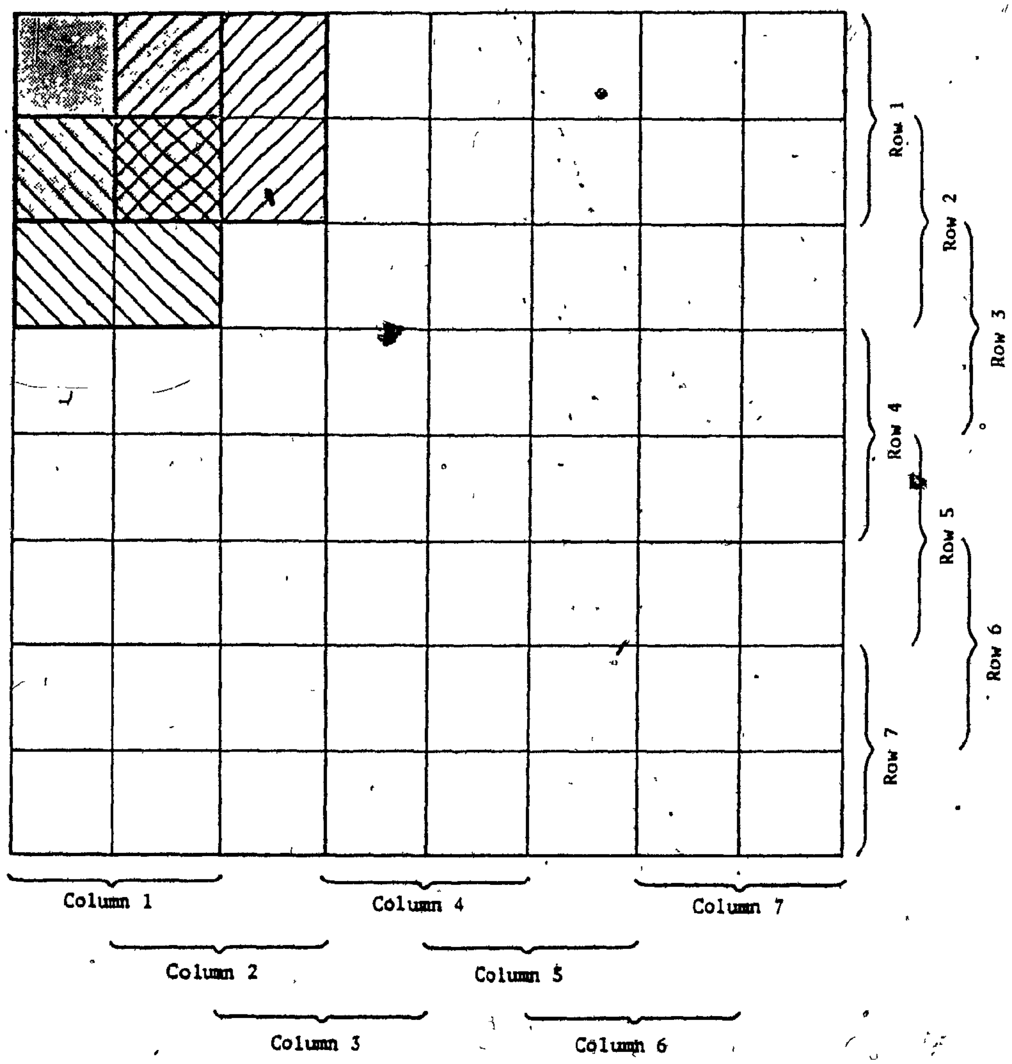
### 8.19.2 The Overlapping Cell Model

A technique in which the size of a cell can be enlarged four times to incorporate a greater geological variability but without a proportionate reduction in the number of reference cells is to overlap adjoining cells. As Figure 19 shows, the first enlarged cell is a combination of cells 1, 2, 9, and 10; the second one is the addition of cells 2, 3, 10, and 11; this can be continued till the end of the first row. The second row has its first enlarged cell consisting of cells 9, 10, 17, 18, and so on as in the first row. In this manner, 49 cells, each four times the initial size, are obtained. The known endowment cells increase to 22 from the original eight because of repetitions with neighbouring cells. This seemed to be a good solution to some of the problems mentioned in Section 8.19.1.

The results obtained on applying regression analysis over the overlapping cells are poor, to the extent that even in the known endowment cells, the predicted estimates are highly erratic. The only explanation lies in a highly increased multicollinearity resulting from overlapping and repetitions of cells. If so many inter-relationships exist among explanatory variables in normally adjoining cells, then the addition of intermediate overlapping cells is bound to

FIGURE 19

OVERLAPPING PROCEDURE FOR 4x1 ENLARGED CELLS



multiply the intercorrelative effect.

Similar problems are incurred when overlapping is attempted using two cells at a time, whether the overlapping is done along the rows or along the columns.

The approach is not considered statistically sound and was therefore not pursued further.

### 8.19.3 Non-Overlapping Enlarged Cells

Figure 20 shows the three different grids that have been attempted to obtain estimates of endowment in the region. They are described below.

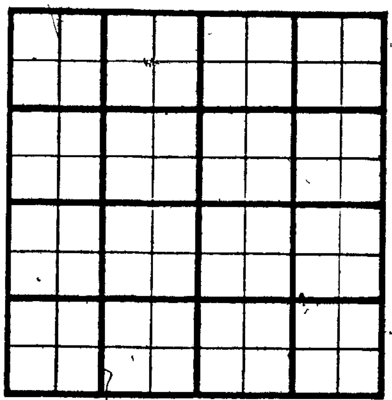
#### The 4x1 Enlarged Cells

Iterative regression analysis was carried out on the sixteen 4x1 enlarged cells. A checkerboard division of enlarged cells is not desirable because it reduces in half the already reduced number of cells, and as a result, further decreases the degrees of freedom for variable input. Besides, it is felt that enlarging the cell size does not result in as much multicollinearity as with cells one-fourth the size. The results shown in Table 14 are obtained after the fifth regression run. The input variables are the same as used in the checkerboard analysis described in Section 8.14.

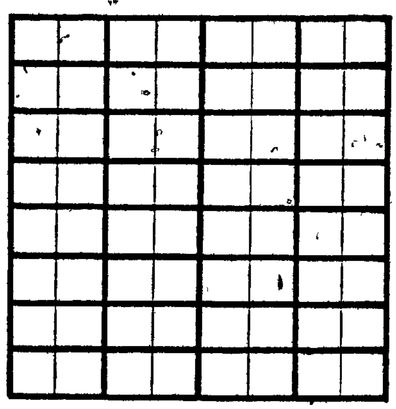
Enlarging cell size has two immediate effects which

# FIGURE 20

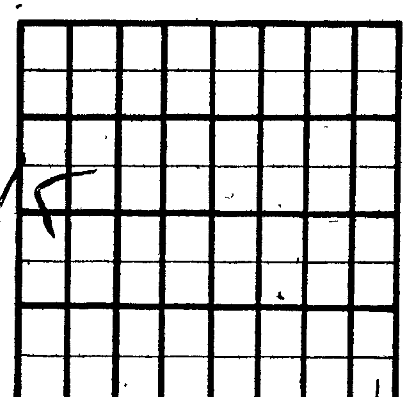
## ENLARGED NON-OVERLAPPING CELLS



4x1 cells



2x1 Horizontal Cells



2x1 Vertical Cells

Enlarged Cell Outline

Regular Cell Outline

TABLE 14

ENDOWMENT FORECAST IN 4x1 ENLARGED CELLS USING ITERATIVE REGRESSION ANALYSIS

| Cell # | Known Endowment ($x10^6) | Estimated Endowment ($x10^6) |
|---|---|---|
| 1 | - | 41.784 |
| 2 | - | 0.0 |
| 3 | 18.810 | 5.113 |
| 4 | 74.480 | 65.247 |
| 5 | - | 8.662 |
| 6 | - | 0.0 |
| 7 | 1,368.500 | 940.023 |
| 8 | 27.800 | 27.800 |
| 9 | 43.480 | 50.913 |
| 10 | - | 0.0 |
| 11 | 2,109.280 | 1,816.557 |
| 12 | 34.880 | 39.314 |
| 13 | - | 0.0 |
| 14 | - | 0.872 |
| 15 | - | 0.0 |
| 16 | - | 0.0 |
| | $R^2$ = | 0.995 |
| | S.E. = | $3.56x10^6 |

[1]See figure 20.

[2]Results shown are obtained after the fifth regression run.

bear on regression analysis. The first is that the decreased number of cells reduces the number of degrees of freedom; the second is that the proportion of known endowment cells increases. Both of these improve the rate of convergence and the coefficient of determination. The predicted estimates, however, are affected differently. Since in iterative regression analysis, the response variable value of known endowment cells is kept constant, and convergence attempted on these values, the total predicted endowment in the region is likely to be reduced. This is because previously a cell of 6.25 square miles area is a known endowment cell, but enlarging it four times makes the new known endowment cell equal to 25 square miles, yet with exactly the same response variable value as the original smaller cell.

As is shown in Table 14, it is not possible to obtain convergence in most known endowment cells, particularly, cells 3 and 7. In general, 80 percent of the total known endowment is predicted. The only significantly valuable cell is #1 worth over 41 million dollars. This enlarged cell is the equivalent of cells 1001, 1002, 1009, and 1010 of the checkerboard analyses in which a total endowment estimate of 25.6 million dollars is obtained for the four cells.

The fact that convergence is not fully attained, stresses again the effect of intercollinearity amongst the explanatory variables; in spite of this, the low endowment or barren cells are predicted in reasonable terms. When using enlarged cells, the significance of smaller cells that jointly comprise the enlarged cells is suppressed. For example, the 4x1 cells analysis estimates a zero endowment in cell 2. This cell is composed of cells 1003, 1004, 1011, and 1012. With the checkerboard analysis, cells 1003, 1004, and 1011 are predicted as barren. However, cell 1012 is forecast as relatively favourable worth 16 million dollars. Obviously, when the four cells are combined into one enlarged unit, the favourable geological relationships present in one cell are suppressed by the unfavourable relationships in the other three cells. Such a situation can vary, of course, depending upon the relative strengths of the favourable and unfavourable relationships present in the smaller cells that are combined to form the enlarged cell.

## The Rectangular Cells

The use of 2x1 rectangular cells has been made to observe any relationship between the shape of a cell and the directionally oriented geological data. Iterative regression analyses was therefore performed separately on checkerboard data of 2x1 vertical and 2x1 horizontal cells.[1]

A comparative study of regression results and the

---

[1]See Figure 20 for explanation of vertical and horizontal cells.

geology of the region shows that for the Rouyn-Noranda region itself, the directionally oriented cells do not apparently show any particular trend, except in individual cases where each of the two original cells that join together and form a rectangular cell, has a highly endowment-correlated variable, such as rhyolite (AREA 3), or NS to NW trending dykes (DYKE 3). When such explanatory variables strike generally parallel with the longer direction of the cell, a disproportionate weightage is attached to them which tends to rapidly increase during iterations. Rectangular cells are therefore subject to an increased bias and increased multicollinearity depending upon the nature and orientation of the explanatory variables. Further, it has not been possible to achieve convergence on the known endowment rectangular cells despite the checkerboard approach. In view of these reasons, it is felt that for the present study region, rectangular cells are neither efficient nor effective for the combination of variables, and under the conditions used in the tests.

## 8.20 Concluding Statement

Of all the models attempted under regression analyses, the 6.25 square miles square cells give the most effective results under the checkerboard technique. This statement should hold for the Rouyn-Noranda region evaluated in this study. However, a model can always be improved with additional information or improved techniques. Maki and Thompson (1973, p. 418) sum up the situation as follows:

> A statistical evaluation of the accuracy of a model is usually carried out with the use of a standard measure of the discrepancy between the predicted and the observed data. In this way one obtains a numerical measure of the goodness of fit for each model. Naturally, if one model gives a consistently better fit than any other model, then this model will be accepted and the others rejected. However, it often happens that one model will be the best to explain and fit certain sets of data, while another model will be the best at explaining and predicting other sets. Neither model can be rejected, since each is better under certain circumstances. Likewise, neither model should be completely accepted since in certain cases each model is not the best available. They can be conditionally accepted, studied, and used in those circumstances where they are the appropriate choice. Naturally a scientist would like to have a single model which is the best at explaining all the known experimental results. However, such a model is not always available, and the scientist must work with the models at hand until better ones are developed.

## 8.21 Regression Summary

(1) Regression analysis is a powerful and effective method for determining relationships amongst explanatory geological variables, and for applying these relationships in making forecasts of endowment within the geological system under investigation.

(2) The forecast values are calculated with respect to the existing known endowment within the system. Since the

known endowment is defined as the sum of past production and reserves in the region, it represents the minimum possible endowment. The forecast values too, therefore, are the minimum possible values. However, these forecast values should be assessed in terms of their absolute forecast estimates as well as their comparative values within the system.

(3) As a result of iterative regression, the relationships determined and the response variables calculated are in a state of sensitive equilibrium within the geological system of the region. If a new discovery is made in the region or if there is an increase in the known endowment with additional reserves, the regression model will have to be updated and the system re-equilibrated. This would result in increased forecasts because any addition to the known endowment will modify response variable — explanatory variables relationships.

(4) The selection of an optimal set of explanatory variables is an extremely important aspect in regression analysis. While factor analysis is an efficient technique of understanding underlying relationships amongst explanatory variables, the final judgment in their selection has to be made in the light of accepted geological theories on cause and effect relationships. Should these theories change, both the basis and interpretation of the model may have to be revised in the light of this new information.

(5) While geological concepts are the basis of understanding and interpreting regression analyses, the reverse is also true in that regression analyses help understand geological relationships in quantitative terms. The particular variables determined to be the most significant by regression analysis can then be given a greater attention at the time of geological mapping. This improves the cost-effectiveness of mapping and results in a better data base for future exploration.

(6) The standard square cells used in their checkerboard form give more efficient results than the larger 4x1 square cells or the 2x1 rectangular cells. The square checkerboard cells are less prone to bias or multicollinearity up to a certain stage depending upon the cell size and the geological "grain" in the region.

(7) In descending order of value, the following cells are predicted by regressing the checkerboard set of data to be the most favourable ones from amongst the 56 cells in the region with no known endowment.

| CELL # | PREDICTED: $\$X10^6$ |
|---|---|
| 1033 | 62.32 |
| 1012 | 16.02 |
| 1014 | 15.08 |
| 1002 | 13.78 |
| 1007 | 12.40 |
| 1010 | 9.81 |
| 1009 | 2.01 |

The total potential predicted by regression analysis is 131 million dollars. This indicates a near exhaustion of base-metal endowment in the region when evaluated under current geological information and concepts. Newer concepts will have to be evolved and additional information, particularly as regards the depth dimension, will be required for exploration to be successful. Geophysical information can be usefully incorporated in a model for evaluating the depth dimension.

# CHAPTER 9

## DISCRIMINANT ANALYSIS[1]

### 9.1 Introduction

Discriminant analysis is analogous to regression analysis in that both techniques attempt to predict a "best fitting" line or plane. For this reason, Mather (1976) defines discriminant analysis as being equivalent to the regression of intergroup mean differences on the explanatory variables. Discriminant analysis, however, disregards relationships obtaining between continuously measured classes by using group membership as the criterion and making all comparisons between the groups and none within the groups.

