A SPATIAL APPROACH TO MINERAL POTENTIAL MODELLING USING DECISION TREE AND LOGISTIC REGRESSION ANALYSIS

PAULINE HONARVAR

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

A SPATIAL APPROACH TO MINERAL POTENTIAL MODELLING USING DECISION TREE AND LOGISTIC REGRESSION ANALYSIS

by

©Pauline Honarvar

A thesis submitted to the

School of Graduate Studies

in partial fulfilment of the

requirements for the degree of

Master of Science

Department of Geography
Memorial University of Newfoundland

October 2001

St. John's Newfoundland

ABSTRACT

Logistic regression analysis and classification methods using decision tree analysis were used to generate two quantitative mineral potential maps for the Lake Ambrose area (NTS 12A/10) of central Newfoundland. The response variable consisted of 47 surface mineral occurrences plus 49 randomly selected sites representing nonmineral occurrences. Mineral deposit models and regional exploration methods were used to choose a set of predictors consisting of geology, fault proximity, till and lake sediment geochemistry, and surficial geology. A spatial weighting function predictor was developed to account for the clustering of the mineral occurrences.

The predictors were analyzed and recoded to derive a set useful in developing the quantitative models. The categorical geology predictor was converted into two binary predictors; felsic volcanics and mafic volcanics. Fault proximity was analyzed by the weights of evidence method to determine the optimal buffer threshold to convert the continuous distance values to a binary measure 'close to faults' versus 'far from faults'. The optimal thresholds were the 400 m and 1000 m buffers. Principal components analysis was applied to the till and lake sediment geochemistry to derive component summary variables. Three component predictors were added to the database: till component 2 (TPC2) representing base metals and gold, lake sediment component 2 (LPC2) representing base metals and lake sediment component 4 (LPC4) representing gold and its pathfinder elements. The till geochemistry predictors (Au, Cu, Pb, Zn and

TPC2) were analyzed for spatial autocorrelation and an interpolated surface was derived using kriging techniques. The lake sediment geochemistry predictors (Au, Cu, Pb, residual Zn, LPC2 and LPC4) were converted to a surface by mapping their values on the catchment basins in which they were sampled.

The decision tree analysis indicated the spatial weighting function, felsic volcanics and the 400 m binary fault proximity predictor were significant predictors of mineral potential. Logistic regression analysis indicated that the spatial weighting function, felsic volcanics, the 1000 m binary fault proximity predictor and copper in till were significant predictors of mineral potential. The agreement, at the 96 sample sites, between these two modelling methods was 84.3%. The decision tree and logistic regression raster mineral potential maps were compared using Yule's α. A value of 0.54 indicates good agreement between the maps. Both models correctly classified approximately 79% of the 96 mineral/nonmineral occurrences. Due to the sparseness of the dataset, accuracy could not be measured as there were not enough samples to set aside a test dataset.

Mineral potential reliability maps were generated using the mutually exclusive and exhaustive regions from the decision tree analysis and the joint probability model for the logistic regression analysis. The mineral potential and reliability maps were combined (multiplied) to form a favourability map. The favourability maps from the decision tree and logistic regression analyses were combined to indicate overall zones of high mineral potential and high reliability. Mineral exploration claims cover much of the study area and only a minor part of the high favourability areas were not claimed as of September 2000.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Alvin Simms, for his advice, guidance and suggestions in analyzing this multifaceted problem. I would also like to thank my committee members, Dr. Peter Davenport and Dr. Norm Catto, for their suggestions, review of the initial proposal and review of later versions of this thesis.

I would not have embarked on a career in GIS if it were not for the support and encouragement of Peter Davenport as well as Larry Nolan and Gerry Kilfoil of the Newfoundland and Labrador Department of Mines and Energy. Their enjoyment of the subject inspired me to pursue it.

The Department of Mines and Energy provided in-kind support, including digital data and computer support. I would like to thank many people at the Department who assisted me in countless ways, including Scott Swinden, David Liverman, Stephen Colman-Sadd, Baxter Kean, Greg Stapleton, Jan Smith, Harjit Missan, Joe Atkinson and Keith Parsons, among others.

And last, but not least, I want to thank my husband, Mark Wilson, for being my emotional support during all these years, and taking care of everything else while I completed this odyssey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAS – Atomic Absorption Spectrophotometry
amsl – Above Mean Sea Level
CV – Coefficient of Variation
EDA – Exploratory Data Analysis
DEM – Digital Elevation Model
DTA – Decision Tree Analysis
DTMS – DeskTop Mapping System
GIS – Geographical Information System
ICP-AES – Inductively Coupled Plasma – Atomic Emission Spectrometry
INAA – Instrumental Neutron Activation Analysis
IV – Independent Variable
KMO – Kaiser-Meyer-Olkin (measure of sampling adequacy)
LOI – Loss-on-Ignition
LRA – Logistic Regression Analysis
MD – Mahalanobis Distance
MODS – Mineral Occurrence Data System
NAD – North American Datum
NTDB – National Topographic DataBase
NTS – National Topographic System
PCA – Principal Components Analysis
RMS – Root Mean Square
SC – Studentized C
SWF – Spatial Weighting Function
VLG – Victoria Lake Group
VMS – Volcanogenic Massive Sulphide
WOE – Weights of Evidence

LIST OF APPENDICES

CHAPTER 1

Introduction

1.1 Background

Mineral potential mapping is a significant tool in developing new mineral exploration targets as well as providing information to aid in the assessment of boundaries for new parks and native land claim areas. Objective quantitative methods of calculating mineral potential are becoming easier through the use of Geographic Information Systems (GIS), geostatistical and statistical programs. The statistical programs are used to determine the degree of association between mineral occurrences and various predictors without regard for the spatial associations. GIS and geostatistical programs provide the tools to assess and display these spatial associations.

An appropriate set of predictors for mineral potential mapping can be determined by reviewing the mineral deposit models applicable to the study area. Mineral deposit models are general descriptions of the formation of mineral deposits based on theoretical and empirical data. They include such information as the geological setting, lithogeochemical signatures and spatial characteristics (Kirkham *et al.*, 1993).

In addition to predictors determined through mineral deposit models, regional geochemical data (such as till and lake sediment geochemistry) have proved to be excellent predictors in providing regional mineral exploration targets. In many areas of Canada, extensive glacial cover and the presence of numerous lakes and swamps (many of them

glacially derived) overlie and hide surface mineral occurrences. To stimulate mineral exploration activity in this difficult terrain, Canadian federal and provincial geological surveys provide the results of these regional geochemical surveys at little or no cost (Davenport *et al.*, 1993).

The combination of data types and data distributions resulting from the varied predictors limit the types of statistical analyses which can be performed. Common parametric statistical tests require normally distributed numerical data. Mixtures of categorical (*e.g.* geology) and continuous (*e.g.* geochemistry) data with a binary dependent variable (mineral occurrence) reduce the number of useful statistical analyses to only a few.

1.2 Purpose

The purpose of this study is to develop an inductive, quantitative model for assessing mineral potential by determining the statistical relationship between mineral occurrences and geological, geochemical and spatial factors. Mineral deposit models and traditional exploration methods are used to indicate an appropriate set of predictors to assess (*e.g.* geology, fault proximity, till and lake sediment geochemistry). In addition to standard exploration predictors, a spatial weighting function has been developed to provide a spatial measure of proximity between mineral occurrences to account for the grouping of the mineral occurrences.

The mixture of data types (continuous and categorical) and data distributions (non-normal) being analyzed limits the statistical methods which can be used to determine the

quantitative association between the mineral occurrences (as a binary response variable) and the geological, geochemical and spatial factors (as the predictor variables). The sparse mineral occurrence dataset also imposes limitations on the statistical methods. Logistic regression analysis (LRA) and classification methods using decision tree analysis (DTA) are two methods that can handle various data types and data distributions as well as sparse datasets. Due to the sparse dependent dataset, accuracy analysis cannot be accomplished by splitting the data into a training dataset and a test dataset. Therefore, a comparative analysis of the results of the two modelling methods (LRA and DTA) is used to provide an indication of their agreement. The reliability of the two individual models is also assessed, by combining estimates of the errors for the individual predictors. The mineral potential maps and reliability maps are combined to provide a favourability map. Areas which have a high mineral potential with a high reliability are most favourable for mineral exploration.

A better understanding of the statistical and spatial relationships between mineral occurrences and associated significant predictors helps to determine if these quantitative methods can benefit mineral potential modelling. These results may provide more insight and a savings of time and money in the search for new exploration targets and assessments of boundaries for land use planning.

CHAPTER 2

Background Information and Literature Review

The statistical analysis of spatial data requires a good understanding of the factors involved in their genesis and spatial distribution. Statistical analysis methods are dependent on data types and the statistical distribution of the data. This chapter will review the nature of some common geochemical media and provide a summary of statistical methods used in similar studies in the past.

2.1 Exploration Methods

As a part of a mineral exploration program, till and lake sediment samples provide the means to determine areas which have anomalous concentrations of either the primary metals of interest (*e.g.* gold and copper) or pathfinder elements which are often associated with the primary metals (*e.g.* arsenic and antimony). The assumption commonly held is that anomalous till or lake sediment samples should indicate the proximity of anomalous source rocks and possibly mineral occurrences or deposits.

To properly assess the results of till and lake sediment geochemistry the variation (anomalous versus background values) in their geochemical values, the sampling methodology and techniques of chemical analysis in the laboratory need to be understood. Variations in the value of each element are based on numerous factors inherent in sample genesis, location, collection, preparation and analysis.

The location of previously known mineral occurrences are of primary interest to exploration geologists because deposits and occurrences often occur in groups or clusters (mining 'camps'). Obvious surface occurrences have been found by prospectors and geologists so information from these known occurrences, along with information from mineral deposit models, can be used to indicate the potential of new areas.

2.1.1 Till Sampling Surveys

Till sampling programs have been an integral part of many mineral exploration programs in Canada due to the extensive glacial sediment cover over much of the bedrock. Therefore, a good understanding of glacial processes and the specific glacial history of an area is essential in order to analyze the results of a till sampling program.

It is the aim of regional till sampling programs to define areas of exploration interest. Tills, which are sediments produced exclusively from glacial erosion, transportation and deposition, provide an excellent geochemical exploration tool. Theoretically, till samples containing anomalous values of an element (*e.g.* copper) can be traced back to their bedrock source by following the dispersion train in the up-ice flow direction (Shilts, 1976). Dispersion trains resulting from a single source have been modelled by negative exponential curves (Figure 2.1; Shilts, 1976; Strobel and Faure, 1987). This model indicates that a basal till sample is predominantly composed of local-provenance material that gradually decreases in proportion to new material being added and original material being deposited down-ice. The shape of the negative exponential

curve is dependent on factors such as the lithology, structure and topography of the source and dispersal areas, the proximity of ice divides, and the size of the ice sheet.

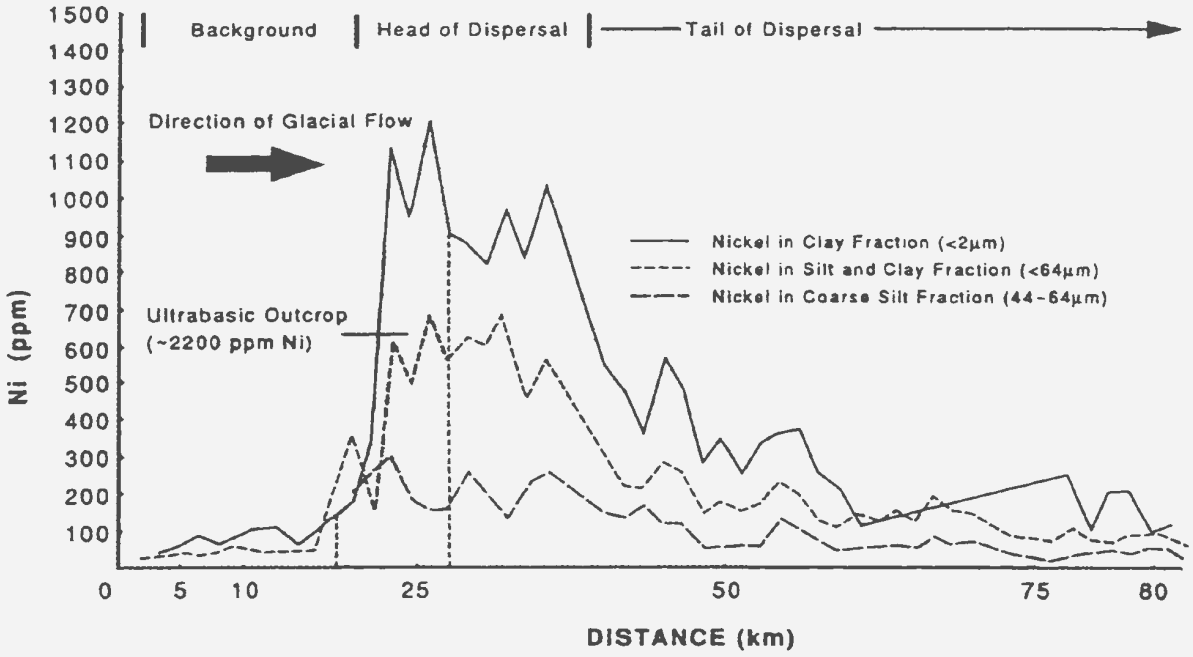

Figure 2.1 : Negative exponential dispersal curve indicating similar patterns for all three sediment fractions (after Shilts, 1993).

The method of analysis of till samples plays an important part in determining anomalous samples and also in determining their provenance. Till samples are initially separated into size fractions grading from pebble-size clasts to sand, silt and clay size fractions. The lithologies of the pebble-size clasts are analyzed in an attempt to trace provenance. Finer fractions, such as the silt and clay size fractions, are the preferred fractions for geochemical analysis (Shilts, 1993).

To determine the most favourable fraction for geochemical analyses (*i.e.* a fraction which produces the highest signal-to-noise ratio), it is necessary to understand the partitioning of the minerals into the different size fractions. The partitioning is based on the 'characteristic terminal mode' of the source rocks (Shilts, 1993). Glacial erosion and abrasion of the minerals composing the rock will result in each mineral being selectively partitioned into a characteristic grain size, due to such factors as the hardness or structure of the mineral. Analyzing each grain size fraction will result in a geochemical signature reflected by the minerals that dominate that fraction.

The most popular grain size for trace element analysis in Newfoundland is the "silt+clay" (<63 μm) fraction (Batterson, 1989). This size fraction is favoured for analysis because compared to larger size fractions it contains higher concentrations of metallic elements, as opposed to inert silicate minerals, and therefore produces a higher signal-to-noise ratio, making it easier to detect anomalies. The "silt+clay" fraction can be easily separated from the bulk sample using a sieve and is more cost effective than the clay fraction which requires settling or centrifuging techniques.

Klassen (1994) sampled upper C or lower B horizon soils developed on tills throughout the Buchans-Robert's Arm and Victoria Lake Group areas. Over 800 samples were collected at a sampling density range of 4 to 100 km^2. The conclusions reached were that till geochemical patterns essentially reflected the bedrock composition. Base metal values were elevated in the tills of the Lake Ambrose area compared to tills in the other areas. The tills above the Tally Pond and Tulks Hill volcanics were geochemically distinct from each other.

Batterson *et al.* (1998) sampled tills from the Grand Falls-Mount Peyton area in a systematic manner with an approximate sampling density of 1 sample per 3 km^2 for a total of about 800 sites. They sampled the BC or C horizon tills. A cursory analysis of the data using principal components analysis indicated inter-element associations reflecting mafic lithologies, peralkaline granite affinity, light rare earth element affinity, and gold and its pathfinder elements (*i.e.* arsenic and antimony). These four components accounted for about 60% of the total variance in the data. From graduated dot plot maps of the elements, visual analysis indicated that measurable geochemical dispersion trains commonly extended less than 5 km whereas field observations indicated clasts of distinct geological units were found up to 10 km away.

2.1.2 Lake Sediment Sampling Surveys

Lake sediment sampling surveys are a cost-effective means of reconnaissance mineral exploration to determine favourable areas as follow-up targets (Coker *et al.*, 1979). These surveys have been carried out in Canada since the early 1960s by private exploration companies and since the early 1970s by the federal and provincial geological surveys.

A lake sediment sample is considered to be geochemically representative of the catchment basin it resides in. This model is based on two concepts: 1) the clastic portion of the lake sediment is a composite representation of the catchment basin as a result of physical weathering (Levinson, 1980; Rogers, 1988), and 2) the fine-grained particles of the lake sediment (*e.g.* clays and organic material) adsorb metal ions which are a product

of chemical weathering (Levinson, 1980). There may also be a component of metals from groundwater recharge into the lake (Levinson, 1980). Limnological studies of Nova Scotian lakes by Ogden (1986, in Rogers, 1988) indicate only limited groundwater-lakewater interaction occurring in the perilimnion zone (zone of water recharge; Figure 2.2), with no groundwater movement in the tardelimnion zone (zone below recharge zone). Therefore, in most lakes surficial processes play a much larger role than previously anticipated.

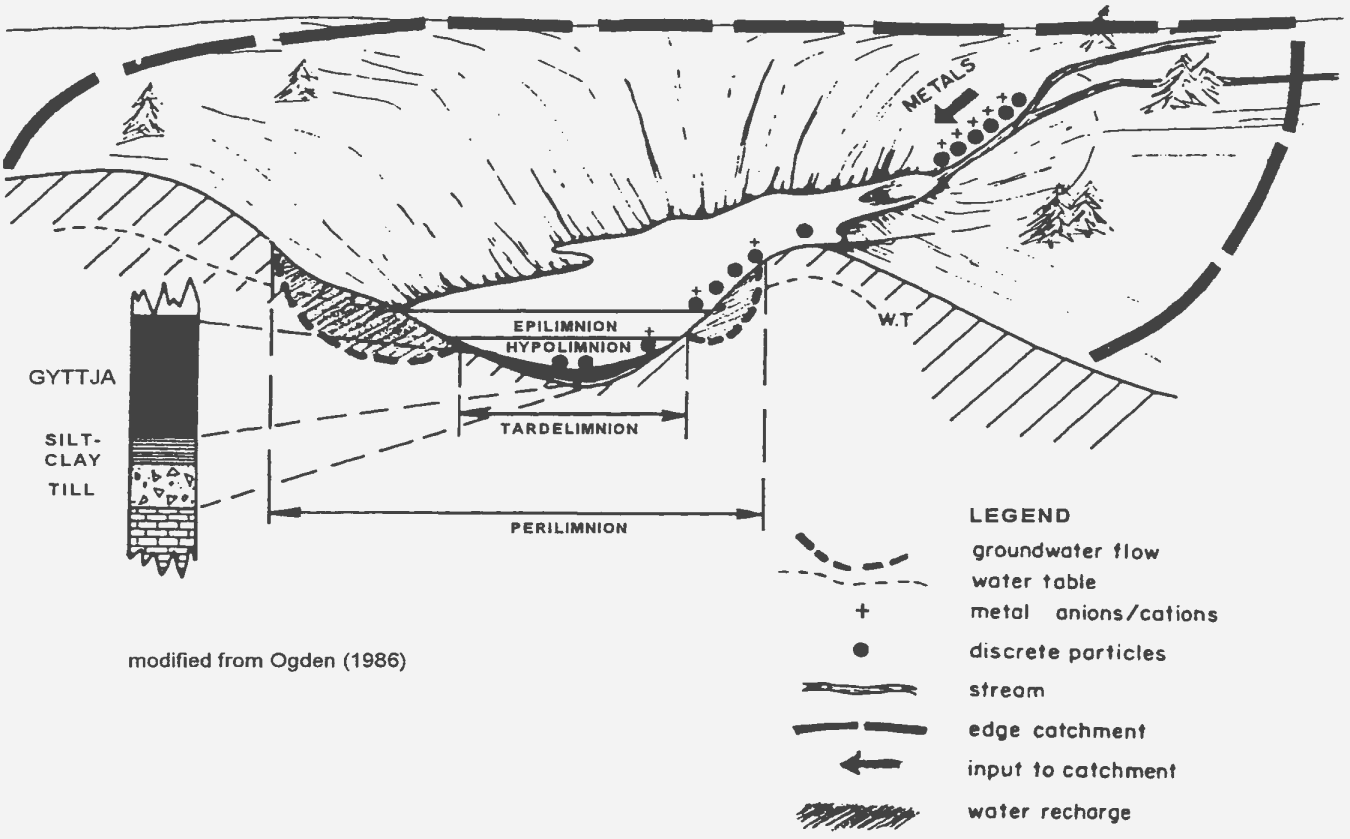

Figure 2.2 : The lake catchment exploration model (after Rogers, 1988). Sediments and metal ions are derived from the catchment area.

Lakes within a catchment basin containing a metallic mineral occurrence are more likely to contain anomalous metal values than similar lakes in barren basins. Many factors

may combine to mask true anomalies or create false anomalies. According to Levinson (1980) and Hornbrook *et al.* (1975), factors that may need to be determined in order to effectively understand the variation in geochemistry of a lake sediment sample include:

1) the prevalent type of weathering of the bedrock and sediments,
2) whether ground or surface waters are the most likely method of trace element transport,
3) the most suitable lake sampling sites,
4) the size and depth of the lake,
5) mineralogy and terminal mode particle size (*i.e.* most likely size for each mineral) of the bedrock and surficial sediments,
6) effects of adsorption by iron and manganese hydroxides and organics,
7) the rate of erosion and sedimentation in the catchment basin (related to such factors as topography and vegetation),
8) contamination by anthropogenic sources, and
9) the oxidizing or reducing conditions of the material sampled.

The change in the sediment geochemistry due to some of these factors (*e.g.* oxidized materials) can be reduced by the selective sampling of a consistently similar material throughout the study area. Other effects (*e.g.* lake depth or adsorption by organics) can be negated by using residual values derived from linear regression techniques (Davenport *et al.*, 1974).

Analysis of lake sediment samples collected in Newfoundland since the late 1970s has indicated element correlations with the size and depth of lakes, as well as with iron, manganese and loss-on-ignition values (an indication of the amount of organic material

present). Through the use of stepwise linear regression, false anomalies caused by these factors have been removed (Davenport *et al.*, 1974). The residual values (expressed as normalized Z values) have provided a clearer picture of the spatial distribution of anomalous element values in lake sediments.

2.1.3 Mineral Occurrences

'Mineral occurrence' is a generic term given to all locations where there is the presence of minerals of potential economic interest. These occurrences range from minor mineral indications (*i.e.* the presence of economic minerals on an outcrop or a minimum assay value of a metallic element; Stapleton, 1999) to producing deposits. Since the early 1970s, mineral occurrences have been mapped and tabulated in Newfoundland and Labrador (Stapleton *et al.*, 2000). The digital Mineral Occurrence Data System (MODS) consists of detailed descriptions of all mineral occurrences and an abbreviated tabular database. The descriptions of each mineral occurrence were summarized from industry assessment reports as well as Department of Mines and Energy geology reports. The tabular database is a summary of selected items from the descriptive data and includes information such as the mineral deposit name, location (UTM coordinates), major and minor commodities (*e.g.* gold, copper, zinc), rock type and a short description of the mineralization.

MODS is used extensively by the mineral industry to assess the potential and simplify the compilation of data for new areas of interest. This information provides geologists with the type and style of mineralization that occurs in an area and indicates the

rock types in which the mineralization occurs. Given this information, the exploration geologist can focus on an exploration method best suited to finding the specific mineral deposits of interest (Swinden, 1992).

2.2 Statistical Analysis Techniques

The analysis of exploration geochemical data is a multistage process that includes spatial and nonspatial techniques. Univariate descriptive statistical analyses, such as the mean and standard deviation and methods used in exploratory data analysis (EDA) such as box-and-whisker plots, provide a first-order indication of the variability in the geochemistry. Trace element concentrations coupled with analytical detection limits can produce a limited range in variance of the geochemical values. This limited variance can reduce the significance of statistical tests. Bivariate statistical analyses, such as correlation analysis, are useful in indicating inter-element associations. Unusually strong associations may indicate adsorption of metals by organic materials or by iron and manganese hydroxides, causing false anomalies. Multivariate analyses within a dataset, such as principal components analysis (PCA), are useful in reducing a large dataset to a few components that better represent key summary factors of the data (*e.g.* geochemistry of sediments). Multivariate analyses among datasets, including linear regression techniques, nonlinear logistic regression analysis, and decision tree analysis, are useful in indicating associations between different types of variables.

Adding a spatial factor to statistical analysis techniques provides a representation of real-world systematics to field data (*e.g.* mineral occurrence locations). Geostatistical

techniques, such as variogram analysis and Moran's I calculations, can help determine whether point sample values are spatially related. This provides important information on whether the data can be represented as an interpolated surface, if autocorrelated, or as polygonal units, if it does not show spatial autocorrelation. The following sections provide a brief overview of the types of statistical analyses mentioned above.

2.2.1 Univariate and Multivariate Analysis within a Dataset

The initial analysis of geochemical data consists of confirmatory statistical analysis (*e.g.* mean, median, standard deviation, skewness or coefficient of variation). Information provided by confirmatory statistics indicates the variation in element values as well as the data distribution (Davenport *et al.*, 1994). Single element or multi-element plots using contours or symbols to represent classes of data provide a view of the spatial distribution of the element values (Hornbrook *et al.*, 1975; Klassen, 1994; Cook *et al.*, 1995).

An alternative to confirmatory statistical methods is exploratory data analysis (EDA) that provides a set of resistant techniques (Sibley, 1991). Resistant techniques are better suited to the analysis of geochemical data which do not exhibit the characteristics of a gaussian distribution. O'Connor *et al.* (1988) used EDA techniques, such as box-and-whisker plots, frequency histograms and cumulative probability plots, to determine class intervals of stream sediment sample analyses. Symbol plots were visually assessed to determine inter-element correlations as well as correlations between the elements and the geology.

Symbol plots assume that the geochemistry of a sample is representative of the 'point' at which the sample was obtained. To determine the actual zone of influence around lake sediment sample points, Bonham-Carter and Chung (1983) used autocorrelograms and kriging methods to determine that most of the spatial autocorrelation was due to the lithological units. Other studies (*e.g.* Bonham-Carter and Goodfellow, 1984) indicated that after the effects of lithology had been removed, no spatial autocorrelation structure remained in the residual values of many elements. Therefore, rather than representing the lake sediment data as points or as interpolated surfaces using kriging, Bonham-Carter and Goodfellow (1984, 1986) determined that a catchment basin model was a better zone of influence for lake and stream sediment samples (Bonham-Carter *et al.*, 1987; Wright *et al.*, 1988; Rogers, 1988). The catchment basin is represented by a polygon that can be subsequently coded to represent different element values.

Spatial autocorrelation of polygonal data, such as the geochemistry of the catchment basins, can be tested using Moran's I coefficient, which is based on comparing the values of neighbouring polygons (Cliff and Ord, 1981; Chou *et al.*, 1990):

$$I = \frac{n\sum_i \sum_j W_{ij}(x_i - \bar{x})(x_j - \bar{x})}{S_0 \sum_i (x_i - \bar{x})^2}$$

where n is the number of polygons in the study area, x is the variable being studied (*e.g.* the variable 'Burn', where x=1 for polygons which have been burned and x=0 otherwise),

W is the spatial weight where W=1 if the ith and jth polygons are contiguous, and S_0 is the sum of the spatial weights $\sum_i \sum_j W_{ij}$.

Chou *et al.* (1990) tested various alternatives to the contiguity weight used in the Moran's I coefficient. Chou *et al.* (1990) determined alternative weighting factors to defining a 'neighbour' such as the length of the boundary between adjacent polygons. The authors determined that the contiguity weight alone was actually the best method of defining the neighbours. Based on this finding they developed a spatial weighting function that defined the contiguity between neighbouring polygons and used this function as a spatial predictor in subsequent regression analyses.

A similar approach was developed by Kvamme (1990) but he applied the Moran's I coefficient for data distributed at specific points rather than polygons (Cliff and Ord, 1981). Instead of using a weight defined by spatial contiguity, Kvamme used the inverse distance between two points as the weight. Therefore, where data cannot be logically converted from points to areal representations, spatial autocorrelation can be tested using weights based on inverse distance measurements.

Given the large volumes of attribute data produced by modern laboratory analytical techniques, principal components analysis (PCA) has been used to reduce the numerous attributes to a few key groups of elements (*i.e.* components). These components may show strong association with certain rock types or secondary alteration (Lindqvist *et al.*, 1987; George and Bonham-Carter, 1989). PCA is based on the covariance between pairs of elements and a transformation onto a new set of axes that are

by definition uncorrelated (Daultry, 1976). The transformed principal component scores can be used as attributes in further analysis.

2.2.2 Multivariate Analysis Between Datasets

Comparison of data between datasets often involves dealing with differences in data types (continuous versus categorical data) and conforming to assumptions necessary for classical statistical analyses (*e.g.* normality, linearity etc.). For example, multiple linear regression and discriminant analysis techniques require multivariate normality and equal covariance matrices for all groups (Norušis, 1990). For mixtures of data types with a good possibility of non-normal and non-linear distributions, methods such as logistic regression analysis and decision tree analysis are preferable (Tabachnick and Fidell, 1996).

2.2.2.1 Linear Multiple Regression

Linear multiple regression has been used extensively to determine factors influencing the variation in geochemistry of till, lake sediment, stream sediment or soil samples. The technique has been used to study inter-element associations in datasets to detect false anomalies caused by Fe and Mn hydroxide adsorption (see section 2.1.2). Linear multiple regression has also been used extensively to analyze a wide range of data from different datasets where the data is commonly continuous and conforms to statistical assumptions.

Wright *et al.* (1988) tested ordinary and stepwise multiple linear regression to determine the best linear combination of lake sediment elements (*i.e.* gold, arsenic, tin and

tungsten) associated with gold mineral occurrences. The lake sediment catchment basin polygons were used to associate the lake sediment site with mineral occurrences occurring within the same basin. The strength of the results were measured using the squared multiple correlation coefficient (R^2), and indicated that of the four elements, gold is the strongest predictor of gold mineral occurrences.

Bonham-Carter *et al.* (1987) ran a number of regression experiments to determine the effects of iron, manganese, pH, areal proportion of rock types, presence/absence of rock types and co-occurrence of various rock types among other effects affecting the variation of stream sediment geochemistry. There is an association (spatial autocorrelation) between a stream sediment sample upstream and its neighbouring sample located downstream. Therefore, statistical tests of significance could not be used because the samples were not independent. But regression analysis can still be run to determine associations between various predictor variables and the dependent variable. The results indicated that iron, manganese and lithologic effects explained a very high proportion of the variance in the stream sediment geochemistry. Residual element content in each catchment basin was calculated once the background associations (*i.e.* variation due to lithology, adsorption by organics or Fe and Mn hydroxides etc.) were determined (Bonham-Carter *et al.*, 1987). A residuals map for each element of significance can be compared to the spatial distribution of mineral occurrences by running a regression similar to Wright *et al.* (1988) as described above.

2.2.2.2 Decision Tree Analysis

Decision tree analysis (Kass, 1980; Breiman *et al.*, 1984) is an inductive method of classifying observations (*i.e.* independent variables such as till copper geochemisty or geology) into homogenous subsets used to predict or indicate the best relationship to a response variable. The results are output as a set of 'IF-THEN' rules stipulating the values of predictor variables used to predict the best response event. The primary advantages of decision tree analysis over other multivariate statistical techniques used to classify data (*e.g.* multiple regression analysis and discriminant analysis) are the lack of assumptions concerning data distribution (*e.g.* linearity, normality, heterogeneity), the lack of restrictions on measurement types (*e.g.* mixtures of continuous, categorical and binary data are acceptable) of both response and predictor variables and the acceptance of small sample sizes. Significances are tested using the chi-squared and F-tests, which are more robust to skewed and non-normal distributions, especially for larger datasets. DTA, as discussed below, is based on methods used in the KnowledgeSeeker® program (Angoss, 1993).

All the data is initially entered into the system, with one variable identified as the dependent variable. Categorical data can be analyzed as alphanumerics rather than being recoded as numbers (as required by logistic regression as indicator or dummy variable encoding; Norušis, 1990). Continuous data is automatically subdivided into a number of discrete categories approximately equal in size (the default is 10). The number of categories can be altered as can the class boundaries. Defining the cluster type for each variable is important in order to arrive at the appropriate output groupings. Monotonic

clustering is assigned to ordered variables (ordinal, interval or ratio measurement types) for which adjoining values will be grouped. Similarly, the floating cluster type is assigned to ordered variables which have missing values. Free clustering is assigned to categorical variables (nominal measurement type) which places no restrictions on how categories are combined into groups.

Two methods are available to split the independent data: cluster and exhaustive. The cluster method (Kass, 1975; Kass, 1980) compares all the values of a predictor with the response variable and groups the predictor values such that the within-group similarity compared to the between-group difference is maximized at a chosen significance level (default $\alpha = 0.05$). Increasing the default value (*e.g.* $\alpha = 0.10$) will result in more branches to the tree. The variable(s) with the highest significance test will be used to split each group until a threshold is reached (Biggs *et al.*, 1991). This method is not considered exhaustive because once a value is grouped with other values it is not considered again.

The exhaustive method (Biggs *et al.*, 1991) uses the maximum statistical significance for each split with the response variable. Initially, a contingency table is made for each predictor category (c) versus each response variable category. Each pair of predictor categories is tested, allowing for monotonic, floating or free clustering, and those which are statistically similar are combined until only two compound groups remain (Kass, 1980; Biggs et al., 1991). Compound categories then need to be checked for similarity and may be split apart again. The most significant predictor grouping overall (with k groupings) will be chosen to subdivide the response variable into $k \leq c$ optimal categories. The process is repeated for all predictors at each new node until the

significance is reduced past a critical value or a stopping value (defined by a minimum number of cases).

The chi-squared statistic is used to test associations between categorical predictors and response variables and the F statistic is used to test associations which are a mixture of categorical and continuous (Biggs *et al.*, 1991). Setting the Bonferroni adjustment to the 'adjusted' significance setting has the effect of adjusting (lowering) the significance threshold level to counteract the effects of re-testing and choosing the 'best' grouping (Angoss, 1993). Another Bonferroni adjustment is used to adjust for the number of predictors which may be highly correlated. Setting the 'filter' to the 'exploration' setting sets the adjusted error rate to 20% and provides a method of reviewing the data for patterns which may be missed on a more rigorous setting (Angoss, 1993). The 'prediction' setting provides an adjusted error rate of 5%, which is more in line with standard statistical decision making levels of significance.

On the final analysis, the branches of the resultant tree should be checked for overfitting. With a large dataset, pruning the tree can be accomplished through the use of a random subset of the data not used in building the tree (cross-validation). With smaller datasets, specifying a minimum threshold case size or stop size (*i.e.* branches cannot contain fewer cases than the specified stop size) will help to reduce overfitting. Terminal nodes can also be checked to see if they actually add to the results by checking the accuracy rate. If terminal nodes do not significantly increase the accuracy rate then they are removed. Note that for smaller datasets, when the same data is used to build and test the tree, the accuracy is overestimated (Angoss, 1993). KnowledgeSeeker outputs the

results as a set of mutually exclusive and exhaustive statistically significant 'IF-THEN' decision rules which can be mapped as a set of conditions.

Reddy and Bonham-Carter (1991) used decision tree analysis to determine the best group of data (consisting of geological and geophysical independent variables) to predict the presence of mineral occurrences. They compared two methods of analysis: one using binary predictors (optimized using weights of evidence modelling) and the other using continuous predictors (*i.e.* distance from contacts and geophysical values) converted into categorical classes. The resulting DTA 'IF-THEN' rules provided a summary of the independent variable values used to predict the response variable and the accuracy of that prediction. Overall, the continuous predictors provided a better prediction accuracy of mineral occurrences than the binary predictors optimized by weights of evidence modelling.

2.2.2.3 Logistic Regression Analysis

Logistic regression techniques are used to determine the probability of a discrete event occurring as well as to provide information on the independent variables (IVs) best suited for predicting the discrete event (Tabachnick and Fidell, 1996). Definitions and methods discussed in this section are predominantly based on the logistic regression techniques applied by SPSS™ (Norušis, 1990 and Anonymous, 2000).

The main advantages in using logistic regression over other regression techniques are its lack of assumptions regarding the distribution of the predictors (*i.e.* normal distributions, linear relationships and homoscedasticity) and the lack of constraints on the

predictor measurement type (*i.e.* any combination of continuous or categorical variables; Norušis, 1990; Tabachnick and Fidell, 1996). Other statistical techniques, such as discriminant analysis, can also be used to predict discrete response variables but these may show increased association between response variables and dichotomous predictors (Hosmer and Lemeshow, 1989). Another advantage of logistic regression is that the output can be stated in terms of probabilities with values between 0 and 1 (Figure 2.3; Norušis, 1990).

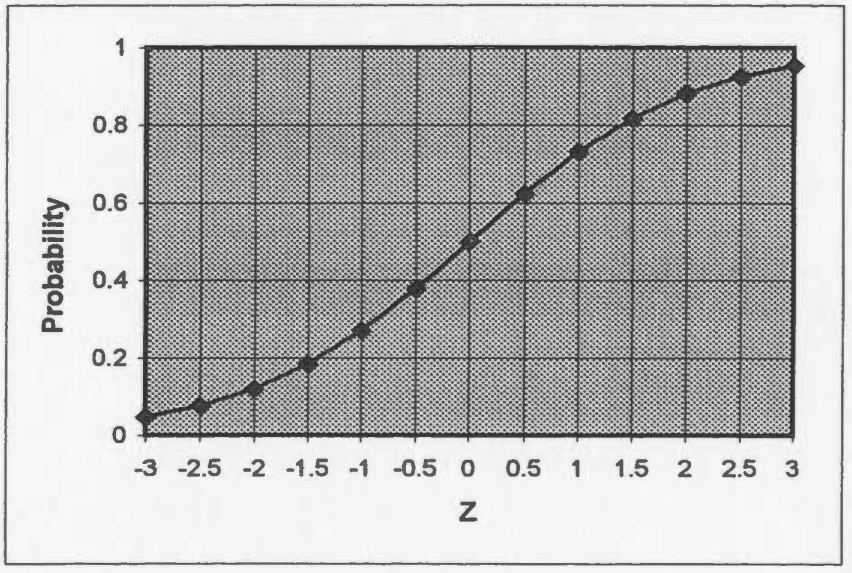

Figure 2.3 : Logistic Regression S-shaped curve. Note that probability values for the curve are between 0 and 1 for any standard normal deviate, Z.

Figure 2.3 indicates the S-shaped (nonlinear) nature of the logistic regression curve that is mathematically defined as:

$$P(\text{event}) = \frac{e^u}{1+e^u} \qquad \text{(Norušis, 1990)}$$

where P(event) is the probability of an event occurring, and u is a linear sum of coefficients and independent variables similar to a linear multiple regression equation:

$$u = B_0 + B_1X_1 + B_2X_2 + \ldots + B_nX_n$$

Rewriting the logistic regression equation as the log of the odds (*i.e.* logit) indicates the nature of the equation:

$$\ln \left[\frac{P(event)}{1 - P(event)} \right] = u \qquad \text{(Tabachnick and Fidell, 1996)}$$

Therefore, the linear sum of coefficients and predictors, u, is equal to the $\log_e$ of the ratio of the probability of an event occurring versus the probability of it not occurring. Equations of this type are solved using iterative calculus techniques, using the maximum likelihood method to determine the best linear combination of predictors (Tabachnick and Fidell, 1996).

A first step in any logistic regression analysis is to check if a model with all predictors (the 'full' model) improves the prediction of an event compared to a model containing just the constant. If the difference between these two models is not significant (*e.g.* α=.05) then it is not likely that the independent variables chosen to predict the response variable are adequate.

Applying stepwise analysis and checking the difference in the log likelihoods can help determine the best statistical model, that is, the model with the fewest number of predictors which best predicts the event (Tabachnick and Fidell, 1996). Due to problems with partial correlations, Hosmer and Lemeshow (1989) suggest increasing the probability value to enter the model, from the default of 0.05 to a value such as 0.20. Similarly, the

probability value to be removed from the model should be increased from the default of 0.10 to, say, 0.30. Using these values, each predictor can be tested individually and in a stepwise analysis to see the effects of predictors on each other and on the outcome.

The best set of predictors to use in the model can be based on a number of statistical tests. The Wald test is defined as:

$$W = \left(\frac{B}{SE}\right)^2 \qquad \text{(Norušis, 1990)}$$

where B is the coefficient and SE is its standard error. For 1 degree of freedom and for large sample sizes, this tests the significance that each coefficient is 0 (Norušis, 1990). This test is distributed as chi-square. Problems with the Wald test occur if the absolute value of the coefficient is very large, producing large standard errors. In this case the predictor values can be altered (transform by logging or removing a standard value, to reduce the large values) or other tests can be applied to judge the significance of the predictor.

Coefficients which prove to be significant in the model provide information on the change in the odds of an event occurring given a single unit change in that predictor (Tabachnick and Fidell, 1996). This is provided by the odds ratio (e^B) where B is the predicted coefficient. Therefore, if the coefficient is greater than 1, e^B will be greater than 1 and the odds of an event occurring are increased if that predictor is included in the model (Norušis, 1990).

Other information provided in the output for the coefficients includes the partial correlation (R) of each predictor with the response variable, given that all other predictors are included. The partial correlation is given by:

$$R = \sqrt{\frac{Wald - 2K}{-2\log likelihood}}$$

where K is the number of parameters estimated and the log likelihood is of the model containing only the constant (Norušis, 1990).

Using a chi-squared goodness-of-fit test, each predictor can, in turn, be tested by including it and removing it from the model. The test used is the difference in the log likelihood (ll) for the model including the predictor versus the model without the predictor:

$$\chi^2 = -2[ll(\text{bigger model}) - ll(\text{smaller model})]$$

The difference is multiplied by two to obtain a chi-squared distribution (Tabachnick and Fidell, 1996). This same method can also be used to test the significance of a larger model against a smaller model, where the smaller model is a subset of the larger model.

As a check on the final model, output of the standardized residuals, Z, indicate those cases that are outliers to the solution:

$$Z_i = \frac{residual_i}{\sqrt{(pred.prob._i)(1 - pred.prob._i)}} \quad \text{(Norušis, 1990)}$$

Examination of outliers may lead to increased understanding of the predictors in the model. Removal of outliers and repeating the analysis may have the effect of changing the relative importance of predictors (Tabachnick and Fidell, 1996; Norušis, 1990).

Chou *et al.* (1990) described a logistic regression method to determine the probability of a wildfire occurring within a polygon *i*. Using the basic logistic regression model:

$$P(\text{occurrence})_i = \frac{e^{u_i}}{1+e^{u_i}}$$

$P(\text{occurrence})_i$ was the probability of a wildfire occurring in polygon *i* and u_i was defined as:

$$u_i = B_0 + B_1AREA_i + B_2PERI_i + B_3ROTA_i + B_4BLDG_i$$
$$+ B_5CAMP_i + B_6ROAD_i + B_7TEMP_i + B_8RAIN_i + B_9SWF_i + e_i$$

where B_0, B_1, B_2,..., B_9 are the logistic regression coefficients which will be estimated from a logistic regression program, AREA and PERI are the area and perimeter of the ith polygon, ROTA is the fire rotation weight, BLDG, CAMP, and ROAD are the areal proportions within range of human influence as defined by buildings, campgrounds and roads, TEMP is the average July maximum temperature, RAIN is the annual precipitation, SWF is the spatial weighting function (representing the local spatial autocorrelation or neighbourhood effect) of the ith polygon, and e_i is the random error term (Chou *et al.*, 1990). The results of the logistic regression analysis indicated that the SWF significantly improved the results of forecasting wildfires in a polygon. This indicated that knowledge of wildfire history was the most significant factor in predicting future wildfires.

2. 3 Summary

In summary, there have been numerous approaches undertaken to statistically compare spatial data and determine significant associations. Based on the literature, confirmatory and exploratory statistical methods and data plots have been chosen to provide a preliminary statistical and spatial assessment of the data. Correlation, linear regression and PCA provide information on how to reduce the number of attributes. Spatial autocorrelation analysis provides information on the spatial distribution of attributes. LRA and DTA determine those independent variables that best predict the mineral occurrences to provide a quantitative mineral potential map of the study area.

CHAPTER 3

Study Area and Data

3.1 Study Area

3.1.1 Location, Access and Physiography

The characteristics of a good study area for mineral potential modelling include the availability of geological, geochemical and mineral occurrence data at a scale, density and variability to provide reasonable comparative analysis. The area chosen for this study is in central Newfoundland, south of Red Indian Lake, and consists of the Lake Ambrose 1:50,000 scale map sheet (NTS 12A/10; Figure 3.1). This area is approximately 15 km south of the Buchans mining camp which extracted copper, lead and zinc ores and is presently undergoing a resurgence in exploration activity. The Lake Ambrose map area has an important mining camp in its own right, consisting of the Victoria Mine prospect (copper-zinc-lead-sulphides in felsic volcanic rocks) and numerous mineral occurrences within proximity (Kean and Jayasinghe, 1980). There are also many mineral occurrences (predominantly consisting of copper, lead, zinc, and gold) throughout the map area. The bedrock and surficial geology are varied, as are the geochemical results from till and lake sediment samples.

Much of the Lake Ambrose map area is not easily accessible due to the presence of wetlands (covering approximately 10% of the area), lakes (covering approximately 13% of the area) and dense forest (covering approximately 70% of the area). However, since

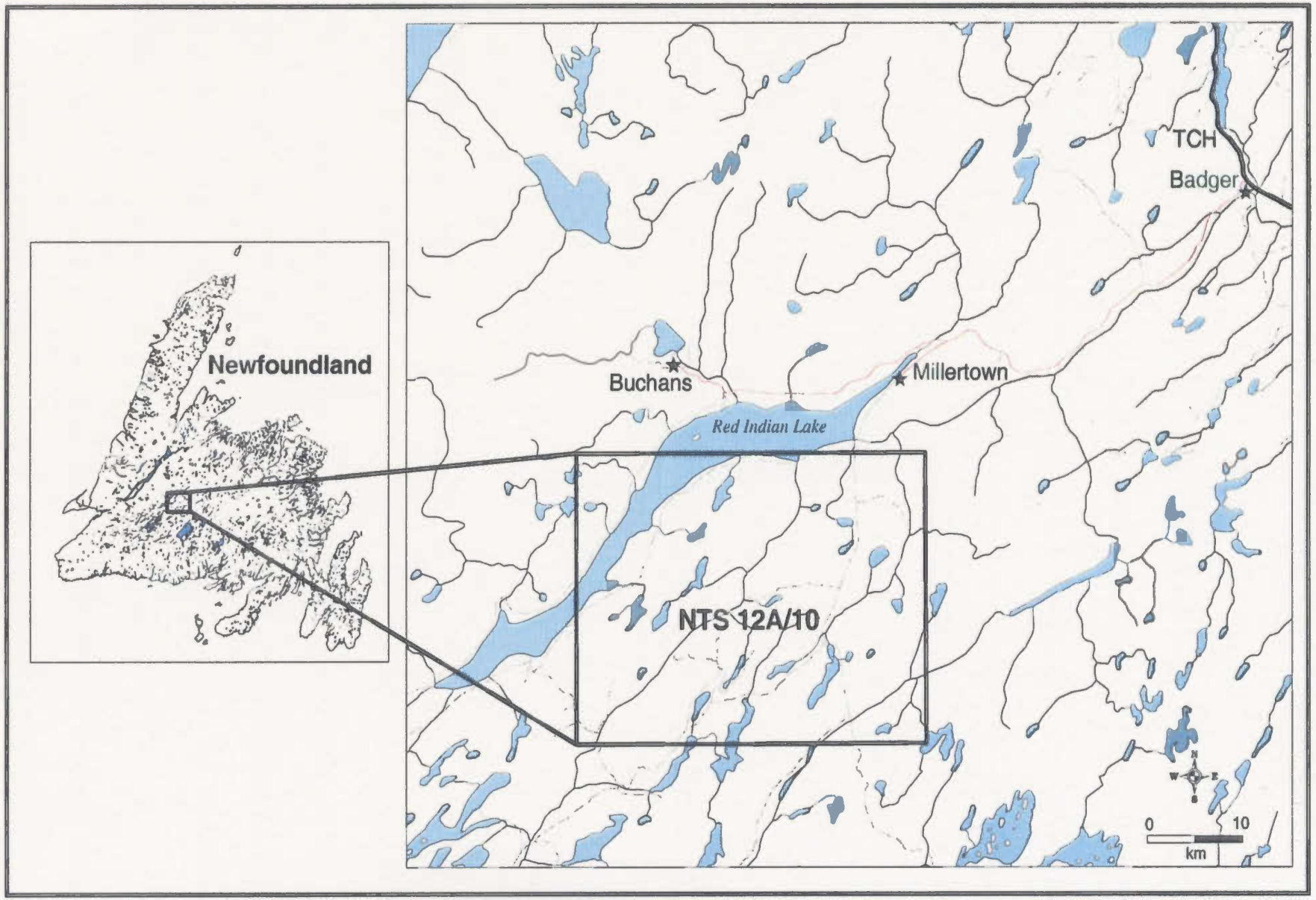

Figure 3.1 : Location of the Lake Ambrose map area (NTS 12A/10).

the early 1900s, the construction of forestry roads has provided increased access (Neary, 1981). The main road from the Trans-Canada highway leads from Badger southwest to Millertown, the closest town to the study area (Figure 3.1).

The Lake Ambrose map area is characterized by an undulating relief with numerous small hills widely scattered throughout (Figure 3.2). The elevation ranges from a high of 480 m above mean sea level (amsl) northwest of Red Indian Lake to a low of 157 m amsl at Red Indian Lake. South of Red Indian Lake, the maximum elevation occurs to the north and west of Lake Ambrose where two hills reach 420 m amsl. The central part of the study area is comparatively flat lying (Figure 3.2) and extensively covered by wetlands. Bedrock exposure is poor due to the glacial till cover (Evans *et al.*, 1990). The lakes vary in size and occur randomly throughout the area. The lakes are predominantly elongate, with a NE-SW trend (Figure 3.2), parallel to the local structural and bedrock trend.

3.1.2 Bedrock Geology and Mineral Occurrences

The geology of the Lake Ambrose map sheet is predominantly composed of the Victoria Lake Group (VLG; Figure 3.3). The VLG is part of the Exploits Subzone, occurring to the south of the Red Indian Line structural lineament, which goes through Red Indian Lake (Williams *et al.*, 1988). The following geological description and age dates are based on Evans *et al.* (1990) unless otherwise stated.

The VLG consists of pre-Caradocian volcanic and sedimentary rocks. There are two assemblages of volcanic rocks in the map area, consisting of the Tulks Hill volcanic

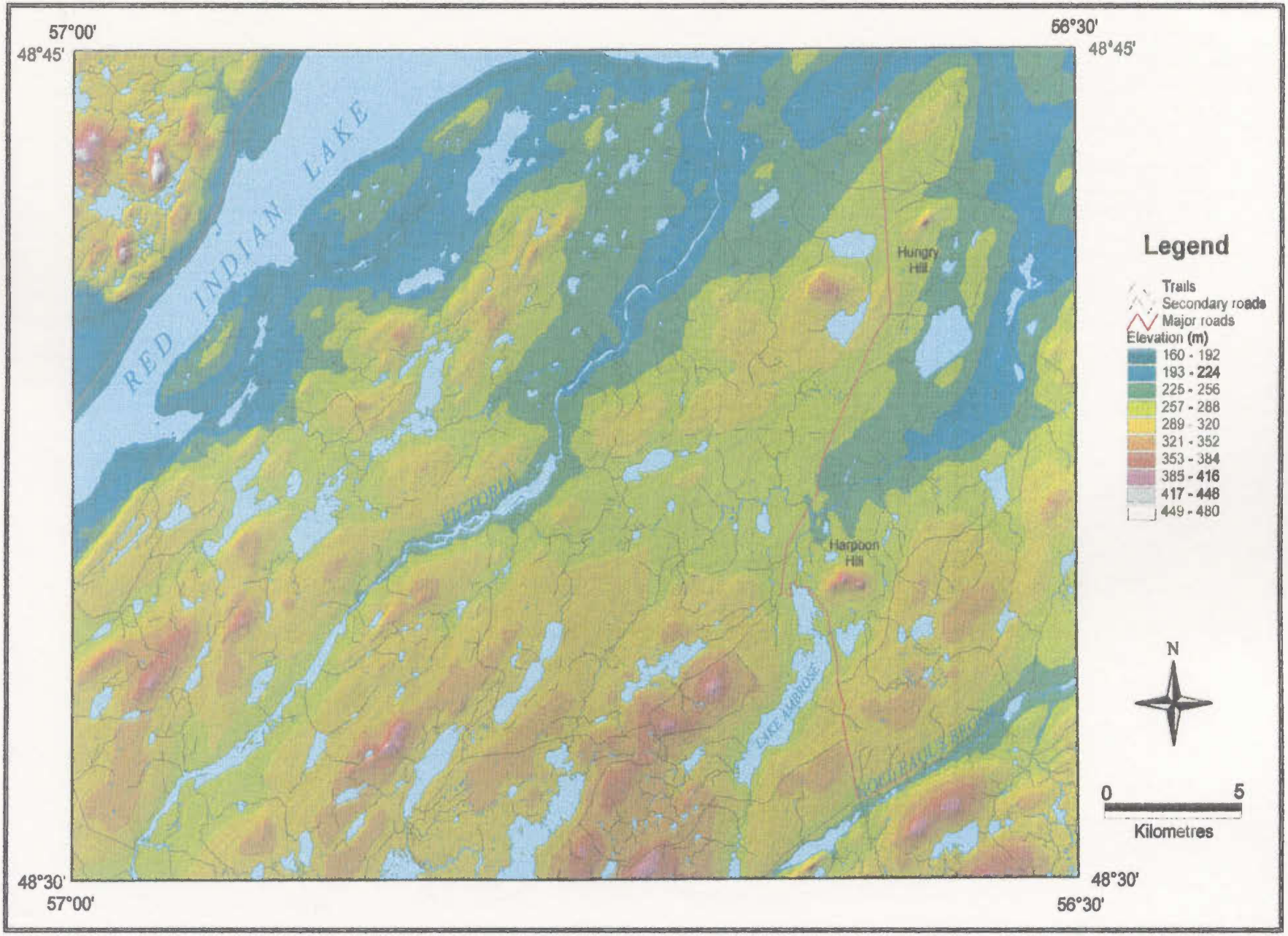

Figure 3.2 : The topography of the Lake Ambrose map area (NTS 12A10). The elevation is gridded on a continuous scale from 160 m to 480 m.

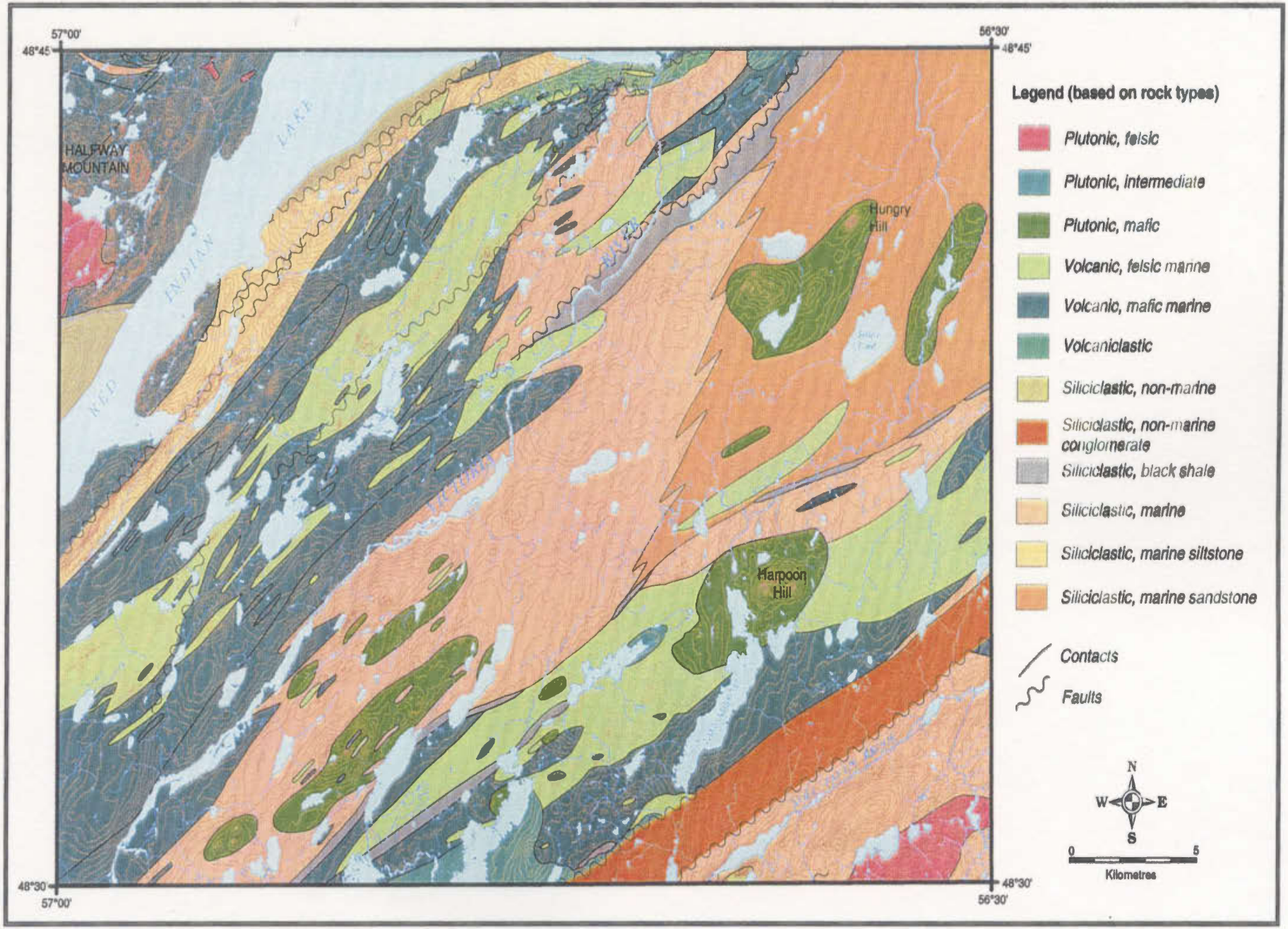

Figure 3.3 : Geology of the Lake Ambrose map area (NTS 12A/10; after Evans et al., 1994).

rocks (494-504 Ma) to the southwest of the map area and the Tally Pond volcanic rocks (511-515 Ma) to the southeast. Both volcanic suites consist of felsic pyroclastic units intercalated with mafic flows, pillow lava, tuff, agglomerate and breccia. The Tally Pond volcanic rocks tend to be more mafic in composition whereas the Tulks Hill volcanic rocks are characterized by more intense deformation. Lower greenschist facies metamorphism is prevalent through much of the map area except along the southern margin where rocks grade to lower-amphibolite facies. Chlorite and sericite define a regional foliation in the VLG. The sedimentary rocks occur predominantly to the northeast and are interpreted as a turbidite sequence derived from the volcanic rocks. The sedimentary rocks include greywacke and interbedded siltstone, shale, argillite, conglomerate and some limestone. Intrusive rocks in the VLG (362-443 Ma; Evans *et al.*, 1994) consist of quartz monzonite, granite, granodiorite, diorite and gabbro and form the major hills in the area.

The VLG is unconformably overlain by the Rogerson Lake Conglomerate (418-443 Ma) in the southeast corner of the map sheet (Figure 3.3). In the area of the Red Indian Lake basin (north and west sections of the map sheet), the VLG is conformably overlain by siltstone and sandstone of the Harbour Round Formation (458-504 Ma). To the west, Devonian and Carboniferous (345-362 Ma) sediments occur consisting of conglomerate, sandstone, shale and siltstone.

According to Evans and Kean (1987), two main types of mineralization occur within the VLG; volcanogenic massive sulphide (VMS) and epigenetic gold (Table 3.1). VMS mineralization occurs predominantly within the Tally Pond and Tulks Hill volcanic belts and consists of copper, zinc and lead mineralization with minor silver and gold

Table 3.1 : Mineral deposit types (from Stapleton, 1999) of the surface mineral occurrences in NTS 12A/10.

Deposit Type	Description
0	Insufficient data to classify
	Stratabound Mineralization: Seafloor (volcanogenic) Sulphide Association (110-169)
	Deposits of the marine volcanic association formed through sub-seafloor hydrothermal processes: includes both volcanogenic massive sulphides and volcanogenic stockworks.
	Deposits associated with sequences of mixed mafic - felsic volcanic rocks (130-139)
130	Undivided volcanogenic sulphide deposits in thick, mixed mafic/felsic volcanic/epiclastic sequences
131	Massive sulphide (± stockwork)
	Deposits associated with dominantly felsic volcanic rocks that are part of thick volcanic/epiclastic sequences (140-149)
140	Undivided volcanogenic sulphide deposits in thick, felsic-dominated volcanic/epiclastic sequences
141	Massive sulphide (± stockwork)
	Deposits hosted by marine sedimentary rocks that are nonetheless part of a dominantly volcanic association (150-159)
151	Clastic host (e.g. "Besshi-type" massive sulphides)
	Hydrothermal, Structurally-Controlled Mineralization (300-399)
	Deposits for which the controlling mechanisms are dominantly structural (e.g. shear zones, faults, fold hinges) rather than stratigraphic.
300	Undivided hydrothermal, structurally-controlled deposits
	Structurally-controlled vein systems with base or precious metals (310-329)
310	Undivided hydrothermal vein systems

(Swinden *et al.*, 1989). Gold mineralization in the VLG appears to be epigenetic and spatially related to major fault zones, lineaments and alteration zones (Evans *et al.*, 1990).

The genetic model for the formation of VMS mineral deposits was first outlined by Oftedahl in 1958 (Franklin, 1993). These deposits are a syngenetic accumulation of sulphides that are deposited from rapidly cooled hydrothermal fluids escaping through fracture and fault zones below sea floor vents (Figure 3.4). Alteration of the host rocks (commonly submarine volcanic rocks), including silicification, sericitization and chloritization, occurs along the faults and fracture zones, as does copper-rich sulphide (stockwork-zone; Figure 3.4). Large accumulations of copper-zinc-lead sulphides form mounds and bedded ores immediately around the seafloor vents.

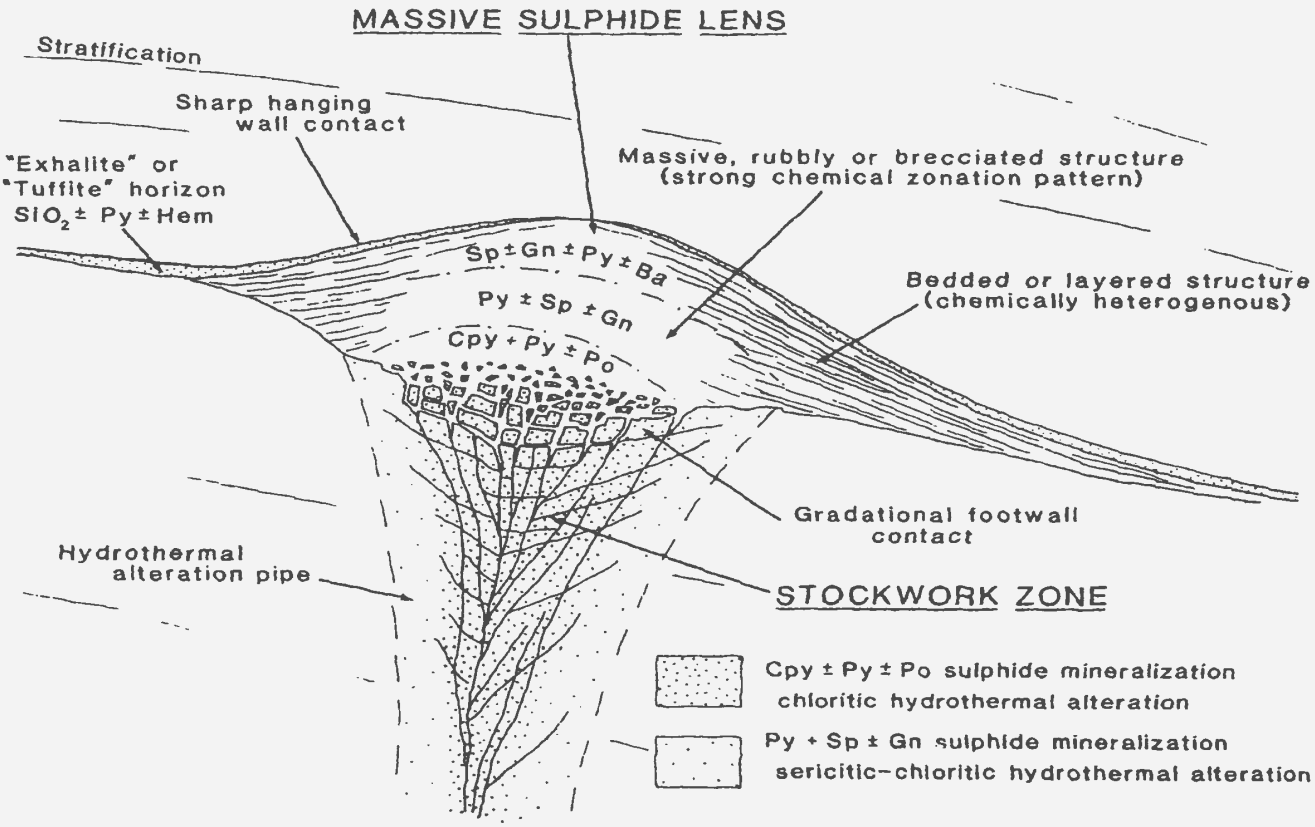

Figure 3.4 : Volcanogenic massive sulphide model indicating zonation of mineralization (after Lydon, 1988).

3.1.3 Glacial History and Surficial Geology

The glacial history in the area of Red Indian Lake is complex with numerous and changing ice flow directions. From the work of Murray (1955), Grant and Tucker (1976), James and Perkins (1981), Vanderveer and Sparkes (1982) and Sparkes (1985, 1987) the following glacial history has been summarized by Klassen (1994).

During the Late Wisconsinan, the Island of Newfoundland was glaciated by local ice caps (Grant, 1974) centred on high altitude areas around the province. Klassen (1994) described four phases of ice flow based on the work of previous geologists and from his own work on glacial striations and dispersal trains of red granite and red micaceous sandstone. The oldest ice flow directions, possibly pre-Wisconsinan, were centred north of Red Indian Lake and flowed south (Phase Ia and Ib, Figures 3.5a and 3.5b). James and Perkins (1981) noted glacial dispersal trains of Buchans-type ore up to eight kilometres southwest of Buchans. The next phase of glacial ice flow (Phase II, Figure 3.5c) was a regional ice flow with an ice cap centred between Lake Ambrose and Victoria Lake to the southwest of the NTS 12A/10 area (Sparkes, 1985). In the study area, the ice flow was to the northeast, as indicated by red micaceous sandstone erratics found to the northeast of Red Indian Lake and also Gullbridge-type ore erratics found northeast of Gullbridge (O'Donnell, 1973). The youngest phases, Phase III and IV (Figure 3.5d), indicate another south to southeast ice flow direction with a final set of flows that tended to be topographically controlled, such as that along the Red Indian Lake basin (Grant, 1975). Klassen (1994) concluded by stating that despite the complex glacial history, the till geochemistry largely reflected the composition of the underlying bedrock geology.

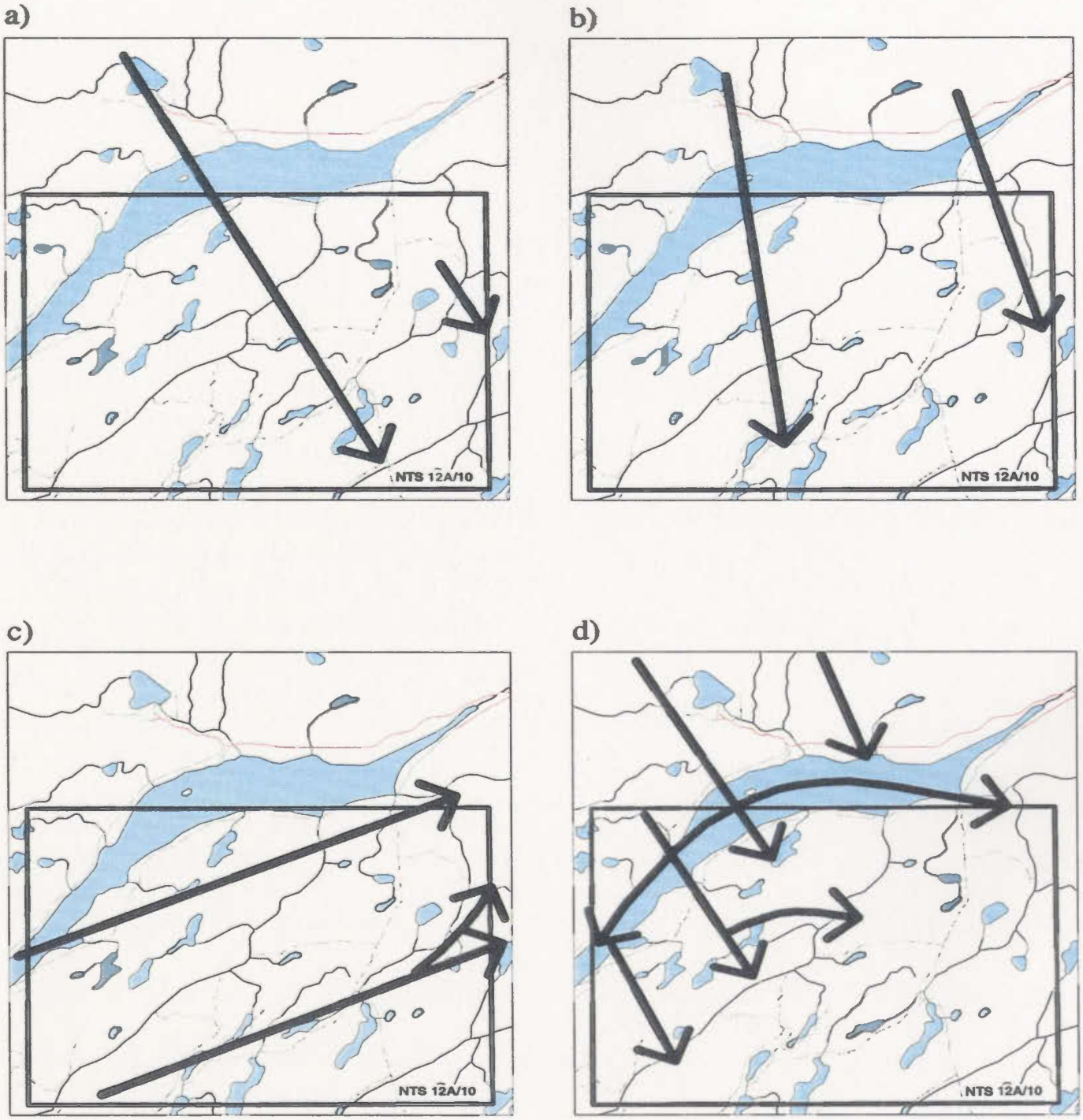

Figure 3.5 : Ice flow history of the Red Indian Lake area (modified from Klassen, 1994). a) Phase I - oldest flow, b) Phase I - younger flow, c) Phase II - older flow to the NE, younger flow to the NNE, d) Phase III - older flow to the SE, Phase IV - youngest flow followed topography.

Glacial sediments in the area are predominantly composed of varying thicknesses of till and ice-contact stratified drift (Figure 3.6). An early compact gray basal till, composed predominantly of local volcanic rocks, is associated with the southerly flow of Phase I (Sparkes, 1985). Other sediments in the area consist of organic material (peat), alluvial deposits, and outwash deposits. Alluvial and outwash deposits predominantly occur along the banks of rivers except for a large outwash deposit to the east of the Victoria River where it enters Red Indian Lake.

In the Victoria Mine area, Mihychuck (1985) determined that the maximum drift thickness was 7.0 m. Three till units were identified in the area:

1) a grey-brown basal till, probably related to the northeasterly flow, containing subangular volcanic clasts in a silty-sandy matrix,

2) a light-grey lower till, probably related to an earlier flow, containing subrounded to subangular clasts in a silty overconsolidated matrix, and

3) a brown-red overlying till, probably of local derivation, containing many angular clasts in a sandy-silty matrix.

Mihychuck's (1985) work in the Tally Pond area, east of the southeast corner of the NTS 12A/10 map sheet, indicated drift cover averaging 5 m in thickness. Only one till unit was identified in this area. The till was light brown, containing subrounded to subangular clasts in a silty, highly overconsolidated matrix. This unit was interpreted as a late stage basal till from ice flow to the northeast.

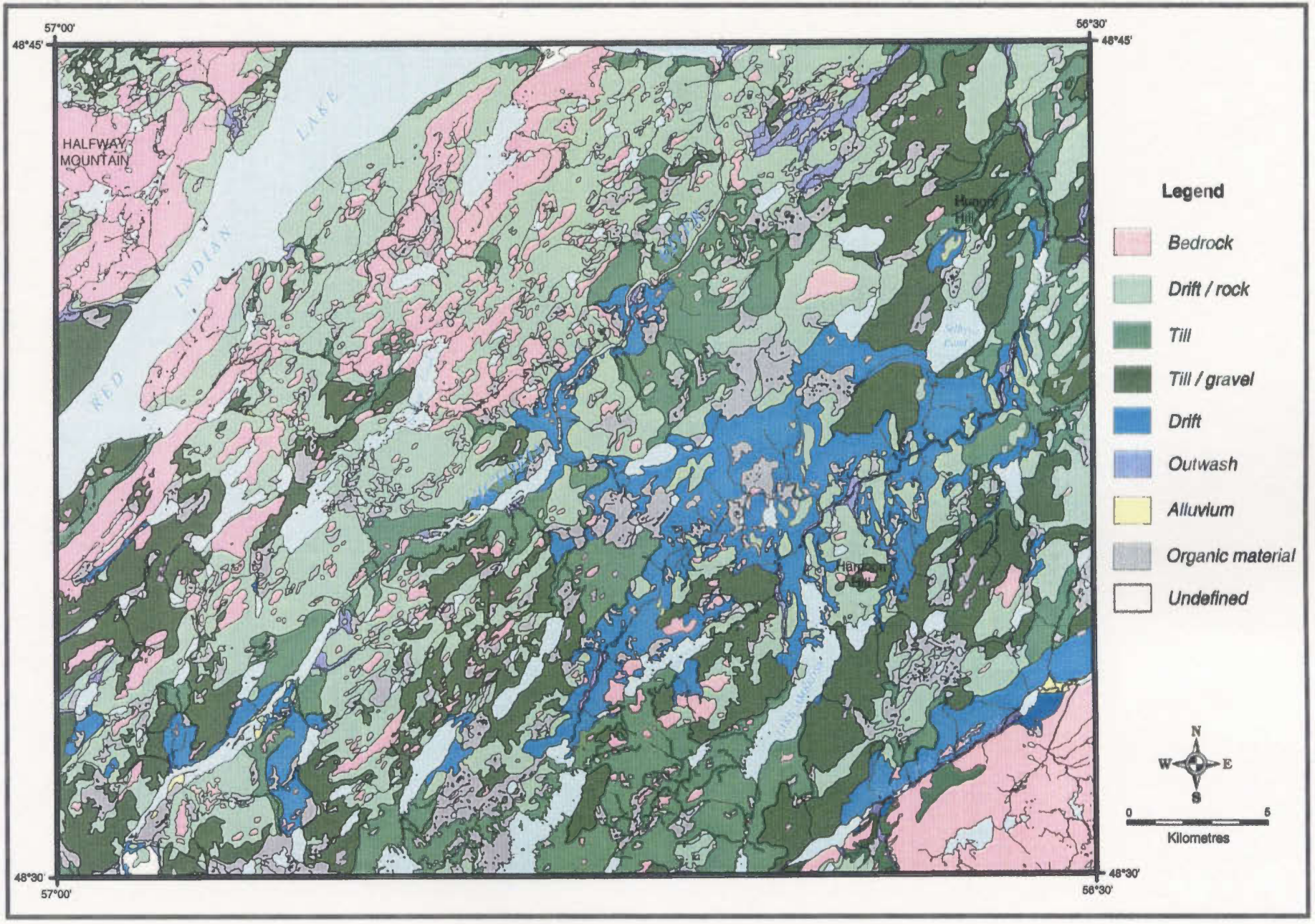

Figure 3.6 : Surficial geology of the Lake Ambrose map area (NTS 12A/10; after Klassen, unpublished data).

3.2 Data

Most of the data used in this study was available in digital format. Analog data were converted to digital format as described below. Issues related to data accuracy and the representation of data types for different types of analyses will be addressed in Chapter 4.

3.2.1 Topographic Data

Topographic data provides a general locational framework on which to base visual analysis (relative location of lakes, rivers, roads etc.). The topographic data also provides information about the drainage system (lakes, rivers, bogs and elevation data). This information may be useful in the analysis of the till and lake sediment geochemistry.

3.2.1.1 Topographic Base Map

The digital topographic data for the NTS 12A/10 map sheet (1:50,000 scale) was obtained from the Geological Survey (Newfoundland and Labrador Department of Mines and Energy) who had acquired it from the Surveys and Mapping Division (Newfoundland and Labrador Department of Government Services and Lands). The data were provided in CARIS® vector format based on the North American Datum (NAD) 83 and was converted to NAD27 because most of the other data and paper maps were in NAD27. The data is based on 1986 aerial photography, with an estimated planimetric accuracy of 10 m and an estimated altimetric accuracy of 5 m (from metadata file for NTS 12A/10 supplied by the Surveys and Mapping Division, Department of Government Services and

Lands, 1998). Vector data codes are based on National Topographic Database (NTDB) standards. Neatline (surrounding the map sheet), lakes, rivers, wetlands, contours, and roads were extracted as separate layers from the database and reformatted for use in other GIS/DTMS software (*i.e.* ArcView® and MapInfo™). Lakes, double-sided rivers and wetlands were polygonized and coded appropriately. Elevation values (in metres amsl) were included as attributes on contour lines and rivers and occur as spot heights on local topographic high points as well as the surface of large lakes. Benchmarks, which are useful for checking the accuracy of digital elevation models (DEM), are not present in the NTS 12A/10 area.

3.2.1.2 Catchment Basin Delineation

The delineation of catchment basins can be used to determine the areal extent that each lake sediment sample represents. There are two methods commonly used to delineate catchment basins; manual and automated techniques. The automated technique requires the development of a DEM.

The DEM was based on the elevation information coded on the contours, rivers and lakes included with the 1:50,000 digital topographic data. The NTS 12A/10 study area was covered with a systematic sample of 30,000 points which were assigned the elevation value from the closest contour, lake or river and the distance to this vector. Points with distances greater than 25 m (*i.e.* the resolution of the final raster DEM) were deleted from the database in order to remove unnecessary points in flat areas. The DEM was created from a TIN model in the program SURFER® (Keckler, 1995). The DEM was

checked against 48 stratified random sample spot heights measured off the paper topographic map. The RMS error calculated from these 48 spot heights was 2.3 m. The altimetric error cited for this map sheet is 5m. Combining the two vertical errors (*i.e.* $\sqrt{5^2 + 2.3^2}$) resulted in a cumulative error of 5.5 m. Therefore, the DEM did not add much error (*i.e.* only 0.5 m) to the inherent error in the map sheet.

The DEM was used to delineate catchment basins using the Spatial Analyst HYDROLOGIC functions in ArcView 3.1® (ESRI, 1996) and the WATERSHED module in IDRISIFW® 1.0 (Eastman, 1993). Both of these methods resulted in poor delineation of the basins when visually compared with a few catchment basins delineated manually. These methods may work better in areas with higher relief (more mature drainage) and a better resolution than the 25 m resolution of this DEM. Due to the poor results obtained from the ArcView® and IDRISI® modules, the catchment basins were manually delineated based on the elevation data on a paper copy of the 1:50,000 scale topographic map. These basins were digitized, polygonized and coded based on the lake sediment samples they contained (Figure 3.7). The data were imported into ArcView® and MapInfo™ for comparative analysis with other data layers.

3.2.2 Mineral Occurrences

The MODS information for the NTS 12A/10 area was extracted and converted to dBase and Lotus 123 formats for viewing and plotting (Figure 3.8; Table 3.2). Prior to

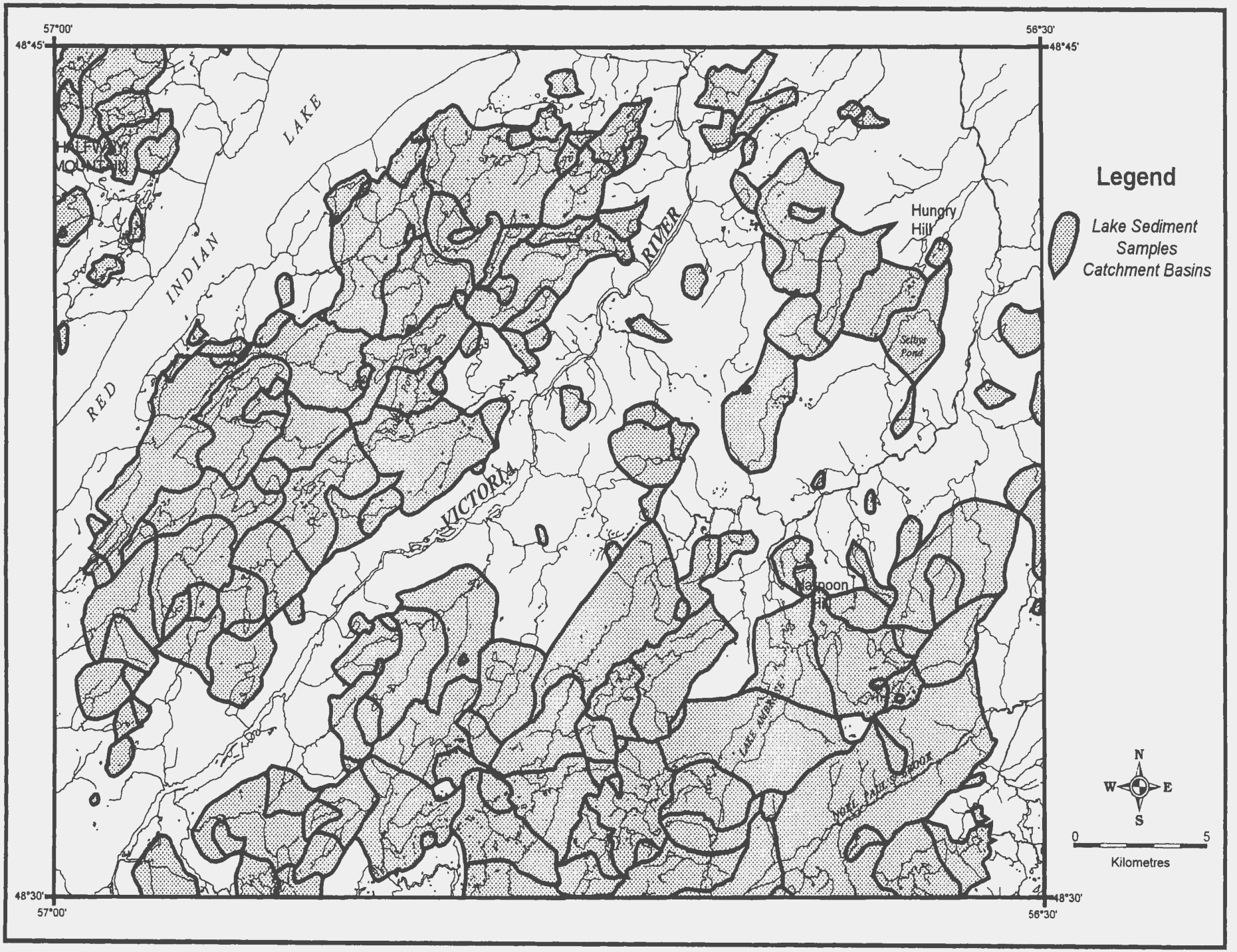

Figure 3.7 : Catchment basins corresponding to the lake sediment samples.

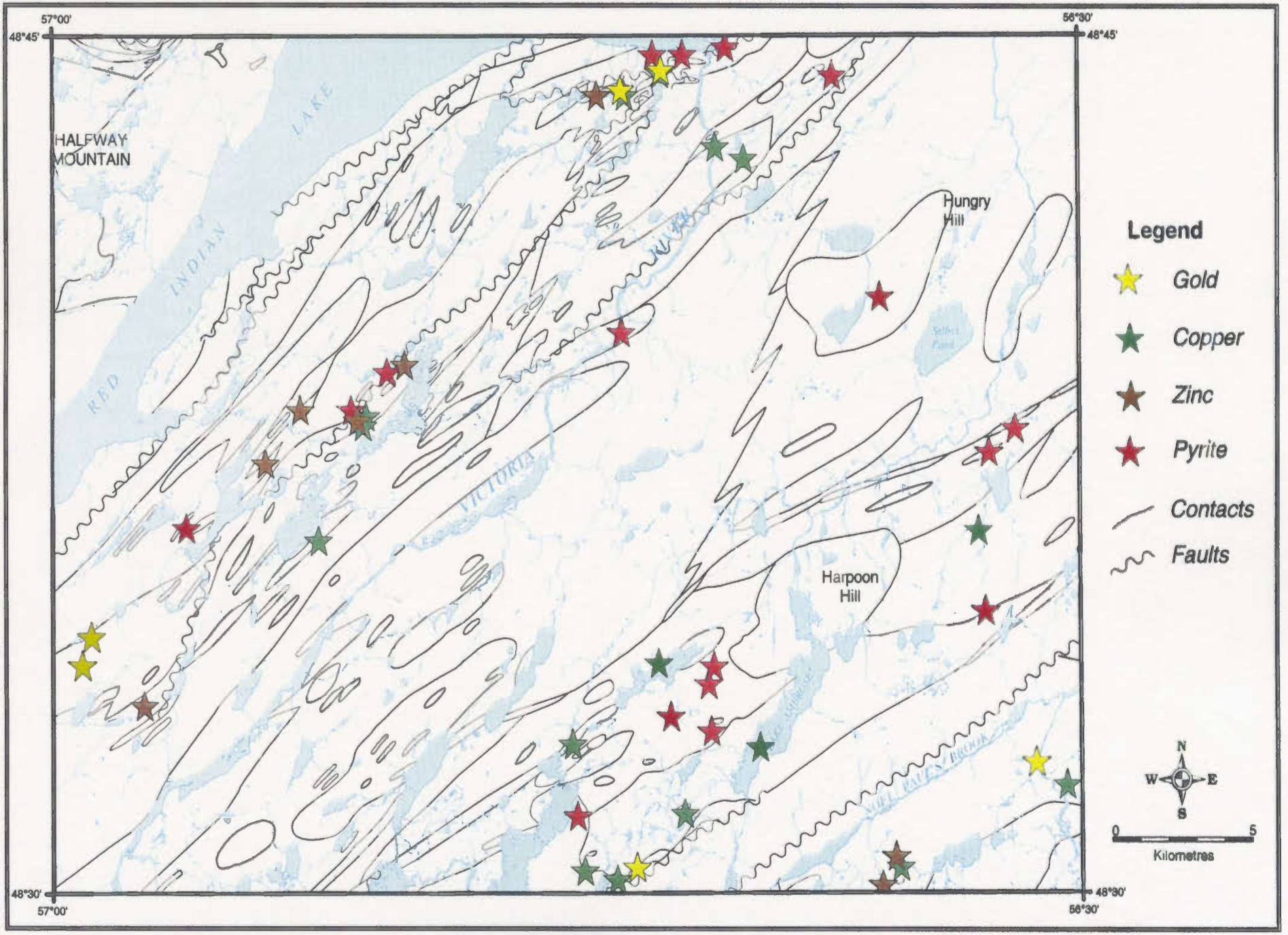

Figure 3.8 : Surface mineral occurrences colour coded by most significant metal or, in the absence of metals of economic significance, pyrite. Geological contacts and faults have been added for reference.