The application of discriminant analysis lies both in understanding differences between two or more sets of groups and in classifying new cases to the most relevant groups. The technique involves first determining statistically significant differences in the multivariable measurements among the given groups. These are then forced to become maximally distinct through a unique weighted linear combination of discriminatory variables selected. The maximization of discrimination is

---

[1]Discriminant analysis as used will refer to both the two-group cases and the multiple group cases.

achieved by spreading apart the group means while simultaneously compressing the spread of individual values about their respective group means (Bibb and Roncek, 1976). The discriminant function formulated can then be used for classifying new cases in their respective groups.

Cacoullos and Styan (1973) compiled a bibliography of publications on the theory and techniques of discriminant analysis. A more recent bibliography is available in Lachenbruch (1975). In the field of earth sciences, the technique has found applications in petrology, geochemistry and paleontology (Griffiths, 1966; Burnaby, 1966; Klovan and Billings, 1967; Link and Koch, 1967; Cameron et al., 1971; and Lenthall, 1972). In the field of mineral resource evaluation, the work of Harris (1965) was, perhaps, the earliest. Other related applications of discriminant analysis in resource evaluation include those of DeGeoffroy and Wignall (1970), and Rose (1972).

## 9.2 The Discriminant Model[1]

The general form of a discriminant function is:

$$y = V_1x_1 + V_2x_2 + \ldots V_px_p$$

where $V_1, V_2, \ldots V_p$ are the optimal weighting coefficients

---

[1] This section of the thesis has been drawn from Tatsuoka (1970) and Bibb and Roncek (1976).

assigned to each of the p original explanatory variables, $x_1, x_2, \ldots x_p$.

In a multi-group discriminant analysis, the objective is to find the set of weights which will maximize the discriminant criterion, i.e., the eigenvalues $\lambda$ and their associated eigenvectors.

Mathematically,

$$\lambda = \frac{\text{Between group variance, } SS_b}{\text{Within group variance , } SS_w} = \frac{V' B_v}{V' W_v}$$

where B is a p x p matrix of <u>among</u> group variances-covariances, i.e., $B = (\overline{X}_g - \overline{X})' (\overline{X}_g - \overline{X})'$ and W is the <u>within</u> group variance-covariance matrix, i.e., $W = (\overline{X} - \overline{X}_g)\ (\overline{X} - \overline{X}_g)'$.

In the above formulas,

$V$ = Vector of weighting coefficients

$\overline{X}_g$ = Vector of predictor variables means in the gth group

$\overline{X}$ = Vector of means of predictor variables

$X$ = Vector of variables in the gth group

When there are three or more groups present, the derivatives of $\lambda$ are set to zero, and following simplification, the matrix equation $(B - \lambda W)\ V = 0$ is obtained. This is the basic equation of discriminant analysis and can be re-written as:

$$(W^{-1}B - \lambda I)\ V = 0 \quad \text{where} \quad |W| \neq 0$$

In the above equation, I is a p x p identity matrix, and 0 is a p dimensional null vector. As a first step, the eigenvalues of the matrix $W^{-1}B$ are determined. These are always positive or zero because of the nature of the matrices W and B. The number of positive eigenvalues, r, is always equal to the lesser of either the number of predictor variables or one less than the number of groups. Each eigenvalue $\lambda_i$, $i = 1,2,...r$ has a unique eigenvector $V_i$ which satisfies the equation $(W^{-1}B - \lambda_i I) V_i = 0$. In this equation, the matrix $(W^{-1}B - \lambda_i I)$ is now known.

After ordering the eigenvalues in a descending order, the successive eigenvalues and their associated eigenvectors impart the following properties to the discriminant functions. The first discriminant function is that single weighted combination of measurements which has the maximum possible variance between groups relative to the variance within groups. The second discriminant function is that weighted combination of the measurements which of all possible weighted combinations independent of the first discriminant function accounts for a maximum of the remaining group differences. And so on for the third and successive discriminant functions, the maximum number of which is equal to the number of positive eigenvalues. In practice, however, the first few functions are the most important, much like the factors in factor analysis.

The calculated discriminant function is assessed for statistical significance to determine whether or not the between-group differences are real. Some variables may have to be added; others may have to be deleted to improve the discriminatory power of the function so that future observations are suitably classified. The error rate of the discriminant function is also estimated as this rate will affect future observations (Lachenbruch, 1975). Because these aspects are related to resource potential evaluation in the Rouyn-Noranda region, they are discussed in the following sections.

## 9.3 Variables Selection and Assumptions

The selection of variables for discriminant analysis can be done in much the same way as in regression analysis. To be most effective, the selected discriminatory variables should have directly or indirectly made a positive or negative contribution to the relationships that caused ore formation within the concepts accepted for the study under investigation. For an efficient discriminant analysis, Tatsuoka (1970) suggests that:

(a) The number of variables included should be more than the number of groups being compared;

(b) The total sample size should be at least two and preferably three times the number of variables used;

(c) The size of the smallest group should not be less than the number of variables used.

With the data available, it is possible to accommodate the first two suggestions. The cells containing known endowment are too few for the third suggestion to be accommodated except in a two-group discriminant function.

Discriminant analysis, unlike regression, assumes that the measurements have a multivariate normal distribution with equal variance-covariance matrices within the several samples. These are the two basic assumptions in discriminant analysis.

When the assumption of multivariate normal distribution and equal dispersion matrices do not hold, the calculated discriminant function will not be optimal or efficient. How crucial the equality of dispersion matrices is, remains a matter of dispute (Bibb and Roncek, 1976). It is likely that with unequal dispersion matrices, the resultant discriminant weight will be biased towards the group having the larger variance. The full impact of non-normal distribution characteristics is not clear either. Lachenbruch (1975, p. 36) states that "finding the distribution of linear combinations of non-normal variables is a difficult and as yet unsolved problem". Mather (1976) believes that moderate departures from the above conditions do not have a serious effect on the results. And Klecka (1975, p. 435) notes that the "technique is very robust and these assumptions need not be strongly adhered to". The situation is summed up by Link and Koch (1967, p. 12) as follows:

> For multivariate data, little is known about the effect on analysis of departures of the data from these two assumptions.[1] The analysis of variance is relatively insensitive to departures of the data from normality and homogeneous variance; by analogy, it can be hoped that multivariate analysis is also.

Since the measurements made in this study are not normally distributed, the effectiveness of discriminant analysis can be viewed in comparison with the results obtained for regression analyses. Another method for testing effectiveness is the "leaving one out" technique similar to that suggested by Lachenbruch (1975). In this technique, one known endowment cell at a time is assigned an arbitrary zero value, or a value equal to that of the lowest valued group, and discriminant analysis performed keeping the rest of the data as such. If the cell with the assumed zero value can be properly classified to its group, then it can be accepted that the absence of ideal conditions did not impair the discriminant function. When such is not the case, it is possible that some other explanation exists requiring another selection of the variables, a revision of the number of groups for discrimination, or even the re-evaluation of the discriminant model itself. A misclassification will also result if all the cases are not drawn from the same population. Such a situation can occur in geology where the variables are a result of both evolutionary

---

[1] I.e., multivariate normality and equality of dispersion matrices.

and interruptive processes, and in addition, subject to additional changes following post-ore processes.

9.4 Methodology Used

When discriminant analysis is applied in resource evaluation, the selection of optimal[1] groups can be as important as the selection of discriminating variables. In general, group selection should be a matter of pragmatism as long as all cases are drawn from the same parent population. This means that geological variables quantified should all be part of the same geological system. This information is only subjectively and often, incompletely known prior to the application of discriminant analysis.

A two-group discriminant analysis based on the presence or absence of known endowment is the simplest approach to distinguishing between areas of favourable and unfavourable potential. The advantage of the two-group analysis is that in small sample studies such as the present one, the number of known endowment cases per group increases because they are all classified into one group. The calculated discriminant function can thus take into account a greater variety of geological relationships associated with the individual known

[1] Optimal in terms of the range of values in individual groups, and the total number of groups.

endowment cases. However, beyond indicating the "presence" or "absence" potential, the two group function will not discriminate between different levels of endowment values.

An increased number of groups permits a better defined evaluation of various levels of endowment estimates. However, an increased number of groups can reduce the probability of correct classification because there are more chances of erroneous assignments (Lachenbruch, 1975). The optimal number of groups lies in between the minimum two and the maximum equal to the total number of cases with known endowment. Further, when the response variable distribution is skewed the selection of group value ranges will be affected by the distribution characteristics. Because of a highly skewed endowment value distribution in the Rouyn-Noranda region, there are more groups of the lower range of endowment values and few of the higher.

The following groups are selected for discriminant analysis. Their explanation and results are discussed separately:

- a two-group set based on the presence or absence of known endowment;
- two five-group sets;
- a seven-group set.

A stepwise discriminant procedure is used with both a pre-determined input order of variables, and an input with equal inclusion level for all variables such that the selection order is determined according to the selected criterion used. The variables selected for input are essentially the same as used in regression analysis, and are believed to be pertinent in accordance with the current volcanogenic thinking on massive sulphide formation (Sangster, 1972), and to some extent with the classical hydrothermal theory of Lindgren (1933). Of the criteria available in the S.P.S.S. programme Discriminant, both the Wilks and Mahalanobis methods[1] are used. In the former, the selection criterion is the maximization of the overall multivariate F-ratio for the test of differences among group centroids, and in the latter it is the maximization of Mahalanobis' distance between two closest groups. It should be pointed out here that the ranges used are somewhat arbitrary, in particular for those groups that have only one reference cell in them. When an unknown case is predicted as belonging to a certain group, it is done so with reference to the average endowment existing in that group rather than the actual range used. And for groups with only one reference cell in them, it becomes a single point situation regardless of the range used. If an unknown endowment cell is predicted to fall in the group with only one known endowment reference cell, then the prediction made is for an endowment value equivalent to that contained in the reference cell.

---

[1]These are defined in the next section.

## 9.5 Discriminant Results

### 9.5.1 The 2-Group Model

If the mere presence or absence of known mineral endowment is assumed to be a dichotomous function of a quantified geological environment, a two-group discriminant function can be calculated from the endowment-environment relationship to distinguish between potentially favourable and unfavourable areas within that environment. This assumption circumvents the postulate used in regression and multi-group discriminant analysis that both the presence and value of mineral endowment are a function of geological relationships. When the sample size is small, a large number of relationships can be jointly incorporated in a single two-group discriminant function, but without the ability to distinguish between various levels of endowment richness. In such analyses, the endowment if forecast will be at least equivalent to that contained in the cell with the minimum known endowment. In any case, in terms of the presence-absence of endowment, the results from the two group function should be compatible with those obtained by regression and multiple discriminant analyses.

The two group function is also an effective means of determining the commonness of ore forming processes as present in the endowment bearing cases. This knowledge can then be used in evaluating and explaining the results obtained in a multigroup analysis.

Discriminant analysis was first run using the presence-absence of endowment as the criterion against the following set of variables:

| | | |
|---|---|---|
| AREA | 2 | area of tuff, agglomerate |
| AREA | 3 | area of rhyolite |
| AREA | 4 | area of andesite, basalt |
| AREA | 8 | area of diorite, gabbro |
| AREA | 11 | area of granite, granodiorite |
| CNTL | 6 | contact length between AREA 2 & AREA 3 |
| CNTL | 10 | contact length between AREA 2 & AREA 11 |
| CNTL | 11 | contact length between AREA 3 & AREA 4 |
| CNTL | 13 | contact length between AREA 3 & AREA 8 |
| CNTL | 15 | contact length between AREA 3 & AREA 11 |
| FOLT | 1 | fault length, EW to NE |
| FOLT | 4 | fault length, NW to EW |
| DYKE | 3 | dyke length, NS to NW |
| DYKE | 4 | dyke length, NW to EW |

It should be noted that the above list contains a number of variables[1] that have not shown any significant relationship with endowment value, either in factor or in regression analysis. They have been included to compare results using the four options[2] available in the S.P.S.S. programme

[1]I.e., AREA 2, AREA 4, AREA 11, DYKE 4.

[2] (i) Direct Method: all variable are entered concurrently regardless of their individual discriminating power.

(ii) Wilks Method: the variable which maximizes the F-ratio and thus minimizes Wilks Lambda, a measure of group discrimination is entered first.

Discriminant. For this purpose, no inclusion level is specified for the variables so that each option can utilize its own criterion in the selection of variables for the discriminant function. This is then used as guide for selecting the most effective option.

The direct method uses all variables simultaneously. These are listed below in decreasing order of their standardized discriminant function coefficients. The coefficients have been converted into percentages of their absolute values.

| Variable Name | Standardized Coeff % | Variable Name | Standardiest Coeff % |
|---|---|---|---|
| CNTL 10 | -15.7 | FOLT 1 | -3.7 |
| DYKE 3 | -14.9 | AREA 8 | -3.5 |
| AREA 3 | -13.6 | DYKE 4 | 3.3 |
| CNTL 15 | 12.5 | AREA 4 | 1.5 |
| CNTL 13 | 8.5 | AREA 11 | -0.04 |
| FOLT 4 | -8.0 | | |
| CNTL 6 | -5.5 | | |
| AREA 2 | -5.4 | | |
| CNTL 11 | -4.1 | | |

(iii) Mahalanobis Method: the distance between the two closest groups is maximized.

(iv) Rao Method: the variable selected for inclusion is the one that contributes the largest increase in Rao's V, resulting in the greatest overall separation of the groups.

In a two group analysis, there is only one discriminant function accounting for 100 percent discrimination. The standardized coefficients of a function are measures of the discriminatory powers of the variables. This statement is amplified later in this chapter.

The Wilks, Mahalanobis and Rao methods use their particular criteria in the selection and inclusion of independent variables in a stepwise procedure. However, for the two-group analysis, the calculated discriminant function is exactly the same in each method. This should not be taken as a general rule. When the inclusion levels are specified "a priori", and when there are more than two groups, the forecast results by the three stepwise methods may not be the same. In the present case, the following variables are included in the discriminant function by the stepwise method. These are listed below in descending order of their absolute standardized contributions to the discriminant function. The figures below represent percentages of the total for better comparison.

| Variable Name | Standardized Coefficient Value (%) |
|---|---|
| AREA 3 | 24.3 |
| DYKE 3 | 21.6 |
| CNTL 10 | 19.4 |
| CNTL 15 | -15.3 |
| FOLT 4 | 11.2 |
| CNTL 13 | -8.3 |

As a result of the stepwise procedure, only six variables are included in the function instead of the original 14 used in the Direct method. It is easier to observe the relative contributions of the variables in a reduced space.

In Table 15, results are shown for the predicted presence of mineral endowment. One, indicates presence and zero, absence. The table includes the known endowment cells, but does not include those cells where neither a known nor a forecast endowment exists.

From the table, it is seen that despite different discriminant functions, the results obtained with both the direct and the stepwise methods are comparable. The only difference is that an endowment occurrence is forecast in cell 1050 by the direct method, but not the stepwise method.