Table 3.2 : Partial attribute file for all the mineral occurrences in the study area.

Deposit Name	Min Occurrence Number	Major Commodity	Minor Commodity	Deposit Type	UTMEAST	UTMNORTH
Harpoon Brook	012A/10/asb001	asbestos		706	535125	5392750
Jacks Pond	012A/10/Cu 001	copper	Pb Zn pyr	131	501240	5379700
Bobbys Pond West Copper #1	012A/10/Cu 002	copper	pyr Au	130	511150	5387150
Victoria Mine	012A/10/Cu 003	copper	Pb Zn pyr	131	521710	5398300
Skidder Prospect	012A/10/Cu 004	copper	Zn pyr	121	504750	5396450
Rogerson Lake	012A/10/Cu 005	copper	pyr	130	518650	5376525
Loon Lake	012A/10/Cu 006	copper	po pyr	151	530350	5372600
Rogerson Lake East #1	012A/10/pyr001	pyrite		130	518725	5374210
Rogerson Lake East #2	012A/10/pyr002	pyrite	Cu	130	522590	5374300
Lake Ambrose	012A/10/pyr003	pyrite	Cu	130	525300	5376450
Beaver Lake	012A/10/pyr004	pyrite	Cu	130	521650	5379150
Beaver Lake South #1	012A/10/pyr005	pyrite		130	522090	5377425
Lake Ambrose West	012A/10/pyr006	pyrite		130	523550	5376950
Beaver Lake East #1	012A/10/pyr007	pyrite		130	523450	5378475
Beaver Lake East #2	012A/10/pyr008	pyrite		130	523650	5379050
Chickadee Lake West	012A/10/pyr009	pyrite		130	533400	5380890
Gill's Pond South Pyrite	012A/10/pyr010	pyrite	Cu	130	533150	5383500
Gill's Pond North Pyrite #1	012A/10/pyr011	pyrite		130	533525	5386050
Gill's Lake North Pyrite #2	012A/10/pyr012	pyrite		151	534450	5386825
Island Pond	012A/10/pyr013	pyrite		300	529800	5381100
Loon Lake Northeast	012A/10/pyr014	pyrite	po Cu	151	536310	5375275
Victoria River East #1	012A/10/pyr015	pyrite		130	527900	5396250
Red Indian Lake East	012A/10/pyr016	pyrite		130	524075	5399150
Red Indian Lake Southeast #1	012A/10/pyr017	pyrite		151	522525	5396800
Red Indian Lake Southeast #2	012A/10/pyr018	pyrite		130	521450	5396890
Victoria River #1	012A/10/pyr019	pyrite	Cu	130	523700	5395925
Victoria River East #2	012A/10/pyr020	pyrite	Cu	130	524725	5395550
Victoria River #2	012A/10/pyr021	pyrite		130	520325	5396900
Bobbys Pond West Pyrite	012A/10/pyr022	pyrite		130	510625	5387350
Beaton's Pond	012A/10/pyr023	pyrite	Cu	130	508450	5383100
Harbour Round Pond	012A/10/pyr024	pyrite		130	504700	5383500
Havens Steady	012A/10/Zn 011	zinc	Au Ag Cu Pb pyr	141	530180	5372940
Hoffs Pond	012A/10/Zn 001	zinc	Pb Cu	141	514630	5389860
Balloon Pond Northeast	012A/10/Au 001	gold	pyr	300	535230	5376000
Halfway Mountains	012A/10/Au 002	gold	Au Ag Pb pyr	140	503800	5392250
Spencers Pond	012A/10/Au 003	gold	pyr	140	520860	5372550
Jack's Pond West	012A/10/Au 004	gold	pyr Pb	310	500968	5379040
Bobbys Pond West Copper #2	012A/10/Cu 007	copper	Pb Zn	0	511040	5386830
Gills Pond South Copper #1	012A/10/Cu 008	copper	pyr Pb Zn	120	535050	5382810
Gills Pond South Copper #2	012A/10/Cu 009	copper	Pb Zn	110	534100	5382920
Victoria Mine West	012A/10/Cu 010	copper	pyr	140	520430	5397570
Spencers Pond East Copper	012A/10/Cu 011	copper	Pb Au	0	520150	5372150
Spencers Pond West Copper	012A/10/Cu 012	copper	Pb Au Ag pyr	0	518980	5372430
North Pond	012A/10/pyr025	pyrite		141	511900	5386580
Cathys Pond	012A/10/pyr026	pyrite	Au Cu	141	501280	5379960
Bobbys Pond West Sulfur	012A/10/S 001	sulphur		140	510520	5386900
Side of the Hill	012A/10/Zn 002	zinc	Pb Cu pyr	140	503190	5377730
Bobbys Pond East Zinc #1	012A/10/Zn 003	zinc	pyr Cu	140	514630	5388300
Bobbys Pond East Zinc #2	012A/10/Zn 004	zinc	Cu Au	140	514500	5388000
Bobbys Pond East Zinc #3	012A/10/Zn 005	zinc	pyr	140	514140	5387420
Bobbys Pond West Zinc #1	012A/10/Zn 006	zinc	Pb	140	510650	5387000
Bobbys Pond West Zinc #2	012A/10/Zn 007	zinc	Cu Pb	140	508790	5387340
Bobbys Pond West Zinc #3	012A/10/Zn 008	zinc	Pb Ag Cu	140	512530	5388820
Daniels Pond	012A/10/Zn 009	zinc	Ag Pb Cu Au pyr	141	507550	5385800
Sutherlands Pond East	012A/10/Zn 010	zinc	Pb Cu asp Fe pyr pph Au	130	519440	5397570
Havens Steady (Road Zone)	012A/10/pyr027	pyrite	Zn	141	529700	5372000
Victoria Mine Gold	012A/10/Au 005	gold		310	521750	5398400
Victoria Mine West Inco	012A/10/Au 006	gold		310	520300	5397750
Hoff's Pond Gold	012A/10/Au 007	gold		310	514630	5389860

1999, the most recent information for the NTS 12A/10 area was compiled in 1994, which excluded confidential data for the previous three years. Therefore, 1991 data were the most recent information available. A recent check on the database (July 2000) indicated the addition of only two more mineral occurrences, both of which occur in the vicinity of previously mapped mineral occurrences and are not surface occurrences. Therefore, they were not added to the database.

3.2.3 Bedrock Geology

The bedrock geology (Figure 3.3) was compiled at a 1:50,000 scale in digital form by S. Colman-Sadd of the Geological Survey Branch, Newfoundland Department of Mines and Energy (Ash and Colman-Sadd, 1997). The geological linework was digitized from the most accurate geology maps (*e.g.* Evans *et al.*, 1994) and composited together in CARIS . The final data were obtained from the Geological Survey in MapInfo format. Separate layers were available for geological polygons, geological contacts and faults. Attribute information was included with the geology polygons (Table 3.3; Colman-Sadd, 2000). Outcrop locations, for use in subsequent reliability analysis, were digitized from Map 94-223 (Evans *et al.*, 1994).

3.2.4 Surficial Geology

The glacial history, surficial geology and striation patterns of an area provide a framework on which to base the interpretation of the geochemistry of till samples and possibly lake sediment samples.

Table 3.3 : Partial attribute file for the geology map. Each row represents a separate polygon.

Rock Type	Group	Formation	Lower Age	Upper Age	Reference
siliciclastic, non-marine		(012A/10/0686 red sst; Botwood Gp?)	443	362	Evans et al., 1994a
plutonic, intermediate		(012A/10/0686 SD intrusions)	443	362	Evans et al., 1994a
siliciclastic, marine		Spruce Brook Formation	550	475	Colman-Sadd and Russell, 1988
volcanic, felsic marine	Lake Douglas terrane		545	458	Colman-Sadd, 1987
volcanic, marine	Buchans Group		476	467	Williams et al., 1985
volcanic, mafic marine	Buchans Group	Sandy Lake Formation	476	467	Thurlow and Swanson, 1981
volcanic, mafic marine	Victoria Lake Group	Tulks Belt	504	494	Kean, 1979a
siliciclastic, non-marine		(012A/10/0686 red sst; Botwood Gp?)	443	362	Evans et al., 1994a
volcanic, mafic marine	Victoria Lake Group	Tulks Belt	504	494	Kean, 1979a
volcanic, felsic marine	Victoria Lake Group	Tulks Belt	504	494	Evans et al., 1994a
siliciclastic, marine siltstone		Harbour Round Formation	504	458	Kean, 1979a
plutonic, felsic		Skidder basalt	545	471	Evans et al., 1994a
plutonic, felsic		Skidder basalt	545	471	Evans et al., 1994a
volcanic, mafic marine	Buchans Group	Sandy Lake Formation	476	467	Thurlow and Swanson, 1981
plutonic, felsic		(012A/10/0686 SD intrusions)	443	362	Evans et al., 1994a
siliciclastic, non-marine		Shanadithit Formation	362	345	Evans et al., 1994a
volcanic, mafic marine	Buchans Group	Sandy Lake Formation	476	467	Thurlow and Swanson, 1981
volcanic, felsic marine	Victoria Lake Group	Victoria Bridge sequence	467	458	Evans et al., 1994a
volcaniclastic	Victoria Lake Group	Victoria Bridge sequence	467	458	Evans et al., 1994a
volcanic, felsic marine	Victoria Lake Group	Tulks Belt	504	494	Evans et al., 1994a
plutonic, mafic		(012A/10/0686 SD intrusions)	443	362	Evans et al., 1994a
siliciclastic, marine	Victoria Lake Group		515	460	Evans et al., 1994a
volcanic, felsic marine	Victoria Lake Group	Tulks Belt	504	494	Evans et al., 1994a
volcanic, mafic marine	Victoria Lake Group	Tulks Belt	504	494	Kean, 1982
plutonic, intermediate		(012A/10/0686 SD intrusions)	443	362	Evans et al., 1994a
volcaniclastic	Victoria Lake Group	Victoria Bridge sequence	467	458	Evans et al., 1994a
plutonic, mafic		(012A/10/0686 SD intrusions)	443	362	Evans et al., 1994a
volcanic, mafic marine	Victoria Lake Group	Tulks Belt	504	494	Kean, 1979a
plutonic, mafic		(012A/10/0686 SD intrusions)	443	362	Evans et al., 1994a
volcanic, felsic marine	Victoria Lake Group	Tulks Belt	504	494	Evans et al., 1994a
siliciclastic, marine sandstone	Victoria Lake Group		515	460	Evans et al., 1994a
siliciclastic, black shale	Victoria Lake Group		515	460	Kean, 1982
siliciclastic, marine	Victoria Lake Group		515	460	Evans et al., 1994a
plutonic, intermediate		(012A/10/0686 SD intrusions)	443	362	Evans et al., 1994a
plutonic, mafic		(012A/10/0686 SD intrusions)	443	362	Evans et al., 1994a
volcanic, mafic marine	Victoria Lake Group	Tally Pond Belt	515	511	Evans et al., 1994a
volcanic, felsic marine	Victoria Lake Group	Tally Pond Belt	515	511	Evans et al., 1994a
siliciclastic, non-marine conglomerate		Rogerson Lake Conglomerate	443	418	Evans et al., 1994a
volcanic, felsic marine	Victoria Lake Group	Tally Pond Belt	515	511	Evans et al., 1994a
plutonic, intermediate		(012A/10/0686 SD intrusions)	443	362	Evans et al., 1994a
plutonic, mafic		(012A/10/0686 SD intrusions)	443	362	Evans et al., 1994a
siliciclastic, marine	Victoria Lake Group		515	460	Evans et al., 1994a
volcanic, mafic marine	Victoria Lake Group	Tulks Belt	504	494	Kean, 1982
volcanic, mafic marine	Victoria Lake Group	Tulks Belt	504	494	Kean, 1979a
siliciclastic, black shale	Victoria Lake Group		515	460	Kean, 1982

3.2.4.1 Ice Flow Data

A province-wide striation database is available in digital format through the Newfoundland and Labrador Department of Mines and Energy (Figure 3.9; Taylor *et al.*, 1993). This database of point locations of striations includes information on the relative ages of the striae and whether they have a directional component. Ice flow indicators were also recorded by Klassen (1994) at the time of regional till sampling and were incorporated in the ice flow history of the area (Figure 3.5). Other ice flow indicators (*i.e.* pebble lithologies) have been collected over the years but their use was beyond the scope of this study.

3.2.4.2 Surficial Geology Map

The surficial geology, at the 1:50,000 scale, was mapped by R.A. Klassen of the Geological Survey of Canada (GSC) in 1991 and 1992. Klassen updated information from previous work by Vanderveer and Sparkes (1982) and Sparkes (1985 and 1987). The preliminary GSC map (Figure 3.6) was digitized but has not been published (Klassen, 1997). There were eight units mapped in the area. The units consist of bedrock, drift/rock, till, till/gravel, drift, outwash, alluvium, and organics. Klassen (pers. comm., 2001) uses the term 'till' to refer to sediment deposited directly from glacial ice with no reworking. 'Drift' is a sediment very much like till but may have had minor reworking. 'Outwash' refers to poorly sorted sand, gravel and boulder gravel deposited by glacial meltwaters, whereas 'alluvium' is composed of silt, sand, gravel and boulder gravel

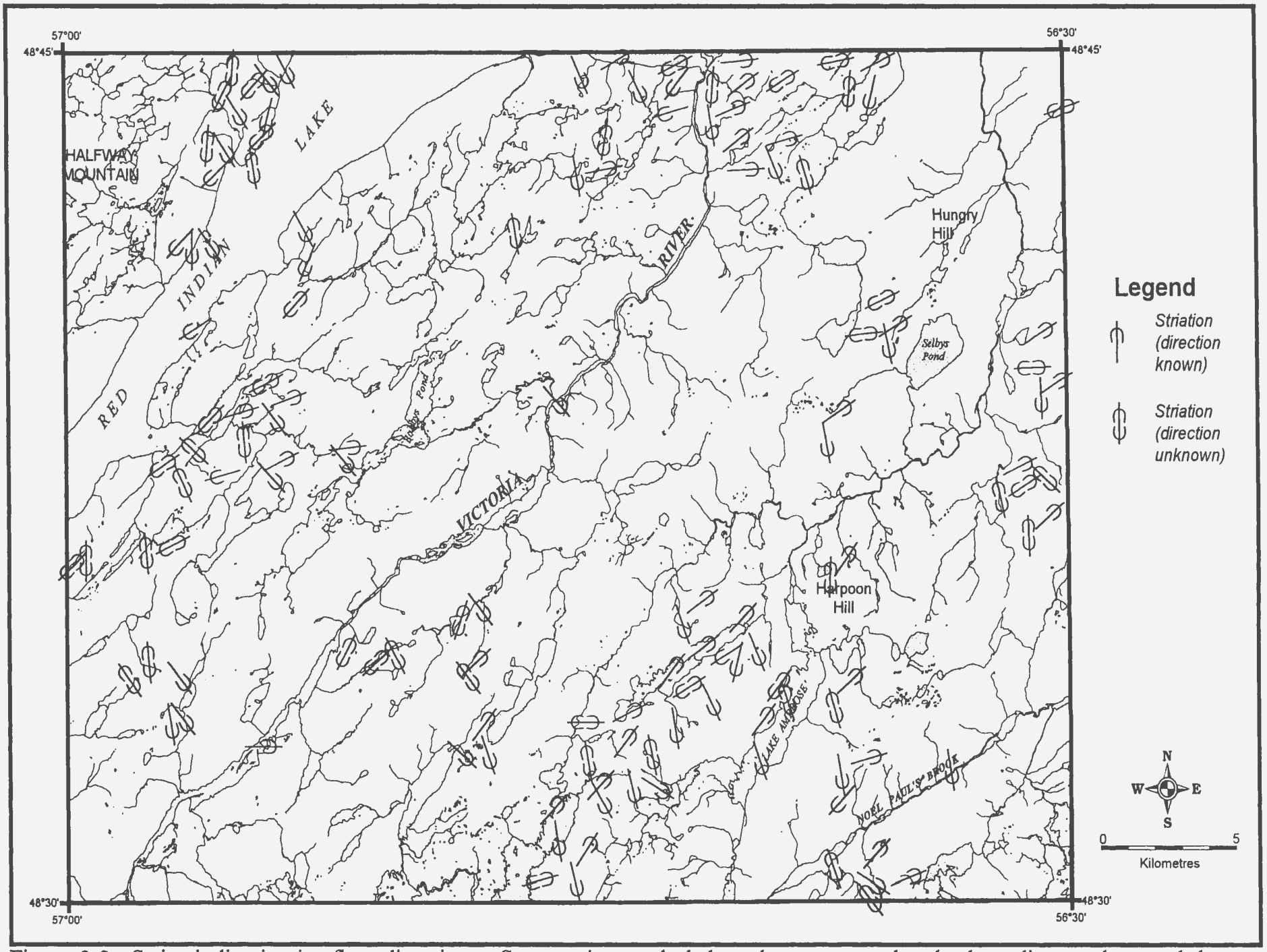

Figure 3.9 : Striae indicating ice flow directions. Some striae symbols have been removed and others dispersed around the sample site for clarity.

deposited by rivers and streams unrelated to glacial meltwaters (*i.e.* more modern in origin).

3.2.5 Till Sampling and Chemical Analysis

The most recent till sampling program to include the 12A/10 map sheet area was completed by Klassen in 1992 (Klassen, 1994) and subsequently included as part of the Buchans-Robert's Arm multidisciplinary project (Honarvar *et al.*, 1996). Approximately 250 till samples were collected in the Lake Ambrose map area, predominantly along woods roads and old railway lines (Figure 3.10). Unweathered upper C or lower B soil horizon (till) samples, representative of the original parent material, were the preferred sampling material (Klassen, 1994). In addition to sample numbers and UTM coordinates, field descriptions included sample colour, soil horizon and depth of sample below the surface.

Two geochemical datasets were prepared from the till field samples. A clay fraction (< 2 μm) and a "silt+clay" fraction (< 63 μm) were obtained from each field sample. Both sets of samples were analyzed (at Chemex Ltd., Vancouver) by inductively coupled plasma-atomic emission spectrometry (ICP-AES) following a hot acid digestion. Hot nitric acid-hydrochloric acid liberates loosely adsorbed metals and metals from most sulphides but does not decompose most silicates (Levinson, 1980) and so is considered a 'partial' analysis. The samples were analyzed by ICP (denoted by the suffix '_C') for the following elements: Ag, Al, As, Ba, Be, Bi, Ca, Cd, Co, Cr, Cu, Fe, Ga, Hg, K, La, Mg, Mn, Mo, Na, Ni, Pb, Sb, Sc, Sr, Ti, Tl, U, V, W, and Zn. The "silt+clay" fraction was

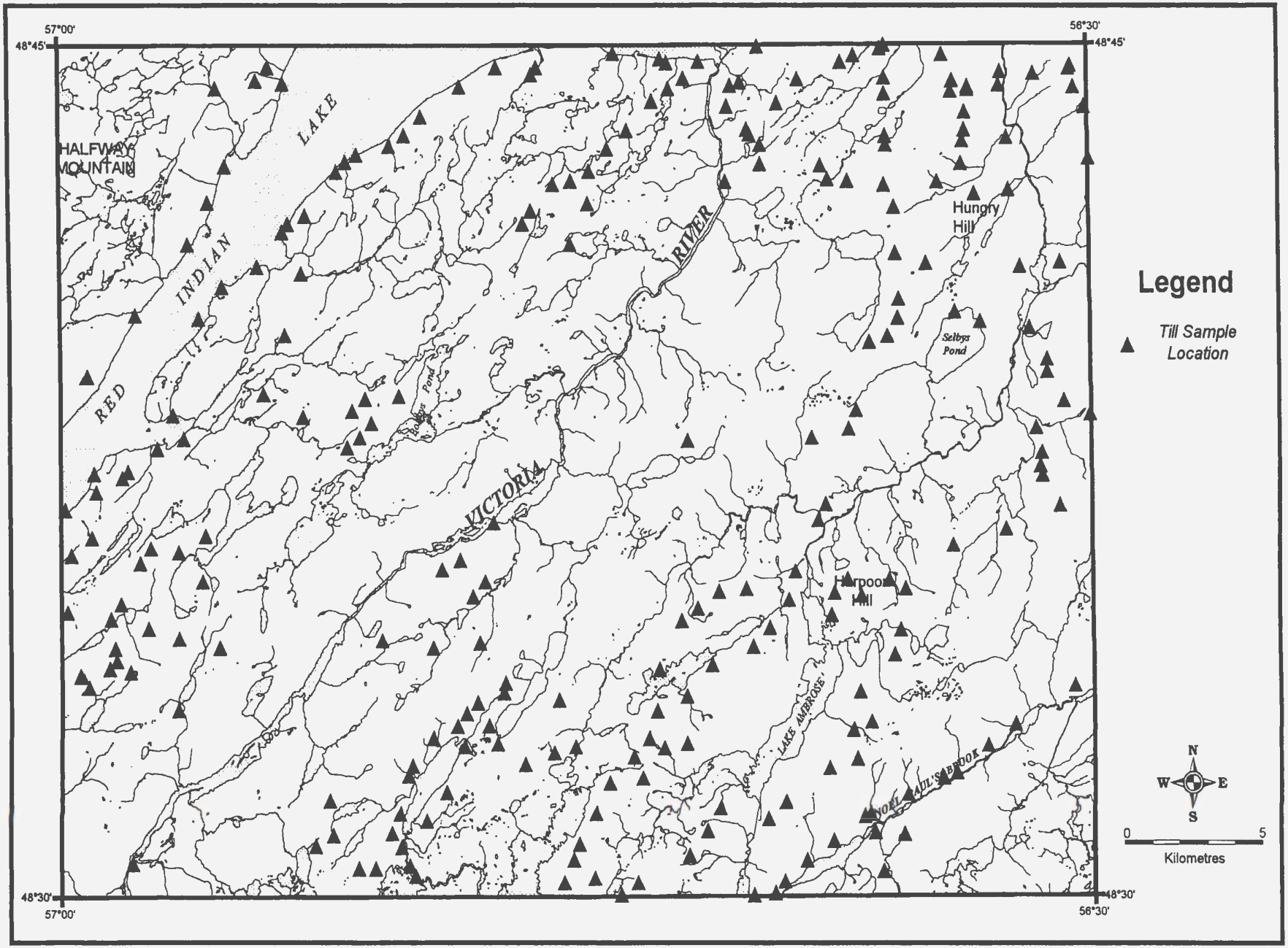

Figure 3.10 : Till sample locations.

also analyzed by instrumental neutron activation analyses (INAA, by Neutron Activation Laboratories) which provides a total analysis of the sample. The samples were analyzed by INAA (denoted by the suffix '_N') for the following 35 elements: Ag, As, Au, Ba, Br, Ca, Ce, Co, Cr, Cs, Eu, Fe, Hf, Hg, Ir, La, Lu, Mo, Na, Nd, Ni, Rb, Sb, Sc, Se, Sm, Sn, Sr, Ta, Tb, Th, U, W, Yb, and Zn. Analytical precision was monitored by the analysis of laboratory duplicates (Klassen, 1994).

A comparison of Klassen's data with till data collected and analyzed by the Newfoundland Geological Survey laboratory ('Liverman's dataset') for the Buchans-Robert's Arm study indicated slight differences due to different analytical techniques (Davenport *et al.*, 1996). To plot the two datasets on one map, the data were levelled (*i.e.* the differences were removed or reduced) using regression techniques as described by Davenport *et al.* (1996). The INAA and ICP methods used by Liverman were both 'total' analysis methods. Therefore, to obtain a final dataset of 'total' analyses for this study, those elements from Klassen's INAA (also a 'total' analysis) dataset which showed a well-defined linear relationship with Liverman's data were retained (see Table 15, Davenport *et al.*, 1996). In addition, those elements from Klassen's ICP method (a 'partial' method) which showed a well-defined linear relationship with Liverman's data were obtained from the combined (levelled) dataset (essentially a 'total' analysis). The final dataset reflects 'total' element concentrations of the till samples and contains 250 samples. Three samples were duplicates, resulting in 247 samples for statistical analysis.

Other regional samples of till have been taken in the NTS 12A/10 area and may be useful in filling in some of the sparse areas in Klassen's sampling. These consist of 226

regional samples taken by Vanderveer and Sparkes (1979) and 158 detailed samples taken by Mihychuck (1985) on two 200 m interval sampling grids around the Victoria Mine and the Tally Pond showing (northeast of Tally Pond). These samples were analyzed by AAS at the Geochemical Laboratory of the Department of Mines and Energy for the following elements: Cu, Pb, Zn, Co, Ni, Ag, Mn, Fe, Mo, and U.

The Klassen (1994) and Vanderveer-Sparkes (1979) datasets were compared to determine if they could be combined to provide a more complete till coverage of the 12A/10 area. Using ANOVA to compare the datasets, it was determined that more than half the elements had unequal means, therefore, the datasets could not be easily combined. Moreover, the Vanderveer-Sparkes data has not yet been released to the public. Therefore, only Klassen's data were used in this study.

3.2.6 Lake Sediment Sampling and Chemical Analysis

Regional lake sediment sampling in the NTS 12A/10 area was carried out during 1977. Almost 200 samples were collected (Figure 3.11), using a helicopter, at an approximate sampling density of one site per 6 km^2 (Davenport *et al.*, 1990a). The samples were dried and sieved through a 180 micron stainless-steel sieve. Analyses on the fine fraction were carried out using instrumental neutron activation analysis (INAA – 'total' analysis, at Becquerel Laboratories Inc., Mississauga, Ontario) for 27 elements (Au, As, Ba, Br, Ce, Co, Cr, Cs, Eu, Fe, Hf, La, Mo, Na, Ni, Rb, Sb, Sc, Se, Sm, Ta, Tb, Th, U, W, Yb and Zn). Uranium was also analyzed by delayed neutron counting analysis (DNA). Following a partial hot acid digestion using 4M HNO_3 - 1M HCl (Davenport *et*

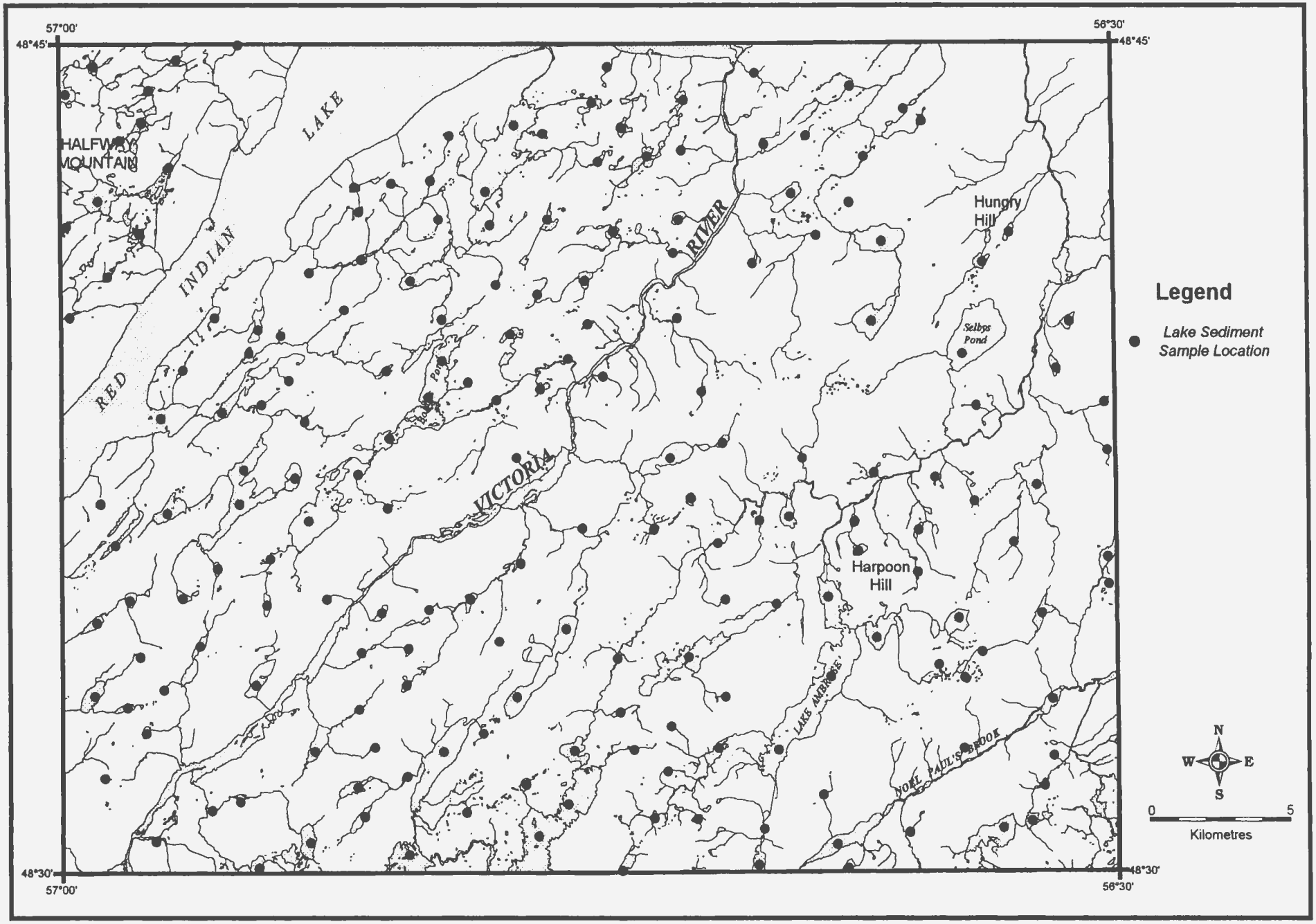

Figure 3.11 : Lake sediment sample locations.

al., 1990a), 7 elements (Cu, Pb, Zn, Co, Ni, Mn, and Fe) were analyzed by atomic-absorption spectrophotometry (AAS) at the Geochemical Laboratory of the Department of Mines and Energy. Molybdenum was also analyzed by AAS, but the digestion was in concentrated HNO_3. To indicate the different analyses a numbered suffix is added to each element symbol; '1' indicates INAA, '3' indicates AAS, '5' indicates AAS for the molybdenum digestion, and '8' indicates DNA for uranium. Loss-on-ignition (LOI), a measure of organic-carbon content, was determined by weighing before and after a three hour ashing at 500°C. Duplicate samples were taken from approximately 5% of the lakes to provide a measure of field precision (noise). Five percent of the samples were split in the lab to provide a measure of laboratory precision (added noise). Accuracy was measured by analyzing one standard sample for every 20 field samples. Information on the duplicates and standards is discussed by Davenport (1990a).

CHAPTER 4

Methodology

A quantitative model to reliably predict mineral occurrences must contain a set of appropriate predictor variables. This chapter outlines the methods used to prepare the response and predictor variables as well as describing the development of the quantitative models using decision tree analysis (DTA) and logistic regression analysis (LRA). Methods used to compare and estimate the reliability of the two models are also discussed.

A series of programs were used to explore and analyze the data. Confirmatory statistics, exploratory data analysis (EDA), principal components analysis (PCA), and LRA were processed through SPSS™. Cumulative frequency plots were output through the program UNISTAT (Nolan, 1990). Spatial statistics (variogram analysis and kriging) were analyzed and output using GS^{+}™ (Robertson, 1998). DTA was analyzed through KnowledgeSeeker® (Angoss, 1993). Various GIS and desktop mapping systems, including CARIS®, Arc/Info® , ArcView® , MapInfo Professional™ , and IDRISI®, provided spatial analysis (*e.g.* proximity analysis) as well as visual display and output of the results. Microsoft Excel® was used for spreadsheet calculations in the Weights of Evidence modelling and the Moran's I calculations for point data. In addition, a few programs (see Appendix A) were written in Microsoft QuickBASIC® to reformat or analyze simple sets of data (*e.g.* determination of spatial weighting function based on point distance table output from Arc/Info®).

4.1 Data Preparation

4.1.1 Response (Dependent) Variable

To arrive at an effective response variable, the mineral occurrences in the study area were screened and only surface occurrences and specific deposit types were chosen. The surface mineral occurrences were more likely to be reflected in the till and lake sediment samples and can be related to the surface geology whereas mineral occurrences noted at depth in drill core can not easily be related to surface phenomena. The prevalent deposit type in the study area is volcanogenic massive sulphide (VMS) with economic and pathfinder minerals consisting of pyrite (an iron sulphide), chalcopyrite (a copper/iron sulphide), galena (a lead sulphide), sphalerite (a zinc sulphide), and gold. Of the 59 mineral occurrences in the study area, 39 are surface occurrences of the VMS deposit type (deposit type codes 130 to 151; see Table 3.1). To increase the number of occurrences in the database, 5 occurrences of the hydrothermal/structurally-controlled deposit types were included (coded 300-310; see Table 3.1). The mineralization in this deposit type consists of the same economic minerals of interest as the VMS deposit type. There are also 3 surface mineral occurrences that did not contain enough information to classify (deposit type code 0; Table 3.1) but they contain the same economic minerals of interest as the above and so were included in the database. This resulted in a total of 47 surface mineral occurrences. As a check on the validity of this pooling, an additional attribute (VMS) was added to the database and coded '1' to represent the VMS deposit type plus all the nonmineral occurrences and '2' to represent the gold and unknown deposit types. The

DTA and LRA were repeated on the VMS group (*i.e.* VMS= 1) to determine if the results were significantly different from the grouped (*i.e.* VMS= 1 and 2) results.

The DTA and LRA were based on an attribute file consisting of mineral and nonmineral occurrences. The 47 mineral occurrences were coded '1' in the attribute file (variable MINOCC) indicating the presence of mineralization. A similar number of sites were chosen to represent the 'absence' of mineralization. These sites were chosen using the stratified random sampling method of IDRISI's® SAMPLE module (Eastman, 1997). The stratified random sampling method reduces bias and provides a good coverage of the study area. Of 50 points located by IDRISI®, 49 fell within the study area and did not coincide with the mineral occurrence point locations. The locations (UTM easting and northing) of these 49 points were added to the attribute file and coded '0' (variable MINOCC) indicating the absence of mineralization. Therefore, the attribute file contained a total of 96 data points using a binary coding scheme to represent the dependent variable MINOCC. The locations of these 96 data points were used to extract data for each independent predictor variable (*e.g.* geological rock type).

Classical statistical tests require that data be independent and randomly distributed. To test that the 96 data points were randomly distributed the nearest neighbour index (Hammond and McCullagh, 1978) was calculated.

The nearest neighbour index, R, is a ratio of the observed mean-distance (D_{obs}) between the nearest neighbour sites to the expected mean-distance (D_{exp}) of the same number of sites randomly distributed.

$$R = \frac{D_{obs}}{D_{exp}}$$

where $D_{obs} = \Sigma (d_{nn}) / N$, $D_{exp} = 1/2\sqrt{N / area}$, d_{nn} is the distance between nearest neighbours and N is the number of points. A z-score can be calculated to determine the significance of the result R (Hammond and McCullagh, 1978).

$$z = (D_{obs} - D_{exp},) / \sigma_{Dexp}$$

where $\sigma_{Dexp,} = 0.26136 / \sqrt{n^2 / area}$. In the test of significance the null hypothesis tests for a random point distribution. Z-scores less than 2.58 ($\alpha = 0.01$) indicate the null hypothesis is acceptable and the points are randomly distributed.

4.1.2 Predictor (Independent) Variables

An effective mineral potential model will contain the fewest and most reliable predictors that best explain the presence/absence of mineralization. The predictors chosen for this study were based on information from previous studies, ore deposit models for the study area and factors that may influence other predictors (*e.g.* the effects of wetlands on till geochemistry).

As a general exploration tool, till and lake sediment geochemistry have been shown to indicate regionally favourable areas for exploration. Ore deposit models associated with the types of mineral occurrences in this study area (*e.g.* VMS) can help in defining useful predictors. A review of the VMS deposit models indicated the presence of felsic volcanic rocks, mafic volcanic rocks, various alteration types and proximity to faults

(Franklin, 1993) were positive factors to include in a study. Rock alteration was not systematically mapped or compiled in the study area and therefore not included. Other predictors that may have an effect on the geochemistry of till and lake sediments are surficial geology and wetlands. And a measure of proximity or neighbourhood effects of the mineral occurrences should be included. Overall, about 20 predictors were included in the analysis (*i.e.* 3 geology predictors, 3 fault proximity predictors, 5 till predictors, 6 lake sediment predictors, surficial geology, wetlands and a spatial weighting function) and will be discussed in more detail below.

4.1.2.1 Till and Lake Sediment Geochemistry

The review and analysis of both the till and lake sediment geochemistry databases, follow a similar procedure with some minor individual analyses unique to each database. The general procedure is outlined in Table 4.1.

A review of the database using EDA techniques consisting of box-and-whisker plots, histograms and cumulative frequency plots provided an overview of the variation in the geochemistry values as well as indicating elements that may be poorly distributed or truncated (usually caused by instrumental detection limits at the low end of the scale). Poorly distributed variables may cause problems in statistical analyses due to deflated correlations (Tabachnick and Fidell, 1996). These variables were removed from the geochemistry databases. Univariate statistics, consisting of the mean, standard deviation, median, geometric mean, log standard deviation and range, were tabulated as a summary

Table 4.1 : Procedure to review and analyze the till and lake sediment databases.

Database Review:
a) EDA
b) univariate Statistics
c) visual Spatial Distribution
Linear Regression Analysis:
-lake sediments only – used to remove lake effects
Principal Components Analysis:
a) review factorability
b) review outliers
c) estimate number of components
d) judge results
Produce Surface:
Tills–
Spatial Autocorrelation –
a) variogram and correlogram analysis
b) kriging
c) CV reliability surface
Lake sediments –
Catchment basin assignment
Extract Values from surface and input into attribute table

overview of the data. To assess normality of the distributions, the shape of the distributions were evaluated using box-and-whisker plots and histograms. Skewness and kurtosis were not used because the number of samples was too large (*e.g.* greater than 150 to 200) for a reliable test (Tabachnick and Fidell, 1996). All values were logged to bring all element values into the same scale range (Brower and Merriam, 1990). To maintain consistency of scale and for interpretation purposes all the units were changed to

ppm values (*i.e.* Au ppb was changed to ppm; Fe, Na, Mg, and Ca percent was changed to ppm).

A visual assessment of the spatial distribution (as graduated dot plots) of element concentrations in till and lake sediment samples can indicate preliminary associations between the elements and other spatial factors (*e.g.* bedrock geology, faults).

To reduce false anomalies caused by lake effects (change in element concentrations due to the lake environment, such as adsorption on hydroxides and organics, as well as lake area and sample depth) all elements were regressed against logFe, logMn, LOI_pct, loglarea and logsdpth (Davenport *et al.*, 1974). The data were checked for univariate and multivariate outliers prior to running the linear regression. Extreme outliers were removed from the database so their values would not adversely affect the linear regression results.

Linear regression was used to remove lake effects on those elements which correlated (using $r>0.5$; Davenport, 1974) with at least one independent variable (IV; *i.e.* logFe, logMn, LOI_pct, loglarea and logsdpth). Therefore a preliminary stepwise linear regression analysis was run with each element as the dependent variable against all 5 IVs. The results of this regression run were checked and the change in R^2 values was noted. Only those IVs that enter the equation causing a greater than 2% change in R^2 were retained. The next stepwise linear regression was run using only the retained IVs and the computed residuals were saved. These residuals were reviewed and cases which were outliers (*i.e.* cases where the standardized residual value was greater than ±3.29) were removed from the final linear regression run. These outliers were cases that could not be well predicted by the solution. To include the 'outlier' cases in the database (as they may

be anomalous and of significance for mineral exploration), their residuals were manually calculated using the final regression equations and input back in the database (P. Davenport, pers. comm., 2001).

It is not practical to review the statistical and spatial distribution of every element in a large database. PCA is an excellent method of reducing the database to a small set of orthogonal components that will provide insight into the underlying nature of the data. Outliers and poorly distributed variables can greatly affect the results of PCA because it is based on correlations. Variables that were poorly distributed were already removed from the database at the initial phase of data screening. To check the effect of outliers in the database, PCA was run before and after removal of univariate and multivariate outliers and the results compared. If the results were essentially the same then the outliers would be left in the database. If the analysis was not robust to outliers then the outliers would be removed and the data reanalyzed. As mentioned above, outliers may be anomalous and of significance for mineral exploration. Therefore, if outliers were removed their scores were manually computed using the results of the PCA analysis on the nonoutlier cases (P. Davenport, pers. comm., 2001).

The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was calculated to determine if the data was suitable for PCA. A KMO value greater than 0.6 indicates that the dataset can be factored (Tabachnick and Fidell, 1996). The next step was to determine the number of components to extract by reviewing the eigenvalues and the scree plot. The number of components with eigenvalues greater than 1.0 and the component number at the change in slope of the scree plot provide an indication of the approximate number of

components to extract. Other factors to take into account when assessing the number of components are that the components should make sense with respect to their geochemical consistency, they should not be composed of just one or two elements and the maximum loadings in each component should be fairly high (*e.g.* > 0.7; Tabachnick and Fidell, 1996). Examination of communalities and the residuals matrix also indicate the effectiveness of the number of components chosen (Tabachnick and Fidell, 1996). Removal of elements with low communalities may change the loadings of the elements in different components. This is an iterative technique.

To determine if a 'simple' method of PCA analysis was robust, Spearman's rank correlation matrix (a nonparametric technique), calculated on the raw, unlogged data was used as input in the PCA instead of data that had been checked for normality, linearity, outliers etc. If this nonparametric method of PCA analysis provides results similar to the parametric method then PCA can be easily calculated and incorporated into more exploration analyses.