It is of particular interest at this stage to observe that no endowment is forecast for cells 1013 and 1042, each of which is a reference cell with a known endowment. The validity of a discriminant model immediately is suspect if the known endowment cells cannot be predicted. It is possible that the cells discriminated do not necessarily all come from the same population. In other words, in spite of the relative smallness of the study region and the belief that all volcanogenic activity related to massive sulphide formation was confined therein, the possibility cannot be ruled out that individual cells do not all exhibit the same geological environment of ore genesis. Another possibility is that the set of variables

## TABLE 15

PRESENCE OF ENDOWMENT AS FORECAST WITH A 2-GROUP DISCRIMINANT FUNCTION USING DIRECT AND STEPWISE METHODS

| Cell No. | Known | Direct Method Forecast | Stepwise Method Forecast |
|---|---|---|---|
| 1003 | 0 | 1 | 1* |
| 1012 | 0 | 1 | 1* |
| 1013 | 1 | 0 | 0 |
| 1016 | 1 | 1 | 1 |
| 1021 | 1 | 1 | 1 |
| 1023 | 1 | 1 | 1 |
| 1027 | 0 | 1 | 1* |
| 1029 | 1 | 1 | 1 |
| 1031 | 0 | 1 | 1* |
| 1038 | 1 | 1 | 1 |
| 10[illegible] | 1 | [illegible] | 1 |
| 1041 | 0 | 1 | 1* |
| 1042 | 1 | 0 | 1 |
| 1049 | 0 | 1 | 1* |
| 1050 | 0 | 1 | 0* |
| 1053 | 0 | 1 | 1* |

*Cells indicating endowment potential by direct and/or stepwise method.

initially selected are not the best.

To clarify the anomalous situation of cells 1013 and 1042 referred to above, eight stepwise discriminant runs were made as before with the same set of variables, using the Wilks and Mahalanobis criteria separately. However, in each run, only one reference cell at a time was input as endowment bearing, and all others were arbitrarily assumed to have zero values. The Wilks and Mahalanobis methods give exactly the same results. These are shown in Table 16.

It is seen from the table that each known endowment cell when input as such is able to be predicted as endowment bearing with a probability close to one, except cell 1029 which has a probability of 0.86 and cell 1042 which has a probability of 0.58. The reference cells assigned a zero endowment value which are predicted by the other reference cells are as follows:

| | | |
|---|---|---|
| Cell 1038 | predicted by cell | 1016 |
| Cell 1029 | predicted by cell | 1021 |
| Cell 1021 | predicted by cell | 1029 |
| Cell 1013 | predicted by cell | 1042 |
| Cell 1021 | predicted by cell | 1042 |
| Cell 1029 | predicted by cell | 1042 |

The prediction of endowment in cells with no known endowment is not relevant at this stage of discussion.

When a particular cell is assumed to be barren but has endowment predicted by a known endowment cell, the

TABLE 16

PRESENCE OF ENDOWMENT AS FORECAST USING ONE REFERENCE CELL AT A TIME IN A 2-GROUP DISCRIMINANT ANALYSIS

| Cell No. | Known Endowment | Forecast Endowment Using All 8 Reference Cells | Particular reference cell used in discriminant function: 1013 | 1016 | 1021 | 1023 | 1029 | 1038 | 1039 | 1042 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1003 | 0 | 1 | - | - | - | - | - | - | - | 1(.509) |
| 1012 | 0 | 1 | - | - | - | - | - | - | - | - |
| 1013 | 1 | 0 | 1(.999) | 0(1.00) | 0(.981) | 0(1.00) | 0(.996) | 0(1.00) | 0(1.00) | 1(.559) |
| 1016 | 1 | 1 | 0(1.00) | 1(1.00) | 0(1.00) | 0(1.00) | 0(1.00) | 0(1.00) | 0(1.00) | 0(.645) |
| 1021 | 1 | 1 | 0(.985) | 0(1.00) | 1(.997) | 0(1.00) | 1(.883) | 0(1.00) | 0(1.00) | 1(.734) |
| 1023 | 1 | 1 | 0(.998) | 0(1.00) | 0(.988) | 1(1.00) | 0(1.00) | 0(1.00) | 0(1.00) | 0(.645) |
| 1027 | 0 | 1 | - | - | 1(1.00) | - | 1(.831) | - | - | 1(.749) |
| 1028 | 0 | 0 | - | - | - | - | - | - | - | 1(.695) |
| 1029 | 1 | 1 | 0(.977) | 0(1.00) | 1(.995) | 0(1.00) | 1(1.00) | 0(1.00) | 0(1.00) | 1(.864) |
| 1031 | 0 | 1 | 1(.705) | - | - | - | - | - | - | - |
| 1032 | 0 | 0 | 1(1.00) | - | - | - | - | - | - | - |
| 1035 | 0 | 0 | - | - | 1(.854) | - | - | - | - | - |
| 1038 | 1 | 1 | 0(1.00) | 1(.970) | 0(.999) | 0(1.00) | 0(1.00) | 1(1.00) | 0(1.00) | 0(.598) |
| 1039 | 1 | 1 | 0(.924) | 0(1.00) | 0(1.00) | 0(1.00) | 0(1.00) | 0(1.00) | 1(1.00) | 0(.645) |
| 1041 | 0 | 1 | - | - | - | - | - | - | - | 1(.645) |
| 1042 | 1 | 0 | 0(.995) | 0(1.00) | 0(.973) | 0(1.00) | 0(.985) | 0(1.00) | 0(1.00) | 1(.580) |
| 1049 | 0 | 1 | - | - | - | - | - | - | - | - |
| 1050 | 0 | 0 | - | - | - | - | - | - | - | 1(.636) |
| 1053 | 0 | 1 | - | - | - | - | - | - | - | 1(.738) |
| 1054 | 0 | 0 | - | - | - | - | - | - | - | 1(.538) |
| 1059 | 0 | 0 | - | - | - | - | - | - | - | 1(.512) |

TABLE 16

(CONTINUED)

Notes:

(a) The table includes results for known endowment reference cells and for cells in which the presence of endowment is indicated by a discriminant run. For comparison, the table includes results of discriminant analysis in which all reference cells were input as endowment bearing.

(b) The presence of endowment is indicated by one and its absence by zero. The figures in parentheses are the probabilities associated with the occurrence forecast of endowment.

(c) A stepwise discriminant analysis is used in which the Mahalanobis and Wilks criteria are used separately. The results obtained are the same in each case. The variables input are: AREA 2, AREA 3, AREA 4, AREA 8, AREA 11; CNTL 6, CNTL 10, CNTL 11, CNTL 13, CNTL 15; DYKE 3, DYKE 4; FOLT 1, FOLT 4.

similarity of discriminant function scores between the cells is indicated. When the discriminant function applied includes variables that are pertinent in accordance with accepted ore genesis concepts, and if these variables are relatively heavily weighted, then the function is an important one with valid results. On the other hand, if the discriminant function carries one or more heavily weighted variables that may be peculiar to the known endowment cell used as a "present" input, but whose role is not believed to be important in ore genesis, then both the function and its predictions require a deeper examination to determine if the uniqueness is fortuitous or represents a different geological environment, or if there is some other explanation demanding additional investigation.

It is against this background that the eight discriminant functions obtained separately in each run described above are discussed. Further, the validity of the unknown endowment cells forecast to be endowment bearing will be evaluated accordingly.

The standardized discriminant function coefficients calculated for each run with one known endowment cell input as a "present" case are shown in Table 17. As before, the discriminant coefficients have been converted into percentages.

In the eight functions shown in Table 17 the most commonly selected variable is AREA 3 (5 times), followed by CNTL 15 and DYKE 3 (4 times each), and CNTL 6, CNTL 11, and CNTL 13 (3 times each). The remaining variables, regardless

## TABLE 17

RELATIVE CONTRIBUTIONS OF VARIABLES TO THE 2-GROUP DISCRIMINANT FUNCTIONS[1]
USING ONE REFERENCE CELL AT A TIME

| Variables Input | All Reference Cells Input | Particular reference cell used as showing endowment presence | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1013 | 1016 | 1021 | 1023 | 1029 | 1038 | 1039 | 1042 |
| AREA 2 | - | - | - | - | - | - | - | 9.0 | - |
| AREA 3 | 24.3 | 23.7 | 42.2 | - | -8.1 | - | -29.9 | 11.9 | - |
| AREA 4 | - | - | - | - | - | 17.5 | - | 9.0 | - |
| AREA 8 | - | - | - | - | - | - | 9.4 | - | - |
| AREA 11 | - | - | - | 38.8 | - | - | - | - | - |
| CNTL 6 | - | - | -8.0 | - | - | - | 7.8 | -21.7 | - |
| CNTL 10 | 19.4 | - | - | - | 68.0 | - | - | - | - |
| CNTL 11 | - | - | - | - | 11.8 | - | 16.8 | -20.5 | - |
| CNTL 13 | -8.3 | -59.0 | -29.5 | - | -8.2 | - | - | - | - |
| CNTL 15 | -15.3 | - | -11.4 | -21.9 | 3.8 | - | 13.6 | - | - |
| DYKE 3 | 21.6 | 17.3 | - | 39.3 | - | 82.5 | - | - | 100.0 |
| DYKE 4 | - | - | - | - | - | - | -5.2 | 8.4 | - |
| FOLT 1 | - | - | -8.9 | - | - | - | -7.4 | - | - |
| FOLT 4 | 11.2 | - | - | - | - | - | -9.9 | -19.5 | - |

[1] See Table 16 for discriminant results, and text for details.

of their weightage are local to the particular known endowment cell used in the analysis.

Each known endowment cell when discriminated individually has its own discriminant function. While the same variables may be included in several of the functions, the discriminant coefficient varies. Thus each variable makes a unique contribution to each function. When all the known endowment cells are simultaneously input as showing the presence of endowment, the resulting function is affected by the strength of the discriminant function coefficients as indicated in their individual cases. In addition, some variables may be eliminated even though locally in the concerned individual function, they are heavily weighted. This explains why cells 1013 and 1042, both endowment bearing, are not forecast as such by the two-group discriminant function when all reference cells are input as showing the presence of endowment.

Cell 1013 is forecast as belonging to the "absent" endowment group with a probability of 0.684, and to the "present" endowment group with a probability of 0.316. The probability figures for cell 1042 are 0.654 and 0.346 respectively.

It is seen from Table 17 that CNTL 13 contributes a predominant 59 percent to the cell 1013 function. However, the contribution of this variable is only 8.3 percent in a discriminant function based on all eight reference cells.

The larger[1] discriminant function also includes other variables not present in the cell 1013 function. This therefore results in the absence of a forecast of endowment in this cell, although with only a moderate probability because the larger discriminant function also includes all the variables present in the cell 1013 function.

The situation with cell 1042 is rather extreme in the sense that its discriminant function is composed of just one variable, DYKE 3, the NS to NW trending dyke length which accounts for 100 percent of its discrimination. The larger discriminant function which includes the joint contribution of the relationships of all other endowment bearing cells also includes DYKE 3, but the weightage attached to it is only 21.6 percent of that for all the included variables. For this reason, it is not possible for cell 1042 to be predicted by the larger discriminant function.

Also, since the presence of endowment of the type exhibited in cell 1042 is forecast on the basis of the presence of one variable alone,[2] therefore, every cell that has a measure of this variable commensurate with its score in cell 1042 will show the presence of endowment, even though under

[1]The larger discriminant function refers to that obtained using all eight reference cells. The results obtained with the function are also included in Table 17.

[2]I.e., DYKE 3.

theoretical geological concepts, this should not be possible. The predicted endowment in cells 1003, 1013, 1021, 1027, 1028, 1029, 1041, 1042, 1050, 1053, 1054, and 1059 using cell 1042 are all a result of the presence of DYKE 3. DYKE 3 is a post ore event, and as has been stated earlier its role has not been fully resolved. There is no doubt, however, that most of the known ore deposits in the region lie adjacent to this variable, or conversely, this variable appears to be spatially associated with most of the ore deposits in the region. The correlation of DYKE 3 with endowment is so high that to avoid fortuitous predictions, its role has to be controlled in any stepwise analysis.

Another variable that deserves attention at this stage is CNTL 10, the contact length between AREA 2 and AREA 11. This feature is unique to endowment occurrence in cell 1023, carrying a 68 percent weightage of the variables in the concerned function. The relationship of this variable is so strong with endowment in cell 1023, that it becomes almost impossible to make a prediction of the presence of endowment in that cell without including this variable. And because this variable extends into cell 1031 on the south, the strong relationship results in a prediction of endowment there.

Cell 1021's function is dominated by DYKE 3 and AREA 11. Both these features are local to the cell, in particular, AREA 11. This cell predicts presence of endowment in the known endowment cell 1029, a cell that also has a function dominated by DYKE 3. The unknown endowment cells predicted to contain endowment by cell 1021 are, cells 1027 and 1035, both rich in AREA 11 and containing DYKE 3.

The conclusion from the above set of runs is that the known endowment cells do not exhibit exactly the same geological environment and therefore they may not all belong to the same population in a strict statistical sense. It is possible that some of the ore deposits in the region are associated with different statigraphic levels of basically the same type of rocks, indicating an interruptive type of eruptive geological activity. What is required therefore is some kind of discriminating function that can predict all the known endowment cases or at least most of them, so that a greater credence can be attached to the predictions made in the unknown endowment cells. To do this requires manipulating the forced inclusion of the variables believed to be fundamentally associated with ore occurrence as part of the discriminant function.

The above set of runs also makes the analyst aware of the spurious predictions that are likely to be made, e.g., cells 1027 and 1035, or cell 1031. The role of the responsible variables in such cases can therefore be controlled.

### 9.5.2 Other 2-Group Discriminant Functions

A set of three two-group discriminant analyses was performed using pre-specified inclusion levels of variables in two and an equal inclusion level in the third. The variables selected for discrimination are believed to be the most pertinent to the endowment in the region, and their inclusion order is based on subjective judgment in line with the accepted thought on massive sulphide ore genesis in the region. The analyses were made using the following variables.

| | |
|---|---|
| First Analysis: | AREA 3, CNTL 11 |
| | CNTL 6, CNTL 13, AREA 8, FOLT 4 |
| | CNTL 15 |
| | CNTL 10, DYKE 3 |
| Second Analysis: | AREA 3, CNTL 11 |
| | CNTL 6, CNTL 13, AREA 8, FOLT 4 |
| | CNTL 15 |
| | CNTL 10 |
| Third Analysis:[2] | AREA 3, AREA 8, CNTL 6, CNTL 11, |
| | CNTL 13, CNTL 15, FOLT 4 |

In the above listings, the top row is assigned the highest inclusion level followed by decreasing inclusion levels in the lower rows. Where two or more variables are shown to be in the same row, they have an equal inclusion level, and the order in which they will enter the discriminant function is based upon the Wilks or Mahalanobis[1] criterion selected. The standardized discriminant function coefficients

---

[1]The results obtained are identical in both cases.

[2]An equal inclusion level is specified for all variables in the third analysis.

are shown in Table 18 for the three analyses. These are expressed in terms of percentages of their total absolute value. Of the five variables constituting the discriminant function in the first analysis, the highest contribution is made by DYKE 3 in spite of the fact that this variable had been assigned to the lowest inclusion level along with CNTL 10, which also makes a high contribution, almost as much as that of the second highest contributor, AREA 3.

DYKE 3 is omitted from the second analysis, and as such, there appears to be an increased contribution by the remaining four variables, the highest contributor being CNTL 10, and the maximum increase in value being shown by FOLT 4, a feature that is probably, evidence of its relationship with DYKE 3.