To determine if till geochemistry is important as a predictor of mineral occurrences, it is necessary to add the till geochemistry to the mineral occurrence attribute database by extracting the geochemical values at each mineral/nonmineral occurrence site. This requires that the till point data be represented as a surface. Only the ore-forming elements (*i.e.* Au, Cu, Pb, and Zn, as logged values) and the component that most represented the ore-forming minerals were considered, to determine if the individual elements or the component was a better predictor of mineral occurrences. Variograms

and correlograms were analyzed to determine if the variables showed positive spatial autocorrelation and could be interpolated as a surface.

The variograms were reviewed with and without outliers to see if the presence of outliers caused significant changes in the variogram model parameters. If outliers did cause significant changes then they were removed. The best variogram model was chosen based on the minimum value of the reduced sum of squares (RSS) which indicates how well the model fits the data using the lag parameters specified (Robertson, 1998). The isotropic and anisotropic models were compared to determine which best fit the data.

Correlograms were also reviewed to determine the distance over which positive autocorrelation exists. Based on Moran's I calculations, the correlogram graphically shows the distance over which positive spatial autocorrelation occurs.

If the till data was determined to be spatially autocorrelated then the point data can be interpolated using block kriging methods. Input parameters in the block kriging procedure consist of the number of nearest neighbours, the local grid size, the search radius, and the best variogram model (which is automatically input based on decisions made in the variogram analysis section). The number of nearest neighbours to include in the kriging and the local grid size were iteratively tested by comparing the minimum and maximum of the output data with the input minimum and maximum. The cross-validation plot of observed versus predicted values provides an indication of whether the data contains outliers that are causing a bias in the variogram. These checks ensured an interpolated surface which best represented the measured points.

Once the best kriged surface was calculated, an error surface (in standard deviations) was also calculated to provide an indication of the reliability of the interpolated surface based on the distribution and variation of the point data (Robertson, 1998). A coefficient of variation (CV) surface was calculated by dividing the standard deviation surface by the mean surface. The mean surface was represented by the kriged surface (for which each point represents the mean in that immediate area). The CV surface was multiplied by 100 to provide an indication of the percent variability around each point (Berry, 1993) An increase in variability indicates an increase in uncertainty (Berry, 1995).

When the kriged and CV surfaces had been prepared for each till element and PCA component of interest, their values at each of the 96 mineral/nonmineral occurrence site were extracted and added to the attribute database.

The lake sediments were assumed to be representative of the catchment basin from which they were collected (see Chapter 2.1.2). Therefore, rather than kriging the lake sediment data, the catchment basin polygons were assigned the value of each ore-forming element (*i.e.* Au, Cu, Pb and Zn as logged or residual values) or component. These values were extracted at the mineral/nonmineral occurrence site and added to the final attribute database.

4.1.2.2 Geology

There are 12 rock types in the study area (Table 3.3). The alphanumeric rock types were extracted from the digital geology map at the 96 mineral/nonmineral occurrence sites and added to the attribute database (variable GEOLOGY). Numeric codes are often easier to work with and are necessary for some analyses (*e.g.* SPSS™ boxplot factor list requires numeric codes). Therefore, the alphanumeric GEOLOGY descriptors were recoded to numeric values 1 to 12 (variable GEOLCODE). Deposit models for VMS deposits indicate some deposit types favour felsic volcanic rocks while other deposit types favour mafic volcanic rocks (Franklin, 1993). Therefore, the 12 rock types were recoded to binary values where '1' represents felsic volcanics and '0' represents non-felsic volcanics in the variable VOLCFELS and '1' represents mafic volcanics and '0' represents non-mafic volcanics in the variable VOLCMAF.

To provide an indication of agreement between the regional 1:50,000 scale geology (Colman-Sadd, 2000) and the detailed geology from the MODS database, the rock types (variable HOST_ROC) were extracted from MODS for the 47 surface mineral occurrence sites and added to the attribute database. The rock types in HOST_ROC and GEOLOGY were compared to indicate if the scale of geological mapping produces significantly different results overall.

4.1.2.3 Fault Proximity and Weights of Evidence Modelling

VMS and gold deposit models indicate fault proximity is important as an aid to mineral exploration (Franklin, 1993). Therefore the digital fault vectors were rasterized

using a 200 m cell resolution, based on the ±100m uncertainty of many of the mineral occurrence locations (Stapleton and Smith, 1999).

Three faults occur adjacent to the NTS 12A/10 map sheet but do not cross onto the map sheet in the southeast corner. These faults may extend into the map sheet area but were not mapped by the geologists due to a lack of evidence in this area (*e.g.* due to glacial debris cover). Therefore, to account for the possibility that these faults may exist on the NTS 12A/10 map sheet, they were extended slightly into the map sheet so buffering will indicate their presence. The proximity to the faults was determined using IDRISI's® DISTANCE module where 0 metres was recorded on a fault and distance (in metres) increased away from the faults. The continuous fault proximity surface was sampled (variable FLTDST) at the 96 mineral/nonmineral occurrence sites and added to the attribute database.

In LRA, variable coding affects the direction of the odds ratio and the sign of the B coefficient. If predictors are coded such that the higher values are most likely positively associated with the response variable then the resulting logistic regression parameter estimates will be positive and interpretation will be simplified (Tabachnick and Fidell, 1996). The values in FLTDST range from 0 m, on the fault, to over 8000 m furthest from the faults. The VMS and gold deposit models indicate that close proximity to faults is a positive factor in the exploration for mineral occurrences. Therefore, FLTDST was recoded into the variable NFLTDST by normalizing the distances:

$$\text{NFLTDST} = \frac{\max(\text{FLTDST}) - \text{FLTDST}}{\max(\text{FLTDST})}$$

This results in a continuous variable (NFLTDST) ranging from 1 (representing a point on a fault) to 0 (representing a point far removed from a fault). The distribution of NFLTDST is negatively skewed. The normalized distances were arcsine transformed to produce a continuous variable (ARCSFLT) which has a more normal distribution but a comparable range to NFLTDST.

A binary measure of fault proximity simplifies the proximity measurement to those areas close to faults (coded '1') and those areas far from faults (coded '0'). Weights of evidence modelling (see below) was used to determine the optimal threshold distance which represents areas close to faults and favourable for mineral occurrences.

Weights of evidence (WOE) modelling is a method of combining binary maps to produce an output binary map with an optimal measure of spatial association between two input maps (Bonham-Carter, 1994). One map is considered the response map (*e.g.* mineral/nonmineral occurrences) and the other is the predictor map (*e.g.* fault proximity). By changing the input predictor map (*e.g.* increasing buffer distances), the map with the best spatial association to the input response map can be determined. In this study, the spatial association was based on a density measurement of the number of mineral occurrences in close proximity to faults versus the number of nonmineral occurrences far from faults. The following procedure is based on Bonham-Carter (1994).

The study area was converted into a raster grid with a resolution of 200 m, based on the uncertainty of the location of many of the mineral occurrences. The subsequent calculations were done on counts of the unit cells. Each unit cell intersecting a mineral

occurrence was coded '1', otherwise it was coded '0'. N(T) represents the total number of unit cells in the study area. N(MO) represents the number of mineral occurrence cells. Therefore, N(MO)/N(T) represents the average density or prior probability, P(MO), of occurrences in the study area given no other information. To determine how the probability of mineral occurrences changed with the addition of new information (*i.e.* a fault buffer), the conditional probability was calculated as:

$$P(MO|FB) = P(MO \cap FB) / P(FB)$$

where P(MO|FB) represents the conditional probability of a mineral occurrence (MO) given the presence of a fault buffer (FB), $P(MO \cap FB)$ represents the probability of mineral occurrences and a fault buffer occurring simultaneously and P(FB) represents the probability of a fault buffer (after Bonham-Carter, 1994). Stating this in terms of counts of units cells results in the following:

$$P(MO|FB) = N(MO \cap FB) / N(FB)$$

where the conditional probability was calculated as the number of unit cells where both mineral occurrences and a fault buffer occurs divided by the total number of unit cells where a fault buffer occurs (after Bonham-Carter, 1994).

The WOE calculations are based on the conditional probabilities, given the presence and absence of the mineral occurrences and fault buffers. The weights were calculated for both the presence (positive weight W^+) and absence (negative weight W^-) of the new information (*i.e.* fault buffers) given the presence and absence of mineral occurrences. The positive weights of evidence is:

$$W^+ = \ln [P(FB|MO) / P(FB|\overline{MO})]$$

where $P(FB|MO)$ and $P(FB|\overline{MO})$ represent the conditional probability of the fault buffer given the presence and absence of mineral occurrences, respectively (after Bonham-Carter, 1994). The negative weights of evidence is:

$$W^- = \ln [P(\overline{FB}|MO) / P(\overline{FB}|\overline{MO})]$$

where $P(\overline{FB}|MO)$ and $P(\overline{FB}|\overline{MO})$ represent the conditional probability of the absence of the fault buffer given the presence and absence of mineral occurrences, respectively (after Bonham-Carter, 1994). Note that positive values for W^+ and negative values for W^- indicate that the new information (*i.e.* fault buffer) adds to the prediction of mineral occurrences. If $W^+ = W^- = 0$ then no correlation is indicated and the fault buffers add no additional information in predicting mineral occurrences. The four conditional probabilities for the above equations were calculated as follows:

$$P(FB|MO) = P(FB \cap MO) / P(MO) = N(FB \cap MO) / N(MO)$$

$P(FB|\overline{MO}) = P(FB \cap \overline{MO}) / P(\overline{MO}) = N(FB \cap \overline{MO}) / N(\overline{MO})$

$P(\overline{FB}|MO) = P(\overline{FB} \cap MO) / P(MO) = N(\overline{FB} \cap MO) / N(MO)$

$P(\overline{FB}|\overline{MO}) = P(\overline{FB} \cap \overline{MO}) / P(\overline{MO}) = N(\overline{FB} \cap \overline{MO}) / N(\overline{MO})$

(after Bonham-Carter, 1994).

The optimal buffering distance (*i.e.* fault proximity value) for converting continuous proximity values to binary values was determined by calculating contrast values. A contrast value was calculated for each fault proximity buffer distance (in increments of 200m from 0 to a maximum of 2400m):

$$C_w = W^+ - W^-$$

The contrast is a measure of correlation between the mineral occurrence map and the data in each of the fault proximity maps (Bonham-Carter, 1994). Contrast values usually range from 0 (indicating spatial independence between the two maps) to 2 (indicating a positive spatial association). A plot of contrast versus distance (where distance is the cutoff for the buffer area around the fault) indicates the best threshold distance to use to recode the continuous variable FLTDST into a binary variable. FLTDST values from 0 m to the threshold value were coded as '1' (representing distances from faults favourable for mineral occurrences) and values greater than the threshold value were coded as '0' (representing distances from faults unfavourable for mineral occurrences).

The variance of the weights and contrasts is calculated as a check on their uncertainty. The variance of the weights and contrasts can be approximated by the following equations, assuming a large number of occurrences (Bishop *et al.*, 1975):

$$S^2(W^+) = [1 \setminus N(FB \cap MO)] + [1 \setminus N(FB \cap \overline{MO})]$$

$$S^2(W^-) = [1 \setminus N(\overline{FB} \cap MO)] + [1 \setminus N(\overline{FB} \cap \overline{MO})]$$

$$S^2(C_W) = S^2(W^+) + S^2(W^-) \text{ and } s(C_W) = \sqrt{S^2(C_W)}$$

Note that the variances are unit dependent and, therefore, only significant in a relative sense (Bonham-Carter, 1994). The standard deviation of the contrast is used to calculate the Studentized value of the contrast (SC), an approximate test of the null hypothesis C=0 (*i.e.* no correlation between the maps):

$$SC = C_W / s(C_W)$$

Values of SC much greater than 2 indicate reliability in the contrast (Bonham-Carter, 1994).

4.1.2.4 Surficial Geology

Surficial geology information may be significant in providing a framework on which to relate till and lake sediment geochemistry. Klassen (1997) mapped 9 surficial

sediment types in the study area (Figure 3.6). The surficial geology descriptors (*e.g.* bedrock, till) were extracted from the digital map (Figure 3.6) at the 96 mineral/nonmineral occurrence sites and added to the attribute database (variable SURFICIA). The 96 alphanumeric descriptors were recoded to numeric values 1 to 9 and added to the attribute database (variable SURFCODE) to simplify analyses and further coding schemes. To determine if a relationship exists between the thickness of tills and their geochemical values, bedrock and the four sediments (*i.e.* drift/rock, till, till/gravel, and drift) were recoded with the mean thickness for the sediment (variable SURFTHIC). The mean thickness was based on the range of thickness assessed by Klassen (1997). Till geochemistry (*e.g.* logCu) was plotted against SURFTHIC using box-and-whisker plots to quickly assess any relationships.

4.1.2.5 Wetlands

The geochemistry of till and lake sediments may be affected if the samples were collected within a wetlands area. Wetlands cover about 10% of the study area and are coded on the topographic map as 'wetlands' (WAWL) and 'string bogs' (WASB). The wetland codes were extracted at the 96 mineral/nonmineral occurrence sites and added to the attribute database (variable BOG). Only 4 of the 96 sites occurred within wetland polygons, so further subdivision (*e.g.* using Peatland Inventory information) of the wetland type was not necessary. The variable BOG was recoded to a binary variable (BOGCODE) where '1' represents wetlands and '0' represents nonwetlands. This is a highly skewed dataset (with the potential for producing low correlations; Tabachnick and Fidell, 1996)

but was still used in nonparametric analysis (*e.g.* box-and-whisker plots) to determine if these 4 samples were related to any other factors.

4.1.2.6 Spatial Weighting Function

A known trait of mineral occurrences is that they often occur in spatial proximity to one another (Lydon, 1988). This trait will be incorporated in the database as a spatial weighting function (SWF) based on the spatial point-distribution of the mineral/nonmineral occurrences (Kvamme, 1990; Chou *et al.*, 1990). In a polygonal analysis of spatial autocorrelation, Chou *et al.* determined that the contiguity of neighbouring polygons represented the most basic form of the spatial relationship and applied a contiguity weight in the SWF. In point analysis of spatial autocorrelation, Kvamme (1990) applied the inverse distance between two points to represent the most basic form of spatial relationship.

Before developing a SWF it was necessary to demonstrate quantitatively that the data (mineral/nonmineral occurrences) show positive spatial autocorrelation and therefore warrant the use of a SWF. The calculations to determine the spatial autocorrelation measure are from Kvamme (1990). The statistical test was set up such that the null hypothesis tested the probability that all sites are equally mineral occurrences (*i.e.* no spatial autocorrelation). The alternative hypothesis was that a site neighbouring a mineral occurrence had a higher probability of being a mineral occurrence than not being a mineral occurrence (positive spatial autocorrelation). The spatial autocorrelation for point data were evaluated using the Moran's I statistic for point data and tested using the

randomization model for Ho, at $\alpha = 0.05$. The randomization model was used because it is not based on any assumptions about the distribution of the x_i whereas the alternative normality model assumes the x_i are normally distributed (Kvamme, 1990). Since only positive spatial autocorrelation is of interest, a one-tailed test was used where the standard normal deviate is $z = [I\text{-}E(I)] / \sqrt{var(I)}$. The spreadsheet and equations used to calculate Moran's I are presented in Chapter 5 (Tables 5.15 and 5.16).

The SWF was based on the distance between the mineral/nonmineral occurrence sites. This distance was calculated using Arc/Info's® POINTDISTANCE algorithm. A search radius of 5000 m provided a list of sites within 5000 m of each other. The list consisted of the identification of the point in question, 'i', the id of the neighbouring point, 'j', and the distance (in metres) between them. If an optimum search radius was later determined to be less than 5000 m then the list could be easily pruned to eliminate the longer distances. The optimum search radius was determined from a correlogram (Moran's I versus Separation Distance) of the 96 mineral/nonmineral occurrence sites. The distance over which Moran's I is positive indicates the neighbourhood (*i.e.* search radius) over which positive spatial autocorrelation occurs.

The spatial weighting function was calculated as follows (see Program A2 in Appendix A):

$$SWF_i = \frac{\sum_{j=1}^{n} W_{ij} * x_j}{\sum_{j=1}^{n} W_{ij}}$$

where the weight, W_{ij}, is the inverse distance between points i and j and x_j equals 1 if the jth point is a mineral occurrence and 0 otherwise. This SWF is very similar to Chou *et al.*'s (1990) SWF, only differing in the weight assigned. The values for the SWF vary between 0 (indicating the point is not close to any mineral occurrences) and 1 (indicating all neighbours around a point are mineral occurrences). The SWF values were calculated for each of the 96 mineral/nonmineral occurrence sites and incorporated into the attribute database.

4.2 Decision Tree Analysis

Decision tree analysis (DTA) is a robust method of classifying (*i.e.* splitting into subsets) dependent variables based on their relationship to the independent variable. The relationship is interpreted as a set of statistically significant 'IF-THEN' rules developed using the procedures in the KnowledgeSeeker® program (Angoss, 1993). See Chapter 2.2.2.2 for a more indepth discussion of DTA.

The KnowledgeSeeker® program requires the dependent variable (*i.e.* MINOCC) to be specified. Each independent variable (IV) was viewed to check the cluster type (*e.g.* continuous variables should be specified as monotonic or floating depending on whether they include missing values; binary variables should be listed as free clustering). The 'grow' method was set to 'exhaustive', which produces statistically reliable splits (Angoss, 1993). The filter level was initially set to the 'exploration' mode. This provides an adjusted error rate of 20% (α=0.20) used in testing the validity of the relationships (Angoss, 1993). The exploration mode was used to review the data and indicate potential

patterns. The filter level was reset to the 'prediction' mode (α=0.05) on the final run to check the results at the more stringent filter setting. Competing significant splits were reviewed at each node in order to assess splits that may make more logical sense but may not be as statistically significant. If there was no reason to override a statistical decision then no splits were forced. The procedure was repeated to test various combinations of IVs (*e.g.* the three different fault proximity variables). The Bonferroni value was adjusted accordingly when correlated IVs (*e.g.* till zinc values and PCA2) were included in the analysis at the same time.

The resulting "IF-THEN" rules are a set of conditions with associated probability values for the binary response variable. A rule-based map was produced using the IVs listed in the decision rules. The probability values, also listed as part of the decision rules, were assigned to each condition to arrive at a map indicating the 'weight' or 'probability', from 0 to 1, of the mineral potential. Since the data set is sparse, it was not practical to separate a test data set to use for accuracy testing. Therefore, accuracy testing was done by comparing the DTA results to the logistic regression analysis results (see below). A reliability map, based on the error or CV maps of each significant independent variable, was calculated for the rule-based map (see below).

4.3 Logistic Regression Analysis

Logistic regression is another robust technique that predicts the probability of a case belonging in one group or another. This technique was used to determine the independent variables that best group cases based on the dichotomous response variable

'mineral occurrences' (*i.e.* which variables best explain the presence or absence of mineralization and, therefore, mineral potential).

A preliminary logistic regression was run, with all IVs entered, to determine whether the full model improved the prediction of the mineral occurrences (the response variable) compared to the constant-only model. If the full model improved prediction of mineral occurrences, this would indicate that at least some of the IVs were useful in predicting mineral occurrences.

With minimal prior knowledge about the relationship between the data and the mineral occurrences, stepwise logistic regression was applied to determine which IVs were most useful in predicting mineral occurrences. The probability value to enter the model was increased from 0.05 to 0.20 and the probability value to be removed from the model was increased from the default of 0.10 to 0.30 to account for the interaction among predictors (Hosmer and Lemeshow, 1989). Each predictor which entered the equation was also tested individually by including it and removing it from the model. The chi-square goodness-of-fit (see Chapter 2) was calculated to determine the significance of adding the new predictor to the model.

Once the significant predictors had been determined, cases with large standardized residuals (*i.e.* >3.29, indicating the case was not well predicted by the model) were selected out and the regression repeated. Classification tables were analyzed at the final stages of model building to determine the percentage of correctly and incorrectly classified cases.

To determine if the assumption of pooling the VMS, gold, and 'unknown' deposit types was justified the LRA was run on just the VMS mineral occurrences (n=39) and all the nonmineral occurrences (n=49). The results for the subset of VMS occurrences versus all the occurrences in the database were compared.

A mineral potential map was created from the logistic regression equation using IDRISI's® IMAGE CALCULATOR. A reliability map was also calculated (see below), based on the error and CV maps of each significant IV in the model equation.

4.4 Comparison of Results from DTA and LRA

The mineral potential maps produced as a result of the DTA and the LRA consisted of two different measurement types. Due to the limited number of rules, the DTA map was composed of discrete probability values whereas the LRA map was composed of continuous probability values. These maps can be compared in a number of ways of which three were used in this study.

The first method was to determine how similar (level of agreement) the results of the DTA and LRA modelling methods were at the 96 sites used in the study. The probability values from the two modelled mineral potential maps were extracted at the 96 sites. These values were converted to binary codes where '0' represented probability values from 0% to just less than 50% and '1' represented probability values from 50% to 100%. The binary probabilities were compared to determine how many sites were classified the same by both methods (a measure of agreement) and how many sites were classified correctly by the two methods. A visual assessment of the spatial distribution of

the correctly and incorrectly classified data was also completed to note any obvious patterns.

The second method of comparing results used the same binary coding as above but, instead of a site analysis, the DTA and LRA mineral potential maps were converted to binary maps. A cross-tabulation of the results plus quantitative map comparison indices (Kappa and Yule's α; Bonham-Carter, 1994) were used to compare the maps.

The third method of comparing results was to summarize a cross-tabulation table of the actual values in the raster maps as opposed to comparing binary values. The cross-tabulation list provided the frequency of each LRA value for the discrete values in the DTA and was summarized by a box-and-whisker plot.

4.5 Reliability and Favourability Analysis

The reliability (*i.e.* certainty) of the models was assessed by combining the reliability maps for each of the significant predictors in the final DTA and LRA models. The predictors of interest consisted of the spatial weighting function (SWF), volcanic felsics (VOLCFELS), the binary fault proximity predictors (FLT400 and FLT1000), and copper in till (TLOGCU). The reliability for the SWF could not be determined because it was a calculated variable. The reliability for the remaining predictors was based on the source of the data for that predictor and will be discussed in more detail below.

Geology (specifically felsic volcanics) was determined to be a significant predictor. The reliability of the geology, in general, was assessed by comparing the geology from the regional 1:50,000 scale mapping to the detailed site-specific geology from the MODS

database. This comparison will give an indication of the usefulness of the 1:50,000 scale geology maps for quantitative analysis.

A reliability map for felsic volcanics was based on the locations of the outcrops because the geology is most reliable at these sites. The outcrop locations were digitized and buffered. Those cells close to outcrops were considered reliable whereas those cells furthest from the outcrops were considered most unreliable. The histogram of the cell proximity values was assessed to determine a reliability coding scheme. The maximum reliability on an outcrop was coded as 90% (S. Colman-Sadd, pers. comm., 2000), due to the less than perfect reliability of assessing the location and rock type, and decreased to 10% for those locations furthest from outcrops. The outcrop proximity map was recoded with the reliability values.

A reliability map for the fault variables was based on the buffers resulting from the WOE analysis. Evans *et al.* (1994) coded the faults in the study area as 'assumed' and 'approximate'. There were no 'defined' faults in the area. This may be due to the extensive forest and glacial cover. Therefore, no areas were considered reliably (*i.e.* 100%) a fault or reliably 'not a fault'. The most reliable areas were in proximity of the faults and also far distant from the faults. These areas were assigned a reliability of 80%. The least reliable areas were the edge of the buffers around the faults (Berry, 1993). Therefore, the buffer boundaries were assigned a reliability of 50% and one cell from the boundaries were assigned a reliability of 65%.

The map of till copper values was determined from an interpolated surface using kriging techniques. The kriging procedure also provided a standard deviation map, from

which a CV map was calculated by dividing the standard deviation map by the till copper map. The CV map was subtracted from 100 to conform to the reliability scale indicated above.

Once the reliability maps for each predictor were determined they were combined based on the method of calculating the DTA or LRA mineral potential maps. The DTA mineral potential map was calculated from rule-based conditions. The summary rules (see Chapter 5, Table 5.19) indicated that each rule was a combination of the SWF and one other predictor (VOLCFELS or FLT400). Since the SWF does not have an associated reliability map, then the reliabilities for the VOLCFELS or FLT400 maps were combined based on the areas they influenced. For example, those areas on the DTA mineral potential map defined by Rule 1 (*i.e.* SWF <0.17 and VOLCFELS=1) were assigned the reliability from the VOLCFELS reliability map for that area. The LRA reliability map was based on a combination of predictors as indicated in the final logistic regression equation (see Chapter 5.4). A simple method of determining the reliability for the LRA was based on the joint probability model as described by Berry (1993). In this method the reliability maps for each predictor in the logistic regression equation were multiplied together to provide a joint probability reliability map.

As a final summary analysis, those areas which have high mineral potential and high reliability will be most favourable for further mineral exploration. To define these favourable areas, the mineral potential maps were multiplied by their reliability maps. The favourability maps were converted to binary and the two final maps, for the DTA and LRA models, were combined (added together) to provide an overall favourability map.

This map was compared with the mineral occurrence sites to determine how many sites coincided with the maximum favourability areas.

4.6 Summary

The preparation of valid response and predictor variables is of primary importance to the modelling procedure. The geochemical databases took time to screen to provide optimal predictors. The varied data types (*e.g.* continuous geochemical values, nominal rock types, binary fault proximity) limited the methods available for quantitative modelling. Both DTA and LRA were able to handle all the data types. DTA and LRA were compared to provide an indication of the reliability of the modelling techniques, since the sparseness of the dataset did not allow for a separate test dataset.

CHAPTER 5

Variable Analysis and Modelling Results

This chapter presents the results of preprocessing of the response and predictor variables for input into the quantitative modelling of mineral potential. The significance of the predictors in determining mineral potential will be analyzed by two methods: decision tree analysis (DTA) and logistic regression analysis (LRA). The results of the analyses will be presented along with an assessment of their reliability. Overall favourability for mineral potential will be calculated by combining the mineral potential and reliability maps for the two models.

5.1 Response Variable

Randomly distributed data is a necessary requirement for classical statistical tests. The nearest neighbour index (Hammond and McCullagh, 1978) was used to test that the response variable, MINOCC, had a random spatial point distribution. Based on the equations presented in Chapter 4.1.1:

$$D_{obs} = \Sigma (d_{nn}) / N = 1.824 \text{ km}$$

$$D_{exp} = 1/2\sqrt{N/area} = 1.633 \text{ km}$$

$$\text{and } R = \frac{D_{obs}}{D_{exp}} = 1.12$$

where d_{nn} is the distance between nearest neighbours and N is the number of points. An index value close to 1 indicates that the points are randomly distributed. The z-score, to determine the significance of the nearest neighbour index is:

$$z = (D_{obs} - D_{exp},) / \sigma_{Dexp} \quad \text{(Hammond and McCullagh, 1978)}$$

$$\text{where} \quad \sigma_{Dexp,} = 0.26136 / \sqrt{N^2 / area} = 0.087$$

$$\text{therefore, } z = (1.824 - 1.633) / 0.087 = 2.19$$

A z-score less than 2.58 (at $\alpha = 0.01$) confirms the index value of 1.12 and indicates an random point distribution.

5.2 Predictor Variables

5.2.1 Till Geochemistry

The till geochemistry database consists of 247 samples with analyses for 34 elements. The first step in data analysis was to remove those elements which had a limited or poor distribution of data. A review of the till database using EDA techniques (*e.g.* box-and-whisker plots) and frequency tables indicated the elements Ag, Ca, Mo, Ta and W had more than 50% of their values less than the analytical detection limit. Also, Cs had 37% of its values less than the detection limit and the remaining Cs values were severely limited in range. These elements were all removed from the database.

Univariate statistics consisting of the minimum, maximum, median, mean and standard deviation, log mean and standard deviation and geometric mean, were tabulated as a summary overview of the remaining 28 elements (Table 5.1). Histograms of the

Table 5.1 : Univariate statistics for elements in till samples in the NTS 12A/10 area.

Element	Unit	N	Minimum	Maximum	Median	Mean	Standard Deviation	Logarithmic Mean	Log Standard Deviation	Geometric Mean
As_N	ppm	246	3.6	180	36.5	43.5	29.74	1.55	0.293	35.2
Au_N	ppb	246	<2	41	6.0	6.4	5.60	0.65	0.396	4.5
Ba_N	ppm	246	<50	620	330	338	88.4	2.51	0.147	324
Br_N	ppm	246	<0.5	81.0	2.6	4.1	6.78	0.25	0.623	1.8
Ce_N	ppm	246	23	110	57.0	58.3	15.69	1.75	0.120	56.2
Co_N	ppm	246	5	62	16.0	16.8	7.22	1.19	0.173	15.5
Cr_N	ppm	246	21	160	50.0	58.6	26.03	1.73	0.169	54.0
Cu_C	ppm	240	15	392	72.1	80.3	40.03	1.86	0.206	72.1
Eu_N	ppm	246	0.4	2.5	1.50	1.45	0.331	0.150	0.107	1.41
Fe_N	pct	246	1.9	9.18	4.32	4.37	1.211	0.624	0.120	4.17
Hf_N	ppm	246	3.0	24.0	5.0	6.0	2.77	0.74	0.156	5.6
La_N	ppm	246	8.8	50	24.5	25.3	6.03	1.39	0.106	24.6
Lu_N	ppm	246	0.29	1.00	0.580	0.589	0.112	-0.238	0.0845	0.578
Mg_C	pct	240	0.52	1.79	0.93	0.94	1.641	-0.040	0.0740	0.91
Mn_C	ppm	240	562	4026	1273	1363	457.7	3.11	0.130	1301
Na_N	pct	246	0.48	2.73	1.66	1.66	2.995	0.212	0.0830	1.62
Ni_C	ppm	240	10	78.1	26.2	28.0	10.15	1.42	0.154	26.3
Pb_C	ppm	240	7	88.9	17.8	20.0	9.46	1.27	0.170	18.4
Rb_N	ppm	246	<5	99.0	31.0	31.9	18.17	1.37	0.426	23.6
Sb_N	ppm	246	0.5	13.0	2.15	2.40	1.410	0.334	0.188	2.16
Sc_N	ppm	246	9.0	28.0	18.0	17.6	3.19	1.24	0.0816	17.3
Sm_N	ppm	246	1.8	8.5	4.90	4.94	1.172	0.681	0.110	4.80
Tb_N	ppm	246	<0.5	1.5	0.90	0.83	0.333	-0.13	0.236	0.74
Th_N	ppm	246	2.6	15.0	6.30	6.65	1.813	0.808	0.112	6.43
U_N	ppm	246	<0.5	11.0	2.10	2.49	1.297	0.340	0.235	2.19
V_C	ppm	240	59	191	107.7	111.8	23.88	2.039	0.0908	109.4
Yb_N	ppm	246	2.0	6.7	3.90	3.96	0.755	0.590	0.0855	3.89
Zn_C	ppm	240	42	731	105.3	111.2	52.70	2.019	0.146	104.4

element values showed a positively skewed distribution. Therefore, all values were logged to normalize the distributions. This had the added benefit of stabilizing the variance (Davis, 1986; George and Bonham-Carter, 1989).

A visual assessment of the spatial distribution of the elements can reveal associations between the elements and spatial factors such as proximity to faults. Therefore, graduated dot plots of many elements were plotted and any interesting spatial associations were noted. Values near the 95^{th}, 85^{th}, 70^{th}, and 50^{th} percentiles were used as quick break point values. The spatial distribution of copper values in till samples (Figure 5.1) shows a distinct difference between samples to the northwest of Victoria River versus samples to the southeast of the river. The distribution may indicate elevated background copper values in the Tally Pond volcanics in the southeast versus the Tulks Hill volcanics in the northwest. The copper values for these two areas were extracted, with the dividing line being the southern extent of the Tulks Hill volcanics (Figure 5.1). A comparison of the box-and-whisker plots (Figure 5.2) for these two areas indicates some overlap in the interquartile range but the 95% confidence interval for the medians (Rock, 1988) for the two areas do not overlap (see below) indicating two separate groups:

$$95\%\ \text{Confidence Interval} = \text{median} \pm 1.58\ (\text{interquartile range}) / \sqrt{n}$$

$$\text{Northwest Area: } 95\%\ CI = 53.9 \pm 1.58\ (24.3) / \sqrt{96} = 53.9 \pm 3.9$$
(actual range = 50.0 to 57.9)

$$\text{Southeast Area: } 95\%\ CI = 89.2 \pm 1.58\ (44.1) / \sqrt{150} = 89.2 \pm 5.7$$
(actual range = 83.5 to 94.9)

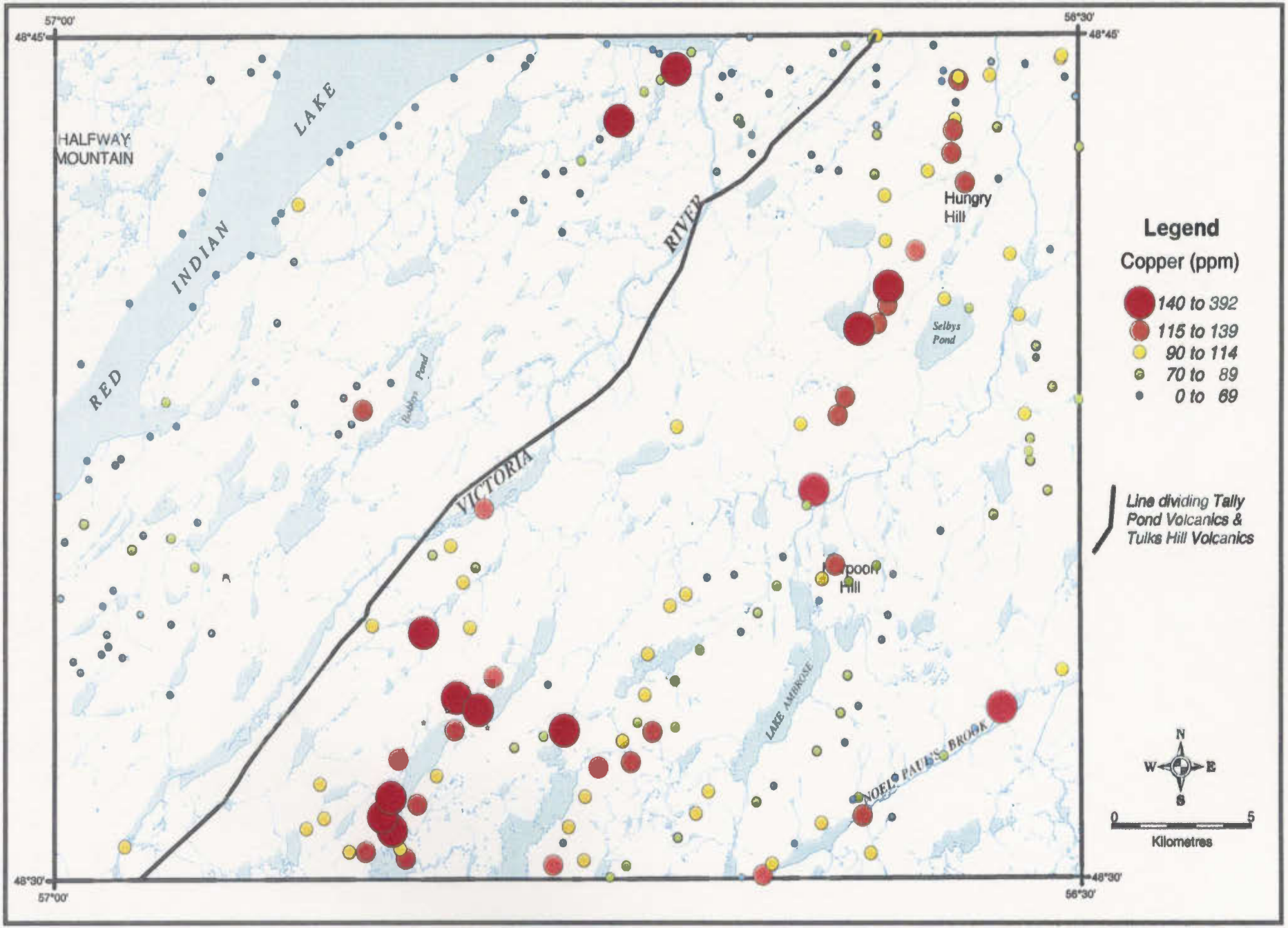

Figure 5.1 : Copper distribution in till samples. Note the higher number of anomalous samples in the southeast half of the map.

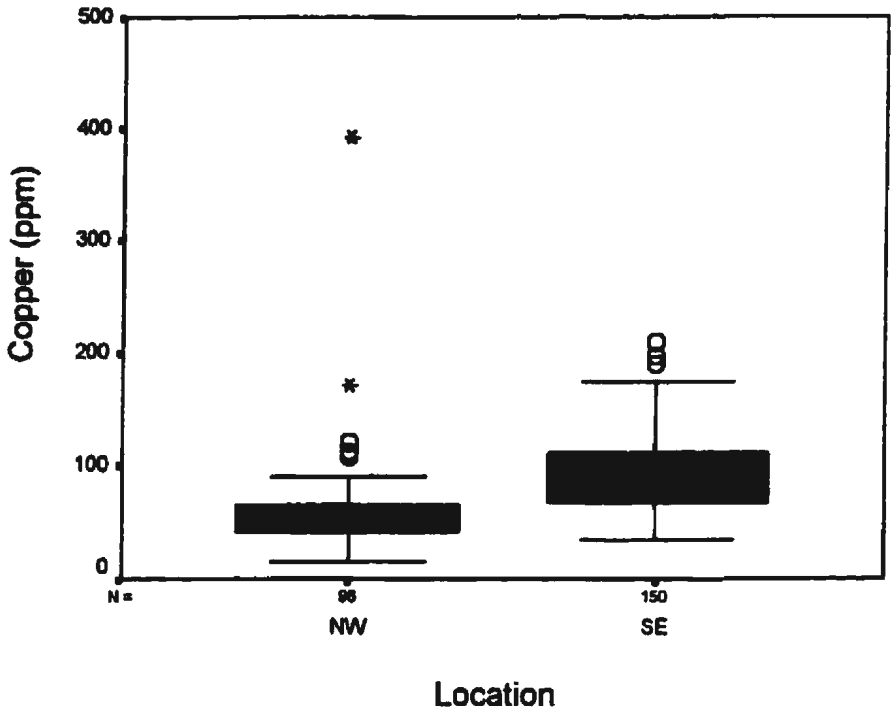

Figure 5.2 : A comparison of copper values in till samples collected from the northwest and southeast of the study area.

A graduated dot plot map of the gold values in the till samples (Figure 5.3) indicates a similar distribution to copper, but graduated dot plots and box-and-whisker plots of lead and zinc do not show any distinct spatial distribution. These results suggest a gold-copper association, which is supported by Spearman's rank correlation coefficient between logCu and logAu (r=0.45; Table 5.2). Spearman's rank correlation coefficient is used instead of the Pearson correlation coefficient because, even though the data is logged, some of the elements do not exhibit a true normal distribution.

Moderate to strong correlations (Table 5.2) exist between most ore elements (logAu, logCu, logPb, and logZn) as well as some pathfinder elements (logAs, logSb).

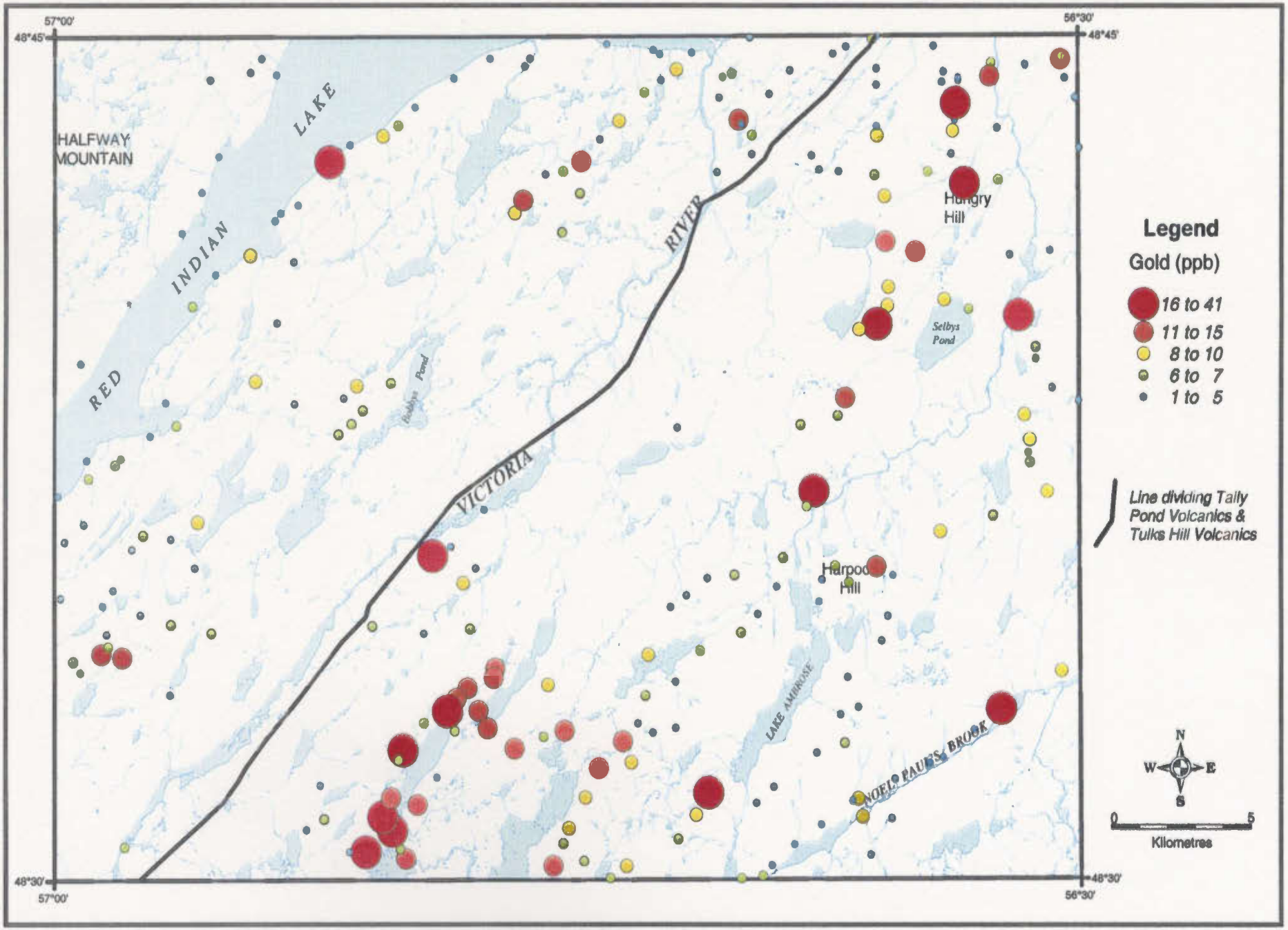

Figure 5.3 : Gold distribution in till samples. Note the higher number of anomalous samples in the southeast half of the map.

Table 5.2 : Spearman's rank correlation coefficient (r) for selected indicator elements from till samples. Note that all correlations are 2-tailed and significant at the 0.01 level.