In the third analysis, both DYKE 3 and CNTL 10 are omitted, and despite the same inclusion level, no fresh variable is included in the equation. AREA 3 shows the highest increase in value and makes the maximum contribution to the function.

The roles of CNTL 10 and DYKE 3 have been discussed previously.[1] CNTL 10 is a very local feature in cell 1023, and since it is absent in all other known endowment cells, it is bound to be highlighted in a discriminant function. DYKE 3 carries a strong spatial correlation with most of the

---

[1]See Section 9.5.1.

# TABLE 18

## STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS IN 2-GROUP ANALYSES

| Variance Name | First Analysis | Second Analysis | Third Analysis |
|---|---|---|---|
| AREA 3 | 22.5 | 29.9 | 47.6 |
| CNTL 10 | 22.2 | 30.4 | - |
| CNTL 15 | -16.4 | -20.6 | -25.6 |
| DYKE 3 | 25.9 | - | - |
| FOLT 4 | 13.0 | 19.1 | 26.8 |

[1]The coefficients are expressed as percentages, indicating their relative contributions to the discriminant function.

known endowment cells. Therefore, its contribution is disproportionately high.

The predicted values obtained by the three runs are shown in Table 19. The results are shown only for those cells that are known to be endowment bearing, and for those unknown endowment cells in which endowment has been forecast.

The primary evidence of a good discriminant function is that it should correctly classify the maximum number of known cells. This is accomplished by the first analysis which correctly classifies all the known endowment cells except cell 1042. However, the omission of variable DYKE 3 in the second analysis results in cell 1013, 1021, and 1029 being misclassified in addition to cell 1042. The results from the second analysis therefore are only with reference to the remaining four known endowment cells, i.e., cells 1016, 1023, 1038, and 1039. In actual effect, the results are with reference to only three cells because cell 1023 is unique in having a very close association with CNTL 10.

The omission of both CNTL 10 and DYKE 3 from the third function results in the restoration of cell 1013 as showing the presence of endowment, but because of the absence of CNTL 10, cell 1023 now shows an absence of endowment.

The above described changes brought about by omitting a certain variable are evidence of the strength and weakness of discriminant analysis. It therefore emphasizes the importance of selecting the "best" variables. The fact that the

TABLE 19

PRESENCE OF ENDOWMENT AS FORECAST BY 2-GROUP DISCRIMINANT ANALYSES[1]

| Cell No. | Analysis No. 1 | 2 | 3 |
|---|---|---|---|
| 1002 | - | 1(.655) | 1(.758) |
| 1003 | 1(.840) | 1(.653) | 1(.756) |
| 1007 | - | - | 1(.555) |
| 1009 | - | 1(.512) | 1(.647) |
| 1012 | 1(.685) | 1(.753) | 1(.829) |
| 1013 | 1(.769) | 0(.550) | 1(.592) |
| 1016 | 1(.527) | 1(.782) | 1(.848) |
| 1020 | - | - | 1(.569) |
| 1021 | 1(.771) | 0(.904) | 0(.816) |
| 1023 | 1(1.00) | 1(.999) | 0(.793) |
| 1027 | 1(.790) | - | - |
| 1029 | 1(.997) | 0(.796) | 0(.664) |
| 1031 | 1(.959) | 1(.969) | 1(.687) |
| 1032 | - | 1(.515) | 1(.649) |
| 1038 | 1(.987) | 1(.976) | 1(.978) |
| 1039 | 1(.989) | 1(.989) | 1(.986) |
| 1041 | 1(.977) | 1(.628) | 1(.738) |
| 1042 | 0(.629) | 0(.828) | 0(.706) |
| 1045 | - | - | 1(.551) |
| 1046 | - | - | 1(.506) |
| 1049 | 1(.578) | 1(.847) | 1(.855) |
| 1053 | 1(.824) | - | - |

[1]One indicates endowment presence and zero, absence. Figures in parentheses indicate probabilities associated with predicted endowment presence.

endowment in certain known cells is classified as being present or absent by the inclusion or omission of some variables indicates that all the members are not drawn from the same parent population in a statistical sense. Cell 1023 is perhaps the best example in this regard, because the endowment associated with the cell is unique in being associated with tuff and agglomerate (AREA 2) and not with rhyolite (AREA 3) as in the other instances. Other implications are discussed later.

In Table 19, all three analyses indicate probability of endowment in cells 1003, 1012, 1031, 1041, and 1049; two of the three runs also predict endowment in cells 1002, 1009, and 1032. None of these cells has any known endowment associated with it; of these, cells 1002, 1009, 1012, 1041, and 1049 have also been predicted as probably favourable cells by regression analysis.[1] However, regression analysis does not predict endowment in cells 1003, 1031, and 1032. Also, cells 1010 and 1033 predicted as very favourable by regression are not denoted as favourable by the three discriminant runs shown in Table 19. This latter situation is better understood when it is seen that of the known endowment cells 1042 is not predicted to contain endowment by any of the three runs, cells 1013 and 1023 by one of the three runs, and cells 1021 and 1029 by two of the three runs. When these cells are not

---

[1]See table 9.

predicted as endowment bearing, then those cells bearing statistical similarity with them, e.g., cells 1010 and 1037, will also not be predicted.

The above criticism should not detract from the utility of a 2-group analysis, particularly when the general environments associated with known ore bearing cells do not deviate to an extent that the discriminant function fails to classify them properly. From this point of view, the two-group discriminant function should work efficiently over areas the geology of which is not complex, particularly in terms of more than one or two cycles of processes. It should also work well at a reconnaissance level where the smaller scale of mapping does not determine the more complex and local features of geology.

### 9.5.3. 5-Group Discriminant Analyses A

The following groups are selected for the first five-group set of analyses. The selection of variables and their input order is made in accordance with their pertinence as explained for 2-group discriminant analyses. The known endowment cells that the groups represent are also shown below:

| Group No. | Range of Value ($\$ \times 10^6$) | Included Cell Nos. |
|---|---|---|
| 0 | < 10 | All unknown endowment cells. |
| 1 | 10- 50 | 1013, 1023, 1039, 1042 |
| 2 | 50- 500 | 1016, 1021 |
| 3 | 500-1500 | 1029 |
| 4 | > 1500 | 1038 |

It has been stated earlier that the range limits are based on pragmatism. For such groups as #3 and #4 which have only one known cell each, the range becomes practically meaningless because all predicted cases falling within these groups refer to the specific endowment known in the single cell composing the group.

The following variables are input in the three analyses to be described. All variables falling in a row have the same inclusion level, while those on the lower rows have respectively lower inclusion levels.

ANALYSIS #1: AREA 3, CNTL 11
CNTL 6, CNTL 13, AREA 8, FOLT 4
CNTL 15
DYKE 3, CNTL 10

ANALYSIS #2: AREA 3, CNTL 11
CNTL 6, CNTL 13, AREA 8, FOLT 4
CNTL 15
CNTL 10

ANALYSIS #3: AREA 3, CNTL 11
CNTL 6, CNTL 13, AREA 8, FOLT 4
CNTL 15

The discriminant function coefficients, standardized and converted to relative percentages, are shown in Table 20 for the three analyses.

The difference in the three analyses is that while the first one includes all the input variables, the second analysis does not include DYKE 3, and the third variant excludes DYKE 3 and CNTL 10. Of the four discriminant functions extracted in each of the three analyses, Table 21 shows results based on the first functions in each case. In the first analysis, the first function accounts for 42 percent of the cumulative eigenvalues, for the second analysis, 51 percent, and for the third analysis, 65 percent. The implication is that the best results are obtained in the third case. All three results are compared in Table 21 for both the known endowment cells, and for the unknown endowment cells that have a predicted endowment.

Referring first to Table 20, it is seen that AREA 3 makes the highest contribution to the discrimination in each of the three analyses, and that the contribution of this variable increases with the omission of DYKE 3 and CNTL 10. Variables CNTL 13 and CNTL 15 remain unaffected with the omission of DYKE 3 and CNTL 10. The roles of AREA 8, CNTL 6, and CNTL 10 are significantly influenced by the removal of DYKE 3.

# TABLE 20

## STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS IN 5-GROUP ANALYSES A[1]

| Variance Name | Analysis No. 1 | 2 | 3 |
|---|---|---|---|
| AREA 3 | 25.0 | 34.1 | 36.3 |
| AREA 8 | 4.1 | -2.5 | -2.6 |
| CNTL 6 | -0.9 | -9.9 | -9.3 |
| CNTL 10 | 11.8 | -5.9 | - |
| CNTL 11 | -0.8 | -15.7 | -15.7 |
| CNTL 13 | -13.9 | -13.5 | -12.7 |
| CNTL 15 | -13.8 | -10.9 | -12.8 |
| DYKE 3 | 21.6 | - | - |
| FOLT 4 | 8.0 | 7.5 | 10.5 |

[1]The coefficients are expressed as percentages, indicating their relative contributions to the discriminant function.

The removal of CNTL 10 in analysis #3 does not appear to influence any other variable, an indication of its highly local and independent nature. Obviously, the predicted presence of an endowment when the variable CNTL 10 is present in the function will become suspect when the same cell shows no prediction on the omission of CNTL 10. The only possible exception is a cell with a known endowment.

Referring now to the predicted endowment shown by the three analyses in Table 21, it is seen that the most consistent results are obtained for cells 1007, 1012, 1032, 1041, and 1045. None of these cells originally had any known endowment associated with it. Cells 1007, 1012, and 1041 have been classified as belonging to group #2, i.e., the $50-500 million range. Since the group range is made to accommodate endowment bearing reference cells 1016 and 1021 worth 74.48 and 244.66 million dollars respectively, the group actually indicates a value closer to $110 million, the mean of the two known cases. These three cells have also been predicted as endowment bearing by the 2-group function as well as by the iterative regression analysis. The predicted values for cells 1007, 1012, and 1041 by regression are $12.4 million, 16 million and $15 million respectively. While these values appear anomalously high amongst the unknown endowment cell predictions, their values appear low when compared to the discriminant function analyses. There are two reasons for this. The first is that the discriminant function compares the group value as such

# TABLE 21

## FORECAST ENDOWMENT GROUPS IN 5-GROUP DISCRIMINANT ANALYSES A

| Cell No. | Known GRP | Analysis No. 1 | 2 | 3 |
|---|---|---|---|---|
| 1002 | - | - | 2(.409) | 2(.356) |
| 1004 | - | - | 2(.589) | 2(.584) |
| 1007 | - | 2(.724) | 2(.805) | 2(.796) |
| 1009 | - | - | - | 1(.609) |
| 1010 | - | - | 3(.710) | 3(.673) |
| 1011 | - | - | - | 1(.772) |
| 1012 | - | 2(.829) | 2(.792) | 2(.781) |
| 1013 | 1 | 0(.513) | 0(.701) | 1(.904) |
| 1015 | - | 2(.616) | 2(.488) | 2(.478) |
| 1016 | 2 | 2(.998) | 2(.995) | 2(.995) |
| 1018 | - | - | 3(.519) | 3(.506) |
| 1021 | 2 | 3(.577) | 0(.485) | 0(.464) |
| 1022 | - | - | 3(.858) | 3(.846) |
| 1023 | 1 | 1(1.00) | 1(1.00) | 0(.625) |
| 1027 | - | 2(.609) | - | - |
| 1029 | 3 | 3(1.00) | 3(.758) | 3(.753) |
| 1031 | - | 1(.886) | 1(.939) | - |
| 1032 | - | 1(.750) | 1(.930) | 1(.995) |
| 1033 | - | - | 3(.676) | 3(.613) |
| 1038 | 4 | 4(1.00) | 4(1.00) | 4(1.00) |
| 1039 | 1 | 1(1.00) | 1(.999) | 1(.999) |
| 1040 | - | - | - | 1(.559) |
| 1041 | - | 2(.949) | 2(.449) | 2(.410) |
| 1042 | 1 | 1(.543) | 0(.451) | 0(.422) |
| 1045 | - | 1(.818) | 1(.677) | 1(.918) |
| 1050 | - | 2(.567) | - | - |
| 1053 | - | 2(.509) | - | - |

[1] Figures in parentheses are the associated probabilities of occurrence in the particular forecast group. The results for the known endowment cells are underlined. See text for details.

ignoring intermediate values between the groups. Regression analysis considers the continuum of forecast values drawing closer to the most likely value, but because of iterative regressions, some of the variance is lost.

While cells 1032 and 1045 are both classified as falling in group #1, i.e., the $10-50 million range, they are not forecast as such by regression analysis.

Cells 1002, 1004, 1010, 1018, 1022, 1031, and 1033 have been favourably grouped as endowment bearing. Of these, cell 1031 can be ignored because its endowment prediction is strictly related to the presence of CNTL 10, a variable of highly local significance, and associated with only one endowment cell, 1023. In the absence of CNTL 10 in the function, no endowment is forecast for this cell, much like the case of cell 1027 which gives a high prediction when DYKE 3 is present, but is reduced to a barren classification otherwise. Of the rest, cells 1002, 1010, and 1033 are very favourably forecast by regression analysis. Cells 1004, 1018, and 1022 are not.

Of the cells that are predicted by only one of the three analyses, cell 1009 is also predicted by regression, but not as favourably as the ones mentioned earlier.

However, the problem still remains that the known endowment cells 1021 and 1042 are not correctly predicted. This is discussed in the next section.

### 9.5.4 5-Group Discriminant Analyses B

The following groups are selected for the second five-group set of analyses:

| Group No. | Range of Values ($\$x10^6$) | Cells Included |
|---|---|---|
| 0 | 0- 10 | All unknown endowment cells |
| 1 | 10- 100 | 1013, 1016, 1023, 1039, 1042 |
| 2 | 100- 500 | 1021 |
| 3 | 500-1500 | 1029 |
| 4 | > 1500 | 1038 |

The difference between this set and the preceding one is that a change has been made in the value range for groups #1 and 2, as a result of which, cell 1016 which originally fell with cell 1021 in group #2 now falls in group #1. The effect of this change is analyzed in this section.

The set of variables used, and their inclusion levels, is the same as for the 5-group set A described previously. As before, three runs are made, the first one including all input variables, the second with DYKE 3 omitted, and the third with both DYKE 3 and CNTL 10 omitted.

The highest discriminant function of the four functions extracted in each of the three analyses, accounts for 45 percent, 64 percent, and 76 percent of the cumulative eigenvalues in the first, second, and third analysis respectively. The standardized coefficients of this function, converted into relative percentage are shown in Table 22. DYKE 3 predominates in the first analysis while AREA 3 does in the others. In general, the variables most contributing to the first discriminant function in each analysis are AREA 3, FOLT 4, CNTL 15, and CNTL 11. The omission of DYKE 3 results in a significant increase in the contribution of CNTL 11, AREA 3, AREA 8, and FOLT 4. The omission of CNTL 10 does not affect any variable except CNTL 11.

The predicted estimates of the three analyses are shown in Table 23. The first observation in the table is that with the exception of known endowment cell 1023 in the third analysis, all reference cells are predicted as endowment bearing. The case of 1023 is not unexpected because of the absence of the variable CNTL 10.

A comparison of Tables 21 and 23 indicates the antithetic relationship between the centroids of multivariate relationships in the known endowment cells 1016 and 1021. In the 5-group analyses A described previously, both these cells are included in group #2, and the result is that cell 1021 cannot be predicted. The explanation is also clear when the coefficients of discriminant functions are compared with the variables present in these cells. While cell 1016 is rich in AREA 3 it is devoid of CNTL 15. Both these variables make significant contributions to the functions; and both of those

TABLE 22

STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS IN 5-GROUP ANALYSES B

| Variable Name | Analysis No. 1 | 2 | 3 |
|---|---|---|---|
| AREA 3 | 21.3 | 34.7 | 39.7 |
| AREA 8 | 0.6 | -7.6 | -8.5 |
| CNTL 10 | 18.6 | 13.1 | - |
| CNTL 11 | -0.8 | -12.3 | -20.4 |
| CNTL 15 | -16.2 | -15.5 | -15.3 |
| DYKE 3 | 32.5 | - | - |
| FOLT 4 | 10.7 | 16.8 | 16.1 |

[1]The coefficients are expressed as percentages, indicating their relative contributions to the discriminant function.

# TABLE 23

## FORECAST ENDOWMENT GROUPS IN 5-GROUP DISCRIMINANT ANALYSES B

| Cell No. | Known GRP | Analysis No. 1 | 2 | 3 |
|---|---|---|---|---|
| 1001 | - | - | 3(.465) | 3(.398) |
| 1002 | - | - | 3(.505) | 1(.474) |
| 1003 | - | - | - | 1(.521) |
| 1007 | - | - | 1(.465) | 1(.735) |
| 1009 | - | - | 3(.677) | 3(.517) |
| 1010 | - | - | 3(.490) | 3(.476) |
| 1011 | - | - | 3(.395) | 3(.375) |
| 1012 | - | - | - | 1(.604) |
| 1013 | 1 | 2(.493) | 3(.676) | 3(.582) |
| 1015 | - | - | - | 1(.645) |
| 1016 | 1 | 1(.620) | 1(.673) | 1(.841) |
| 1017 | - | - | 2(.393) | 2(.378) |
| 1018 | - | - | 2(.551) | 2(.539) |
| 1020 | - | - | - | 1(.468) |
| 1021 | 2 | 2(.941) | 2(.507) | 2(.494) |
| 1022 | - | - | 2(.572) | 2(.560) |
| 1023 | 1 | 1(1.00) | 1(1.00) | 0(.427) |
| 1025 | - | - | 3(.576) | 3(.494) |
| 1027 | - | 2(.924) | - | - |
| 1029 | 3 | 3(.943) | 3(.483) | 3(.449) |
| 1031 | - | 1(.927) | 1(.900) | - |
| 1032 | - | 1(.694) | 1(.697) | 1(.865) |
| 1033 | - | - | 3(.605) | 3(.488) |
| 1038 | 4 | 4(1.00) | 4(1.00) | 4(1.00) |
| 1039 | 1 | 1(.997) | 1(.997) | 1(.997) |
| 1041 | - | 2(.591) | 3(.490) | 1(.444) |
| 1042 | 1 | 2(.593) | 3(.415) | 3(.388) |
| 1045 | - | - | 2(.517) | 2(.484) |
| 1049 | - | - | 1(.570) | 1(.716) |
| 1050 | - | 2(.860) | 2(.364) | 2(.347) |
| 1051 | - | - | 3(.361) | 3(.340) |
| 1052 | - | - | 2(.491) | 2(.479) |
| 1053 | - | 2(.871) | - | - |

[1] Figures in parentheses are probabilities of occurrence in the particular group. The forecast values of known endowment cells are underlined. See text for details.

are present in cell 1021 which is rich in CNTL 15 but rather poor in AREA 3.

A similar example is that of cell 1042. This known endowment cell cannot be so predicted in the 5-group A analyses #1 and 2. But the transfer of cell 1016 from group #2 to group #1 in the new grouping B results in the correct endowment prediction for this cell by all three analyses.

It is essential, therefore, that the variables selected be pertinently "best", and that an optimal selection of groups be made so that known endowment cells included in a group are not antithetic with one another in terms of multivariate relationship. The ideal situation is that all cases should be drawn from the same population. However, as long as the known endowment cells themselves are predicted as such, the predictions for the unknown endowment cells should be credible.

In this regard, a set of eight discriminant runs was made over the 64-cell data, with each known endowment cell being assigned to the zero group in turn. The objective was to see if the value of this cell could be predicted from the remaining seven reference cells. In this manner, the general relationships amongst such cells can also be observed. The variables used are AREA 3, AREA 8, CNTL 6, CNTL 11, CNTL 13, CNTL 15, and FOLT 4. Since a known endowment cell is assigned a zero value in each run, the discriminant functions obtained are different in each case. Only one function is extracted when cell 1038 is omitted, and this function includes one variable, CNTL 11. In all other cases, there is a general uniformity in the variables included and their coefficients. For the highest function in the remaining seven runs, AREA 3 makes the highest contribution, followed by CNTL 11, FOLT 4, CNTL 15, and AREA 8. CNTL 6 and CNTL 13 are included in only three runs each.

The results of omitting one endowment cell at a time are shown in Table 24. It is seen that cells 1013, 1023, and 1039 do not receive a predicted endowment. The lack of prediction in cell 1023 is easily understood because of the absence in the input variables of CNTL 10, the variable most locally related to endowment in this cell. And while cell 1013 is predicted to belong to group zero with a probability of 0.700 by the first discriminant function which accounts for 71 percent of the cumulative eigenvalue, it is predicted to belong to group #2 with a probability of 0.254 by the second discriminant function which accounts for 19 percent of the total eigenvalues. Similarly, cell 1039 is predicted to belong to group #1 with a probability of 0.409 by the second discriminant function of that run which accounts for 12 percent of the eigenvalues. The first function in that run accounts for 83 percent of the eigenvalues.

Cells 1016, 1021, and 1042 have been predicted to belong to a higher group than known. Obviously, this is also a measure of the similarity of these cells to the centroid of

## TABLE 24

FORECAST ENDOWMENT GROUPS FOR KNOWN ENDOWMENT CELLS ASSUMED TO HAVE ZERO ENDOWMENT, ONE AT A TIME IN EACH DISCRIMINANT RUN[1]

| Cell No. | Known | Predicted from All Cells In | KNOWN ENDOWMENT CELL OMITTED | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | PREDICTED GROUPS | | | | | | | |
| | | | 1013 | 1016 | 1021 | 1023 | 1029 | 1038 | 1039 | 1042 |
| 1013 | 1 | 3(.582) | 0(.700) | 1(.867) | 1(.574) | 3(.617) | 1(.373) | 1(1.00) | 3(.473) | 3(.641) |
| 1016 | 1 | 1(.841) | 1(.943) | 3(.663) | 1(.529) | 1(.910) | 1(.834) | 1(1.00) | 1(.717) | 1(.911) |
| 1021 | 2 | 2(.494) | 2(.606) | 2(.590) | 3(.480) | 2(.504) | 2(.716) | 2(1.00) | 2(.459) | 2(.507) |
| 1023 | 1 | 0(.427) | 0(.686) | 0(.526) | 0(.716) | 0(.489) | 0(.584) | 3(1.00) | 1(.344) | 0(.469) |
| 1029 | 3 | 3(.449) | 3(.669) | 3(.669) | 3(.854) | 3(.472) | 2(.506) | 3(1.00) | 3(.390) | 3(.475) |
| 1038 | 4 | 4(1.00) | 4(1.00) | 4(1.00) | 4(1.00) | 4(1.00) | 4(1.00) | 3(1.00) | 4(1.00) | 4(1.00) |
| 1039 | 1 | 1(.997) | 1(.999) | 1(.999) | 1(.998) | 1(.999) | 1(1.00) | 0(.250) | 0(.538) | 1(.999) |
| 1042 | 1 | 3(.388) | 2(.405) | 2(.405) | 3(.445) | 3(.406) | 2(.494) | 0(.989) | 3(.341) | 3(.409) |

[1]The results shown above are based on the 5-Group B analyses. Underlined figures indicate predicted endowment in known endowment cells assumed barren for the particular run. Figures in parenthesis indicate probability of belonging to the predicted group.

cell 1029, the only cell belonging in group #3 in which the above three cells have been classified. Cell 1029 which, on the basis of its known endowment falls in group #3 is classified in group #2. Cell 1038, which originally belongs to group #4, has now been predicted to belong to group #3. The cases of cells 1029 and 1038 can be understood when it is observed that there is only one cell, 1029, which constitutes group #3, and only one cell, 1038 which constitutes group #4. So when these very cells are assigned a zero group, there no longer remains any case of group #3 or #4, depending on what cell is omitted. Therefore, the predicted classification of these cells is in the highest grouping available next to the actual original groups.

But the point to be made here, is that the results prove the effectiveness of discriminant analysis in spite of a lack of compliance with all the assumptions. And more importantly, the results are evidence of quantitative relationships existing between predictor variables as selected, and the endowment in the region.

When all known endowment cells are included in the analysis, cell 1016 is classified in its own group, but cells 1013 and 1042 are still classified in higher groups. And while the second discriminant function classifies cell 1013 in its own original group #1 with a probability of 0.161, this is not the case with cell 1042, which the second discriminant function classifies to group #2 with a probability of 0.307.

Thus, cell 1042 is expected to contain a greater endowment than that known.

Of the cells with no known endowment, the most consistent results obtained in the three analyses shown in Table 23 are for cells 1032, 1041, and 1050. Cell 1032 is not predicted as having endowment by the iterative regression analysis. This cell received a favourable prediction by about all runs of discriminant analysis. The remaining two cells, 1042 and 1050, the former in particular, are assigned high values by the regression model.

In analysis #1, the prediction of endowment in cells 1027 and 1053 is based only on the presence of DYKE 3. When the geology of these cells is subjectively evaluated, the results appear to be spurious, particularly when they are no longer favourably predicted on the omission of DYKE 3. Similarly, the prediction in cell 1031 appears to be the result of CNTL 10 only, and thus of little credence. In general, the three analyses appear to overclassify cases in groups #2 and #3. The situation is summarized in Table 25.

In the discriminant analysis performed, it is assumed that all groups have an equal probability of occurrence provided of course that the necessary geological relationships pertaining to the centroid of the group are present in the cases analyzed. Of the eight known endowment cells in the region, there are five that fall in group #1, and one each in groups #2, 3, and 4. There are 56 unknown endowment cells

TABLE 25

5-GROUP ANALYSES B: NUMBER OF CASES CLASSIFIED IN INDIVIDUAL GROUPS

| Group No. | Input Cases | PREDICTED 1st Analysis | 2nd Analysis | 3rd Analysis |
|---|---|---|---|---|
| 0 | 56 | 50 | 37 | 35 |
| 1 | 5 | 5 | 7 | 11 |
| 2 | 1 | 7 | 7 | 7 |
| 3 | 1 | 1 | 12 | 10 |
| 4 | 1 | 1 | 1 | 1 |

which cannot be initially classified, and therefore assigned to group #0. If all these 56 cells could be classified in porportion to ratios existing in the known groups then there should be 35 more cells falling in group #1 and 7 more cells falling in each of groups 2, 3, and 4. But such is not the case, for the region is composed of both favourable and unfavourable parts. Obviously therefore, when seven cells are forecast to fall in group #2 in each of the three analyses (Table 23) and, 12 and 10 cases fall in group 3 in analyses #2 and 3, the situation calls for examination because initially, of the eight known endowment cells, only one cell falls in each of the two groups.

The problem could be partially resolved by adjusting the probabilities of group membership based on "a priori" knowledge of the population distribution of cases. This is not possible in problems of resource evaluation because measurements regarding both geology and endowment are at best incomplete. However, one decision rule that can be applied in the present problem is to accept only those cells classified in groups 2 and 3 which have a probability of belonging to one of them at least equal to that of the individual known endowment cells composing the two groups, i.e., cells 1021 and 1029 respectively. Using this arbitrary rule, the number of cells falling in group #2 is reduced from the original forecast of seven, to one, four, and three in analyses #1, 2, and 3 respectively. Similarly, the number of cells forecast to fall in group 3 is reduced from 12 and 10, to eight and six respectively in analyses #2 and 3. This is more in line with what would subjectively be expected in relation to the ratios of group memberships known.

Table 26 is a reduced form of Table 23 after using the rule mentioned above. However, also included in the table are the classifications based on the second highest probabilities. This table shows results for analysis #3 only, since this analysis does not have the influence of either DYKE 3 or CNTL 10. And as has been stated earlier, the highest discriminant function for this analysis accounts for 76 percent of the eigenvalues followed by the second highest function with 19 percent. While the first discriminant function is dominated by AREA 3, the second is controlled by CNTL 11, both closely related to endowment in the region.

The originally unknown endowment cells retaining this favourable status in both the highest and the next highest probabilities are the following:

1002
1009
1010
1018
1022
1033
1041

All of these cells except 1018 and 1022 have also been predicted as highly favourable by regression analysis.

TABLE 26

5-GROUP B ANALYSIS #3: REDUCED FORM OF TABLE 22

| Cell No. | Known GRP | ANALYSIS #3 PREDICTED GROUPS Highest Probability | Second Highest Probability |
|---|---|---|---|
| 1002 | - | 1(.474) | 3(.316) |
| 1003 | - | 1(.521) | 0(.411) |
| 1007 | - | 1(.735) | 0(.197) |
| 1009 | - | 3(.517) | 1(.286) |
| 1010 | - | 3(.476) | 2(.384) |
| 1012 | - | 1(.604) | 0(.184) |
| 1013 | 1 | 3(.582) | 1(.161) |
| 1015 | - | 1(.645) | 0(.175) |
| 1016 | 1 | 1(.841) | 0(.143) |
| 1018 | - | 2(.539) | 3(.279) |
| 1020 | - | 1(.468) | 0(.368) |
| 1021 | 2 | 2(.494) | 3(.312) |
| 1022 | - | 2(.560) | 3(.368) |
| 1023 | 1 | 0(.427) | 3(.264) |
| 1025 | - | 3(.494) | 0(.171) |
| 1029 | 3 | 3(.449) | 2(.284) |
| 1032 | - | 1(.865) | 0(.083) |
| 1033 | - | 3(.488) | 1(.231) |
| 1038 | 4 | 4(1.00) | - |
| 1039 | 1 | 1(.997) | 0(.003) |
| 1041 | - | 1(.444) | 3(.317) |
| 1042 | 1 | 3(.388) | 2(.307) |
| 1049 | - | 1(.716) | 0(.210) |

[1]See text for details. The results for the known endowment cells are underlined.

Cell 1018 appears to be a case of misclassification because the cell contains only one of the five variables forming the function, i.e., AREA 8. Other cells which are grouped in the zero category in the next highest function, but in which the probability itself is low, are cells 1007, 1012, 1015, 1025, and 1032. Of these cells 1007 and 1012 are also classified as favourable by regression analysis, but the others are not, particularly, cell 1032 which is forecast as completely barren.

The results of the two five-group analyses and the following seven-group analysis are jointly concluded upon at the end of the chapter.