		LOGAS	LOGAU	LOGCU	LOGPB	LOGSB	LOGZN
LOGAS	r	1.000					
	N	246					
LOGAU	r	.447	1.000				
	N	246	246				
LOGCU	r	.497	.450	1.000			
	N	240	240	240			
LOGPB	r	.607	.257	.311	1.000		
	N	240	240	240	240		
LOGSB	r	.651	.397	.482	.412	1.000	
	N	246	246	240	240	246	
LOGZN	r	.584	.408	.667	.450	.415	1.000
	N	240	240	240	240	240	240

The strongest relationship is between logCu and logZn (r=0.667), which supports the relationship between copper and zinc as indicated by the VMS deposit model. Due to the moderately high correlations in Table 5.2 and for the rest of the database, principal components analysis (PCA) should work well to reduce the 28 elements to a smaller number of components.

The KMO measure of sampling adequacy was calculated to determine if the data was suitable for PCA; a value greater than 0.6 indicates good factorability (Tabachnick and Fidell, 1996). The KMO measure of sampling adequacy was 0.827, which indicates that the overall database is factorable.

PCA is sensitive to multivariate outliers. To test the sensitivity of the till database to multivariate outliers, the PCA results were compared before and after the outliers were removed. Multivariate outliers are determined by calculating the Mahalanobis distance (MD), available through the SPSS™ linear regression procedure. For 28 elements the

critical chi-squared value, for $\alpha = 0.001$, is 56.9. Therefore, a MD greater than 56.9 indicates a multivariate outlier case. There were 13 cases greater than 56.9. A box-and-whisker plot of the MD values (Figure 5.4) indicates the gap between the outlier values (*i.e.* 6 cases from 56.9 to 63) and extreme values (*i.e.* 7 cases from 78 to 147).

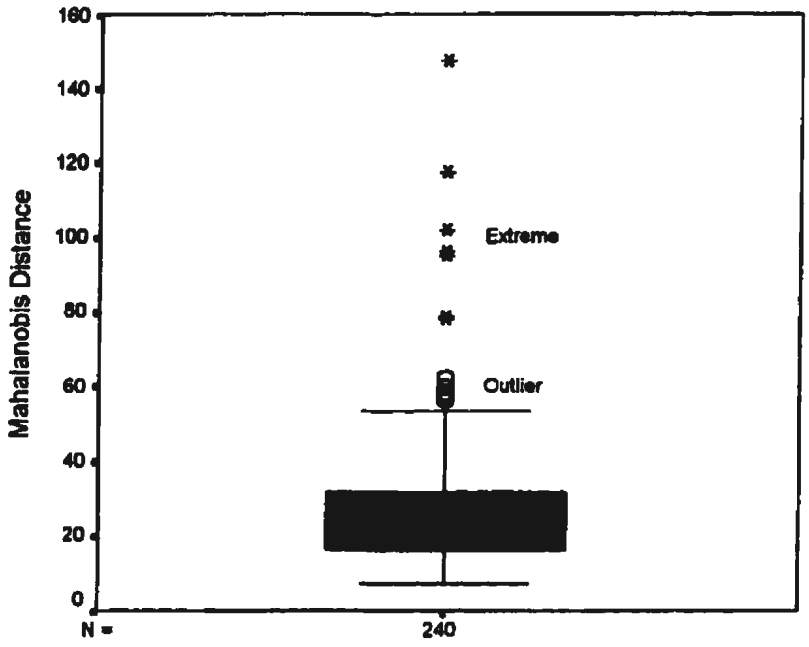

Figure 5.4 : Mahalanobis distance for till samples. Note the gap between the outlier values (57 to 63: circles) and extreme values (78 to 147: stars).

Therefore, the 7 cases greater than 78 were considered multivariate outliers. These 7 cases were temporarily removed from the database. The PCA was repeated, with a varimax rotation, and the results noted. The 7 cases were then included in the database and the PCA was run again. The results (*i.e.* components, variance explained, loadings in the rotated component matrix as well as the component values) were compared with and without the 7 multivariate outliers. The components on both runs consisted of the same

sets of elements. The percent variance explained for each component differed by less than 0.8% and the overall cumulative variance explained differed by 0.5%. The loadings in the rotated component matrix were all comparable. The correlations on the output factor scores were greater than 0.99 for each of the four components. This indicates that, for this dataset, PCA is robust to these 7 multivariate outliers. Therefore, all cases were included in the final PCA.

The next step was to determine the number of components to be extracted. Three measurements provide an indication to the number of components to extract: the number of eigenvalues greater than 1.0, the change in slope of the scree plot and the change in variance explained from one component to the next. Six components had eigenvalues greater than 1.0. The change in slope of the scree plot (Figure 5.5) occurs between 4 and 5 components. The rotated sum of squared loadings percent of variance explained was 19.8, 13.7, 13.5, 13.5, 8.9 and 4.8, respectively by the 6 components. The change in variance explained from the fifth to the sixth component was much greater than the other components. Therefore, 5 components were extracted and analyzed. The fifth component had no loadings greater than |0.7| and consisted of logCr, logBr and logTb. These elements are not geochemically interpretable. Therefore, the next analysis extracted 4 components. Each of the resulting 4 components is geochemically interrelated and has strong loadings. The communalities for logBr and logTb were less than 0.2, indicating that very little variance in these two variables was accounted for by the 4 components. Therefore, logBr and logTb were removed from further PC analysis. LogAu had the next

lowest communality of 0.277 but since gold is a significant element for this study it was kept in the analysis.

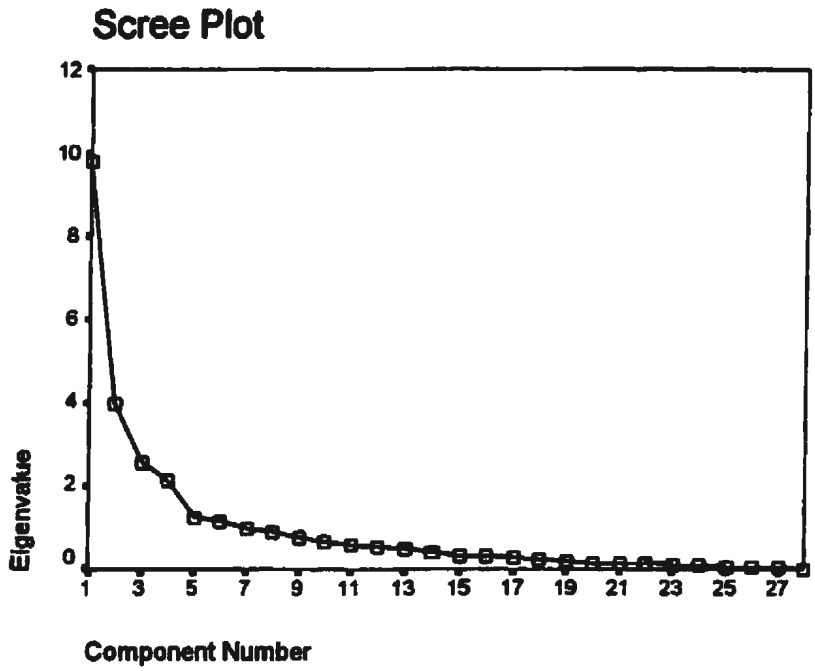

Figure 5.5 : Scree plot for till PCA indicating the decrease in eigenvalues with increasing number of components. To determine the appropriate number of components to extract note the component number at the change in slope.

The PCA was repeated, extracting four components and saving the scores in the till database. The final results are listed in Table 5.3. Each component explains more than 10% of the variance in the data with a total of 69.7% variance explained. The communalities indicate how much of the variation of each element is accounted for by the components (Tabachnick and Fidell, 1996). A summary of the elements characterizing each component and their geochemical affiliation is provided in Table 5.4. The largest communalities are for elements in component 1 (*i.e.* TPC1), indicating that their variance

is very well explained by the components. TPC2 is the most important component for exploration purposes in that it is characterized by the base metals (*e.g.* logZn, logCu, logPb and logCo), gold and gold pathfinder elements (*i.e.* logAu, logAs and logSb).

Table 5.3 : Principal components analysis results for till samples in the NTS 12A/10 area, with varimax rotated component matrix. Only loadings >0.3 are shown. Bold values indicate component element affinities.

	Rotated Component Matrix				
Element	Component				Communality
	1	2	3	4	
LOGYB	**0.905**				0.886
LOGLU	**0.890**				0.822
LOGSM	**0.877**	0.323			0.926
LOGEU	**0.860**	0.348			0.892
LOGLA	**0.699**		**0.563**		0.845
LOGCE	**0.654**		**0.588**		0.797
LOGZN		**0.795**			0.723
LOGAS		**0.757**			0.713
LOGCU	0.326	**0.711**		0.396	0.789
LOGCR		-0.613	0.538		0.692
LOGHF		-0.604		-0.499	0.688
LOGFE	0.491	**0.589**		0.473	0.843
LOGCO	0.447	**0.565**		0.484	0.762
LOGMN	0.391	**0.557**		0.360	0.614
LOGSB	0.479	**0.506**	0.301		0.593
LOGAU		**0.475**			0.278
LOGTH	0.310		**0.840**		0.827
LOGNA	0.308		**-0.659**		0.623
LOGBA			**0.632**	0.324	0.516
LOGPB		**0.571**	**0.581**		0.703
LOGU			**0.569**		0.431
LOGRB			**0.540**		0.313
LOGMG				**0.883**	0.781
LOGV				**0.825**	0.697
LOGSC		0.319		**0.756**	0.763
LOGNI			0.377	**0.562**	0.614
% of variance	21.4	18.5	15.2	14.6	69.7

Table 5.4 : A summary of the geochemical affinity of the four till principal components.

Component	Elements	Geochemical Affinity
TPC1	Yb, Lu, Sm, Eu, La, Ce	Rare Earth Elements
TPC2	Zn, Cu, Pb, Co, Fe, Mn, Au, As, Sb	Base Metals & Gold Association
TPC3	Th, -Na, Ba, Ce, Pb, U, La, Rb	Granophile Elements
TPC4	Mg, V, Sc, Ni	Mafic Elements

Outliers in the resulting component scores indicated those cases for which the factor solution was not appropriate. In the four components there were only three negative outliers (*i.e.* standardized values less than –3.29) and one positive outlier that was only slightly higher than +3.29 (*i.e.* 3.46 in TPC4). TPC2, the component of most interest in this study, had no outliers (range –2.53 to 2.45) indicating that all cases fit this solution well. Therefore, outliers to the solution were not deemed critical and the final PCA was not repeated.

Data screening is required prior to PCA because it is based on the Pearson correlation coefficient matrix, which is affected by skewness, linearity etc. A more robust method of analysis would replace the Pearson correlation with Spearman's rank correlation coefficient in the PCA. The Spearman's rank correlation coefficient has no assumptions about the distribution of the data. The raw geochemical data can be used without tedious data screening requirements. PCA results based on the Spearman's rank

correlation coefficient on the raw data produced essentially the same components as the Pearson correlation on the logged data.

To determine whether the individual ore elements (*i.e.* Au, Cu, Pb, and Zn) or component 2 (*i.e.* TPC2) are significant predictors of mineral occurrences their values must be extracted at the 96 mineral/nonmineral occurrence sites. Therefore, the till point data must be represented as a surface. If the data has a positive spatial autocorrelation then kriging methods can be used to prepare an interpolated surface. Semi-variograms and Moran's I plots were evaluated for the four ore-forming elements (Au, Cu, Pb, and Zn) and for TPC2 (representing the base metals and gold pathfinder elements). Outliers were changed to values just above the lower neighbouring value (Tabachnick and Fidell, 1996). For example, logCu had one high outlier (2.59) which was adjusted to 2.33 (next lower value is 2.32). Adjustments of outliers had very little effect on the results. Therefore the original values were retained. The variograms and Moran's I plots are presented in Figures 5.6a to 5.10b. A summary of the variogram results are listed in Table 5.5.

The results of the variogram and correlogram analyses indicate that all the variables are positively, spatially autocorrelated and can be spatially interpolated using kriging techniques. The following parameters were determined to provide the best block kriging: a search radius of 8000 m, 3X3 local block grid and the use of 6 nearest neighbours. For example, Figure 5.11 presents the interpolated surface for logCu. In addition to indicating spatial autocorrelation, the correlograms also provide specific information about the individual variables; TPC2 has the highest Moran's I of 0.594 at an

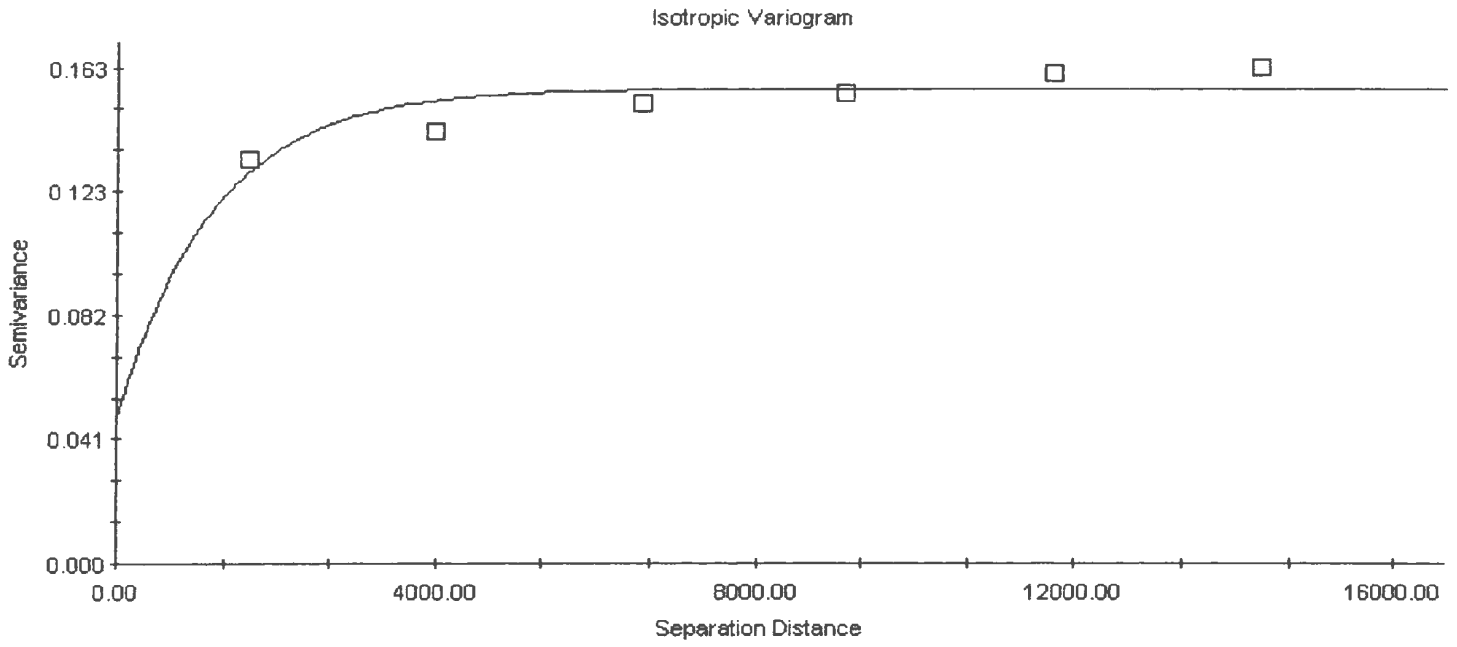

Figure 5.6a : Variogram of logAu in till. The best fitted model is exponential where Co=0.049, Co+C=0.156, Ao=1190, r^2=0.682 and R_{SS}=0.0002.

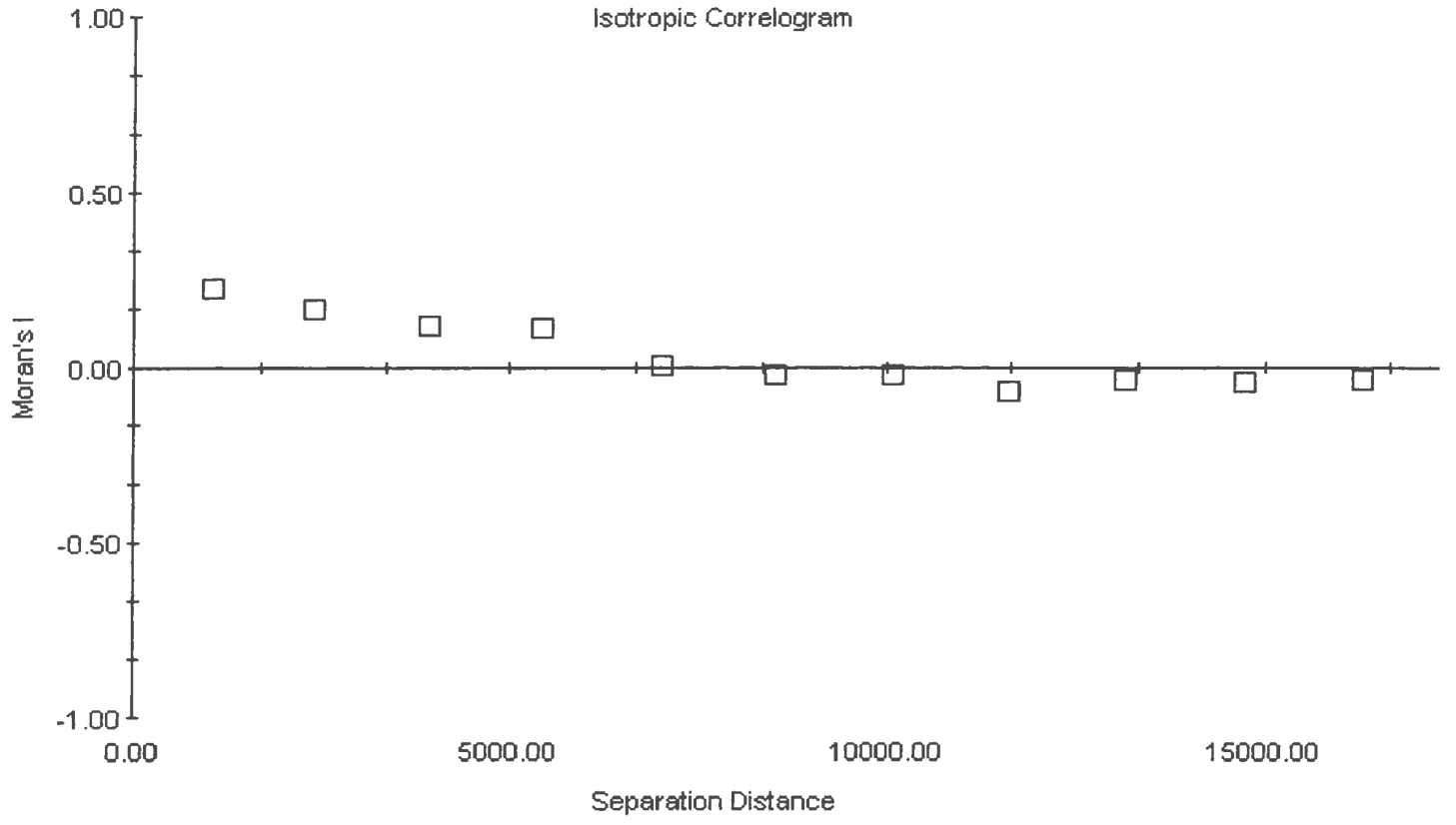

Figure 5.6b : Correlogram of logAu in till. Positive autocorrelation occurs to a distance of about 6,000 m.

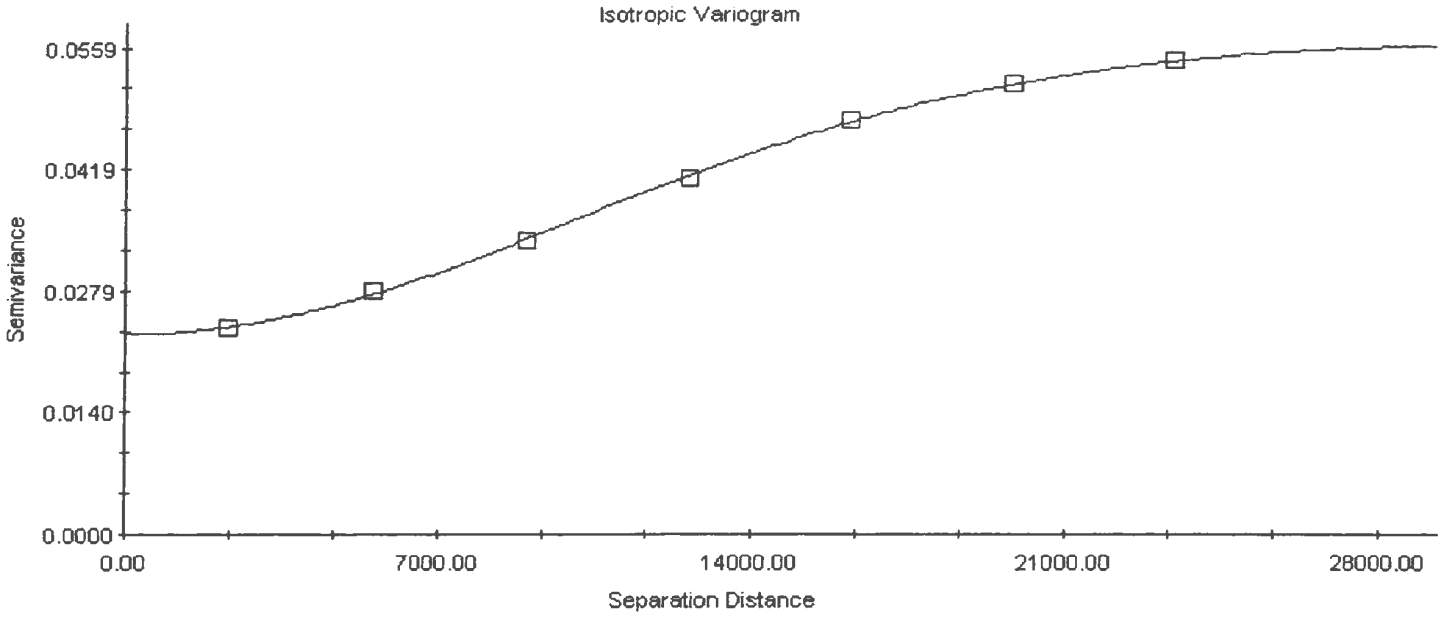

Figure 5.7a : Variogram of logCu in till. The best fitted model is gaussian where Co=0.023, Co+C=0.057, Ao=24740, r^2=1.000 and R_{SS}=0.0000.

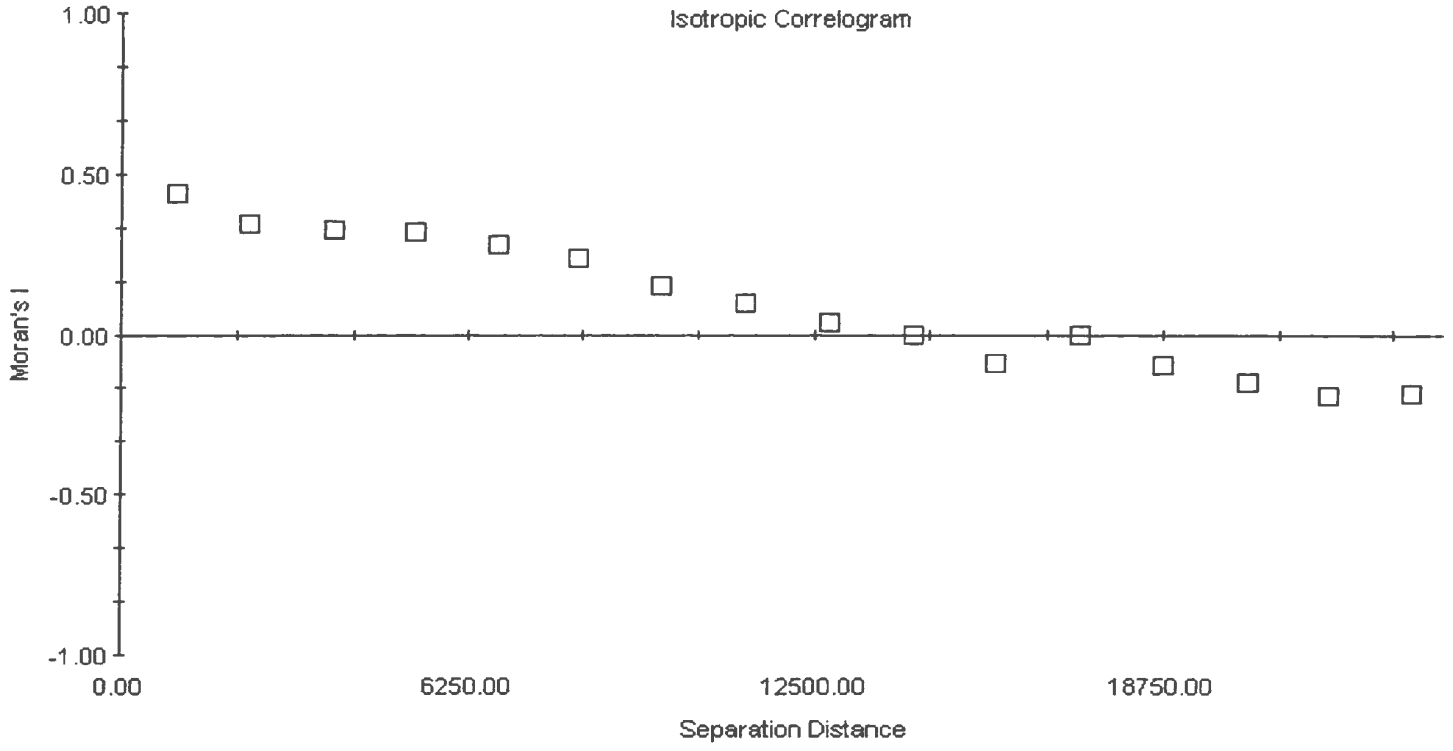

Figure 5.7b : Correlogram of logCu in till. Positive autocorrelation occurs to a distance of about 12,000 m.

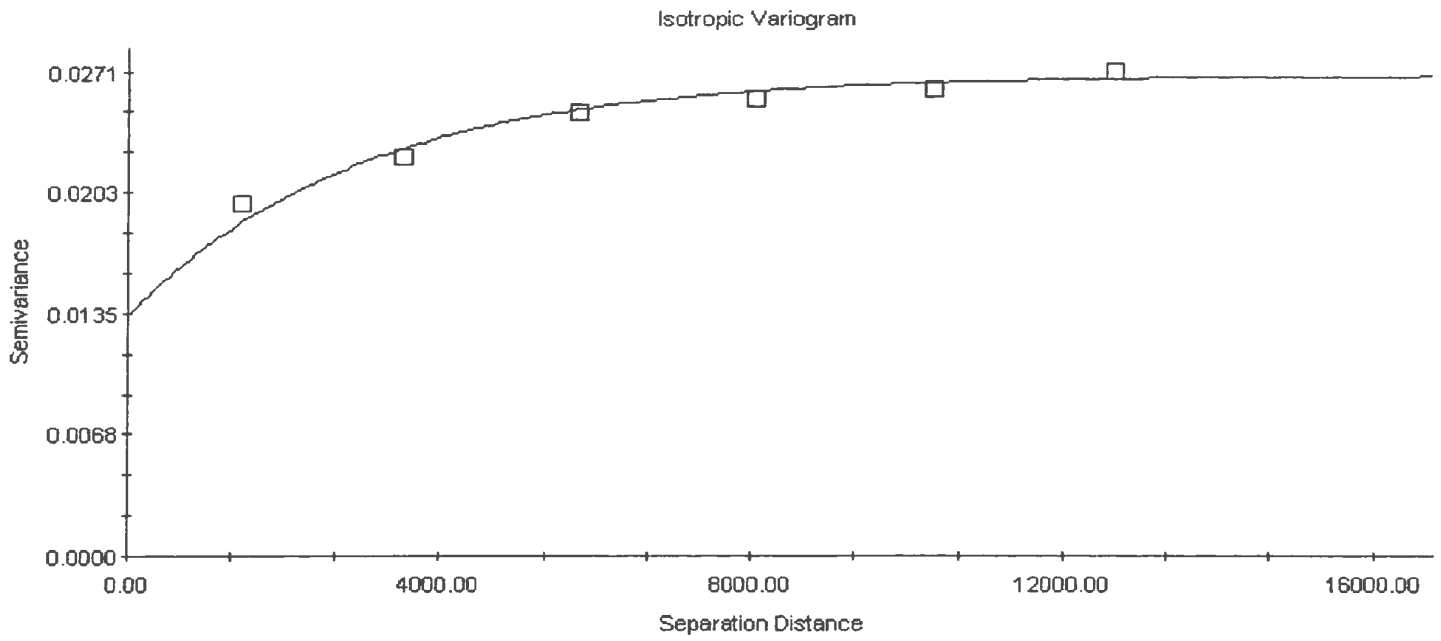

Figure 5.8a : Variogram of logPb in till. The best fitted model is exponential where Co=0.013, Co+C=0.027, Ao=2930, r^2=0.968 and R_{SS}=0.0000.

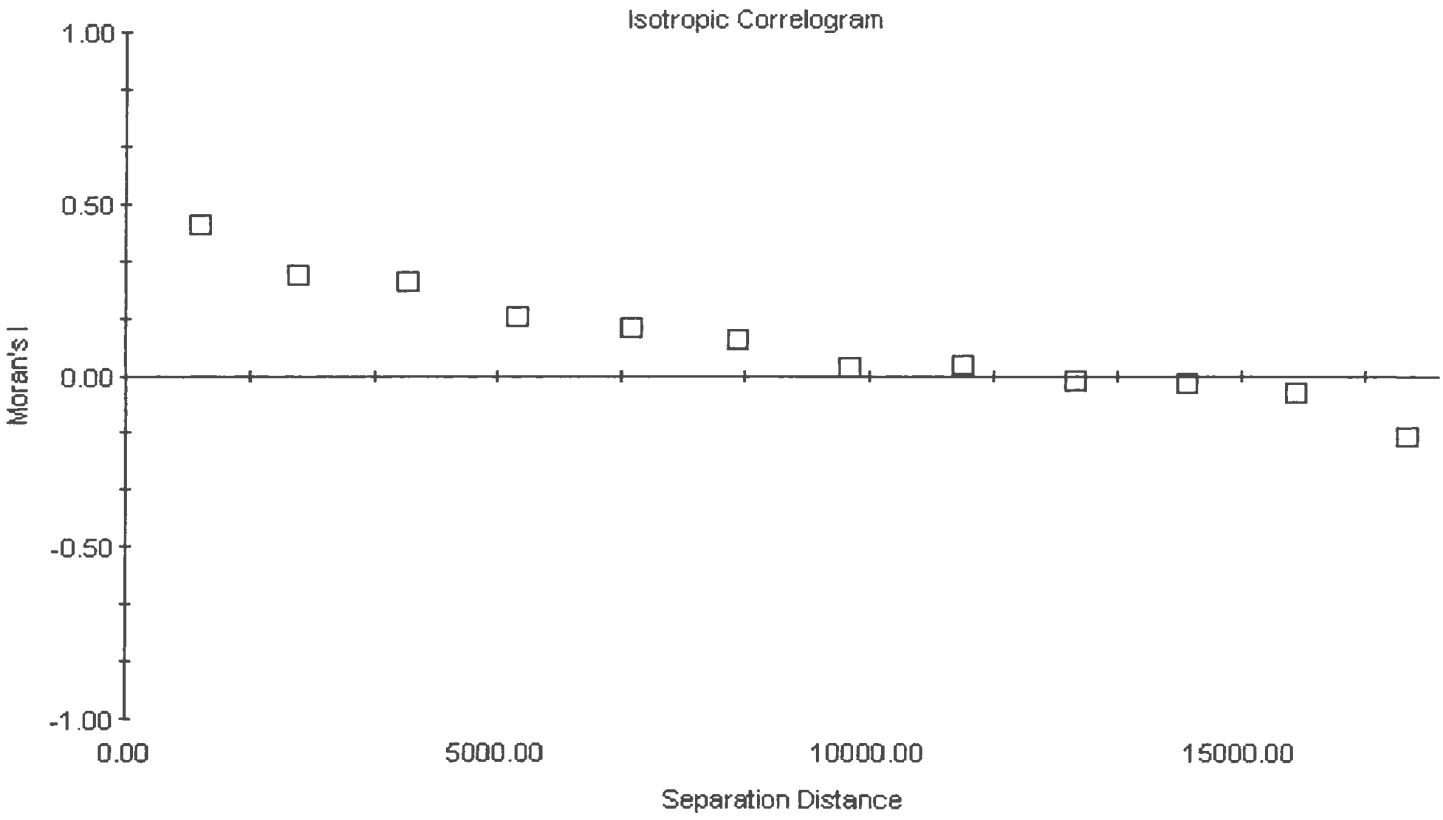

Figure 5.8b : Correlogram of logPb in till. Positive autocorrelation occurs to a distance of about 10,000 m.

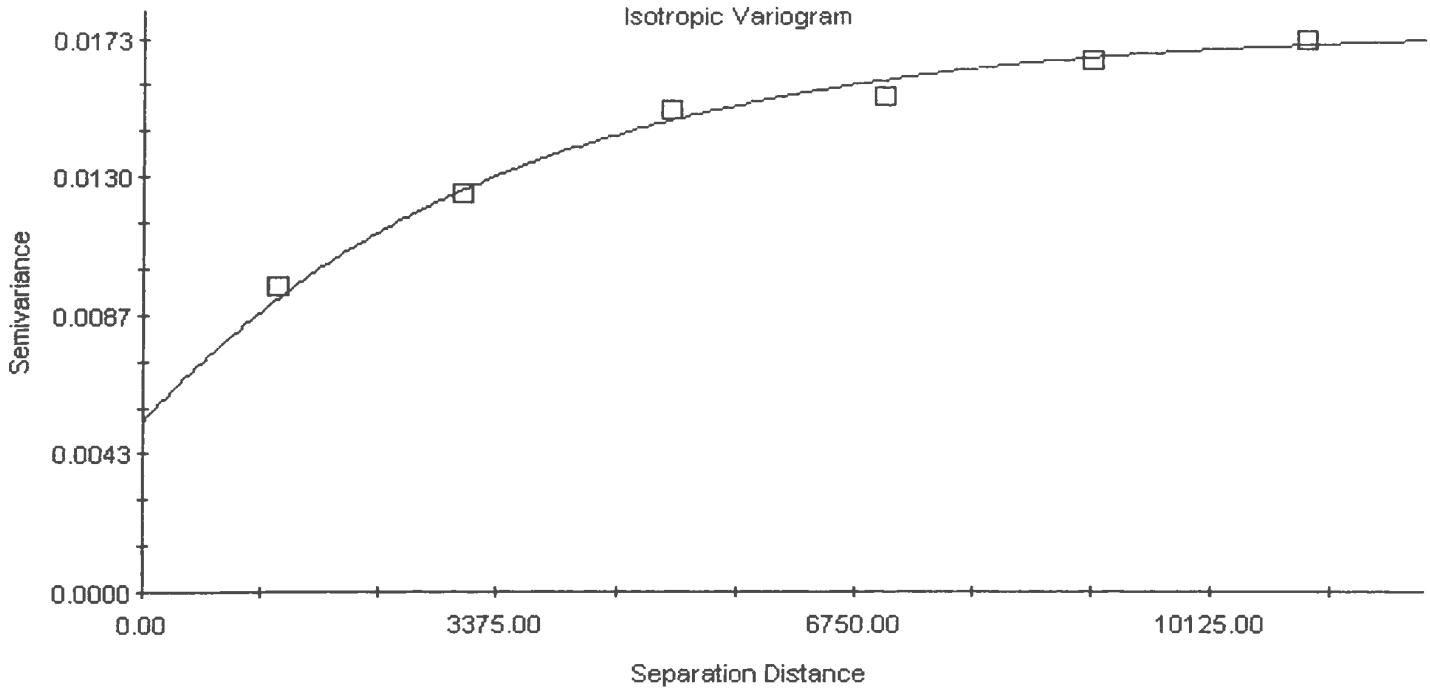

Figure 5.9a : Variogram of logZn in till. The best fitted model is exponential where Co=0.005, Co+C=0.018, Ao=3400, r^2=0.990 and R_{SS}=0.0000.

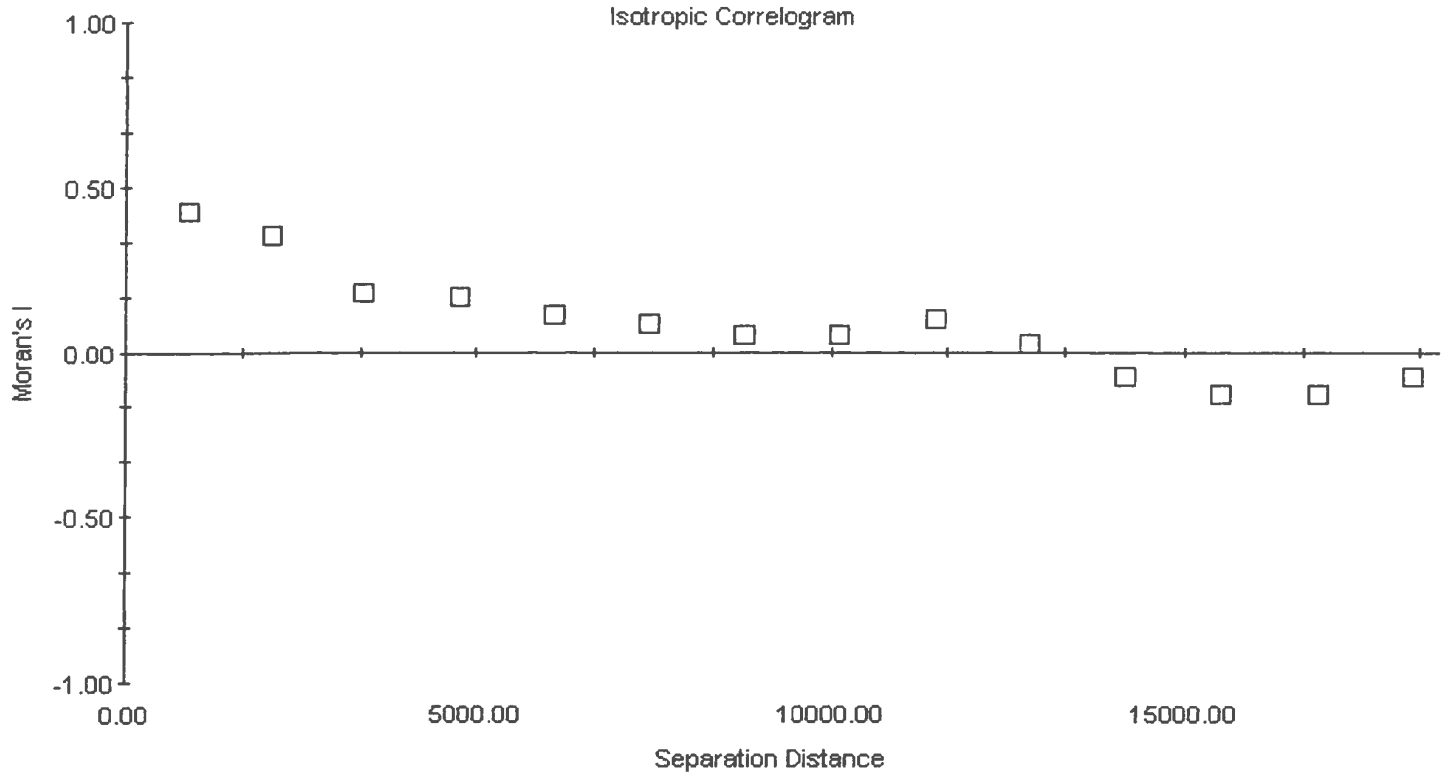

Figure 5.9b : Correlogram of logZn in till. Positive autocorrelation occurs to a distance of about 10,000 m.

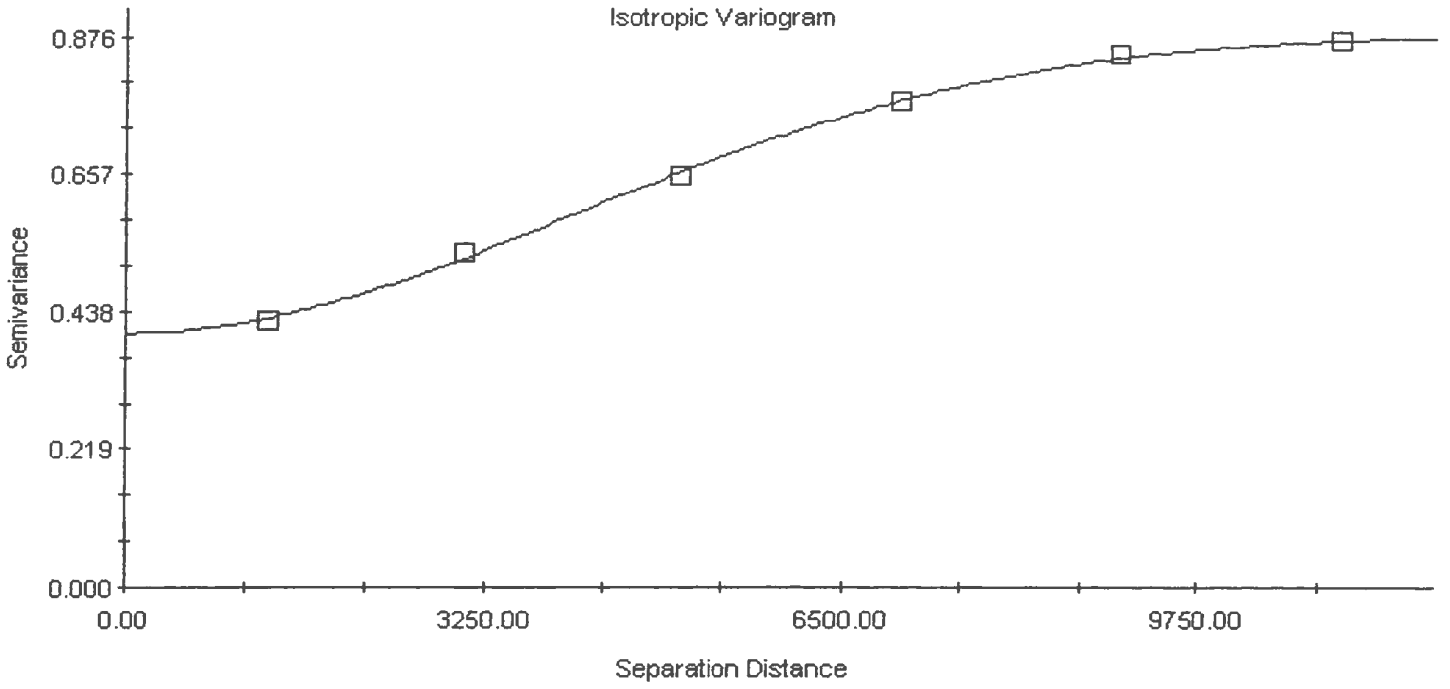

Figure 5.10a : Variogram of the base metal component (TPC2) from till. The best fitted model is gaussian where Co=0.404, Co+C=0.879, Ao=9890, r^2=0.999 and R_{SS}=0.0002.