### 9.5.5 Seven-Group Analyses

Increasing the number of groups in a discriminant analysis serves the purpose of a more relevant classification but with an increased risk of misclassification. In the present study there are only eight known endowment cases and a ninth set of unknown cases. It would therefore appear that dividing these nine categories into seven groups would fail to bring out discriminatory relationships jointly between two or more known endowment cells; it would, however jointly analyse cells that are classified in the zero or barren category. Since in some of the earlier runs, it was observed that a number of endowment cells had shown an antithetic relationship with one another when considered jointly in the same group, e.g., cells 1016 and 1021, the 7-group analyses are made to observe the behaviour of the discriminant function in separating the various groups. The following groups are used in the analyses:

| Group # | Range ($\$x10^6$) | Cells # Included |
|---|---|---|
| 0 | Assumed Zero | All unknown endowment cells |
| 1 | 0- 25 | 1013 |
| 2 | 25- 50 | 1023, 1039, 1042 |
| 3 | 50- 100 | 1016 |
| 4 | 100- 500 | 1021 |
| 5 | 500-1500 | 1029 |
| 6 | > 1500 | 1038 |

The best results as judged from the predictability of the known endowment cells are obtained using the following variables in their order of inclusion:

AREA 3
AREA 8
CNTL 11
CNTL 15
DYKE 3

The results[1] obtained and the associated probabilities

---

[1]The table does not include cells predicted to belong to zero category.

are shown in Table 27. The known endowment cells have been highlighted in the table. It is observed that with the exception of cell 1042 which is forecast to belong to a lower group, all known endowment cells have been properly classified.

Another set of discriminant runs was made using the same variables as before, but assuming a zero input endowment for one known endowment cell at a time. The runs were made over the 64-cell data, but the results shown in Table 28 are for the known endowment cells only because they alone can be used to observe the predictive efficiency of the discriminant function, or of the choice of using seven groups.

The first observation is that of the eight known endowment cells, only cells 1016 and 1023 fail to have any endowment predicted in them. Both these cells, however, are classified in group #2 by the second highest probability.

The probability of cell 1016 belonging to group #0 as given in Table 28 is 0.44. However, the probability that a member of the predicted group zero would be as far from the centroid as cell 1016 is very low, only 0.115; this suggests the possibility that cell 1016 might not belong to the population of cells from which group #0 is drawn. This would appear to be the case because cell 1016 is a known endowment cell. The cell is misclassified because it is located away from the cluster of ore deposits in the region. The cell is devoid of any kind of dyke activity and is mainly pyrite and sphalerite rich. Furthermore, the Mobrun ore deposit contained

# TABLE 27

## FORECAST ENDOWMENT GROUPS IN 7-GROUP ANALYSIS

| Cell No. | Known Group | Predicted Groups[1] |
|---|---|---|
| 1002 | - | 1(.349) |
| 1003 | - | 3(.669) |
| 1007 | - | 2(.805) |
| 1009 | - | 2(.453) |
| 1012 | - | 1(.364) |
| 1013 | 1 | 1(.914) |
| 1015 | - | 2(.944) |
| 1016 | 3 | 3(.985) |
| 1021 | 4 | 4(.900) |
| 1023 | 2 | 2(.498) |
| 1025 | - | 2(.593) |
| 1027 | - | 4(.916) |
| 1028 | - | 2(.534) |
| 1029 | 5 | 5(.932) |
| 1031 | - | 6(.460) |
| 1032 | - | 2(.917) |
| 1033 | - | 2(.760) |
| 1038 | 6 | 6(.995) |
| 1039 | 2 | 2(.731) |
| 1041 | - | 1(.899) |
| 1042 | 2 | 1(.362) |
| 1045 | - | 1(.708) |
| 1050 | - | 4(.687) |
| 1053 | - | 4(.854) |

[1]Figures in parentheses indicate probability of belonging to the predicted group.

## TABLE 28

### 7-GROUP ANALYSES: ONE KNOWN ENDOWMENT BEARING CELL ASSUMED TO HAVE ZERO ENDOWMENT IN EACH DISCRIMINANT RUN

| Cell No. | Known GRP | KNOWN ENDOWMENT CELL OMITTED (ASSUMED BARREN) PREDICTED GROUPS 1013 | 1016 | 1021 | 1023 | 1029 | 1038 | 1039 | 1042 |
|---|---|---|---|---|---|---|---|---|---|
| 1013 | 1 | 2(.42) | 1(.91) | 1(.93) | 1(.89) | 1(.87) | 1(.85) | 1(.81) | 1(.95) |
| 1016 | 3 | 3(.99) | 0(.44) | 3(.98) | 3(.98) | 3(.98) | 3(.92) | 3(.60) | 3(.98) |
| 1021 | 4 | 4(.92) | 4(.90) | 5(.63) | 4(.90) | 4(.94) | 4(.87) | 4(.86) | 4(.90) |
| 1023 | 2 | 2(.52) | 2(.52) | 2(.51) | 0(.87) | 2(.53) | 0(.53) | 0(.60) | 0(.51) |
| 1029 | 5 | 5(.92) | 5(.93) | 5(1.0) | 5(.94) | 4(.97) | 5(.91) | 5(.91) | 5(.93) |
| 1038 | 6 | 6(1.0) | 6(1.0) | 6(1.0) | 6(1.0) | 6(1.0) | 3(.94) | 6(.61) | 6(1.0) |
| 1039 | 2 | 2(.73) | 2(.95) | 2(.75) | 2(.82) | 2(.74) | 3(.45) | 3(.44) | 2(.90) |
| 1042 | 2 | 4(.44) | 1(.36) | 1(.43) | 1(.36) | 4(.32) | 1(.34) | 2(.42) | 1(.43) |

[1] Underlined figures indicate predicted endowment in known endowment cells assumed barren for the particular run. Figures in parenthesis indicate probability of belonging to the predicted group.

in the cell is believed to be stratigraphically higher than the rest of the ore deposits in the region (Dugas, 1977, oral communication).

The case of cell 1023 has been explained before. It contains the West Macdonald deposit, the only deposit in the region that does not have a direct rhyolite association. Its host rock is AREA 2 consisting of tuff and agglomerate.

The predicted endowment in cells 1013, 1021, and 1039 falls in the next group higher than that known when endowment is assumed to be absent in them. This implies a possibility of a greater endowment in these cells than that presently known.

The predicted endowment in cell 1029 falls in group #4 instead of the known #5. Similarly, for cells 1038 and 1042, the classification is in groups #3 and #1 instead of the known groups #6 and #2 respectively.

So far as cell 1038 is concerned, it would be difficult to forecast its original known group when its endowment is assumed zero. The reason is that this is the only cell associated with the highest valued group #6. When the environment of this highest valued cell is assumed to have no endowment associated with it, then the highest valued group does not exist as such, and the predicted value has got to be in relation to the next lower group the centroid of which is close to that of cell 1038.

Cell 1029 comprises the second highest group value, and is the only cell to do so. Its predicted value therefore at the next lower group in the light of explanation given for cell 1038, appears perfectly acceptable.

The case of cell 1042 appears to be somewhat like that of cell 1016 in that it is located away from the main group of ore deposits in the region. Besides, this cell has the only deposit in the region which has no reported zinc in it; only copper. Like cell 1016, this too could be related to a separate eruptive phase of volcanism. The cell is classified in group #1 instead of its known group #2. It is interesting to note that even when all known endowment cells, including cell 1042 are discriminated under the 7-group analyses, its predicted value still is in group #1 instead of #2. The second highest probability classifies this cell in group #4 with a probability of 0.301. The possibility has been expressed earlier that the cell may belong to a higher group.

The results of the 7-group discriminant analyses shown in Table 27 contain three obvious misclassifications, i.e., cells 1027, 1028, and 1031. Cell 1027 is wholly composed of AREA 11, and the only variable present in the cell and also in the discriminant function is DYKE 3.[1] The removal of DYKE 3 from the analysis makes cell 1027 a barren one,

---

[1]DYKE 3 accounts for +54.7% of the highest discriminant function. Other variables contribute as following: AREA 3: -13.2%, AREA 8: +21.0%, CNTL 11: +7.2%, and CNTL 15: -3.7%.

which indeed it should be because in the Rouyn-Noranda region, no massive sulphides are known to be associated with AREA 11, i.e., granites and granodiorites.

In case of cell 1028, the probability of any member from group 2 being as far away from the centroid as this cell, is 0.015, a very low probability. The cell does not appear to belong to the same population as that of group 2 even though it may have been classified as being closest to group 2. This is further borne out by the fact that the second highest probability for this cell is the zero value group.

Similarly, cell 1031 has been forecast to belong in the highest group #6. Yet, the probability of a member from group 6 being as far away from the centroid as cell 1031 is only 0.083. The second highest probability is for the cell to belong in group zero. In all previous runs, whether regression or discriminant, the only basis for an endowment forecast in this cell has been the presence of CNTL 10, the contact length between AREA 2 and AREA 11. CNTL 10 is a variable highly locally related with endowment in cell 1023, and extends into cell 1031. When CNTL 10 is not included in the analysis, the cell is forecast as barren. This too is a result of misclassification.

Cell 1045 too appears to be misclassified in group #1, because the probability of a member from group #1 being as distant from the centroid as this cell is only 0.025.

Cells 1050 and 1053 are both predicted to fall in group 4, but with probabilities less than that of the only reference cell[1] composing the group. These too can be rejected.

The remaining cells appear to conform with the favourable predictions obtained in other discriminant runs. In the following section, the results are jointly discussed and evaluated.

## 9.6 Discriminant Results Review

The efficiency of a discriminant function is affected by the number of groups, by the particular cases falling in individual groups, and by the presence and absence of certain variables. Changes in the number of groups will, for a given set of data, result in modifications to the group centroid resulting from changes in the members composing the groups. The effect can be profound if cells falling in a group show a lack of statistical similarity in terms of multivariate geological relationships. This has been explained previously as the reason for the misclassification of the known endowment cells into a barren[2] category under the two-group and the 5-group A analyses.

The discriminatory power of the function is affected by the variables input, both individually and jointly, depending

---

[1] I.e., cell 1021.

[2] The barren category refers to the zero value group.

on their relative correlations. As Lachenbruch (1975, p. 76) notes:

> (1) If most of the correlations among the poor variates and between poor and good variates are positive and not too large, the joint contribution of the poor variates will be less than in independent cases.
>
> (2) If the correlations are negative, the joint contribution of the poor variates will be more than in the independent case.
>
> (3) Positive correlations have to be quite high if they are to be helpful.
>
> (4) Any variable having negative correlation with the good variate will be helpful.

The role of the variables in resource evaluation is more complex because of the evolutionary nature and variable timings of geological events as inferred by the measurements made. For example, in the 5-group discriminant analysis B,[1] the exclusion of DYKE 3 from the input variables results in the classification of cells 1027 and 1053 as barren. This is because of the high correlation of DYKE 3 with the spatial presence of ore deposits in the region. Similarly, the exclusion of variable CNTL 10 results in cells 1023 and 1031 being classified as barren, even though the former is a known endowment cell. In the same 5-group B, the new discriminant function resulting after excluding DYKE 3 classifies the following additional cells as probably endowment bearing: 1001,

---

[1]See Table 23.

1002, 1007, 1009, 1010, 1011, 1017, 1018, 1022, 1025, 1033, 1045, 1049, 1051, and 1052. All these cells either do not contain DYKE 3, or if it is present, it is insignificant. This underlines the influence of a highly positively correlated variable despite its having been relegated to a lower inclusion level in stepwise discriminant analysis.

A criterion of the effectiveness of the discriminant function is its ability to properly predict known endowment cells. In this regard, both 5-group B, and 7-group analyses have been effective. Table 29 is a comparison of Tables 26 and 27. Cells 1027, 1028, 1031, 1045, 1050, and 1053, already commented upon in evaluating the 7-group analysis are not included in the table. The following is a comparison of the group ranges used in the two analyses:

| Group # | 5-Group Set B ($\$x10^6$) | 7-Group Set ($\$x10^6$) |
|---|---|---|
| 0 | 0- 10 | 0 |
| 1 | 10- 100 | 0- 25 |
| 2 | 100- 500 | 25- 50 |
| 3 | 500-1500 | 50- 100 |
| 4 | > 1500 | 100- 500 |
| 5 | | 500-1500 |
| 6 | | > 1500 |

It will be seen that Groups #1, 2, and 3 of the 7-group set fall within the range of group #1 of the 5 group set B. And groups #4, 5, and 6 of the 7-group set correspond

TABLE 29

COMPARISON OF 5-GROUP B AND 7-GROUP ANALYSES RESULTS[1]

| Cell No. | 5 GRP "B" Known | 5 GRP "B" Predicted Groups and Probabilities | 7 GRP Known | 7 GRP Predicted Groups and Probabilities |
|---|---|---|---|---|
| 1002 | - | 1(.474) | - | 1(.349) |
| 1003 | - | 1(.521) | - | 3(.669) |
| 1007 | - | 1(.735) | - | 2(.805) |
| 1009 | - | 3(.517) | - | 2(.453) |
| 1010 | - | 3(.476) | - | - |
| 1012 | - | 1(.604) | - | 1(.364) |
| 1013 | 1 | 3(.582) | 1 | 1(.914) |
| 1015 | - | 1(.645) | - | 2(.944) |
| 1016 | 1 | 1(.841) | 3 | 3(.985) |
| 1018 | - | 2(.539) | - | - |
| 1020 | - | 1(.468) | - | - |
| 1021 | 2 | 2(.494) | 4 | 4(.900) |
| 1022 | - | 2(.560) | - | - |
| 1023 | 1 | 0(.427) | 2 | 2(.498) |
| 1025 | - | 3(.494) | - | 2(.593) |
| 1029 | 3 | 3(.449) | 5 | 5(.932) |
| 1032 | - | 1(.865) | - | 2(.917) |
| 1033 | - | 3(.488) | - | 2(.760) |
| 1038 | 4 | 4(1.00) | 6 | 6(.995) |
| 1039 | 1 | 1(.997) | 2 | 2(.731) |
| 1041 | - | 1(.444) | - | 1(.899) |
| 1042 | 1 | 3(.388) | 2 | 1(.362) |
| 1049 | - | 1(.716) | - | - |

[1]Figures in parentheses are probabilities associated with the predicted groups. Results for known endowment cells are underlined.

with the #2, 3, and 4 of the 5-group set. The results correspond therefore, for the originally unknown endowment cells 1002, 1003, 1007, 1012, 1015, 1032, and 1041. In terms of the endowment presence forecast, these results agree with those of regression with the exception of cells 1015 and 1032.

Cells 1010, 1018, 1020, 1022, and 1049 forecast by the 5-group functions are not done so by the 7-group discrimination. Of these, only cell 1010 has been forecast as endowment bearing by regression.

Cells 1009, 1025, and 1033, forecast in groups #3 of the 5-group B analysis all fall in group #2 of the 7-group analysis. In this regard, the results of the 7-group set appear to be more reliable because the results are based on comparison with three reference cells constituting group #2 instead of the only one comparing group #3 in the 5-group analysis. Of these, both cells 1009 and 1033 have been favourably forecast as endowment bearing by regression.

Table 30 makes a comparison of the results obtained for the most favourable cells predicted by regression analysis with those of the 5-group and the 7-group analyses. With the exception of cell 1010, the most comparable results are obtained between regression and the 7-group analysis. However, cells 1015, 1025, and 1032, which are not forecast as favourable by regression are predicted as such by both the 5-group and 7-group analyses; their results are summarized in Table 31.