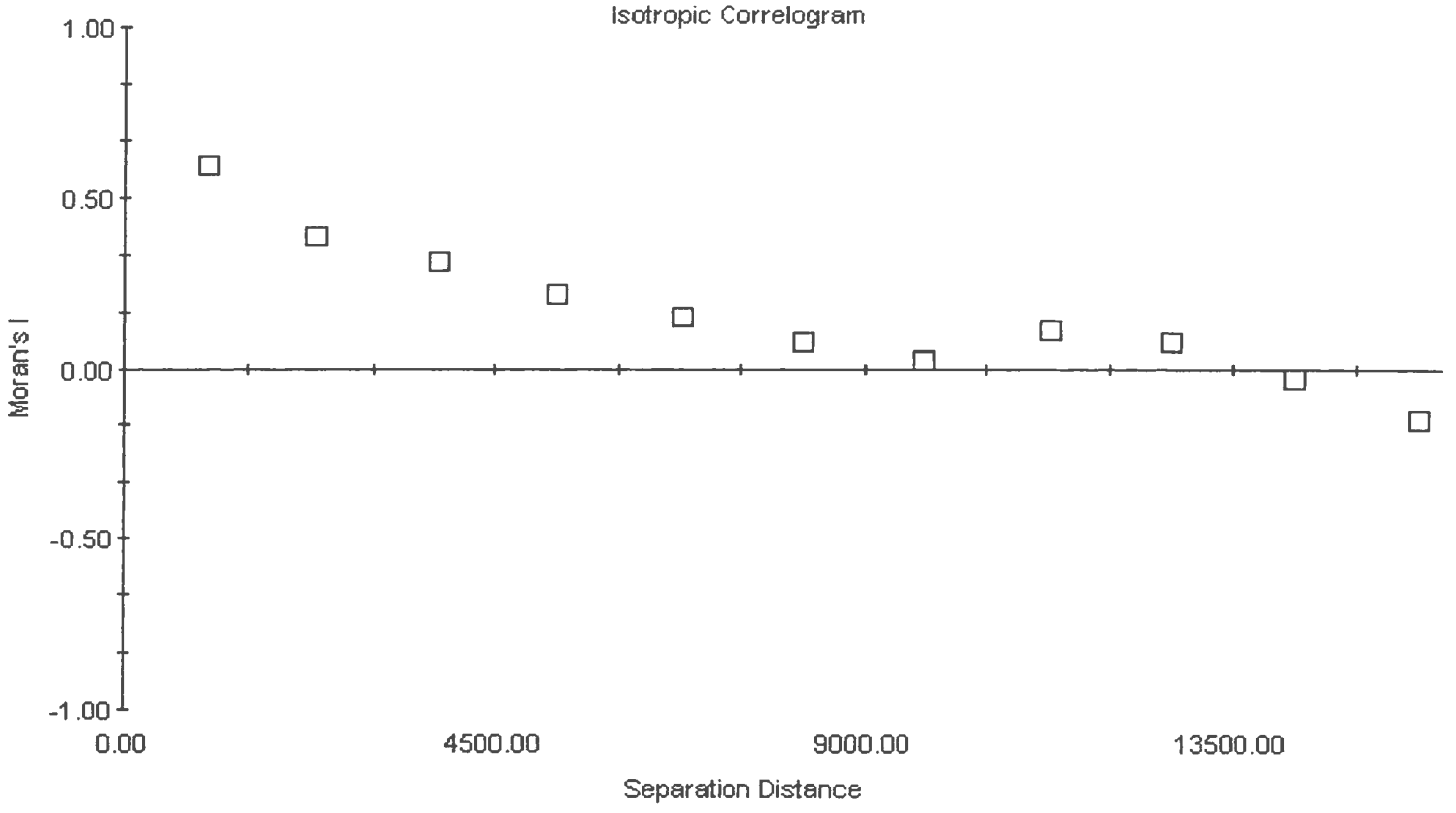

Figure 5.10b : Correlogram of TPC2 from till. Positive autocorrelation occurs to a distance of about 9,800 m.

Table 5.5 : Parameters for variograms of selected elements in till samples.

Element	Active Lag	Lag Class	Nugget Co	Sill Co+C	Range Ao	Effective Range	Model	RSS	R^2
logAu	16000	2600	0.0486	0.1562	1190	3570	Exponen.	$2.19x10^{-4}$	0.68
logCu	28000	3600	0.0229	0.0566	24740	42851	Gauss.	$2.78x10^{-7}$	1.00
logPb	16000	2300	0.0134	0.0268	8790	26370	Exponen.	$1.84x10^{-5}$	0.97
logZn	13500	2000	0.0053	0.0176	3410	10200	Exponen.	$5.11x10^{-7}$	0.99
TPC2	13000	2000	0.4040	0.8790	9890	17130	Gauss.	$1.60x10^{-4}$	1.00

average lag of 1010 m and logCu has the longest positive autocorrelation of 12 km.

As well as providing the interpolated surfaces, the program GS+ also provides a standard deviation surface. This surface indicates the variability of the data around each point. A coefficient of variation (CV) surface was derived from the standard deviation surface by dividing it by the kriged surface (representing the mean). The CV surface for logAu in till ranges from 11.1 to 27.4%. Similarly, logCu and logZn also have good reliability with logCu CV ranging from 7.5 to 18.4% and logZn CV ranging from 8.3 to 26.3%. LogPb had the worst variance with CV ranging from 18.5 to 41.5%. TPC2 CV range from 6.4 to 40.5%. An analysis of the CV values extracted at the 96 mineral/nonmineral occurrence sites indicates the highest values occur at the nonmineral occurrence sites. This may be a reflection of the wider spacing (poorer spatial autocorrelation) occurring around the nonmineral occurrence sites.

5.2.2 Lake Sediment Geochemistry

Similar to the till database, the first step in data analysis was to remove elements in the lake sediment database which had a poor data distribution. EDA techniques indicated

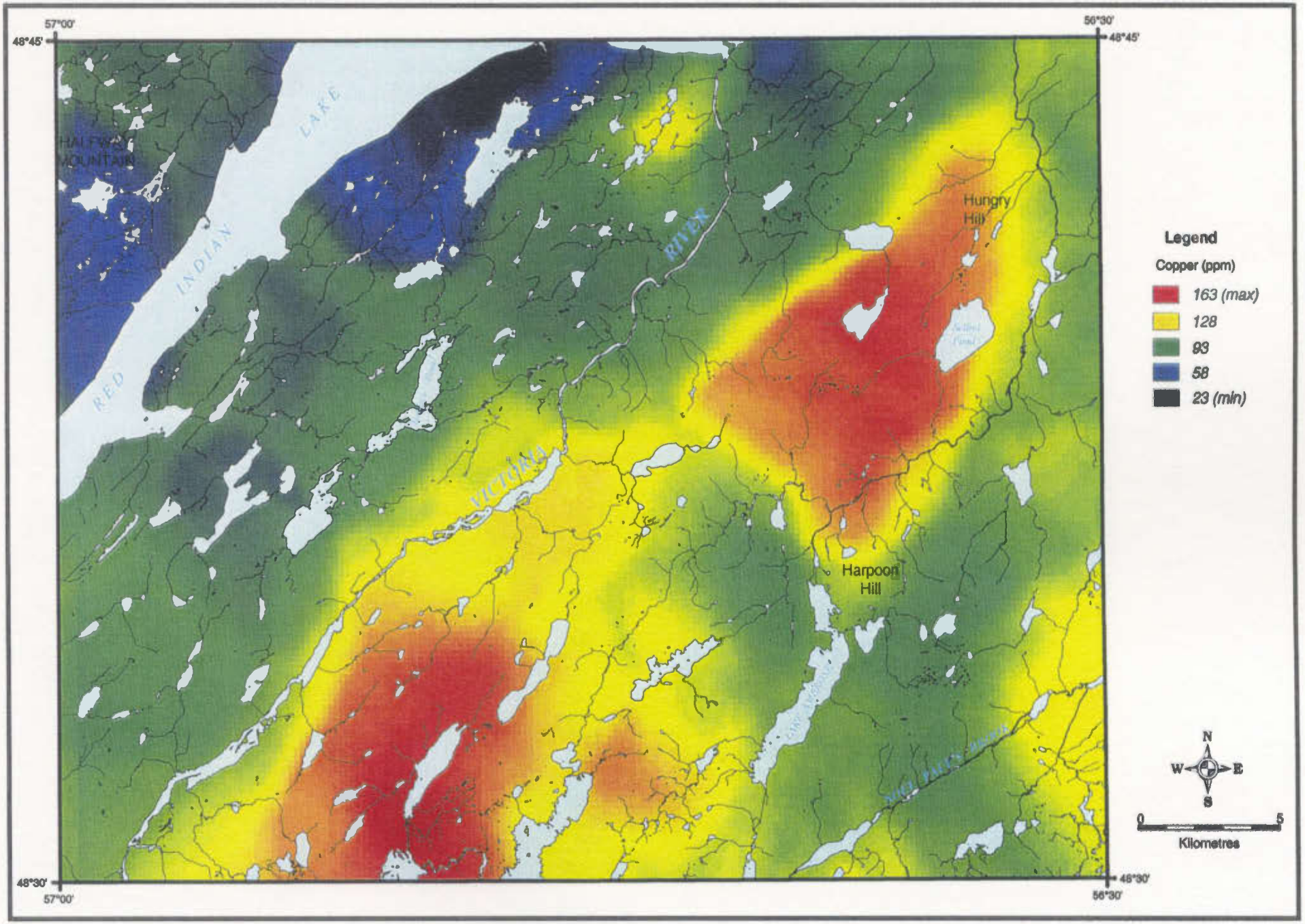

Figure 5.11 : Interpolated surface of copper from tills. Values range from 23ppm to 163 ppm.

the elements Eu, F, Se, Ta, and W all had more than 50% of their values less than the detection limit, and Ag, Cr, Cs, Hf, and Rb all had more than 40% of their values less than the detection limit. These ten elements were removed from the database, leaving 22 elements (*i.e.* As, Au, Ba, Br, Ce, Co, Cu, Fe, La, Mn, Mo, Na, Ni, Pb, Sb, Sc, Sm, Tb, Th, U, Yb, Zn) plus loss-on-ignition (LOI-pct), sample depth (sampdpth), and lake area (lakarea).

The frequency table for LOI indicated 5 samples with LOI values greater than 95%; the next highest value was 79%. These 5 samples had less than 5% sediment (*i.e.* silt plus clay) and so their trace element composition was suspect. These 5 samples were deleted from the database. This reduced the number of lake sediment samples to 194.

Univariate statistics were tabulated as a summary overview of the 22 elements (Table 5.6). Histograms of the element values showed a positively skewed distribution. Therefore, all values were logged which tended to normalize the distributions and stabilize the variance (Davis, 1986; George and Bonham-Carter, 1989).

A visual assessment of the spatial distribution of lake sediment geochemistry can reveal spatial associations. Element values were mapped onto the catchment basins and plots (*e.g.* Figure 5.12) were examined to identify any spatial associations. Values near the 95th, 85th, 70th, and 50th percentiles were used as quick break point values. Unlike the till data, the spatial distributions of the element values in the catchment basins did not show any obvious visual associations with factors such as faults or rock type.

Spearman's rank correlations between selected lake sediment elements (Table 5.7) may indicate associations which provide information on the nature of the geochemical

Table 5.6 : Univariate statistics for elements in lake sediment samples in the NTS 12A/10 area.

							Standard	Logarithmic	Log Standard	Geometric
Element	Unit	N	Minimum	Maximum	Median	Mean	Deviation	Mean	Deviation	Mean
As1	ppm	193	1.2	375	20	36.8	49.81	1.30	0.488	20.1
Au1	ppb	193	<2	129	<2	2.35	9.267	0.152	0.282	1.42
Ba1	ppm	193	<50	2080	72	125	193.0	1.88	0.409	75.4
Br1	ppm	193	2.6	85.4	26	28.7	14.98	1.39	0.270	24.5
Ce1	ppm	193	<2	273	18	24.1	26.07	1.17	0.519	14.7
Co3	ppm	194	<2	102	7.0	9.72	11.40	0.804	0.398	6.37
Cu3	ppm	194	3	80	24	27.7	14.29	1.39	0.221	24.4
Fe3	pct	194	0.13	15.6	1.3	1.98	2.123	0.119	0.394	1.32
La1	ppm	193	<2	81	10	12.4	8.320	1.02	0.258	10.5
Mn3	ppm	194	41	96700	339	1753	7863	2.62	0.559	417
Mo5	ppm	194	<2	22	3.0	4.11	3.015	0.516	0.297	3.28
Na1	pct	193	<0.05	2.36	0.14	0.322	0.4405	-0.775	0.475	0.168
Ni3	ppm	194	<2	66	15	17.2	10.57	1.16	0.258	14.6
Pb3	ppm	194	<2	343	5.0	8.19	24.78	0.687	0.379	4.86
Sb1	ppm	193	<0.05	4.3	0.34	0.471	0.5052	-0.494	0.391	0.320
Sc1	ppm	193	0.5	20	5.1	6.03	3.749	0.704	0.261	5.06
Sm1	ppm	193	0.2	16.8	2.9	3.28	1.887	0.450	0.252	2.82
Tb1	ppm	193	<0.5	2.8	0.55	0.615	0.3438	-0.271	0.230	0.535
Th1	ppm	193	<0.2	17.5	1.5	2.07	1.847	0.208	0.297	1.61
U8	ppm	194	<0.2	16.2	2.3	2.87	2.129	0.364	0.294	2.31
Yb1	ppm	193	<0.5	4.4	1.2	1.36	0.9938	-0.0311	0.428	0.931
Zn3	ppm	194	4	478	96	112	76.37	1.95	0.311	89.4

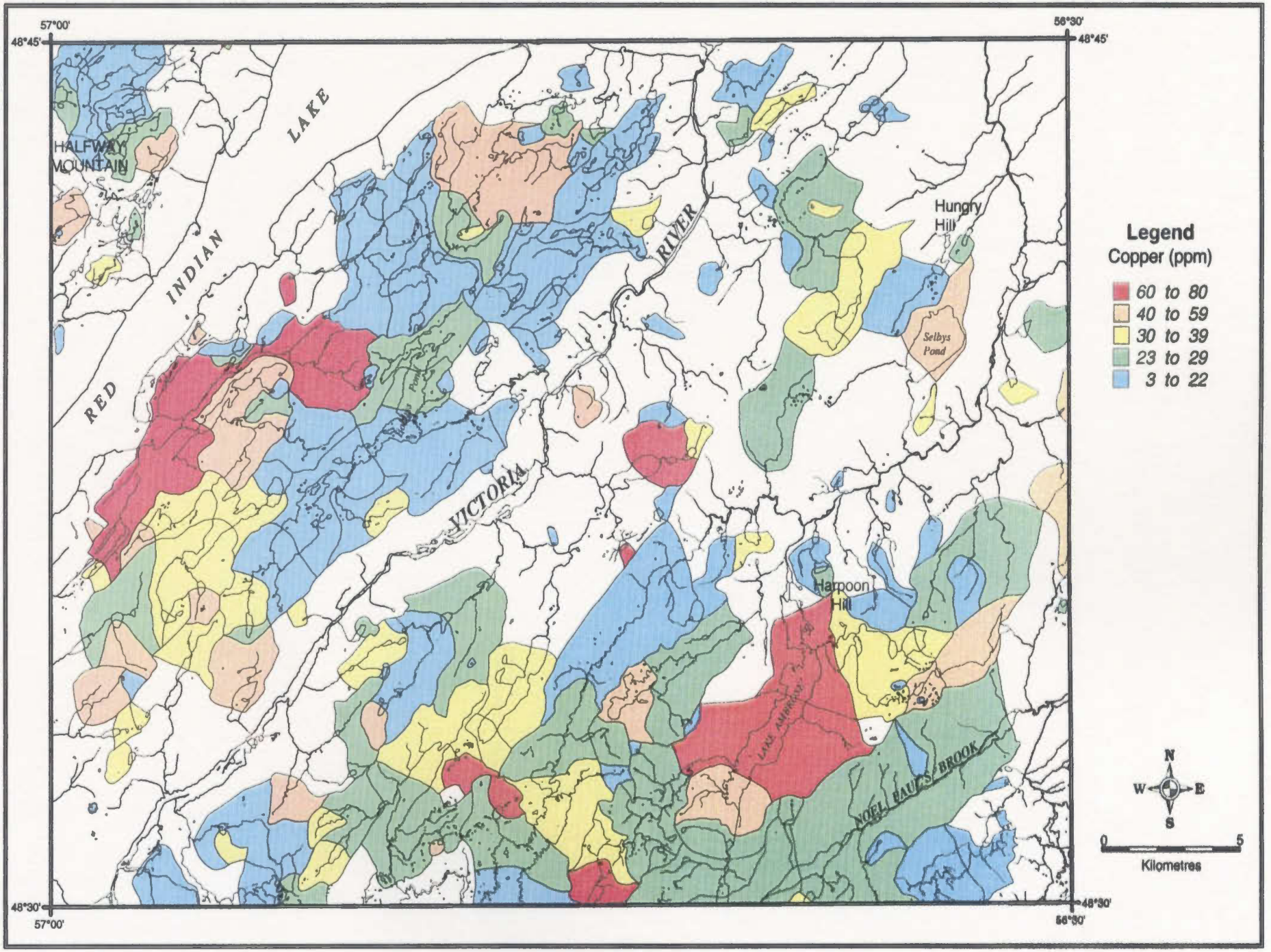

Figure 5.12 : Copper distribution (ppm) in lake sediment catchment basins.

affinities. The highest correlations are between logPb and logSb (r=0.58) and between logZn and logAs (r=0.54). These associations, between base metals (logPb and logZn) and the gold pathfinder elements (logSb and logAs), may indicate the presence of gold in the vicinity of the base metals.

Table 5.7 : Spearman's rank correlation coefficients for selected elements from lake sediment samples.

		LOGAS	LOGAU	LOGCU	LOGPB	LOGSB	LOGZN
LOGAS	Correlation Coefficient	1.000					
	Sig. (2-tailed)	.					
	N	193					
LOGAU	Correlation Coefficient	.218	1.000				
	Sig. (2-tailed)	.002	.				
	N	193	193				
LOGCU	Correlation Coefficient	.227	.227	1.000			
	Sig. (2-tailed)	.002	.001	.			
	N	193	193	194			
LOGPB	Correlation Coefficient	.309	.128	.225	1.000		
	Sig. (2-tailed)	.000	.076	.002	.		
	N	193	193	194	194		
LOGSB	Correlation Coefficient	.498	.286	.361	.577	1.000	
	Sig. (2-tailed)	.000	.000	.000	.000	.	
	N	193	193	193	193	193	
LOGZN	Correlation Coefficient	.541	.039	.397	.454	.464	1.000
	Sig. (2-tailed)	.000	.587	.000	.000	.000	.
	N	193	193	194	194	193	194

Linear regression analysis was performed on the lake sediment database to reduce false anomalies caused by lake effects (see Chapter 2.1.2). Prior to linear regression, it is necessary to remove univariate and multivariate outliers. Univariate outliers were determined by reviewing standardized Z values of all the logged elements. Standardized values greater than an absolute value of 3.29 (for $\alpha = 0.001$) were considered univariate

outliers. Six univariate outliers were temporarily removed from the database prior to linear regression analysis. Multivariate outliers were checked among the independent variables (IVs) in the analysis (*i.e.* logFe, logMn, LOI_pct, loglarea and logsdep; Tabachnick and Fidell, 1996). Multivariate outliers are determined by checking the Mahalanobis distance (MD), available through the SPSS™ linear regression procedure. The critical chi-square value (for $\alpha = 0.001$ and 5 degrees of freedom) is 20.515. Therefore, a MD greater than 20.5 indicates a multivariate outlier case. There were 3 cases greater than 20.5, but the maximum MD was 27.5, which is not much greater than the critical value. Therefore, no cases were removed due to multivariate outliers. The linear regression analysis was run on 187 samples.

Linear regression analysis was performed on those elements where the correlation between the element and lake effects (logFe, logMn, LOI_pct, loglarea or logsdpth) was greater than 0.5 (Davenport *et al.*, 1974; Davenport and Nolan, 1991). All elements except logAu, logCu, logNi, and logPb had a correlation coefficient greater than 0.5 with, at least, some IVs. Stepwise linear regression was run for each element using only the correlated IVs. The computed residuals were saved to the lake sediment database. The residuals were examined and 3 outlier cases (not the same 3 cases as the multivariate outliers) were removed and the linear regression analysis was repeated. Eight elements (logAs, logBr, logCo, logNa, logSc, logTh, logYb and logZn) had significant enough correlations (*i.e.* were strongly influenced by lake effects; see methodology in Chapter 4.1.2.1) to justify calculation of their residual values for use in subsequent analyses. The five lake effects accounted for 41% to 72% of the variance of these 8 elements. The

residual values for the outliers were calculated manually using the final linear regression equations (P. Davenport, pers. comm., 2001).

A comparison of the correlations based on the residual arsenic and zinc values (Table 5.8) as opposed to their log values (Table 5.7) indicated a considerable decrease. For example, the correlation between logZn and logAs was 0.54, which was reduced to 0.11 for rZn and rAs. This indicates the significance of lake effects on element values and how the removal of the lake effects may affect statistical analyses by reducing correlations.

Low correlations among the 22 lake sediment elements, compared to the till elements, may indicate that PCA will not be as effective. Therefore, the KMO measure of sampling adequacy was calculated to determine if the lake sediment data was factorable. The KMO measure of sampling adequacy was 0.810. KMO values greater than 0.6 indicate that the overall database is factorable (Tabachnick and Fidell, 1996).

To reduce the effects of outliers on the PCA, multivariate outliers were determined by calculating the MD. For 22 elements the critical chi-square value (for $\alpha = 0.001$) is 48.3. Therefore, any MD greater than 48.3 indicates a multivariate outlier case. There were 7 cases much greater than 48.3. These cases were temporarily removed and the PCA calculated.

After the outliers were removed the number of components to be extracted in the PCA were determined using eigenvalues, the scree plot and the change in component variance. The number of eigenvalues greater than 1.0 was 7, but the scree plot (Figure 5.13) indicates a change in slope at 4 to 6 components. The percent of variance explained by the varimax rotation had the largest drop from 10.4% to 7.1% from the 4^{th} to the 5^{th}

Table 5.8 : Spearman's rank correlation coefficients for selected lake sediment elements against residual As and Zn. Note the reduced correlations compared to values in Table 5.7.

		RAS	RZN
RAS	Correlation Coefficient	1.000	.111
	Sig. (2-tailed)	.	.124
	N	193	193
LOGAU	Correlation Coefficient	.125	-.061
	Sig. (2-tailed)	.082	.397
	N	193	193
LOGCU	Correlation Coefficient	.033	.163
	Sig. (2-tailed)	.654	.023
	N	193	194
LOGPB	Correlation Coefficient	.231	.268
	Sig. (2-tailed)	.001	.000
	N	193	194
LOGSB	Correlation Coefficient	.198	.134
	Sig. (2-tailed)	.006	.062
	N	193	193
RZN	Correlation Coefficient	.111	1.000
	Sig. (2-tailed)	.124	.
	N	193	194

components, respectively. Therefore, 5 components were extracted on the next analysis.

The communality for rCo was the lowest in the list (0.25), indicating that the factors did not account for much of the variance in rCo, so rCo was deleted. The fifth component had low loadings (< |0.6|) and the variables contributing to the fifth component were not geochemically consistent. Therefore, 4 components were extracted on the next analysis. The fourth component (LPC4) contained low loadings (nothing greater than |0.6|) but the variables were geochemically consistent and significant to this study *(i.e.* gold and its pathfinder elements rAs and logSb). The component scores were saved to the lake sediment database. The component scores were reviewed and 1 score was an outlier to the solution. Since this outlier was not in LPC2 or LPC4 (the two components of interest in this study) the final PCA was acceptable. The PCA results are listed in Table

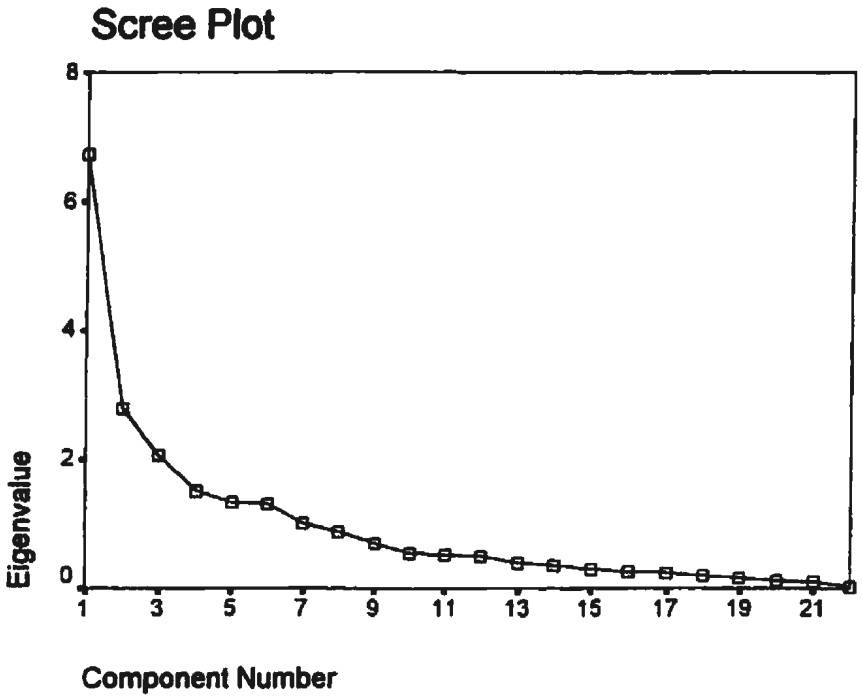

Figure 5.13 : Scree plot for lake sediment PCA indicating the decrease in eigenvalues with increasing number of components. Note the change in slope around 5 components.

5.9. The scores for the 7 multivariate outliers were calculated using the component score coefficient matrix and input back into the database (P. Davenport, pers. comm., 2001).

Each of the first three components explains more than 10% of the variance. Component 4 (LPC4) only explains 9% of the variance but is composed of logAu and it pathfinder elements, logAs and logSb. Component 2 (LPC2) is characterized by the base metals (*e.g.* logNi, logCu, logMn and logMo). Overall, the four components explain 61.9% of the variance. A summary of the elements characterizing each component and their geochemical affiliation is provided in Table 5.10.

Table 5.9 : Principal components analysis results for lake sediment samples in the NTS 12A/10 area, with varimax rotated component matrix. Only loadings >0.3 are shown. Bold values indicate component element affinities.

	Rotated Component Matrix				
Element	Component				Communality
	1	2	3	4	
LogLa	**0.906**				0.918
LogSm	**0.900**				0.898
LogTb	**0.870**				0.827
LogCe	**0.785**				0.632
rTh	**0.584**		0.320	0.386	0.646
LogU	**0.561**	0.336		0.365	0.561
rYb	0.475		0.425		0.514
LogFe	0.511	**0.725**			0.804
LogNi		**0.702**			0.524
LogCu		**0.680**			0.538
LogMn	0.475	**0.672**			0.703
LogMo		**0.648**			0.459
rSc			**0.852**		0.760
rNa			**0.851**		0.780
rZn			**0.603**	-0.378	0.622
LogBa	0.371	0.313	**0.491**		0.518
LogPb	0.305		**0.430**		0.404
LogAu				**0.590**	0.394
rBr				-0.585	0.478
rAs				**0.484**	0.300
LogSb	0.415	0.425	0.365	**0.474**	0.712
% of variance	23.6	15.9	13.4	8.9	61.9

Table 5.10 : A summary of the geochemical affinity of the four lake sediment principal components.

Component	Elements	Geochemical Affinity
LPC1	La, Sm, Tb, Ce, Th, U	Rare Earth Elements
LPC2	Ni, Cu, Mn, Mo	Base Metals
LPC3	Sc, Na, Zn, Ba, Pb	Mafic Volcanics
LPC4	Au, As, Sb	Gold Association

As a test on the effects of the outliers, the PCA results were compared before and after the outliers were removed. Unlike the results from the till PCA, the lake sediment PCA results with and without the outliers were significantly different, indicating the lake sediments PCA is not robust to outliers. This may be because the till outliers were only moderately greater than the MD whereas the lake sediment outliers were much greater than the MD critical value.

5.2.3 Geology

The VMS deposit model indicates that submarine volcanic rocks (*e.g.* felsic and mafic marine volcanics) are the host rocks for the Cu-Pb-Zn mineralization. Table 5.11 presents a summary of the rock types (from the 1:50,000 scale geology map) occurring at the 96 mineral and nonmineral occurrence sites. The largest frequency of occurrence (21 of 96) is of the rock type 'volcanic, felsic marine' and occurs predominantly at the mineral occurrence sites. Marine siliciclastics (an undifferentiated group of marine clastic rocks) are the most common rock type at the nonmineral occurrence sites (12 of 96). Marine sandstone and marine siltstone can also be included in the marine siliciclastic group, occurring at 6 nonmineral occurrence sites.

5.2.4 Fault Proximity and Weights of Evidence Modelling

Both the VMS and gold deposit models indicate that the proximity to faults is a positive factor in the exploration for these types of mineral occurrences. Therefore, the distance from faults may be an important predictor in the mineral potential models.

Table 5.11 : Frequency of 1:50,000 scale rock types at mineral occurrence and nonmineral occurrence sites.

Rock Type (Geolcode)	Frequency at Mineral Occurrences	Frequency at Nonmineral Occurrences
Volcanic, felsic marine (10)	21	6
Volcanic, mafic marine (11)	7	9
Volcaniclastic (12)	8	7
Plutonic, felsic (1)		1
Plutonic, intermediate (2)	2	1
Plutonic, mafic (3)	1	3
Siliciclastic, marine (5)	6	12
Siliciclastic, marine sandstone (6)		4
Siliciclastic, marine siltstone (7)		2
Siliciclastic, non-marine (8)		2
Siliciclastic, non-marine conglomerate (9)		2
Siliciclastic, black shale (4)	2	
Total	47	49

Fault proximity values in the study area range from 0 m to over 8000 m with a very positively skewed distribution. Mineral occurrence distances from the faults range from 0 m to 5622 m whereas nonmineral occurrences range from 0 m to 8109 m.

The normalized fault distance (NFLTDST) was calculated from FLTDST – the distance in metres from the faults. The NFLTDST was very positively skewed. Therefore, the arcsine transform was used to convert it to a more normal distribution in the variable ARCSFLT. ARCSFLT values close to 0 represent areas far from faults and maximum ARCSFLT values are close to faults.

The binary representation of fault distance was calculated by the weights of evidence modelling method (Bonham-Carter, 1994) which is used to determine the optimal threshold or cutoff distance (see Chapter 4.1.2.3.1). The weights of evidence calculations are based on a raster set of fault proximity buffers from 200 m to 2400 m

away from the faults. Most of the mineral occurrences occur within the 2400 m buffer (31 of 47) so it was unnecessary to increase the buffer size beyond 2400 m. The calculations for the positive and negative weights of evidence, along with their variance, are presented in Tables 5.12 and 5.13. Note that all the W^+ are positive and all the W^- are negative, indicating a positive correlation between the fault buffers and the mineral occurrences.

Table 5.12 : Weights of evidence calculations for W^+ (based on Bonham-Carter, 1994).

Buffer	N(B|D)	N(D)	P(B|D)	$N(B\|\overline{D})$	$N(\overline{D})$	$P(B\|\overline{D})$	$P(B\|D) / P(B\|\overline{D})$	W^+	$s^2(W^+)$
0-201	7	47	0.1489	1598	25853	0.0618	2.4095	0.8794	0.1435
0-401	14	47	0.2979	2607	25853	0.1008	2.9539	1.0831	0.0718
0-601	16	47	0.3404	3902	25853	0.1509	2.2555	0.8134	0.0628
0-801	19	47	0.4043	4843	25853	0.1873	2.1580	0.7692	0.0528
0-1001	24	47	0.5106	6248	25853	0.2417	2.1129	0.7481	0.0418
0-1201	26	47	0.5532	7143	25853	0.2763	2.0022	0.6942	0.0386
0-1401	26	47	0.5532	8016	25853	0.3101	1.7841	0.5789	0.0386
0-1601	28	47	0.5957	9144	25853	0.3537	1.6844	0.5214	0.0358
0-1801	29	47	0.6170	10069	25853	0.3895	1.5843	0.4601	0.0346
0-2001	29	47	0.6170	11003	25853	0.4256	1.4498	0.3714	0.0346
0-2201	30	47	0.6383	11650	25853	0.4506	1.4165	0.3482	0.0334
0-2401	31	47	0.6596	12278	25853	0.4749	1.3888	0.3285	0.0323

Note: N(B|D) = number of fault buffer cells intersecting mineral occurrences, N(D) = number of mineral occurrences, P(B|D) = probability of a fault buffer cell given a mineral occurrence = N(B|D)/N(D), $N(B|\overline{D})$ = number of fault buffer cells intersecting nonmineral occurrence, $N(\overline{D})$ = number of nonmineral occurrences, $P(B|\overline{D})$ = probability of a fault buffer cell given a nonmineral occurrence = $N(B|\overline{D}) / N(\overline{D})$, $W^+ = \ln[P(B|D)/P(B|\overline{D})]$, $s^2(W^+) = 1/N(B|D) + 1/N(B|\overline{D})$.

Table 5.14 contains the contrast calculations and the Studentized C values (SC). The SC values were all greater than 2.0 with two-thirds being greater than 3.0, indicating good reliability in the contrast (Bonham-Carter, 1994). The maximum contrast occurs at

Table 5.13 : Weights of evidence calculations for W (based on Bonham-Carter, 1994).

Buffer	$N(\bar{B}\mid D)$	N(D)	$P(\bar{B}\mid D)$	$N(\bar{B}\mid\bar{D})$	$N(\bar{D})$	$P(\bar{B}\mid\bar{D})$	$P(\bar{B}\mid D) / P(\bar{B}\mid\bar{D})$	W	$s^2(W)$
0-201	40	47	0.8511	24255	25853	0.9382	0.9071	-0.0975	0.0250
0-401	33	47	0.7021	23246	25853	0.8992	0.7809	-0.2473	0.0303
0-601	31	47	0.6596	21951	25853	0.8491	0.7768	-0.2525	0.0323
0-801	28	47	0.5957	21010	25853	0.8127	0.7331	-0.3105	0.0358
0-1001	23	47	0.4894	19605	25853	0.7583	0.6453	-0.4380	0.0435
0-1201	21	47	0.4468	18710	25853	0.7237	0.6174	-0.4823	0.0477
0-1401	21	47	0.4468	17837	25853	0.6899	0.6476	-0.4345	0.0477
0-1601	19	47	0.4043	16709	25853	0.6463	0.6255	-0.4692	0.0527
0-1801	18	47	0.3830	15784	25853	0.6105	0.6273	-0.4663	0.0556
0-2001	18	47	0.3830	14850	25853	0.5744	0.6667	-0.4053	0.0556
0-2201	17	47	0.3617	14203	25853	0.5494	0.6584	-0.4180	0.0589
0-2401	16	47	0.3404	13575	25853	0.5251	0.6483	-0.4334	0.0626

Note: $N(\bar{B}\mid D)$ = number of nonfault buffer cells intersecting mineral occurrences, N(D) = number of mineral occurrences, $P(\bar{B}\mid D)$ = probability of a nonfault buffer cell given a mineral occurrence = $N(\bar{B}\mid D)/N(D)$, $N(\bar{B}\mid\bar{D})$ = number of nonfault buffer cells intersecting nonmineral occurrence, $N(\bar{D})$ = number of nonmineral occurrences, $P(\bar{B}\mid\bar{D})$ = probability of a nonfault buffer cell given a nonmineral occurrence = $N(\bar{B}\mid\bar{D})/N(\bar{D})$, $W = \ln[P(\bar{B}\mid D)/P(\bar{B}\mid\bar{D})]$, $s^2(W) = 1/N(\bar{B}\mid D) + 1/N(\bar{B}\mid\bar{D})$.

a buffer distance of 400 m (Figure 5.14), which also had the highest SC value of 4.16. A secondary maximum occurs at 1000 m. From this information, two new variables were added to the attribute table. The variable FLT400 was coded '1' for FLTDST values ranging from 0 to 400 m and coded '0' for FLTDST values greater than 400 m. A second variable, FLT1000, was coded '1' for FLTDST values ranging from 0 to 1000 m and coded '0' for FLTDST values greater than 1000 m.

Table 5.14 : Contrast and Studentized C calculations (based on Bonham-Carter, 1994).

Contrast	Buffer	$s^2(C)$	s(C)	SC=C/s(C)
0.9769	0-201	0.1685	0.4105	2.3797
1.3305	**0-401**	0.1022	0.3196	**4.1627**
1.0659	0-601	0.0951	0.3083	3.4572
1.0797	0-801	0.0886	0.2977	3.6273
1.1861	**0-1001**	0.0854	0.2922	**4.0597**
1.1765	0-1201	0.0863	0.2937	4.0055
1.0134	0-1401	0.0863	0.2937	3.4505
0.9906	0-1601	0.0885	0.2975	3.3296
0.9265	0-1801	0.0902	0.3003	3.0848
0.7768	0-2001	0.0902	0.3003	2.5864
0.7661	0-2201	0.0923	0.3038	2.5216
0.7618	0-2401	0.0949	0.3081	2.4728

Note: Contrast = $W^+ - W^-$, $s^2(C) = s^2(W^+)+s^2(W^-)$.

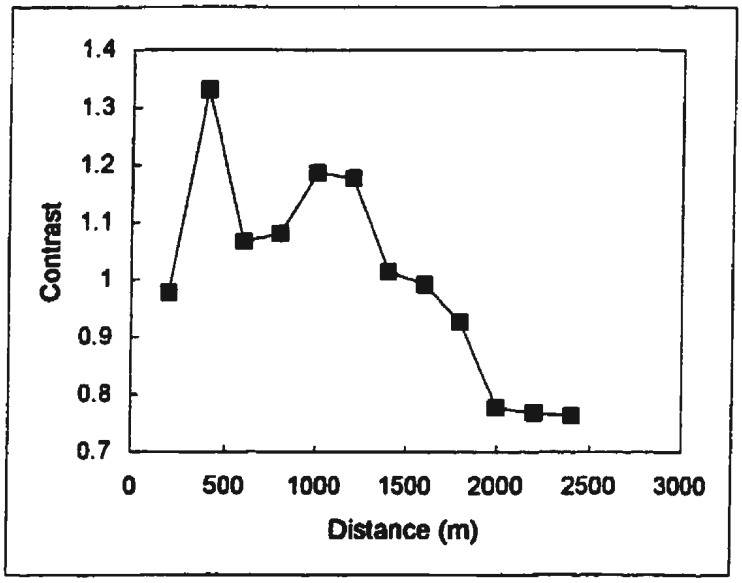

Figure 5.14 : Contrast plot indicating two optimal distances (400 m and 1000 m) to convert fault proximity from a continuous distribution to a binary distribution (based on Bonham-Carter, 1994).

The binary FLT400 and FLT1000 variables will be compared with the continuous fault proximity measure ARCSFLT in the DTA and LRA modelling to determine which variable better predicts mineral occurrences.

5.2.5 Surficial Geology

Table 5.15 presents a summary of the surficial geology units occurring at the 96 mineral/nonmineral occurrence sites. The highest frequency of sites (34 of 96 sites) occurs over drift/rock, which has an average sediment thickness of 1 m. The average thickness of surficial sediments, as estimated by Klassen (unpublished map), was compared with the values of logAu, logCu, logPb, logZn and TPC2 from the till samples. The maximum Spearman's rank correlation coefficient was between the sediment thickness and logCu (r=0.302). The box-and-whisker plots (Figure 5.15) indicate a slight increase in copper values with increasing thickness of sediment, but the reasons for this are beyond the scope of this study.

Table 5.15 : Surficial geology descriptors, codes, frequency, and mean thickness.

Description	Code	Frequency	Mean thickness (metres)
Bedrock	1	18	0
Drift/rock	2	34	1
Till	3	12	3
Till/gravel	4	17	3.5
Drift	5	7	6
Outwash	6	2	-
Alluvium	7	1	-
Organic material	8	3	-
Unlabelled	9	2	-

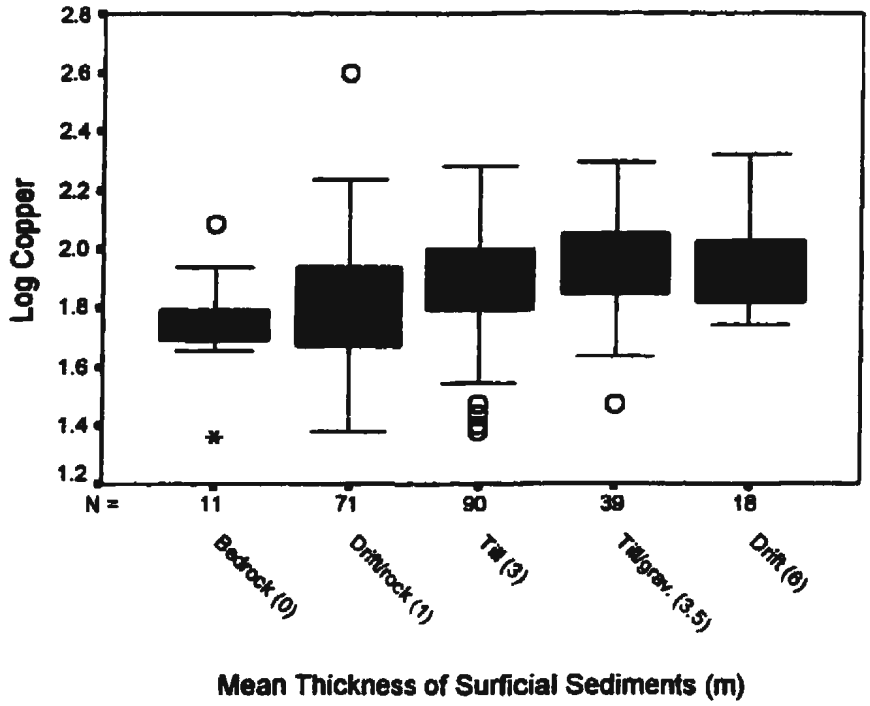

Figure 5.15 : Relationship between copper values (log ppm) and thickness of surficial sediments: bedrock (0.0 m mean thickness of sediment), drift/rock (1.0 m mean thickness), till (3.0 m mean thickness), till/gravel (3.5 m mean thickness) and drift (6.0 m mean thickness).

5.2.6 Wetlands

To determine if there is a relationship between wetlands and the geochemistry of till samples, the wetland type (*i.e.* wetland, stringbog or dryland) was extracted at the 96 mineral/nonmineral occurrence sites. Only 4 sites occurred within wetlands. Box-and-whisker plots of the 4 samples in wetlands compared to the 92 samples on dry land did not indicate any major difference in the values of logAu, logCu, logPb, logZn, and TPC2 from tills. The 95% confidence interval for the medians of the wetland group overlapped the values for the nonwetland group for the five variables. For example, the 95% confidence

interval for the median for logCu in wetands is 1.61 to 1.97, whereas the values are 1.81 to 1.87 for logCu in dryland.