## TABLE 30

COMPARISON OF FORECAST ESTIMATES IN POTENTIALLY FAVOURABLE CELLS IN 5-GROUP B AND 7-GROUP DISCRIMINANT ANALYSES, AND ITERATIVE REGRESSION ANALYSIS

| Cell No. | 5 GROUP B Range | 5 GROUP B Prob. | 7 GROUP Range | 7 GROUP Prob. | REGRESSION |
|---|---|---|---|---|---|
| 1002 | 10- 100 | .474 | 0-25 | .349 | 13.78 |
| 1007 | 10- 100 | .735 | 25-50 | .805 | 12.40 |
| 1009 | 500-1500 | .517 | 25-50 | .453 | 2.01 |
| 1010 | 500-1500 | .476 | - | - | 9.81 |
| 1012 | 10- 100 | .604 | 0-25 | .364 | 16.01 |
| 1033 | 500-1500 | .488 | 25-50 | .760 | 62.32 |
| 1041 | 10- 100 | .444 | 0-25 | .899 | 15.01 |

[1]All figures are in millions of dollars.

# TABLE 31

COMPARISON OF FORECAST ESTIMATES IN POTENTIALLY FAVOURABLE CELLS IN 5-GROUP B AND 7-GROUP DISCRIMINANT ANALYSES[1]

| Cell No. | 5 GRP "B" Range | 5 GRP "B" Prob. | 7 GRP Range | 7 GRP Prob. |
|---|---|---|---|---|
| 1015 | 10- 100 | .645 | 25-50 | .944 |
| 1025 | 500-1500 | .494 | 25-50 | .593 |
| 1032 | 10- 100 | .865 | 25-50 | .917 |

Range = $\$ \times 10^6$

[1]These cells are not favourably forecast by iterative regression analysis.

## 9.7 Summary of Discriminant Analysis

(1) Discriminant analysis is a powerful technique for classifying cells into pre-selected groups. While it can assign a cell to an appropriate group, it cannot define its specific position within the value range of the group. Discriminant analysis is also very useful for explaining and understanding the system being evaluated.

(2) The selection of discriminating variables in discriminant analysis is as important as the selection of explanatory variables for regression. The inclusion of a single highly correlated, but otherwise non-relevant variable can result in a misleading discriminant function. Conversely, this is also true when a relevant variable is omitted.

(3) It is most essential that all cases be drawn from the same population, at least within the groups. In studies using geological data, this information may initially be known only subjectively. To overcome this difficulty, the selection of optimal groups, both in terms of their number and the range of values assigned to them, becomes as important as the selection of the "best" variables.

(4) The cells that are classified as favourable by discrimination are essentially the same as those by regression analysis. As with regression analysis, the forecast values are the minimum possible because the calibrator[1] used

---

[1] I.e., the known endowment.

is the minimum possible.

(5) Discriminant analysis is robust to violation of normality assumption. The validity of the results obtained can be tested by the procedure of "leaving one out" in which, a known endowment cell is assigned to the lowest value group and discriminant analysis performed to observe if that cell can be assigned to the correct group on the basis of the remaining cells.

# CHAPTER 10

## CONCLUSIONS

The objective of this study is to make estimates of undiscovered resource endowment in the Rouyn-Noranda region using multivariate statistical analyses. However, because of the nature of the data base, the study has broadened into an evaluation of the multivariate techniques themselves, and the problems associated in their application to geological data. The following points summarize the essential conclusions reached in this study:

(1) It has been demonstrated in the study that known mineral resource endowment and associated geological characteristics can be quantitatively related for developing a model to forecast unknown resource potential. This is possible in spite of the common shortcomings and problems of geological and resource endowment data.

(2) The quantitative modelling of geological phenomenon draws its guidance and interpretation from the accepted concepts on the processes that created the phenomenon. The more complete the geological explanation available, the better is the interpretation of the quantitative results. However, while a geomathematical interpretation depends fundamentally on the validity of the geological concepts held,

it can also contribute to the evaluation of geological concepts themselves.

(3) It has been shown in the study that a multivariate statistical model developed for a geological system is optimally efficient for the system itself. Its application outside the system can lead to serious errors and should not be attempted except on a reconnaissance level.

(4) In any multivariate statistical application in resource evaluation, the selection of the "best" set of variables and the determination of the "best" order of variables input in stepwise analysis is a critical consideration. Such selection must be guided by the generally accepted concepts on the genesis of the resource being evaluated.

(5) The quantifiable relationships which are shown to exist between endowment and geology, should provide exploration guidelines in terms of the geological variables determined to be the most pertinent, and the subareas estimated to be potentially the most favourable. This supports exploration planning investment and execution.

(6) The application of quantitative methods in geology has certain limitations. Because of the interruptive nature of geological processes, geological data cannot be directly treated like econometric or other "present moment" data. Depending upon the scale of measurements made, the grain and complexity of geology, and the size and shape of the cells adopted, a number of statistical problems will

arise. How critical these problems are will vary with the objective of the study and the statistical approach contemplated. Statistical studies can become meaningless without an awareness of these problems, and attempts must be made to minimize their effects. Furthermore, the results must be interpreted and evaluated against this awareness.

(7) The need for a more objective approach in geological mapping and a more uniform procedure in resource classification has been emphasized in this study. Simultaneously, there is also need for more research in evaluating the role of multivariate techniques in the violation of assumptions and for the development of tests of significance for data that are not normally distributed.

(8) A number of variables in the Rouyn-Noranda region have shown a high positive correlation with known endowment. This is particularly true of the length of north-south to north-west trending dykes. The high correlation may be genetically related with endowment, or may be only spatial and thus spurious. However, even as post-ore features, such variables can have an indirect or consequential relationship with ore formation. This aspect needs additional investigation in the field. The role of the area of diorite, gabbro, requires similar investigation.

(9) The Rouyn-Noranda region is one of the most intensively studied geological regions in Canada. However, there is still debate on the classification of certain rock

types such as rhyolite or tuff and agglomerate. Such lack of uniform terminology is a matter of concern to a quantitative analyst, and this is one of the reasons why extrapolation of a quantitative model outside its own region can be misleading.

(10) In this study, the selection of variables has been based essentially on factor analysis. It is felt that factor analysis can be effectively employed in relating resource endowment and associated geological parameters. It can also be used in determining if more than one mode of ore formation is present in the region. In this manner, factor analysis helps guide exploration. However, it does not provide estimates of undiscovered endowment potential in the region.

(11) Regression analysis and discriminant analysis are two powerful statistical methods of resource potential evaluation. Both these methods have been used in this study and provide comparable results. However, regression analysis is shown to have advantages in terms of greater flexibility and more defined results. Furthermore, this method does not require a normalized distribution for the explanatory variables.

(12) It has been shown that an iterative approach to regression analysis is an effective and efficient approach in resource endowment forecasting for the type of data available for this study. The role of multicollinearity is

minimized using a checkerboard-type division of the data. With these techniques, it is possible to assign the most likely values to response variables in cells that do not have a known endowment value.

(13) In spite of the skewness of the data, factor analysis and discriminant analysis, both of which assume a normal distribution for the data, appear to be robust enough against violation of this assumption. When tests of significance are not applicable because of the absence of a normalized distribution, the results can be validated by the "leaving one out" technique as applied in this study.

(14) Some seven out of the 56 unknown endowment cells in the study region are forecast as potentially favourable for further exploration at different levels of endowment estimates, both by regression and discriminant analyses. Three other cells are forecast as favourable by discriminant analysis alone. The remaining unknown endowment cells are estimated to be barren by both regression and discriminant analyses. From an exploration point of view, it is just as significant to know that an area is barren than that it is favourable, within the hypothesis testing error limits. This, therefore, greatly narrows down the target area for potentially economic exploration effort.

The endowment estimates forecast in favourable cells are of a low order considering the current costs of exploration and mine development. Therefore, these cells should be considered on the basis of both their absolute and their relative estimated values.

(15) The absolute potential of 131 million dollars forecast for the region is insignificant compared to its known endowment. This implies that either the region has reached a state of near exhaustion of endowment or, a state of saturation has been reached in the utility of traditional geological concepts and information in exploration and resource modelling. The region cannot be considered as exhausted when it is considered that the estimated theoretically possible endowment exceeds that known for copper by a factor of 31 and that for zinc by a factor of 222. It is therefore suggested that newer concepts will have to be evolved and new information obtained, particularly for the depth dimension in the region for incorporating in both the forecasting model and in actual exploration. Geophysics can be a valuable tool in this regard.

# CLAIM OF ORIGINAL WORK AND CONTRIBUTION TO KNOWLEDGE

It is claimed that the results and interpretations presented in this thesis are the original work of the author and a contribution to knowledge. The following features of the study have contributed to knowledge:

(1) A quantitative assessment of the undiscovered resource potential of the Rouyn-Noranda region has been made. The application of statistical models to the size of the study area chosen and to the size of individual cells within it, offer advantages that are not possible in "traditional" reconnaissance level studies. The contribution to knowledge is, thus, in solving the problems associated with statistical resource evaluation for a study area and cell size smaller and more detailed than that used in any previous study.

(2) The present study is basic for other similar studies that may have to be undertaken in other mining regions that have a history of intensive exploration, development and mineral production, but which are becoming exhausted. Such studies can further guide exploration investment, decision making, and regional planning.

(3) It has been demonstrated in this study that quantifiable relationships exist between mineral endowment and associated geological characteristics within the Rouyn-

Noranda region, and that this relationship can be applied to make forecasts of unknown endowment within the region.

(4) The problem of selecting a value to assign to a response variable for cells with no known endowment has been resolved in this study by using an "iterative regression" technique.

(5) Although multicollinearity presents a serious problem in the application of statistical models to geological data, the author has been able to control this problem with the "checkerboard" technique.

(6) The author has obtained estimates of resource potential of the Rouyn-Noranda region using regression analysis and discriminant-analysis separately. In doing so, the author has demonstrated the relative strengths and weaknesses of the two techniques.

(7) The use of factor analysis in data reduction and in the selection of variables has been demonstrated. The author has discussed the additional advantage of factor analysis in identifying different sets of geological characteristics in terms of their sympathetic and antithetic associations with known endowment in the region. This knowledge can assist in ore-genesis, and in focussing mineral exploration on the most relevant set of geological characteristics.

(8) The importance of selecting the most appropriate geological variables for statistical analysis has been demonstrated in this study. The advantages of including or

excluding certain variables has been discussed in terms of the spatial and genetic affiliations of each variable with known endowment in the region.

(9) A characteristic of geological processes is that they are evolutionary, interruptive, or both. But geological measurements can only be made at a single point in time. This aspect of geological knowledge has been stressed throughout the thesis. Statistical applications, however sophisticated, can become misleading without a realization of this dichotomy. An understanding of geological concepts is fundamental to the application of multivariate statistical techniques to resource potential evaluation, and to the interpretation of results.

(10) It has been shown that statistical techniques, while drawing their interpretation from geological knowledge, also contribute to the understanding of geological processes to differentiation among different geological environments.

(11) The need for developing distribution-free multivariate techniques and, in particular, tests of significance, has been featured in the study. A leaving one out technique is used in validating the predictive models developed in the study.

(12) It is demonstrated that quantitative relationships developed for a geological system should be confined to that system only. Extrapolation outside the system can result in a loss of resolution and erroneous conclusions.

(13) An up-to-date synthesis of the geology of Rouyn-Noranda region and its base metal sulphide deposits has been presented within the larger framework of the Abitibi belt.

(14) The conclusions arrived at in the present study should significantly contribute to the Quebec government's efforts to expand the province's mineral resources. Particularly, because in April 1977, the government of Quebec[1] announced a 5-year exploration project for the Abitibi region in north-west Quebec. The project will cost between 60 and 80 million dollars. Similar projects will be undertaken in other provincial regions if the Abitibi project proves successful.

(15) The present study is a contribution to the future research needs referred to by Harris (1975, p. 349):

> . . . although progress in the design of geostatistical models for the appraisal of metal resources is apparent, much of this progress is reflected in a greater awareness of the inadequacies of past and present efforts and in a greater ability to delineate the problem and to formulate the questions that will be the substance of future research and achievement.

---

[1]The Northern Miner, April 7, April 14, and April 21, 1977.

# REFERENCES


Agterberg, F.P. (1976). Oral communication.

_______________ et al. (1972). Geomathematical evaluation of copper and zinc potential of the Abitibi area, Ontario and Quebec. G.S.C. Paper 71-41, 55 pp.

_______________ and Robinson, S.C. (1971). Mathematical problems in geology: Bull. of the International Statistical Institute, Proceedings of the 38th session, vol. XLIV, book 1, pp. 567-595.

Allais, M. (1957). Method of appraising economic prospects of mining exploration over large territories: Algerian Sahara case study. Management Science, vol. 3, No. 4, pp. 285-347.

Ambrose, J.W. (1941). Clericy and La Pause Map-Areas, Quebec: Geol. Surv. Can., Mem. 233.

Anderson, C.A. (1969). Massive sulfide deposits and volcanism: Economic Geology, vol. 64, pp. 129-146.

Azis, A., Janakiraman, C., and Werner, A.B.T. (1977). A computer simulation model for the assessment of mineral resources: pp. 1182-1196 IN Proceedings of the 14th symposium, Application of computer methods in the mineral industry, New York, SME of American Inst. of Min., Met. and Pet. Engineers, Inc., 1207 pp.

_______, Barry, G.S., and Haugh, I. (1972). The undiscovered mineral endowment of the Canadian Shield in Manitoba. Mineral Bull. MR 124, Mineral Resources Branch, Department of Energy, Mines and Resources, Ottawa, 42 pp.

Baragar, W.R.A. (1968). Major-element geochemistry of the Noranda volcanic belt, Quebec-Ontario: Can. Jour. Earth Sci., vol. 5, pp. 773-790.

Barnes, H.L., and Czamanske, G.K. (1967). Solubilities and transport of ore minerals, pp. 334-381 IN Barnes, H.L., ed., Geochemistry of Hydrothermal Ore Deposits: New York, Holt, Rinehart and Winston, Inc., 670 pp.

Barry, G.S., and Freyman, A.H. (1970). Mineral endowment of the Canadian Northwest--a subjective probability assessment. Part "B" of Mineral Information Bull. M.R. 105, Mineral Resources Branch, Department of Energy, Mines & Resources, Ottawa, pp. 57-95.

Barth, T.F.W. (1962). Theoretical Petrology: New York, John Wiley and Sons, Inc., 416 pp.

Bates, R.C. (1959). An application of statistical analysis to exploration for uranium on the Colorado plateau: Economic Geology, vol. 54, pp. 449-466.

Bibb, R., and Roncek, D.W. (1976). Investigating group differences: Sociological Methods and Research, vol. 4, no. 3, pp. 349-379.

Blondel, F., and Lasky, S.G. (1956). Mineral reserves and mineral resources: Economic Geology, vol. 51, pp. 686-697.

Botbol, J.M. (1971). Characteristic analysis as applied to mineral exploration: Ninth symposium on decision-making in the mineral industry, C.I.M.M. spec. vol. 10, Montreal, 1970, pp. 92-99.

Brinck, J.W. (1972). The prediction of mineral resources and long-term price trends in the non-ferrous metal mining industry. Proceedings 24th International Geological Congress, Montreal, Section 4, pp. 3-15.

__________ (1967). Note on the distribution and predictability of mineral resources: Euratom Report EUR-3461e, Brussels, 25 pp.