5.2.7 Spatial Weighting Function

The nearest neighbour index (Chapter 5.1) indicated the 96 mineral/nonmineral occurrences are randomly distributed. Prior to developing a spatial weighting function, it was also necessary to show that the 96 occurrences are positively spatially autocorrelated. The spatial autocorrelation was evaluated using the Moran's I statistic for point data (Kvamme, 1990; Cliff and Ord, 1973; see Chapter 4.1.2.6). Sums used in the spatial autocorrelation calculations are presented in Tables 5.16 and 5.17.

Table 5.16 : Calculation of moments (based on Kvamme, 1990). The total number of points is 96 (points 4 to 93 have been removed for presentation purposes), where $x_i = 1$ for the 47 mineral occurrences and $x_i = 0$ for the 49 nonmineral occurrences.

ID	UTMeast	UTMnorth	x	(x-meanx)	$(x-meanx)^2$	$(x-meanx)^4$
1	511150	5387150	1	0.5104	0.2605	0.0679
2	521710	5398300	1	0.5104	0.2605	0.0679
3	518550	5376525	1	0.5104	0.2605	0.0679
:	:	:	:	:	:	:
94	531434	5398226	0	-0.4896	0.2397	0.0575
95	535748	5395629	0	-0.4896	0.2397	0.0575
96	527723	5399466	0	-0.4896	0.2397	0.0575
		sum=	47	0.0000	23.9896	6.0052
		meanx=	0.49			
		b_2 =	1.002			

Note: $^*b_2 = [n\Sigma_i(x_i-mx)^4]/[\Sigma_i(x_i-mx)^2]^2$, where mx represents the mean of x.

Table 5.17 : Weighted calculations between pairs of points (based on Kvamme, 1990). A total of 105 pairs of points were within 2500m of each other (point pairs 8 to 102 have been removed for presentation purposes). Distance between points was measured in metres. The weight (w) is measured as the inverse of the distance.

i	j	distance	w=1/d	w_{ij}^2	x_i	x_j	(x_i-meanx)	(x_j-meanx)	w (x_i-meanx) (x_j-meanx)
1	26	561.81	1.78E-03	3.17E-06	1	1	0.5104	0.5104	4.64E-04
1	33	338.38	2.96E-03	8.73E-06	1	1	0.5104	0.5104	7.70E-04
1	37	1614.75	6.19E-04	3.84E-07	1	1	0.5104	0.5104	1.61E-04
1	40	335.41	2.98E-03	8.89E-06	1	1	0.5104	0.5104	7.77E-04
1	41	2367.64	4.22E-04	1.78E-07	1	1	0.5104	0.5104	1.10E-04
1	42	2166.4	4.62E-04	2.13E-07	1	1	0.5104	0.5104	1.20E-04
2	21	1012.04	9.88E-04	9.76E-07	1	1	0.5104	0.5104	2.57E-04
:	:	:	:	:	:	:	:	:	:
67	68	1026.25	9.74E-04	9.49E-07	0	0	-0.4896	-0.4896	2.34E-04
70	78	2109.56	4.74E-04	2.25E-07	0	0	-0.4896	-0.4896	1.14E-04
75	83	1236.21	8.09E-04	6.54E-07	0	0	-0.4896	-0.4896	1.94E-04
Half matrix sums			9.61E-02	2.11E-04					1.37E-02
Full matrix sums			1.92E-01	4.22E-04					2.75E-02

The spatial autocorrelation measure for Moran's I is:

$$I=n[\Sigma w_{ij}(x_i-mx)(x_j-mx)] / W\Sigma_i(x_i-mx)^2$$

$$= 96*[2.75*10^{-2}]/ 0.1922*23.9896$$

$$=0.5726$$

(where the above values were obtained from Tables 5.16 and 5.17; mx represents the mean of x, $W=\Sigma w_{ij} = 0.1922$).

The expected value of I is:

$$E(I)=-(n-1)^{-1} = -0.0105$$

Using the randomization assumption, the significance test for I is:

$$Var_R(I) = \{[n(P_1+P_2)-b_2(P_3+P_4)] / [(n-1)(n-2)(n-3)W^2]\}-(n-1)^{-2}$$

where $P_1=(n^2-3n+3)S_1 = 7.5397$ ($S_1=2\Sigma w_{ij}^2 = 8.44 \times 10^{-4}$),

$P_2=3W^2-nS_2 = -0.3411$ ($S_2=4\Sigma_i(\Sigma_j w_{ij})^2 = 4.71 \times 10^{-3}$, values from Appendix A: Program A1),

$P_3=(n^2-n)S_1 = 7.6992$, and

$P_4=6W^2-2nS_2 = -0.6823$.

$b_2=n\Sigma_i(x_i-mx)^4 / [\Sigma_i(x_i-mx)^2]^2 = 1.0017$ (values from Table 5.16)

therefore, $Var_R(I) = 0.0222$.

The test for Moran's I using the standard normal deviate, z, is :

$$z=[I-E(I)] / \sqrt{var(I)}$$

$$=[0.5726-(-0.0105)] / \sqrt{0.0222} = 3.91 .$$

The value of the point-pattern Moran's I statistic is 0.573 with an expected value of -0.0105. The variance (under the randomization model) is 0.0222 resulting in a standard normal deviate of 3.91. For a one-tailed test, where $\alpha=0.05$, $Z_c=1.645$. Therefore, the null hypothesis is rejected in favour of the alternative hypothesis that the mineral and nonmineral occurrence points are spatially autocorrelated (*i.e.* neighbouring values tend to be similar). Therefore, the use of a spatial weighting function, representing the potential for mineralization, is warranted as an independent variable.

The spatial weighting function (SWF) calculations were described in Chapter 4.1.2.7. Prior to calculating the SWF, the appropriate search radius (neighbourhood) must be determined. The neighbourhood was determined by plotting a correlogram (Moran's I versus Separation Distance) of the mineral/nonmineral occurrences (Figure

5.16). The correlogram indicates the neighbourhood of positive spatial autocorrelation extends to about 2500 m from the occurrence sites. Therefore, only points within a 2500 m neighbourhood of each site were included in the SWF calculations.

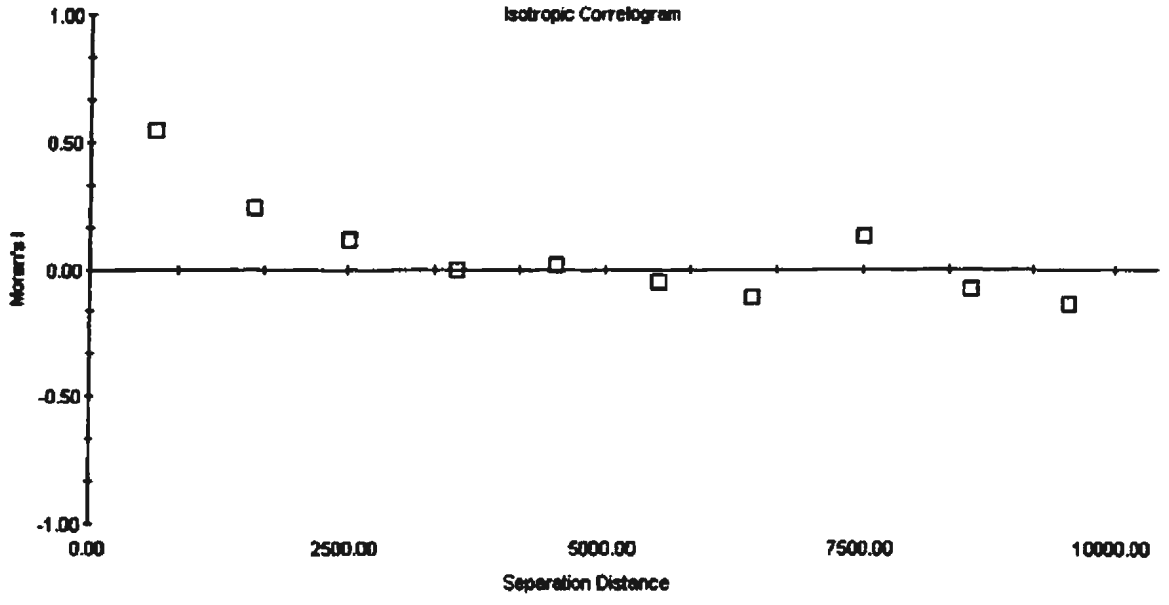

Figure 5.16 : Correlogram of the mineral/nonmineral occurrences, indicating the range (0 to 2500 m) over which spatial autocorrelation is positive.

The equation for the SWF is:

$$SWF_i = \frac{\sum_{j=1}^{n} W_{ij} * x_j}{\sum_{j=1}^{n} W_{ij}}$$

where the weight, W_{ij}, is the inverse distance between points i and j and x_j equals 1 if the jth point is a mineral occurrence and 0 otherwise. The program used to calculate the SWF

is presented in Appendix A: Program A2. The resulting SWF values range from 0 to 1 where 0 indicates the point does not have any mineral occurrence neighbours and 1 indicates all the neighbouring points are mineral occurrences. The median SWF values for mineral and nonmineral occurrences are 0.81 and 0.00, respectively.

5.3 Decision Tree Analysis

DTA was used to classify the 20 predictor variables (consisting of till and lake sediment geochemistry, fault proximity, geology, surficial geology, wetlands and the SWF) based on their relationship to the mineral occurrence response variable, MINOCC.

A preliminary DTA was calculated on the full dataset to determine whether the predictors could explain the response variable significantly more than chance. At this stage the filter level was set to exploration mode (*i.e.* $\alpha = 0.20$) to determine preliminary associations between the variables. The Bonferroni level was set to 3 to reduce spurious groupings by adjusting for correlations within the till, lake sediment, geology and fault predictors (*e.g.* high correlations exist between till logCu, logZn and PC2). The 'grow method' was set to exhaustive and the preliminary decision tree was grown automatically. The resulting accuracy of prediction was 80.2%. This high accuracy indicated that at least some variables predict the mineral/nonmineral occurrences well. Subsequent analyses were used to determine the fewest and best predictors of mineral occurrences.

With the filter level still set to exploration mode ($\alpha = 0.20$) and the grow method set to 'exhaustive' the analysis was repeated more slowly by using the 'find split' method. The most significant split on the dependent variable MINOCC was by the SWF (Figure

5.17) with a significance of 0.001 and an accuracy of 72.9%. This indicates that the neighbourhood effect is the most important predictor. Other significant splits at this level were VOLCFELS, FELSICS, FLT1000, FLT400 and GEOLOGY with significance levels ranging from 0.002 to 0.070, respectively and accuracy ranging from 60% (for FLT400) to 69% (for GEOLOGY). GEOLOGY split the data into 3 groups; with volcanic felsics being the major rock type (27 of 29) of one group. Twenty-three of the 29 cases in this group were mineral occurrences. This validates the importance of the volcanic felsic association with mineral occurrences as indicated in the VMS deposit model (see Chapter 3.1.2). Overall, the SWF was retained as the primary predictor because it seems reasonable to assume that neighbourhood effects are significant in determining mineral potential. Of the 47 mineral occurrences, 39 (83%) are related to the larger SWF values (*i.e.* 0.17 to 1.0), confirming the positive spatial association between mineral occurrences (*i.e.* similarity with their neighbours).

The most significant split off the SWF node [0.17 to 1.0] was by FLT400 (Figure 5.17) with a significance of 0.045. This split did not increase the classification accuracy, which remained at 72.9%. FLT400 split the 57 sites into two groups; 15 sites occur within 400 m of a fault (FLT400=1) and 42 sites occur more than 400 m from a fault (FLT400=0). The FLT400=1 node was further split by the SWF (Figure 5.17). This split characterized 14 mineral occurrences with very high SWF values from 0.64 to 1. This split also classified a single case as near faults but with only moderate SWF values (0.17 to 0.64). The node increased the classification accuracy along this branch from 72.9% to

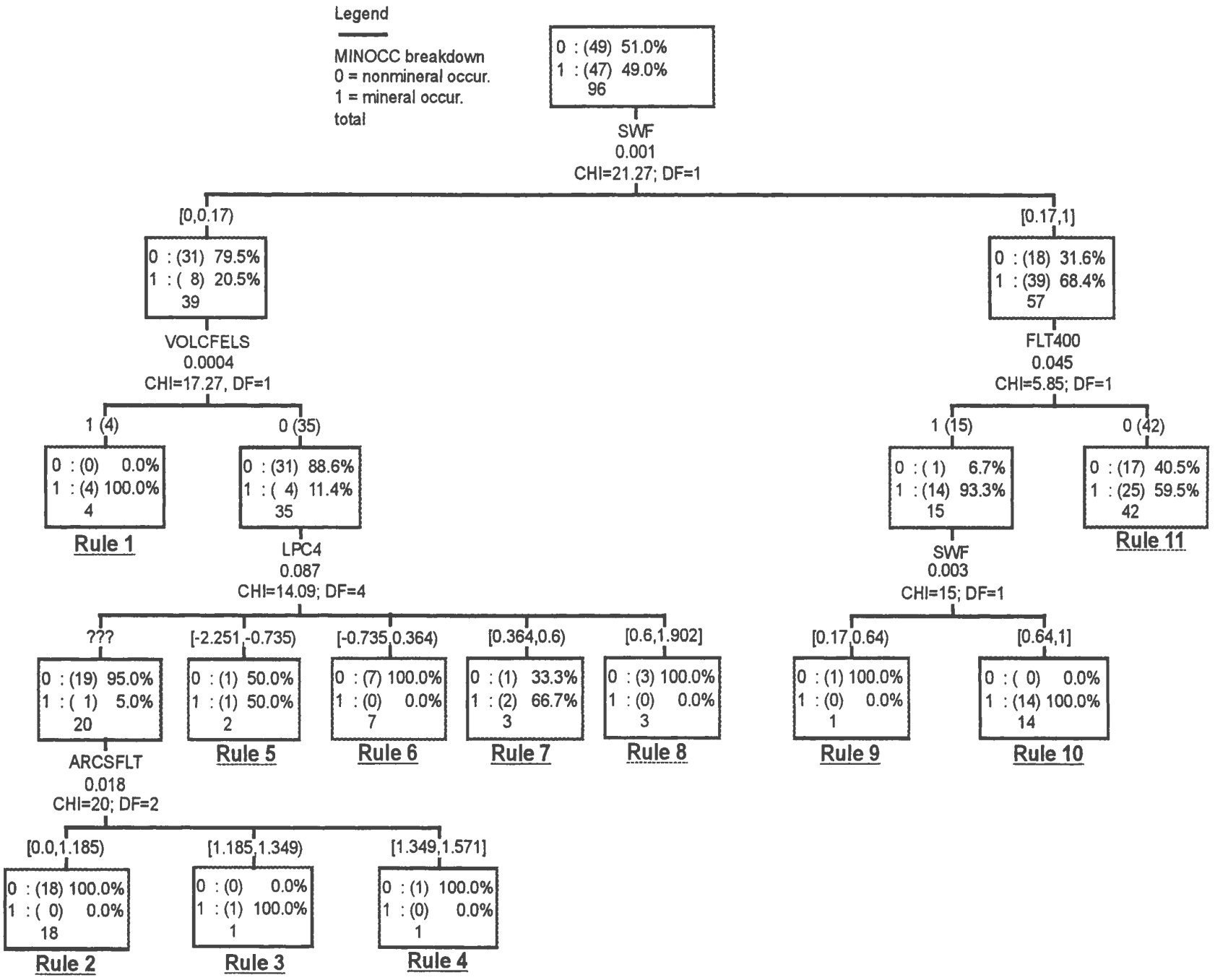

Figure 5.17 : Decision tree analysis results. Eleven rules are defined (see Table 5.18).

74.0%. No further splits were significant along this node at the 'exploration' filter level (α = 0.20).

The most significant split off the SWF node [0,0.17) was by volcanic felsics (Figure 5.17). This split increased the classification accuracy along this node from 72.9% to 77.1%. Alternative splits to VOLCFELS consisted of FELSICS, GEOLOGY and the lake sediment PC variables LPC4 (gold affinity) and LPC2 (base metals affinity). VOLCFELS was retained as the secondary split as there were no other predictors as significant.

The nonvolcanic felsics (*i.e.* VOLCFELS=0) node was split by LPC4 (Figure 5.17), the lake sediment component that represents gold and its pathfinder elements. The majority of these sites (*i.e.* 20 of 35 sites) contain no information (*i.e.* are missing LPC4 values). This node was further split by ARCSFLT (the arcsine transformed fault proximity variable). These last two splits (Figure 5.17) did not provide any interesting information for mineral exploration purposes.

A total of 11 rules were defined by the DTA (Figure 5.17 and Table 5.18) under the 'exploration' filter level (α = 0.20). Rules 2 to 8 do not define factors (or conditions) that are of significance in predicting mineral occurrences. Rules 2 to 8 are also not significant at the 'prediction' filter setting (α = 0.05) and therefore were mapped together as 'SWF=[0,0.17) and VOLCFELS=0'. The final five rules, as summarized in Table 5.19, were mapped by defining the mutually exclusive areas (*i.e.* rules). Note that the five rules are defined by only three predictors: SWF, VOLCFELS and FLT400. VOLCFELS

Table 5.18 : Decision tree results expressed as 'IF-THEN' rules. Refer to Figure 5.17 for the decision tree.

RULE 1: IF SWF = [0,0.17) and VOLCFELS = 1
THEN MINOCC = 1 (probability = 100.0%, n=4)

RULE 2: IF SWF = [0,0.17) and VOLCFELS = 0 and LPC4 = ??? and ARCSFLT = [0.0,1.185)
THEN MINOCC = 0 (probability =100.0%, n=18)

RULE 3: IF SWF = [0,0.17) and VOLCFELS = 0 and LPC4 = ??? and ARCSFLT = [1.185,1.349)
THEN MINOCC = 1 (probability =100.0%, n=1)

RULE 4: IF SWF = [0,0.17) and VOLCFELS = 0 and LPC4 = ??? and ARCSFLT = [1.349,1.571]
THEN MINOCC = 0 (probability =100.0%, n=1)

RULE 5: IF SWF = [0,0.17) and VOLCFELS = 0 and LPC4 = [-2.251,-0.735)
THEN MINOCC = 0 (probability =50.0%, n=1)
MINOCC = 1 (probability =50.0%, n=1)

RULE 6: IF SWF = [0,0.17) and VOLCFELS = 0 and LPC4 = [-0.735,0.364)
THEN MINOCC = 0 (probability =100.0%, n=7)

RULE 7: IF SWF = [0,0.17) and VOLCFELS = 0 and LPC4 = [0.364,0.6)
THEN MINOCC = 0 (probability =33.3%, n=1)
MINOCC = 1 (probability =66.7%, n=2)

RULE 8: IF SWF = [0,0.17) and VOLCFELS = 0 and LPC4 = [0.6,1.902]
THEN MINOCC = 0 (probability =100.0%, n=3)

RULE 9: IF FLT400 = 1 and SWF = [0.17,0.64)
THEN MINOCC = 0 (probability =100.0%, n=1)

RULE 10: IF FLT400 = 1 and SWF = [0.64,1]
THEN MINOCC = 1 (probability =100.0%, n=14)

RULE 11: IF SWF = [0.17,1] and FLT400 = 0
THEN MINOCC = 0 (probability =40.5%, n=17)
MINOCC = 1 (probability =59.5%, n=25)

and FLT400 are represented by binary raster maps. The SWF is a point database of 96 mineral/nonmineral occurrence sites. The SWF values were converted to a continuous surface through the inverse distance squared weighting method. This map was recoded into 3 groups defined by the SWF decision rules (*i.e.* 0 to 0.16, 0.17 to 0.63 and 0.64 to 1.0). The maps representing VOLCFELS, FLT400 and SWF were combined to form the mutually exclusive and exhaustive areas by applying the rule-based conditions in Table 5.19. The probability of mineral occurrences was mapped onto each condition to provide a map of mineral potential (Figure 5.18). For example, the area defined by Rule 1 (SWF<0.17 and VOLCFEL=1) was assigned a probability value of 1.0 (100% probability of mineral occurrences).

Table 5.19 : Probability of mineral occurrences (rounded to the nearest 1%) in relation to significant predictors determined from the DTA. Note that the mapped probabilities are unique conditions.

	Spatial Weighting Function		
	0 to 0.16	0.17 to 0.63	0.64 to 1.0
VOLCFELS=0	P(minocc)=0.11 (Rules 2-8)		
VOLCFELS=1	P(minocc)=1.00 (Rule 1)		
FLT400=0		P(minocc)=0.60 (Rule 11)	P(minocc)=0.60 (Rule 11)
FLT400=1		P(minocc)=0.00 (Rule 9)	P(minocc)=1.00 (Rule 10)

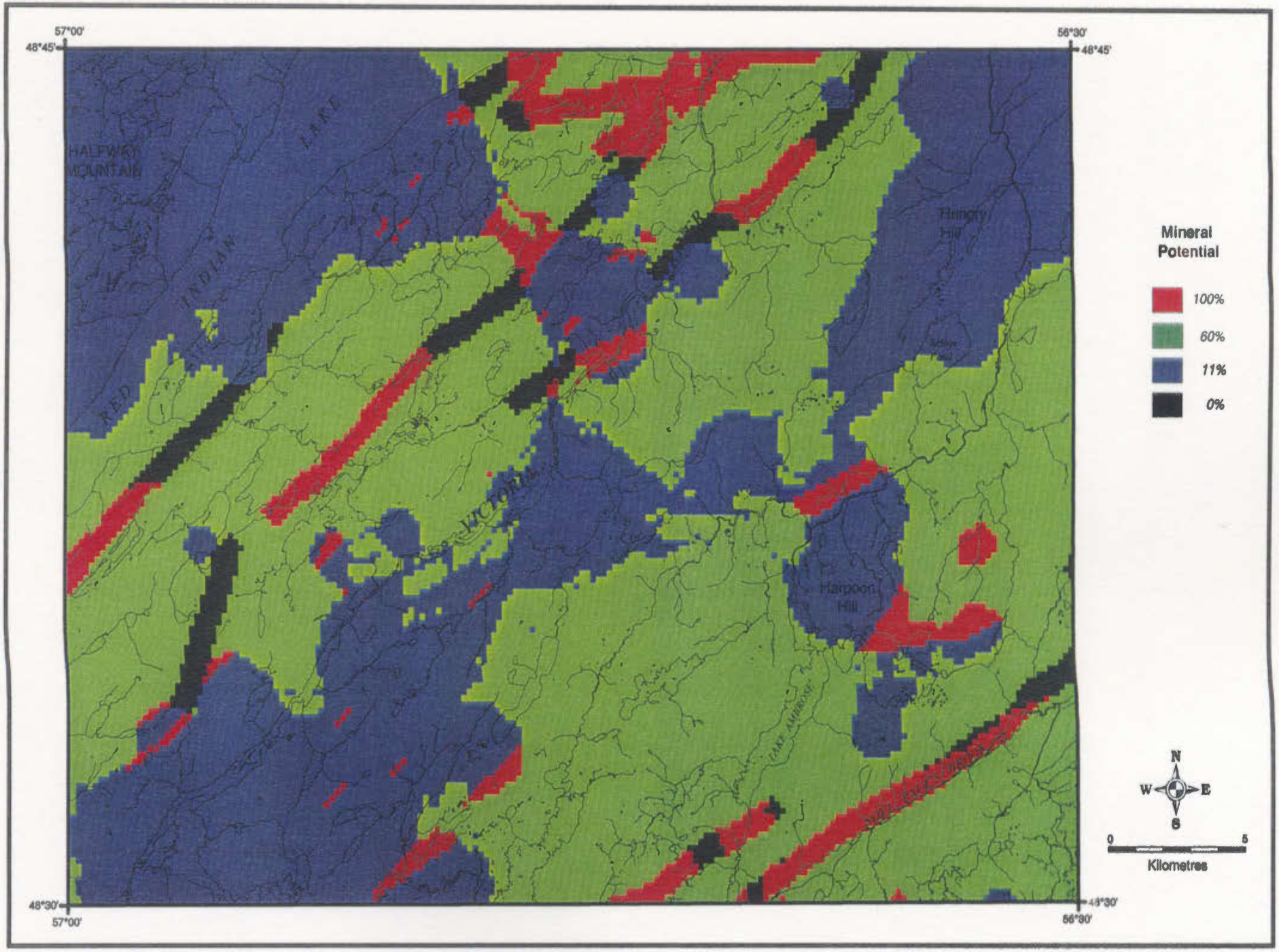

Figure 5.18 : Mineral potential map from the Decision Tree Analysis method. Mineral potential consists of four discrete values from 0% (low mineral potential) to 100% (high mineral potential), using the same colour range as on other figures.

Once the best overall predictors had been determined, DTA was used to determine which predictor, of a set of similar predictors (*e.g.* FLT400, FLT1000 and ARCSFLT), was most significant at directly classifying mineral occurrences. Each predictor was forced to split MINOCC in order to provide a significance and accuracy level to compare with the other predictors in the group. Therefore, the filter setting was specified at α = 1.0 so any significance value, no matter how poor, was displayed. The first comparison was done on the set of fault proximity predictors. FLT1000 had the most significant split off MINOCC with a significance of 0.004 and an accuracy of 64.6%. Aggregating ARCSFLT into 3 groups (by visual inspection) resulted in a split off MINOCC with a significance of 0.010 and an accuracy of 62.5%. FLT400 had the worst significance (0.015) and accuracy (60.4%) compared with the other fault predictors. Overall, FLT1000 provided the best significance and classification accuracy of mineral occurrences directly. This is in contrast to the results with all predictors included in analysis, where the SWF was the primary split and FLT400 was the secondary split.

Comparing the till base metal elements and TPC2 (*i.e.* the base metal principal component), to determine the best classifier of mineral occurrences, indicated that logCu (grouped into 10 classes) provided the best split with a significance of 0.072 and an accuracy of 65.6%. TPC2 did not provide a good split and only had a significance of 0.270. Similarly, for the lake sediment variables (base metal elements, LPC2 and LPC4) residual Zn (grouped into 10 classes) provided the best split off the mineral occurrence node with a significance of 0.070 and an accuracy of 65.6%. LPC4 (grouped into 7 classes) did provide a reasonable split with a significance of 0.191 and an accuracy of

62.5% (which is significant in 'exploration' mode with $\alpha = 0.20$). None of the till or lake sediment geochemical predictors could significantly (*i.e.* $\alpha = 0.05$) classify MINOCC.

A comparison of VOLCFEL with VOLCMAF (volcanic felsics and mafics, respectively) indicated the significance of VOLCFEL in classifying MINOCC. VOLCFEL had a significance of 0.001 and an accuracy of 66.7% whereas VOLCMAF had a very poor significance of 0.653 and an accuracy of only 51% (no increase over chance).

In summary, DTA indicated that the predictors, SWF, VOLCFELS and FLT400, were significant in classifying mineral occurrences. In addition to these three predictors, FLT1000 was significant in classifying MINOCC on an individual basis. The best predictor in the set of till and lake sediment geochemistry was logCu in tills. This predictor was only significant at $\alpha = 0.20$, whereas the other four predictors were significant at $\alpha = 0.05$. The significance of these predictors will be checked through confirmatory analysis using logistic regression.

5.4 Logistic Regression Analysis

To confirm the results of the DTA, LRA was performed on the full set of 20 precictors, using stepwise analysis to determine the best predictors of mineral occurrence.

The first step in LRA is to determine if the full set of predictors improves the prediction of mineral occurrences compared to the constant-only model (Tabachnick and Fidell, 1996). The full model, including 6 lake sediment predictors with missing values,

was not statistically more reliable (χ^2=29.7, df=20, sig=0.075) than the constant-only model, at α=0.05. Due to the missing values, this model only contained 54 cases (*i.e.* 31 mineral occurrences and 21 nonmineral occurrences) and, as a group, did not predict mineral occurrences well. Removing the 6 predictors with missing values and repeating the analysis resulted in a statistically reliable model (χ^2=38.3, df=14, sig=0.000), with a 79.2% correct classification. Therefore, the group of 14 predictors with all 96 cases included, was useful in predicting mineral occurrences.

The stepwise logistic regression procedure was used to determine which predictors were most statistically significant in predicting mineral occurrences. Since the full model (20 predictors, 54 cases) was not more reliable than the constant-only model, it was not analyzed in detail by stepwise logistic regression. A preliminary stepwise analysis did indicate that no lake sediment predictors were significant. Therefore, only the model with 14 predictors and 96 cases was analyzed in detail.

The variables which entered the model on stepwise analysis were SWF, followed by VOLCFELS, FLT1000 and TlogCu, respectively (Table 5.20a). The addition of each variable on subsequent steps changed the parameter values due to their interactions with each other (Table 5.20b). Note the decrease in the exp(B) of the SWF from 7.97, when SWF was the only variable in the model, to 4.22, when the other 3 predictors entered the model.

Table 5.20a : LRA results for the best predictors in the stepwise regression. Parameters are for each IV as it entered the model.

IV	Step Chi-sq.	Sig (Chi)	B	S.E.	Wald	Sig (Wald)	R	Exp(B)	%CC-MO	%CC-NM	%CC-total
SWF	17.84	0.000	2.075	0.523	15.75	0.000	0.322	7.97	74.5	65.3	69.8
VOLCFELS	7.70	0.006	1.464	0.553	7.00	0.008	0.194	4.32	83.0	69.4	76.0
FLT1000	5.15	0.023	1.166	0.521	5.01	0.025	0.150	3.21	74.5	81.6	78.1
TlogCu	2.31	0.129	3.203	2.186	2.15	0.143	0.033	24.61	76.6	81.6	79.2

Note: %CC=percent correctly classified, MO=mineral occurrence, NM=nonmineral occurrence.

Table 5.20b : LRA results for the best predictors in the stepwise regression. Parameters are for the final model, including the constant, with all four predictors entered as a block.

							95% CI for Exp(B)	
IV	B	S.E.	Wald	Sig (Wald)	R	Exp(B)	Lower	Upper
constant	-7.714	4.178	3.41	0.065				
SWF	1.439	0.571	6.35	0.012	0.181	4.22	1.38	12.92
VOLCFELS	1.771	0.594	8.88	0.003	0.227	5.88	1.83	18.83
FLT1000	1.451	0.575	6.37	0.012	0.181	4.27	1.38	13.18
TlogCu	3.203	2.186	2.15	0.143	0.033	24.61	0.34	1786.97

The standardized residuals of the model were checked and only one case had an outlier to the solution (*i.e.* standardized residual = -4.60). Since there was only one outlier to the solution and since all the parameter estimates and standard errors were of a reasonable size it was concluded that the predictors, SWF, VOLCFELS, FLT1000 and TlogCu, provided a good model fit (Tabachnick and Fidell, 1996). A reasonable size for the parameter estimates and standard error also indicates the ratio of cases (96) to variables (14) was sufficient. The final classification indicates that 79.2% of the 96 mineral/nonmineral occurrences were reliably predicted by this model (Table 5.21). Twenty sites (9 mineral occurrences and 11 nonmineral occurrences) were not reliably

predicted. The nonmineral occurrences had a 5% better classification accuracy than the mineral occurrences. The final LRA equation is:

$$\hat{Y} = \frac{e^{u}}{1+e^{u}}$$

where

u = -7.7143 + 1.4393*SWF + 1.7708*VOLCFELS + 1.4511*FLT1000 + 3.2031*TlogCu

This equation was mapped using the IDRISI® IMAGE CALCULATOR. The resulting mineral potential map indicates the percent probability of mineral occurrences (Figure 5.19).

Table 5.21 : Classification table from the logistic regression analysis.

		Predicted		
		Nonmin. Occ.	Mineral Occ.	%Correct
Observed	Nonmin. Occ.	40	9	81.6
	Mineral Occ.	11	36	76.6
			Overall	79.2

A comparative analysis was done for each of the four predictors to determine its individual significance in improving the model (Tabachnick and Fidell, 1996). The models with and without each of the four predictors were compared by using the chi-squared goodness-of-fit test (see Chapter 2). VOLCFELS had the lowest chi-sqaured significance (0.002) when the 3 other variables were considered as a block, followed by FLT1000 (sig = 0.009), SWF (sig = 0.011) and TlogCu (sig = 0.129). These significance values indicate all 4 predictors improved the model fit.

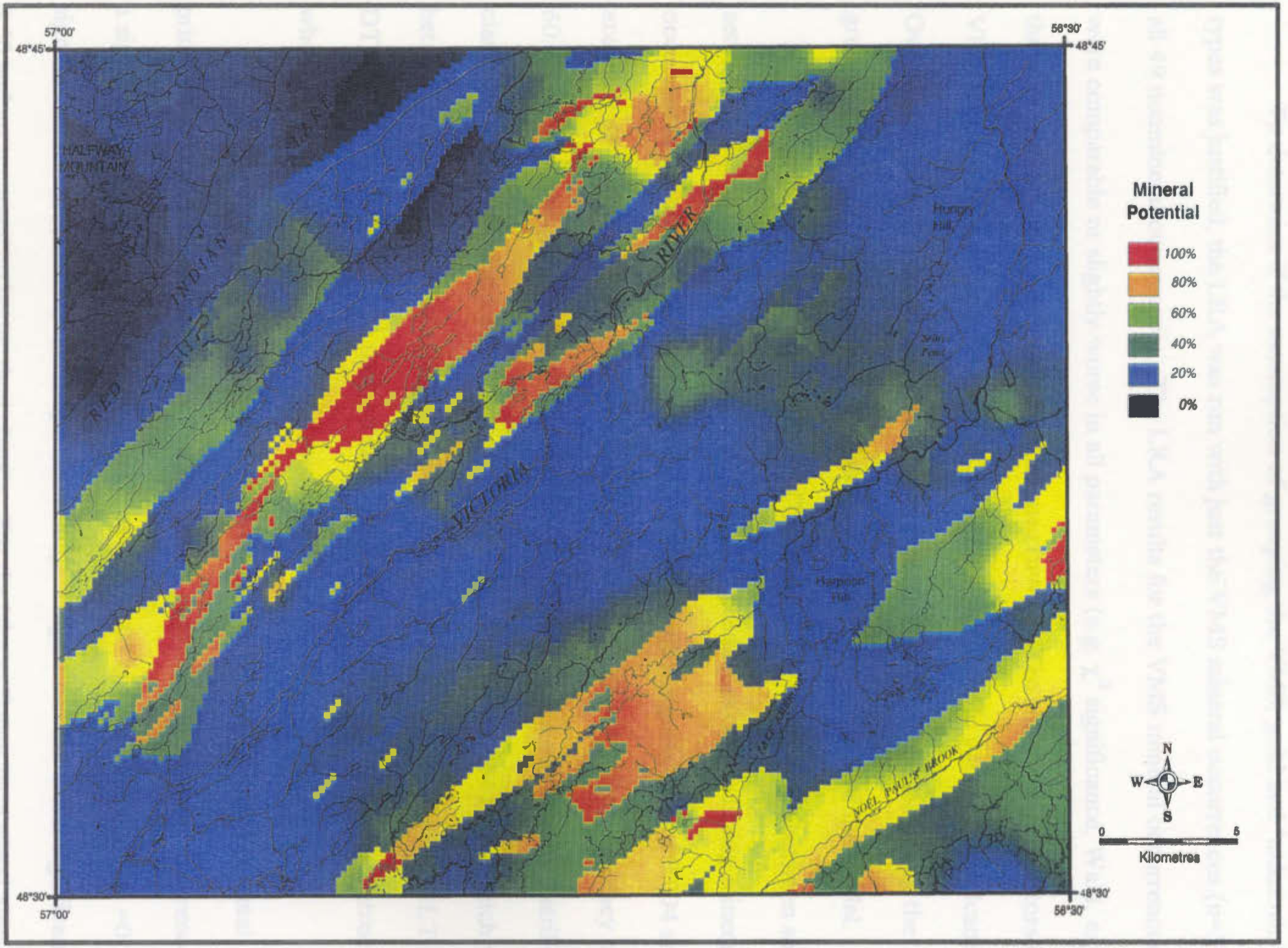

Figure 5.19 : Mineral potential map from the Logistic Regression Analysis method. Mineral potential extends on a continuous scale from 0% (low mineral potential) to 100% (high mineral potential).

To determine if the assumption of grouping the VMS, gold and 'unknown' deposit types was justified, the LRA was run with just the VMS mineral occurrences (n=39) and all 49 nonmineral occurrences. The LRA results for the VMS mineral occurrences (n=39) were comparable or slightly worse in all parameters (*e.g.* χ^2 significance, Wald, exp(B)) than the model with the grouped deposit types (n=47). The significant predictors in the VMS model were the SWF, VOLCFEL and FLT1000; TlogCu was not significant. Overall, the grouped model performed better than the VMS model, indicating the grouping of the deposit types did not detract from the performance of the model.

Similar to the DTA, LRA was used to compare the significance between similar sets of predictors (such as the three fault proximity predictors) in predicting mineral occurrences directly. FLT1000 predicted MINOCC with a significance of 0.004 and accuracy of 64.6%, whereas FLT400 had a significance of 0.015 and an accuracy of 60.4%, exactly the same as the DTA results. ARCSFLT, with a continuous distribution, classified MINOCC with a significance of 0.004 and an accuracy of 61.5%, which was a better significance and comparable accuracy to the ternary grouping of ARCSFLT in the DTA. Overall, FLT1000 was the best fault proximity predictor of mineral occurrences when no other predictors were considered in the model.

A comparison of the till base metal elements and TPC2 (*i.e.* the base metal principal component) indicated that TPC2 was best at predicting mineral occurrences with a significance of 0.081, which is not significant at α=0.05 but is significant at α =0.20 (the significance level set to enter the stepwise model). TlogCu had the worst significance of 0.50 and an accuracy no better than chance. For the lake sediment base metal elements,

LPC2 and LPC4 (where n=54), LlogPb had the best significance of 0.57. LlogAu had a poor significance (0.51) but the best accuracy of 63.0%. All other till and lake sediment predictors, beside LlogAu, had accuracies between 50% and 60%. And finally, a comparison between VOLCFEL and VOLCMAF (volcanic felsics and mafics, respectively) indicated the significance of VOLCFEL to the model. VOLCFEL had a significance of 0.0001 and an accuracy of 66.7% whereas VOLCMAF had a significance of 0.648 and an accuracy of only 51% (no increase over chance).

In summary, LRA indicated that the predictors, SWF, VOLCFELS, FLT1000 and TlogCu, were significant in classifying mineral occurrences. SWF, VOLCFELS, and FLT1000 were significant at $\alpha = 0.05$, but TlogCu was only significant at $\alpha = 0.20$. In addition to these four predictors, ARCSFLT (the continuous fault proximity predictor) was significant (at $\alpha = 0.05$) in classifying MINOCC directly. In the set of till and lake sediment geochemistry, the best predictor of MINOCC directly was TPC2, the till base metal and gold component. This predictor was only significant at $\alpha=0.20$.

5.5 Reliability and Favourability Analysis

One of the most significant predictors resulting from the DTA and LRA modelling was felsic volcanics (VOLCFELS). An indication of the reliability of this predictor can be determined from a comparison of the geology from the regional 1:50,000 scale mapping

versus the detailed site-specific geology from the MODS files. Reviewing the geology from the two sources for the 47 surface mineral occurrence sites, 26 have the same rock types in both detailed and regional geology, 10 are similar and the remaining 11 occurrences have different detailed and regional geology. Therefore, only 36 of 47 (77%) mineral occurrence sites have similar detailed and regional geology. The regional geology, with 12 rock types, was not the best predictor of mineral occurrences; the best geology predictor was the binary presence and absence of volcanic felsics. When the geology data were reduced to a binary code and compared for the two scales, the same rock type occurs at 33 sites, similar rock types at 6 sites and 8 sites are different. Using this binary code, the reliability of the regional geology as compared to the detailed geology increases from 77% to 83% (*i.e.* 39 of 47 sites are similar). This indicates that regional 1:50,000 scale geology can be used for quantitative modelling with good confidence in its reliability.

The reliability analysis above provides an overall impression of the reliability of the geology at the 1:50,000 scale. To assess the DTA and LRA modelling results, the reliability of each of the significant predictors in the models needs to be assessed. Overall, the significant predictors in both modelling results are SWF, VOLCFELS, FLT400, FLT1000, and TlogCu.

The reliability of the SWF was not presented as a separate map because it was based on distance calculations from the mineral/nonmineral occurrences. The uncertainty on the locations of the mineral occurrences (*i.e.* maximum of ±100 m) was incorporated into the other predictors as the basis for the cell resolution of all the raster maps.

Reviewing the reliability of the volcanic felsics in the study area required an assessment of the proximity of the volcanic felsic outcrops (as point locations) to each other and to nonvolcanic felsics. The outcrop locations were digitized and two maps were produced; a felsic volcanic map was coded with a '1' for volcanic felsic outcrops and '0' otherwise and a nonvolcanic felsic map was coded with a '1' for nonvolcanic felsics and '0' otherwise. Since the rock units in the study area trend NE-SW, a small line 3 cells long (representing the maximum length of the smaller outcrops) and trending 050 (average of trends measured from volcanic units) was centred on each outcrop and buffered to produce a proximity map. The buffering of the line provides an elliptical NE-SW shape resembling the shape of the geological units, rather than a circular shape which occurs with point buffering. The felsic volcanic outcrop proximity map was masked to the outlines of the felsic volcanic contacts from the geology map. This provides proximity values from the outcrops to the contacts of the felsic volcanics. Those cells close to outcrops of felsic volcanic are considered reliable whereas those cells furthest from the outcrops but still within the felsic volcanic unit are considered unreliable. The plot of these cell proximity values (Figure 5.20) was assessed and the information was used to derive a coding scheme (Table 5.22). Due to the intrinsic limitations on precisely locating outcrops and determining rock types in the field, the maximum reliability is considered to be 90% (S. Colman-Sadd, pers. comm., 2000). The reliability decreases to 10% for cells greater than 3000 m from a felsic volcanic outcrop.

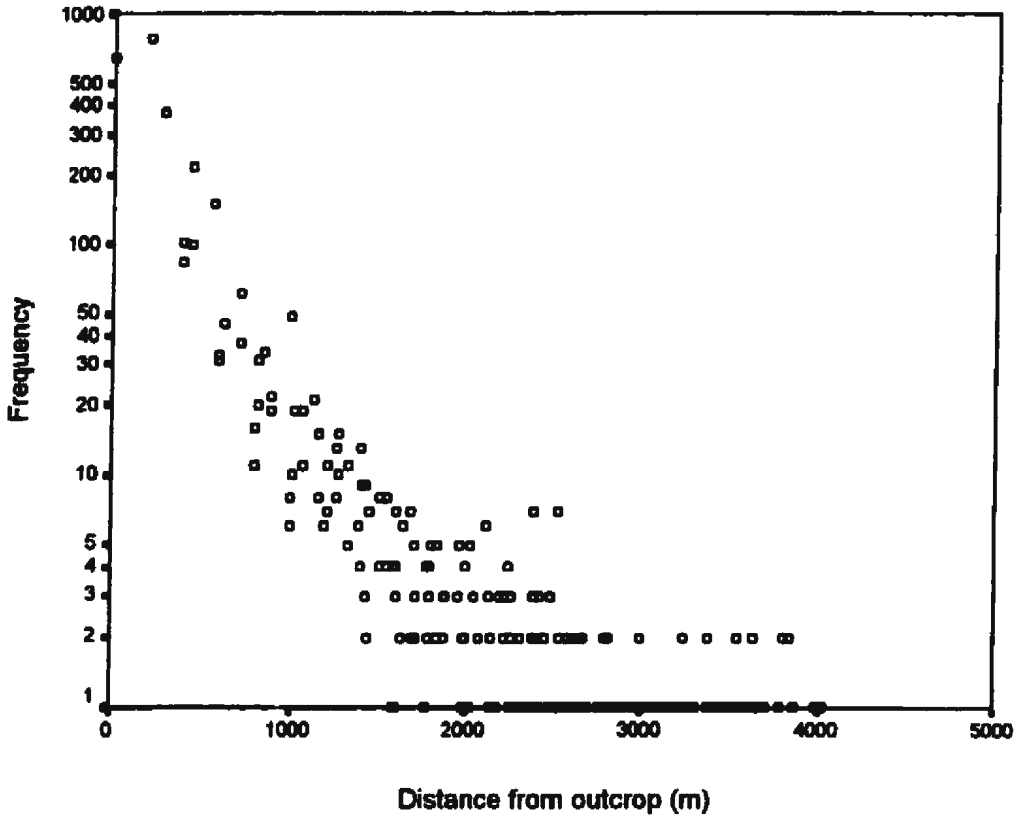

Figure 5.20 : Distance of raster cells from felsic volcanic outcrop in the study area.