Brooks, C., and Gélinas, L. (1977). The relationship of base metal mineralization to chemostratigraphic divisions in volcanic rocks of Canadian Archean greenstone belts: Mineral Exploration Research Institute, Montreal, Research Paper 77-1, 19 pp.

Burnaby, T.P. (1966). Distribution free quadratic discriminant functions in paleontology, IN Merriam, D.F. (Ed.), Computer applications in the earth sciences: Colloquium on classification procedures, computer contribution 7, Kansas Geological Survey, pp. 70-77.

Cacoullos T., and Styan, G.P.H. (1973). A bibliography of discriminant analysis, pp. 375-434 IN Discriminant Analysis and Applications: New York, Academic Press Inc.

Cameron, E.M. et al. (1971). Distribution of ore elements in rocks for evaluating ore potential: nickel, copper, cobalt, and sulfur in ultramafic rocks of the Canadian Shield. IN Boyle, R.W. (Ed.) Geochemical Exploration, C.I.M.M. spec. vol. 11, pp. 298-313.

Campbell, F.A. (1962). Age of mineralization at Quemont and Horne mines, Can. Inst. Min. Met. Trans., vol. 65, pp. 293-296.

Cooley, W.W., and Lohnes, P.R. (1962). Multivariate procedures for the behavioral sciences: New York, John Wiley & Sons, Inc., 211 pp.

Daniel, C., and Wood, J.W. (1971). Fitting equations to data: New York, John Wiley & Sons, Inc., 342 pp.

Davis, J.C. (1973). Statistics and data analysis in geology: New York, John Wiley & Sons, Inc., 550 pp.

DeGeoffroy, J., and Wignall, T.K. (1970). Statistical decision in regional exploration application of regression and Bayesian classification analysis in the southwest Wisconsin zinc area: Economic Geology, vol. 65, pp. 769-777.

_______________, and Wu, S.M. (1970). A statistical study of ore occurrences in the greenstone belts of the Canadian Shield: Economic Geology, vol. 65, pp. 496-504.

Derry, D.R. (1973). Potential ore reserves -- an experimental approach. Western Miner, October 1973, pp. 115-122.

Descarreaux, J. (1973). A petrochemical study of the Abitibi volcanic belt and its bearing on the occurrences of massive sulphide ores: C.I.M. Bull., vol. 66, February 1973, pp. 61-69.

Dimroth, E. (1977). Archean subaqueous autoclastic volcanic rocks, Rouyn-Noranda area, Quebec: classification, diagnosis and interpretation: Geol. Surv. Can. Paper 77-1A, pp. 513-522.

____________, et al. (1974). Preliminary report on stratigraphic and tectonic work on the Rouyn-Noranda area: Quebec Dept. Natural Resources, open-file report DP-246.

____________, et al. (1973). Tectonic and volcanogenic studies in the Rouyn-Noranda area: Quebec Dept. Natural Resources, open-file report GM-28491.

Douglas, R.J.W. (1970). Geology and economic minerals of Canada: Geol. Surv. Can. Econ. Geol. Report, no. 1, 838 pp.

Draper, N., and Smith, H. (1966). Applied regression analysis: New York, John Wiley & Sons, Inc., 407 pp.

Dugas, J. (1977). Oral communication.

__________ (1966). The relationship of mineralization to Precambrian stratigraphy in the Rouyn-Noranda area, Quebec. Geol. Assoc. Can. Spec. Paper No. 3, pp. 43-55.

__________, Latulippe, M., and Duquette, G. (1965). Metallic mineralization in Noranda, Matagami, Val d'Or and Chibougamau areas. Quebec Department of Natural Resources, Mineral Deposits Service, Compilation Map No. 1600 V.

Elliot-Jones, M.F. (1973). Economic forecasting and corporate planning: New York, The Conference Board Inc., 75 pp.

Gélinas, L., and Brooks, C. (1973). Archean quench-texture tholeiites: Can. Jour. Earth Sci., vol. 11, pp. 324-340.

Gilmour, P. (1965). The origin of massive sulphide mineralization in the Noranda district, northwestern Quebec: Geol. Assoc. Can. Proc., vol. 16, pp. 63-81.

Gnanadesikan, R. (1977). Statistical data analysis of multivariate observations: New York, John Wiley & Sons, Inc., 311 pp.

Goodwin, A.M. (1965). Mineralized volcanic complexes in the Porcupine-Kirkland Lake-Noranda region, Canada: Econ. Geol., vol. 60, pp. 955-971.

__________, and Ridler, R.M. (1970). The Abitibi orogenic belt, IN Baer, A.J. (Editor): Symposium on basins and geosynclines of the Canadian Shield: Geol. Surv. Can. Paper 70-40, pp. 1-30.

Griffiths, J.C. (1966). Application of discriminant functions as a classification tool in geosciences, IN Merriam, D.F., (Ed.). Computer applications in the earth sciences: Colloquium on classification procedures, Computer Contribution 7, Kansas Geological Survey, pp. 48-52.

Guild, P.W. (1976). Discovery of natural resources: Science, vol. 191, pp. 709-712.

Harris, D.P. (1975). Geostatistics in the appraisal of metal resources: IN Vogely, W.A. (Ed.), 1975, Mineral Materials Modelling, pp. 287-352, Washington, D.C., Resources for the Future, 404 pp.

__________ (1968). Alaska's base and precious metals resources: a probabilistic regional appraisal. Mineral Resources of Northern Alaska, MIRL Report No. 16, Mineral Industries Research Laboratory, University of Alaska, pp. 60-101.

__________ (1965). An application of multivariate statistical analysis to mineral exploration. Unpublished Ph.D. thesis, The Pennsylvania State University, 261 pp.

__________, and Azis, A. (1970). Future metal requirements for Mexico to 1990 and potential domestic supply: Proceedings of the Council of Economics, New York, AIME, pp. 245-302.

__________, Freyman, A.J., and Barry, G.S. (1970). The methodology employed to estimate potential mineral supply of the Canadian Northwest--an analysis based upon geologic opinion and systems simulation: Part "A" of Mineral Information Bull. MR 105, Mineral Resources Branch, Department of Energy, Mines & Resources, Ottawa, pp. 1-56.

Herfindahl, O.C., and Kneese, A.V. (1974). Economic theory of natural resources: Columbus, Charles E. Merril Publishing Co.

Hutchinson, R.W. (1965). Genesis of Canadian massive sulphides reconsidered by comparison to Cyprus deposits: C.I.M. Bull., vol. 58, Sept. 1965, pp. 972-986.

Jöreskog, K.G., Klovan, J.E., and Reyment, R.A. (1976). Geological factor analysis: Amsterdam, Elsevier Scientific Publishing Co., 178 pp.

Kelly, A.M., and Sheriff, W.J. (1969). A statistical examination of the metallic mineral resources of British Columbia: Proceedings of symposium on decision-making in mineral exploration II, University of British Columbia, February 1969, pp. 221-243.

Klecka, W.R. (1975). Discriminant analysis, pp. 434-467, IN, Nie, et al. (1975).

Klovan, J.E., and Billings, G.K. (1967). Classification of geological samples by discriminant function analysis: Bull. Canadian Petroleum Geology, vol. 15, no. 3, pp. 313-330.

Krauskopf, K.B. (1967). Source rocks for metal bearing fluids, pp. 1-33, IN Barnes, H.L., (Ed.), Geochemistry of Hydrothermal Ore Deposits, New York, Holt, Rinehart and Winston, Inc., 670 pp.

Krogh, T.E., and Davis, G.L. (1971). Zircon U-Pb ages of Archean metavolcanic rocks in the Canadian Shield: Carnegie Inst. Yearbook 70, pp. 241-242.

Lachenbruch, P.A. (1975). Discriminant analysis: New York, Hafner Press, 128 pp.

Larson, L., and Webber, G.R. (1977). Chemical and petrographic variations in rhyolite zones in the Noranda Area, Quebec: C.I.M. Bull., vol. 70, August 1977, pp. 80-93.

Lawley, D.N., and Maxwell, A.E. (1971). Factor analysis as a statistical method: London, Butterworth & Co., 153 pp.

Lenthal, D.H. (1972). The application of discriminatory and cluster analysis as an aid to the understanding of the acid phase of the Bushweld Complex. University of Witwatersrand, Economic Geology Research Unit Information Circular, no. 72, 33 pp.

Lindgren, W. (1933). Mineral deposits: New York, McGraw-Hill Book Co. Inc., 930 pp.

Link, R.F., and Koch, G.S. (1967). Linear discriminant analysis of multivariate assay and other mineral data: U.S. Bureau of Mines, Report of Investigation, no. 6898, 23 pp.

Mackenzie, B.W. (1973). Corporate exploration strategies, pp. 1-8 IN Application of computer methods in the mineral industry, SAIMM, Johannesburg, 441 pp.

__________ (1971). Mineral economics and economic geology: Paper presented at the Central Canada University Geological Conference, Montreal, Oct. 14-17, 1971, 15 pp.

_______________ (1969). A decision-making theory of the mining firm: pp. 321-343 IN Proceedings Council of Economics, AIME, New York, 357 pp.

Maki, D.P., and Thompson, M. (1973). Mathematical models and applications: Englewood Cliffs, Prentice-Hall Inc., 492 pp.

Martin, H.L., Cranstone, D.A., and Zwartendyk, J. (1976). Metal mining in Canada to the year 2000: Requirements for new mines, exploration expenditures, capital investment and labour: Resources Policy, vol. 2, March 1976, pp. 11-24.

Mather, P.M. (1976). Computational methods of multivariate analysis in physical geography: New York, John Wiley & Sons, Inc., 532 pp.

Matheron, G. (1971). The theory of regionalized variables and its applications: Les Cahiers du Centre de Morphologie Mathématique De Fontainebleau, 211 pp.

McKelvey, V.E. (1975). Mineral resource perspectives, 1975. United States Geological Survey Prof. Paper 940, 24 pp.

_______________ (1972). Mineral resource estimates and public policy: American Scientist, vol. 60, Jan.-Feb. 1972, pp. 32-40.

_______________ (1960). Relation of reserves of the elements to their crustal abundance: American Jour. Sci. (Bradley Vol.), vol. 258-A, pp. 234-241.

Nie, N.H., et al. (1975). Statistical package for the social sciences: New York, McGraw-Hill Book Co., 675 pp.

Nolan, T.B. (1950). The search for new mining districts: Economic Geology, vol. 45, no. 7, pp. 601-60 (also, Presidential Address, Society of Economic Geologists, delivered at the El Paso meeting, November 1949).

Oftedahl, C. (1958). A theory of exhalative-sedimentary ores: Geol. Fören. Stocholm Förh., vol. 80, no. 1, no. 492, pp. 1-19.

Overall, J.E., and Klett, C.J. (1972). Applied multivariate analysis: New York, McGraw-Hill Book Co., 500 pp.

Parker, R.L. (1967). Composition of the earth's crust: U.S. Geological Survey Professional Paper 440-D, 17 pp.

Price, P. (1934). The geology and ore deposits of the Horne mine, Noranda, Quebec: Can-Inst. Min. Met. Trans., vol. 37, pp. 108-140.

________ (1948). Horne Mine, IN Structural geology of Canadian ore deposits, Can. Inst. Min. Met., Montreal, pp. 338-361.

Riddell, J.E. (1952). Wall-rock alteration around base-metal sulphide deposits of north-western Quebec: Unpublished Ph.D. thesis, McGill Univ., Montreal.

Ridge, J.R. (1972). Noranda, IN Annotated bibliographies of mineral deposits in the Western Hemisphere: G.S.A. Mem. 131, pp. 130-135.

Roscoe, S.M. (1965). Geochemical and isotopic studies, Noranda and Matagami areas: C.I.M. Bull., vol. 58, Sept. 1965, pp. 965-971.

Rose, A.W. (1972). Favourability of Cornwall-type magnetite deposits in Pennsylvania using geological, geochemical and geophysical data in a discriminant function. Jour. Geochem. Exploration, vol. 1, pp. 181-184.

Sakrison, H.C. (1966). Chemical studies of the host rocks of the Lake Dufault mine, Quebec. Unpub. Ph.D. thesis, McGill University, Montreal.

Sangster, D.F. (1976). Oral communication.

________ (1972). Precambrian volcanogenic massive sulphide deposits in Canada: a review. Geol. Surv. Can. Paper 72-22, 44 pp.

Schanz, Jr., J.J. (1975). Resource terminology: an examination of concepts and terms and recommendations for improvement. Final Report, Research Project 336, prepared by Resources for the Future, Inc., for Electric Power Research Institute, Palo Alto, California, 116 pp.

Sekine, Y. (1963). On the concept of concentration of ore-forming elements and the relationship of their frequency in the earth's crust. International Geology Review, vol. 5, no. 5, pp. 505-515.

Shaw, D.M., Dostal, J., and Keays, R.R. (1976). Additional estimates of continental surface Precambrian Shield composition in Canada. Geochimica et Cosmochimica Acta, vol. 40, pp. 73-83.

Simmons, B.D., et al. (1973). Geology of the Millenbach massive sulphide deposit, Noranda, Quebec. C.I.M. Bull., vol. 66, Nov. 1973, pp. 67-78.

Sinclair, A.J., and Woodsworth, G.L. (1970). Multiple regression as a method of estimating exploration potential in an area near Terrace, B.C.: Economic Geology, vol. 65, pp. 998-1003.

Spence, C.D. (1967). The Noranda area, IN Can. Inst. Min. Met. Centennial Field Excursion Guidebook, pp. 36-39.

____________, and Spence, A.F. de Rosen (1975). The place of sulphide mineralization in the volcanic sequence at Noranda, Quebec: Economic Geology, vol. 70, pp. 90-101.

Spetzler, C.S., and Von Holstein, Carl-Axel, S. (1975). Probability encoding in decision analysis: Management Science, vol. 22, no. 3, pp. 340-358.

Stanton, R.L. (1972). Ore petrology: New York, McGraw-Hill Book Co., 713 pp.

Tatsuoka, M.M. (1970). Discriminant analysis, number 6 IN selected topics in advanced statistics, Champaign: Institute of personality and ability testing, 57 pp.

Tatsumi, T. (Editor) (1970). Volcanism and ore genesis: Tokyo, Univ. of Tokyo Press, 448 pp.

Van de Walle, M. (1972). The Rouyn-Noranda area, IN, 24th Int. Geol. Cong. Field Excursion Guidebook A41-C41, pp. 41-51.

Webber, G.R. (1962). Variation in the composition of the Lake Dufault granodiorite: Can. Inst. Min. and Met. Trans., vol. 65, pp. 55-62.

Wilson, H.D.B., et al. (1965). Archean volcanism in the Canadian Shield: Can. Jour. Earth Sci., vol. 2, pp. 161-175.

Wilson, M.E. (1941). Noranda district, Quebec: Geol. Surv. Can. Mem. 229, 162 pp.

Wilson, M.E. (1962). Rouyn-Beauchastel map areas, Quebec: Geol. Surv. Can. Mem. 315, 140 pp.

Zwartendyk, J. (1975), (Ed.). "Departmental Terminology and Definitions of Reserves and Resources (Metals, Industrial Minerals, and Coal)": Canadian Department of Energy, Mines and Resources, Interim document, Ottawa, January 30, 1975, 42 pp.