Table 5.22 : Coding scheme representing reliability based on proximity to outcrop.

Distance from Outcrop (m)	Reliability (%)
>3000	10
2001-3000	20
1001-2000	30
801-1000	40
601-800	50
401-600	60
201-400	70
2-200	80
0-1	90

Similarly, the nonvolcanic felsics (*i.e.* all other rock types besides volcanic felsics) were masked to provide a proximity map of their outcrops. A plot of these cell proximity values was similar to the felsic volcanics, so the same coding scheme was used for both. The two maps were combined to form a final reliability map for the binary felsic volcanics (Figure 5.21).

The fault proximity predictors of significance in the DTA and LRA modelling are FLT400 and FLT1000. These are the binary fault proximity maps with the buffer thresholds at 400 m and 1000 m from the fault, respectively. Assuming the center-line of the buffer polygon (*i.e.* the location of the fault) is the most reliable area (Berry, 1993), these areas were assigned the highest reliability of 80%. The boundary threshold was assigned the lowest reliability of 50% (Berry, 1993). The reliability map of the FLT400 proximity buffers is presented in Figure 5.22.

The only geochemical predictor of significance was logCu in till. This predictor was only significant in the LRA. The reliability for the till logCu map was calculated as discussed in Chapter 4.5, using the standard deviation map resulting from the kriging. The standard deviation map was divided by the kriged logCu map, to result in a coefficient of variation (CV) map. The CV map was subtracted from 100 to arrive at a reliability map. The reliability map is presented in Figure 5.23. Reliability values range from 75% to 93%, indicating a very reliable kriged logCu map.

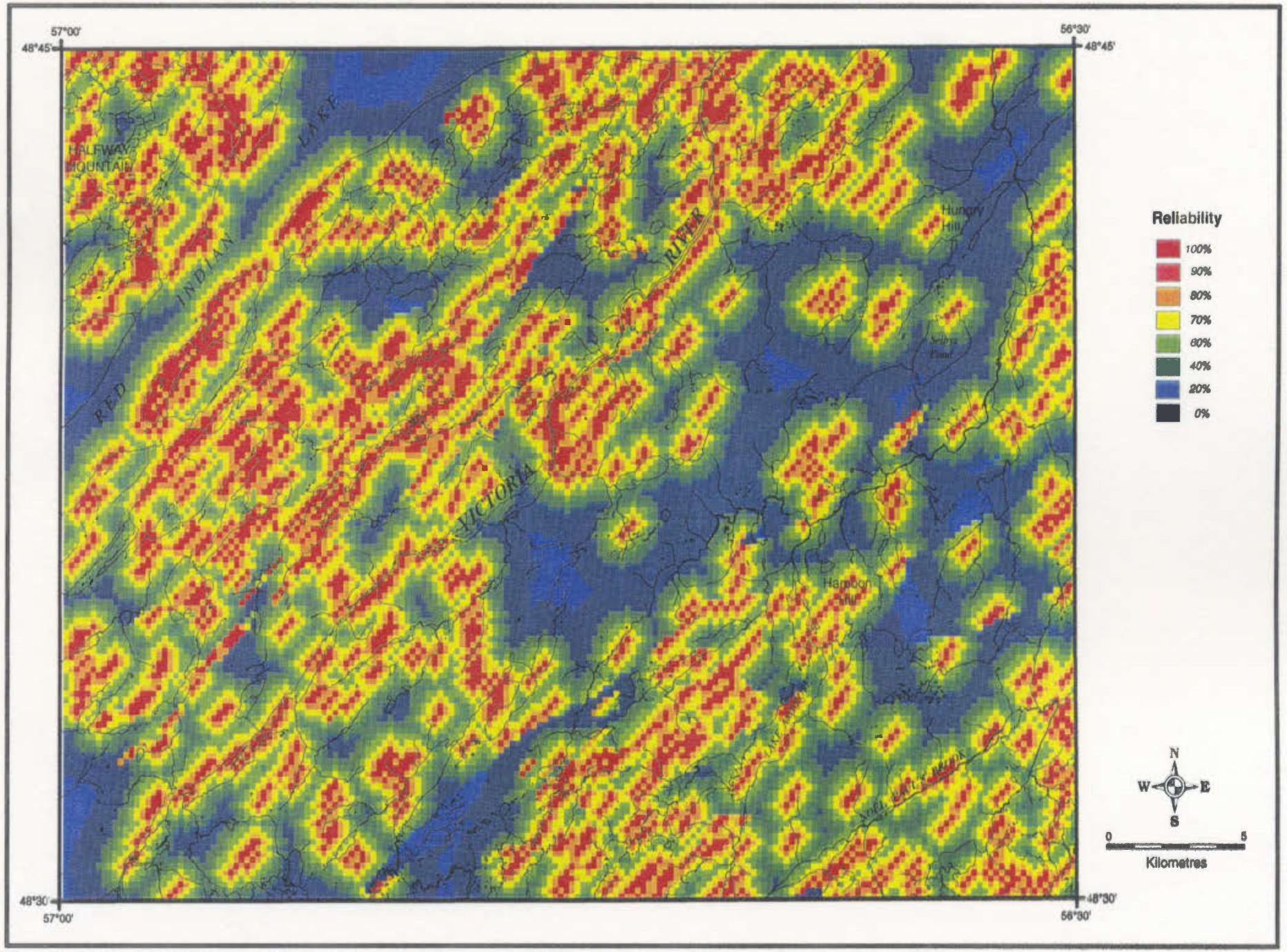

Figure 5.21 : Reliability of the felsic volcanics and rocks other than felsic volcanics. Reliability extends on a continuous scale from 0% (not reliable) to 100% (reliable).

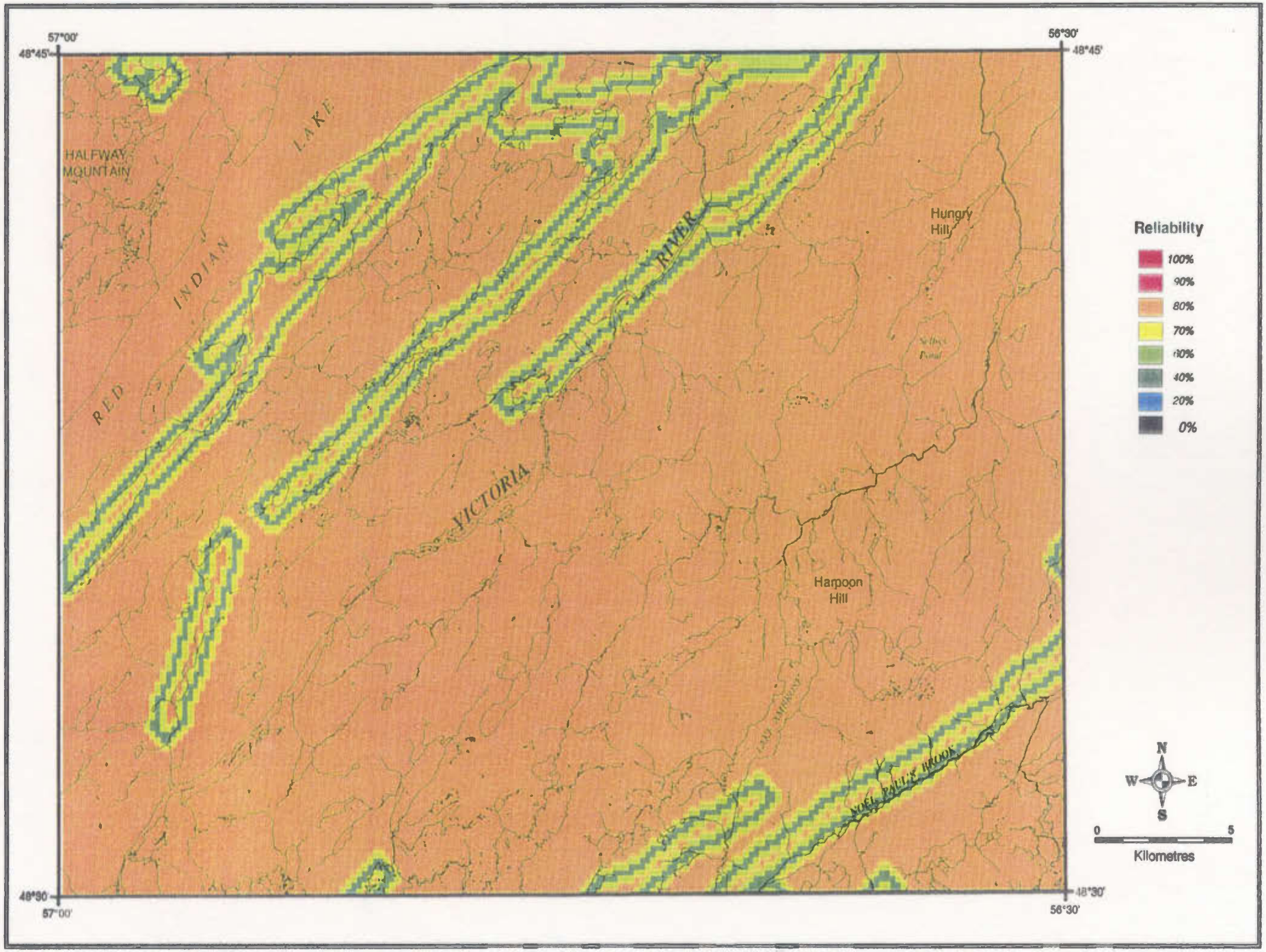

Figure 5.22 : Reliability of the fault proximity 400 m buffer. Reliability values consist of 50% (fair), 65%, and 80% (good).

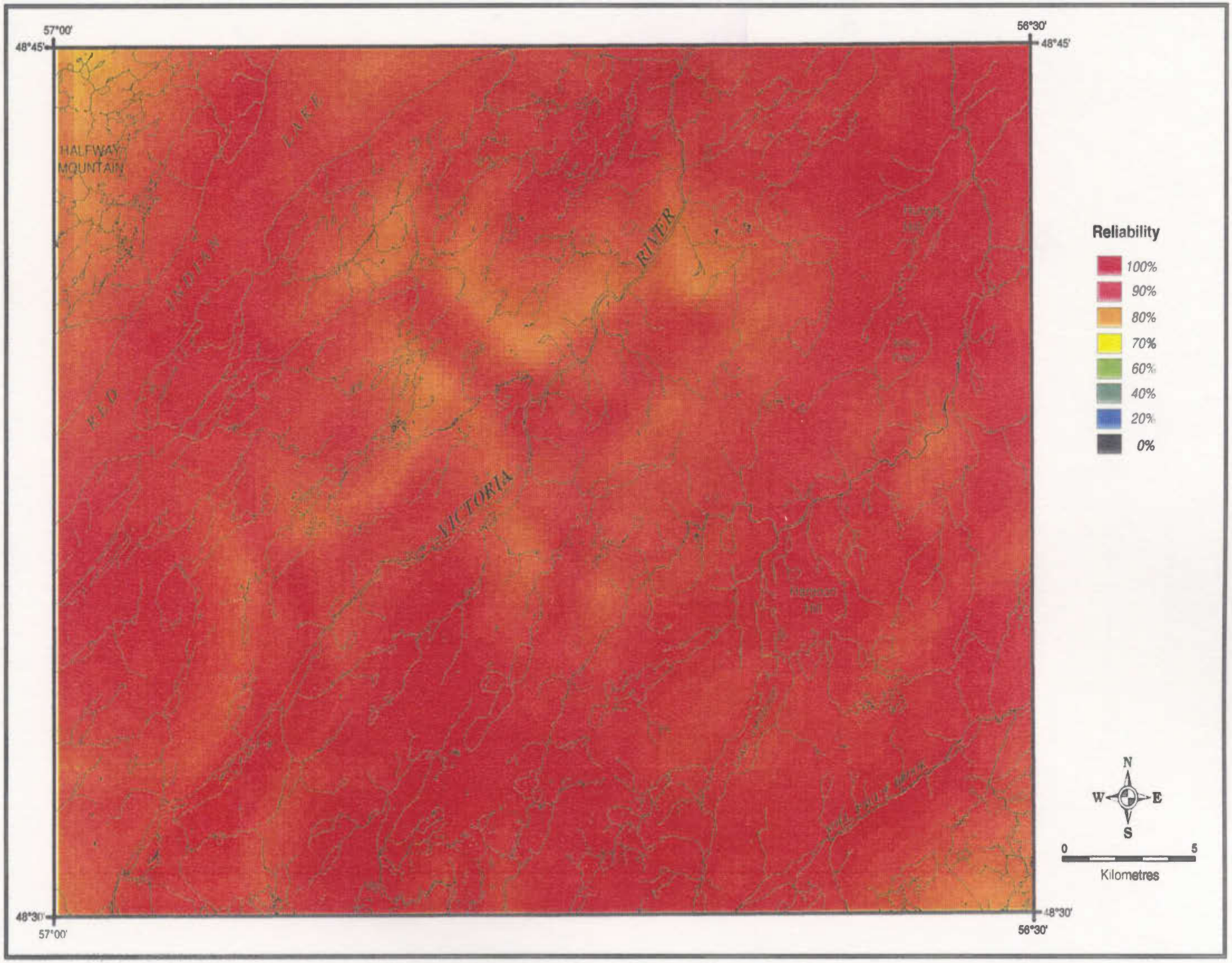

Figure 5.23 : Reliability of copper in till (from kriged surface). Reliability extends on a continuous scale from 0% (not reliable) to 100% (reliable).

With reliability maps prepared for each of the significant predictors, they were combined to provide the reliability for the DTA and LRA mineral potential maps. Due to the nature of the DTA results (Table 5.19), each mutually exclusive area was defined by the SWF and only one other predictor (*i.e.* VOLCFELS or FLT400). Since the SWF does not have a reliability map then only one predictor was mapped in each area. Therefore, the DTA reliability map was pieced together by masking the reliability maps of the individual predictors, VOLCFELS and FLT400, for the areas they influenced. The DTA reliability values range from 10% to 90% (Figure 5.24).

Berry's joint probability method (1993) was used to determine the reliability for the LRA mineral potential map by multiplying the reliability maps for predictors VOLCFELS, FLT1000 and TlogCu. The resulting LRA reliability values range from 4% to 67% (Figure 5.25).

To determine favourable areas for further exploration, the mineral potential map for each model was multiplied by its reliability map and converted to a scale from 0 (indicating low mineral potential with low reliability) to 1.0 (indicating high mineral potential with high reliability). The range of values for the DTA favourability map was 0.00 to 0.90 and the range of values for the LRA favourabiltiy map was 0.00 to 0.62. The histograms of both maps were analyzed to determine how best to classify the results. The DTA histogram has a large peak at 0.48 favourability, due to the large area assigned as 60% mineral potential and 80% reliability. This peak covers the 48^{th} to 94^{th} percentiles. Therefore, these values can only be reclassed as a block. The most interesting favourability (consisting of high mineral potential and high reliability) is the top 6% from

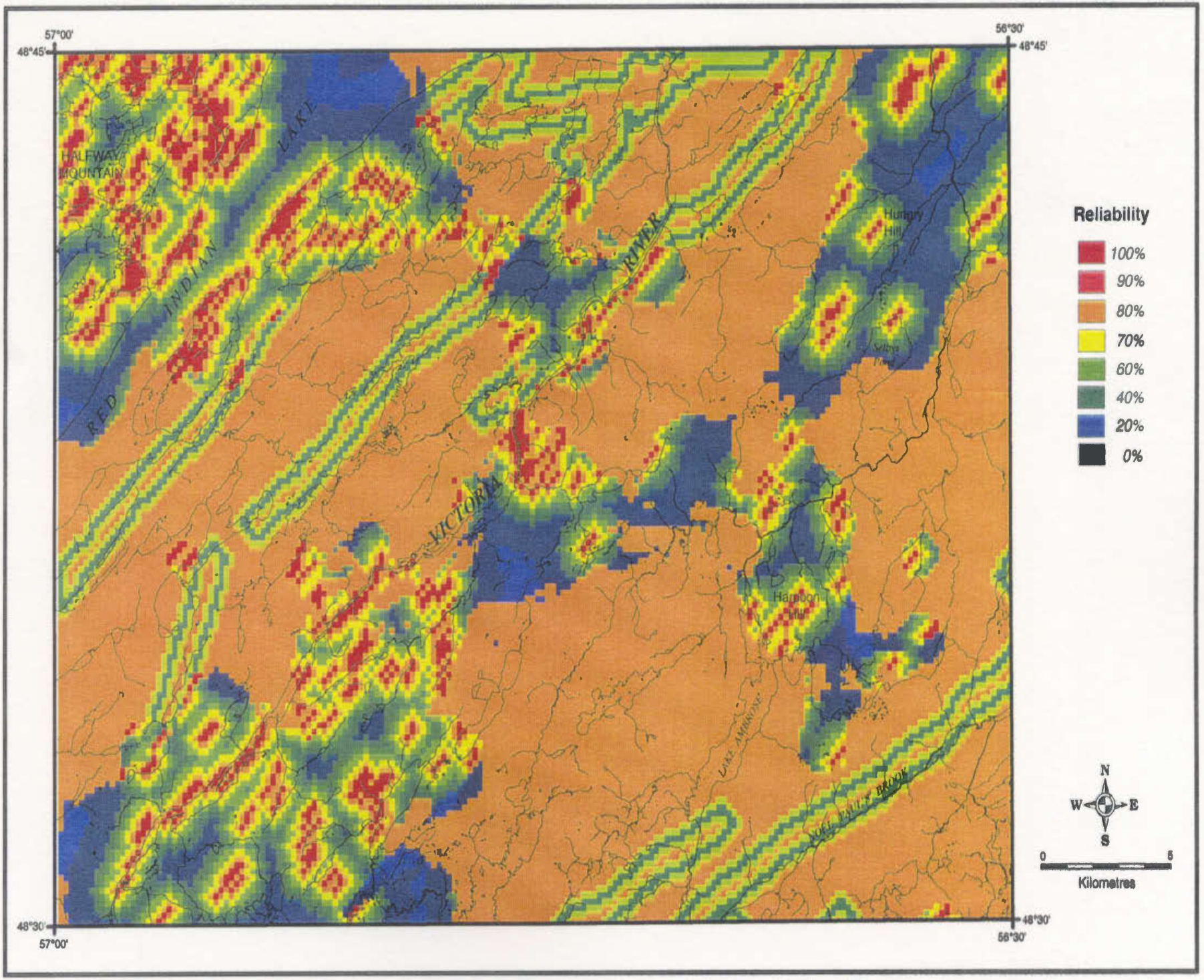

Figure 5.24 : Reliability for the Decision Tree Analysis mineral potential map. Reliability extends on a continuous scale from 0% (not reliable) to 100% (reliable).

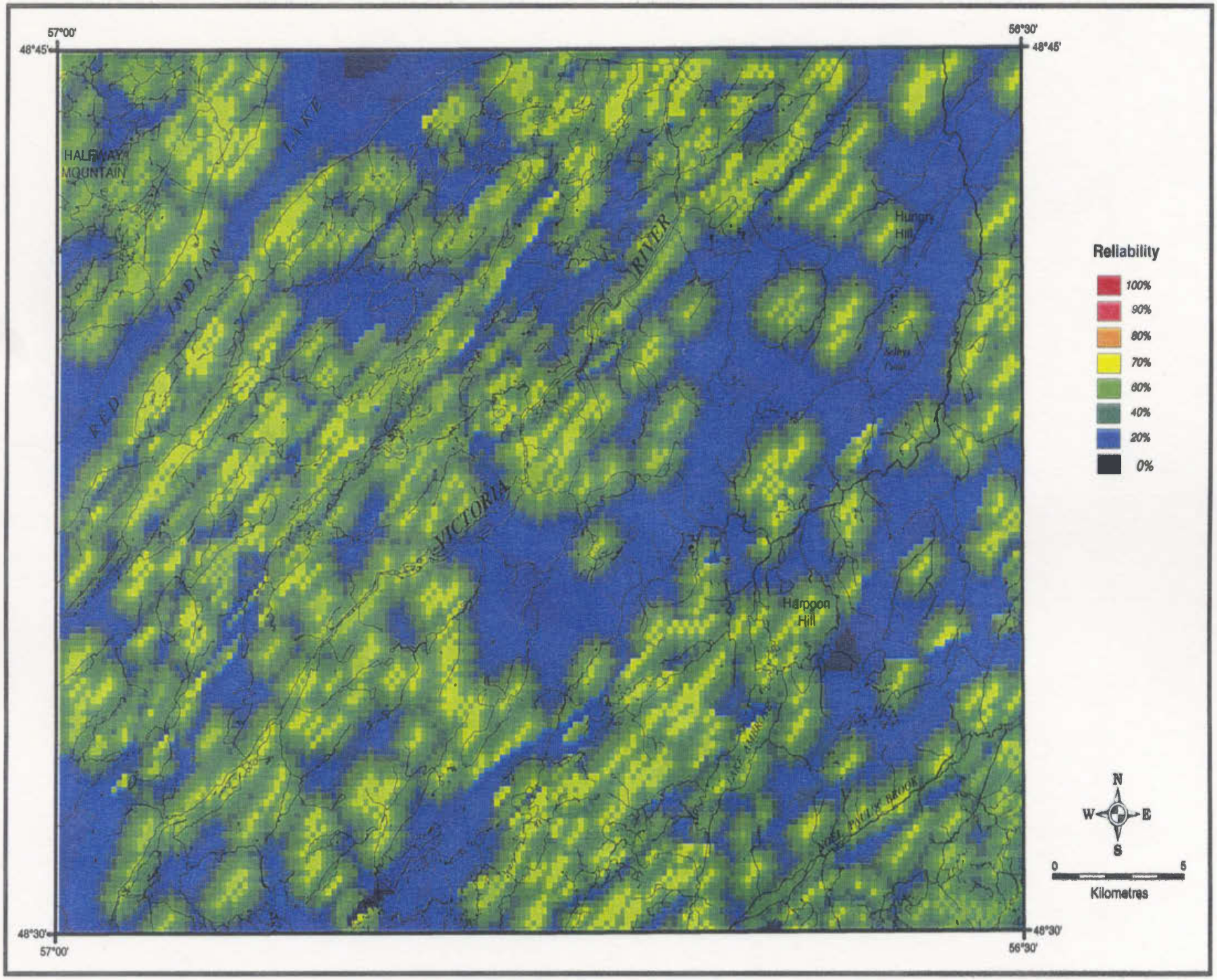

Figure 5.25 : Reliability for the Logistic Regression Analysis mineral potential map. Reliability extends on a continuous scale from 0% (not reliable) to 100% (reliable).

the 95th to 100th percentiles. These values were reclassed as a '1' and values less than the 95th percentile were reclassed as '0'. Therefore, the top 6% of the LRA favourability map was also reclassed to '1' (as highly favourable areas) and all other values were reclassed to '0' (as not highly favourable). The two maps looked somewhat similar, with high favourability extending along the volcanic belts. To provide a final summary map of favourability, the DTA and LRA binary favourability maps were added together resulting in a map where '0' represents areas of no interest in either model, '1' represents areas with high mineral potential and high reliability in at least one of the models, and '2' represents areas with high mineral potential and high reliability in both models (Figure 5.26).

As a final comparison, the locations of the mineral occurrences were compared with the combined favourability map. Of the 47 mineral occurrences, 10 were within areas coded '2', 15 were within areas coded '1' and 22 were within areas coded '0'. The top 6% favourability area defined by DTA contained 17 mineral occurrences and the LRA 6% favourability area contained 18 mineral occurrences. Overall, 53% (25 of 47) of the mineral occurrences were within the top 6% favourability for either the DTA or LRA or both models.

5.6 Summary

After preparation of the individual predictors, two quantitative mineral potential models were developed; a rule-based DTA model and a continuous LRA model. Both models indicated SWF and VOLCFELS were significant predictors of mineral potential. Both models also indicated fault proximity was significant but the DTA favoured a 400 m

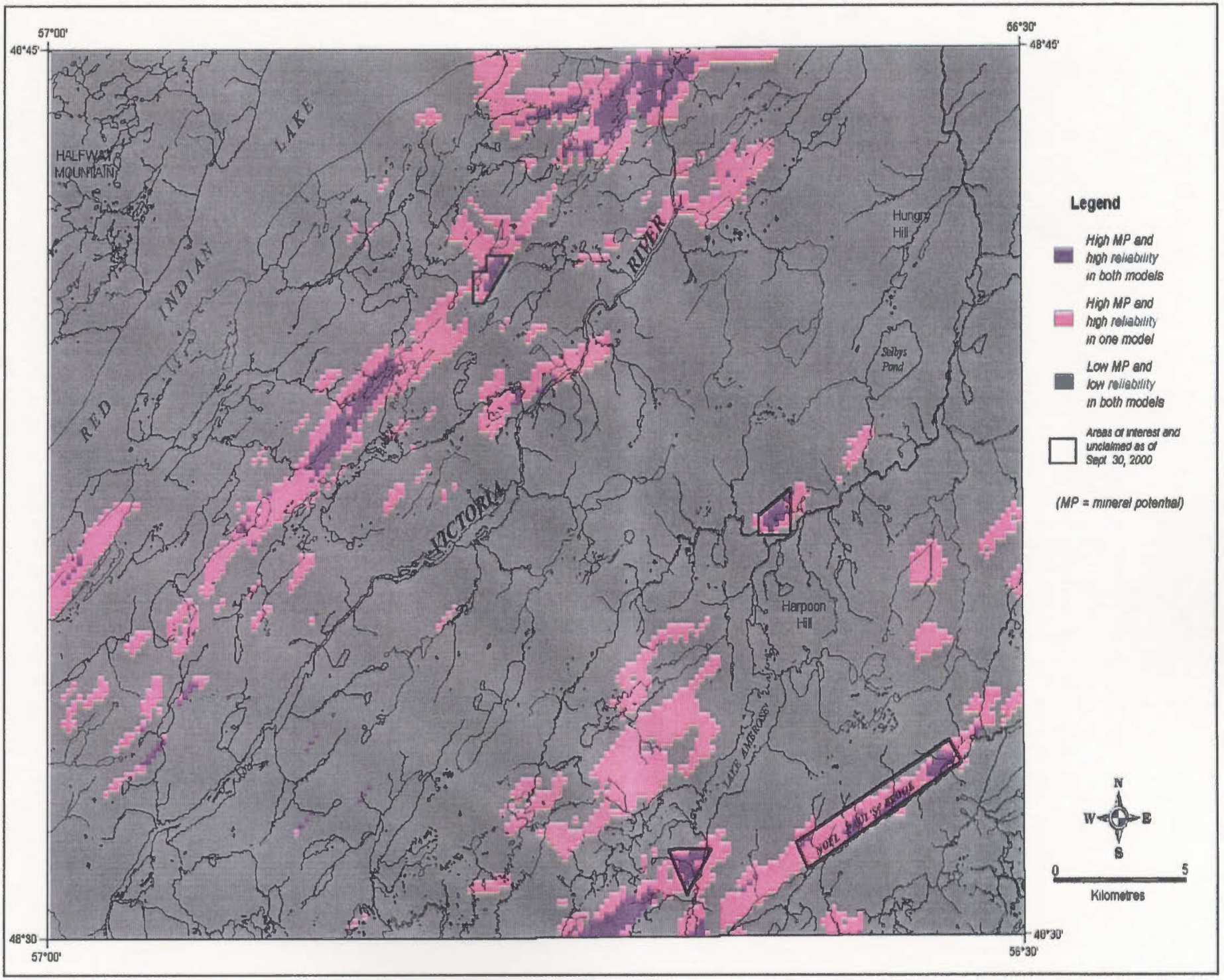

Figure 5.26 : The main areas of interest based on high mineral potential and high reliability. Four unclaimed areas of interest are outlined.

fault buffer whereas the LRA favoured a 1000 m fault buffer. Copper in till was only a significant predictor in the LRA model.

Reliability maps were determined for each of these significant predictors and combined to provide reliability maps for the DTA and LRA models. Favourability maps, for each model, were derived by multiplying the mineral potential by the reliability. Overall favourability was determined by combining the favourability for the two models. A comparison between the overall favourability map and the 47 mineral occurrences indicated that 53% of the mineral occurrences were in the highest 6 percentiles for at least one model.

CHAPTER 6

Discussion and Conclusions

The primary goal of this study was to develop a quantitative model for mineral potential mapping using statistical and geostatistical methods on geochemical, geological and spatial predictors for the Lake Ambrose area in central Newfoundland. The two statistical techniques used for modelling were decision tree and logistic regression analysis. These techniques required prescreening and processing of the geological and geochemical raw data to produce a set of useful predictors for classifying mineral occurrences. A spatial predictor was also calculated from the neighbourhood analysis of the mineral occurrences. The following discussion reviews the predictors and the results of the mineral potential modelling. In particular, the mineral potential maps resulting from DTA and LRA are compared, and their reliability maps are assessed. An overall favourability map, based on the mineral potential and reliability maps, reveals that a few small regions (a few kilometres in size) within the Tally Pond and Tulks Hill volcanic belts, close to faults and known mineral occurrences, have high mineral potential and high reliability.

6.1 Discussion of Predictors

Twenty predictors were initially considered for the mineral potential modelling. Of these, 5 (*i.e.* SWF, VOLCFELS, FLT400, FLT1000 and TlogCu) proved to be the most

useful and the others were not significant. The following discussion focuses on the 5 important predictors, but reviews some key points about the other predictors.

Principal components analysis of till geochemistry resulted in a component characterized by base metals plus gold and its pathfinder elements (*i.e.* TPC2). Since the base metals and gold occur in the same component, it gives strong support to the decision to group the gold and VMS mineral occurrence deposit types as one population in the modelling. This decision was further supported when the VMS mineral occurrences (without the gold and unknown occurrences) were analyzed separately by LRA and the results substantiated the fact that grouping the gold and VMS occurrences produces very similar results to analyzing VMS occurrences on their own.

Of the till geochemical predictors, TlogCu was the only predictor that was significant in predicting mineral occurrences. TlogCu was significant in the stepwise LRA, along with 3 other predictors, and it was also significant as a direct predictor of mineral occurrences in the DTA. This may reflect the fact that copper minerals are more prevalent in the mineral occurrences (present in 23 of 47 occurrences) than minerals of the other elements. TlogAu, TlogPb, TlogZn and TPC2 were not significant in predicting mineral occurrences.

The processing of the lake sediment data to remove lake effects, such as element adsorption on Fe and Mn hydroxides, also removes part of the primary geochemical signal of the elements regressed against them (Davenport, 1990). This changes the correlations among the elements, which, in turn, affects the lake sediment PCA results. This problem is illustrated by residual zinc (rZn) which is unexpectantly associated with component 3

(*i.e.* mafic volcanic lithologies) in the lake sediment PCA, as opposed to component 2 (*i.e.* base metals). The expected association of zinc with the base metals is substantiated in the till PCA where logZn is associated with the base metals component. Therefore, PCA results for the lake sediment data should be treated with caution due to the removal of part of the primary geochemical signal of the regressed elements.

The lake sediment database was problematic because only 54 of 96 mineral/nonmineral sites (*i.e.* 31 mineral occurrences and 23 nonmineral occurrences) occurred within catchment basins containing lake sediment samples. Therefore, 42 cases had missing data. An advantage of the decision tree modelling method was that it incorporated the missing data as a group. The LRA could not incorporate the missing data and removed all pairwise cases with missing data from the analysis (*i.e.* 42 sites not within a catchment basin).

LPC4, the lake sediment gold PCA component, was the only predictor which was somewhat significant (0.087) at predicting mineral occurrences in DTA, in conjunction with SWF, VOLCFELS, and FLT400. Residual zinc in lake sediments was also somewhat significant (0.070) at directly predicting mineral occurrences in DTA. In contrast, none of the lake sediment predictors (*i.e.* LlogAu, LlogCu, LlogPb, rZn, LPC2 and LPC4) were significant in the LRA, as part of the stepwise analysis or as a direct predictor of mineral occurrences. The significance in some of the lake sediment predictors in DTA indicates that there may be fundamental differences between DTA and LRA; specifically, DTA can incorporate missing values as a group and continuous values are grouped into specified discrete bins. When the LRA included the lake sediments, it analyzed only 54 cases,

whereas DTA was able to include the lake sediment predictors and still analyze all 96 cases.

The applicability of the regional 1:50,000 scale geology as a basis for quantitative modelling had to be assessed. A comparison of the regional geology with that specifically reported for the 47 mineral occurrence sites, indicated good reliability (77% agreement) which was increased still further (83% agreement) when the geology was reclassed to binary volcanic felsics. The good agreement indicates that 1:50,000 scale geology, readily available in digital format for most of the province, provides reliable geological information for quantitative modelling analysis.

Based on the literature, the VMS deposit model has felsic and mafic volcanics as the most common hosts to VMS deposits (Franklin, 1993). Both DTA and LRA confirmed that, in the Lake Ambrose area, felsic volcanics (not mafic volcanics) are associated with the mineral occurrences. The mafic volcanics were not at all significant in classifying mineral occurrences in either DTA or LRA. The nominal geology predictor, consisting of 12 rock types, was not significant in predicting mineral occurrences.

Based on the literature, deposit models for VMS and gold deposits indicate that proximity to faults is an important factor in the exploration for these mineral occurrences (Franklin, 1993). The binary predictors FLT400 and FLT1000 were both significant in classifying mineral occurrences in DTA and LRA, respectively. As a direct predictor of mineral occurrences, FLT1000 was the best predictor in both DTA and LRA. FLT400 was a significant predictor in DTA, rather than FLT1000, probably because it only classified a portion of the cases (*i.e.* 57 of 96) off the primary predictor, SWF, whereas, in

LRA, each predictor classifies all 96 cases. The ARCSFLT predictor was not significant in classifying mineral occurrences.

The surficial geology and wetlands predictors were not significant in predicting mineral occurrences. The wetlands data was too skewed (*i.e.* only 4 of the 96 sites intersected wetlands) to provide a reasonable analysis. The surficial geology was mapped on a regional scale and an assessment of its reliability is not available. The deposition of surficial sediments by glacial processes are not directly related to mineral occurrences, except for the possibility that surface mineral occurrences may be more easily found where bedrock is exposed or covered by only thin sediment units (*e.g.* drift/rock, which averages less than 1 m in thickness). However, the thickness of surficial sediments may be related to the till geochemistry. Copper values in till increased with increasing sediment thickness. The reasons for this are beyond the scope of this thesis, but indicate that a combined surficial geology/TlogCu predictor may be of future interest.

The tabulated glacial striation database was included as part of the database for this study but was not assessed as a predictor of mineral occurrences. The striation data are not directly related to mineral occurrences but may play a part in assessing the spatial proximity of anomalous till geochemistry to mineral occurrences. Striation data may be more useful in areas where only one major ice flow direction is recognized, and in areas that have less complex geology such as only a few contrasting lithogeochemical units. This would permit striation data to be correlated with the transport and provenance of till. This information could be used to back-correct the location of the till data (closer to the source) and then the corrected till predictors could be used in mineral potential modelling.

This approach could best be tested in areas with less complex geology than the Lake Ambrose area used in this study. Till and lake sediment geochemistry are presently used by exploration geologists only on an empirical basis at a regional scale to locate areas of potential exploration interest. With corrected till predictors, mineral potential mapping may be able to provide more focused areas for exploration.

The most significant predictor in both the DTA and LRA was the spatial weighting function. The SWF values were calculated based on the proximity to neighbouring mineral occurrences and represent the potential for mineralization. This predictor quantifies and spatially defines the exploration geologists "rule-of-thumb" that new mineral occurrences are often to be found "within sight" of known mineral occurrences.

6.2 DTA and LRA Modelling Results

Due to the sparseness of cases in the database, a test dataset could not be partitioned for use in determining the accuracy of the results. Therefore, two modelling techniques, decision tree analysis and logistic regression analysis, were chosen to act as a comparative test in assessing the relative accuracy of their resulting mineral potential maps. The resulting maps were compared in a number of ways, including a visual inspection, a site comparison (based on the 96 mineral/nonmineral occurrence sites) and a comparison of the two raster mineral potential maps.

The DTA indicated that SWF, VOLCFELS, and FLT400 were the 3 predictors which best classify mineral occurrences. The model correctly classified 78.1% of the 96 mineral/nonmineral occurrences. The mineral potential map resulting from the DTA

modelling consists of four mutually exclusive areas (Figure 5.18) as defined by the rule-based conditions. Most of the map area is defined by two probabilities for mineral potential; 11% and 60%. A probability of 100% is defined by two rules; i) where the number of mineral occurrences in a neighbourhood is low (SWF<0.17), the presence of volcanic felsics defines a high mineral potential area, and ii) where the number of mineral occurrences in a neighbourhood is high (SWF>0.64), the proximity to faults is important. An unusual case occurs for areas close to faults but with a moderate number of neighbourhood mineral occurrences (0.16<SWF<0.64). These areas are defined as having 0% mineral potential and were classified based on only 1 case. Therefore, a larger mineral occurrence database may help to reduce this type of spurious classification. Also, pruning the decision tree of splits defined by too few cases (where the threshold number of cases is defined by the analyst) will help to define more robust rule-based conditions.

The LRA indicated that SWF, VOLCFELS, and FLT1000 and TlogCu were the 4 predictors which best classify mineral occurrences. The model correctly classified 79.2% of the 96 mineral/nonmineral occurrences; 5% more nonmineral occurrence sites were correctly classified than mineral occurrence sites. The mineral potential map resulting from the LRA modelling consists of a continuous probability surface of mineral potential (Figure 5.19) as defined by the logistic regression equation. Most of the map area is defined by less than 30% probability of mineral potential. Two NE-SW trending belts contain greater than 70% mineral potential. These two belts coincide with the Tally Pond and Tulks Hill volcanic belts where most of the mineral occurrences are located and where the faults and anomalous copper in till occurs.

The DTA and LRA mineral potential maps were compared in a number of different ways. The first method was to determine how similar (level of agreement) the results of the DTA and LRA modelling methods were at the 96 sites used in the study. The probability values from the two modelled mineral potential maps were extracted at the 96 sites. These values were converted to binary codes where '0' represented probability values from 0% to just less than 50%, and '1' represented probability values from 50% to 100%. Eighty-one of the 96 sites (84.3%) were predicted the same by both methods (whether correctly classified or not); 5 mismatches occurred at mineral occurrence sites and 10 occurred at nonmineral occurrence sites. Therefore, there is good agreement for the 96 sites between the two methods, indicating that they do produce comparable results for the data provided.

The above mismatch test only compares the models with one another (*i.e.* test of agreement) and does not indicate whether the sites were predicted correctly (*i.e.* test of accuracy). Probability values greater than 50% were classified as mineral occurrences, and values less than 50% were classified as nonmineral occurrences. The LRA predicted 80.9% of the mineral occurrence sites correctly and 81.6% of the nonmineral occurrence sites correctly (Table 6.1). The DTA predicted 91.5% of the mineral occurrence sites correctly but only 65.3% of the nonmineral occurrence sites were correctly classified. Twelve sites were not correctly classified by either method due to predictor values that were not consistent with the VMS model (*e.g.* mineral occurrences in nonvolcanic felsic rock types). Spatially, the 12 sites are distributed in all sections of the map area.

Table 6.1 : A comparison of the classification accuracy of the DTA and LRA modelling methods for the mineral occurrence (MO) and nonmineral occurrence (NMO) sites.

	Predicted			
Observed	DTA		LRA	
	MO	NMO	MO	NMO
MO	43	17	38	9
NMO	4	32	9	40

The second method of comparing results used the same binary coding as above (using 50% probability as the cutoff) but, instead of a 96 site analysis, the DTA and LRA mineral potential maps were recoded to binary maps. A cross-tabulation of the results (Table 6.2) was used to compare the maps. Overall, 62.1% of the cells are in agreement. The binary DTA results indicate 57.0% of the area has a "good" mineral potential (*i.e.* >50% probability) whereas the binary LRA indicates that only 23.5% of the area has a "good" mineral potential. The overlap of areas with "good" mineral potential between the two maps is 21.3%. Within this 21.3% area there are 37 of the 47 mineral occurrence sites (78.7%).

Table 6.2 : Cross-tabulation (number of grid cells) of the results of DTA and LRA converted to binary maps. Mineral occurrences=1 and nonmineral occurrences=0.

		Binary DTA Results		
		0	1	Total
Binary LRA Results	0	10576 (40.8%)	9248 (35.7%)	19824 (76.5%)
	1	563 (2.2%)	5513 (21.3%)	6076 (23.5%)
	Total	11139 (43.0%)	14761 (57.0%)	25900 (100%)

Measures of map similarity provide a quantitative assessment of the spatial association between maps. Two measures of spatial association were determined from the cross-tabulation data in Table 6.2. The kappa coefficient of agreement (Bonham-Carter, 1994) compares the total observed agreement (p_{ii}) with the total expected agreement due to chance (q_{ii}):

$$\kappa = \frac{\sum_{i=1}^{n} p_{ii} \sum_{i=1}^{n} q_{ii}}{1 - \sum_{i=1}^{n} q_{ii}}$$

Kappa values range from –1 (disagreement) to +1 (agreement). For the DTA and LRA mineral potential maps kappa = 0.29, indicating a moderate level of agreement. But, values of kappa may be underestimated due to the overestimate of the expected agreement (Foody, 1992). Therefore, Yule's α may provide a more appropriate map comparison. Yule's α can be calculated using the following equation (Bonham-Carter, 1994):

$$\alpha = \frac{\sqrt{T_{11}/T_{21}} - \sqrt{T_{12}/T_{22}}}{\sqrt{T_{11}/T_{21}} + \sqrt{T_{12}/T_{22}}}$$

where T_{mn} are the values in the m^{th} row and n^{th} column of the cross-tabulation table (Table 6.2). For the DTA and LRA mineral potential maps $\alpha = 0.54$, which indicates a good level of agreement.

The third method of comparing results, from the DTA and LRA mineral potential maps, was to summarize the cross-tabulation table of the actual values in the raster maps rather than converting the values to binary. A cross-tabulation list of this data provided the cell frequency of each LRA value (from 4% to 96%) for the 4 unique values in the

DTA (*i.e.* 0%, 11%, 60% and 100%). A summary using a box-and-whisker plot (Figure 6.1) indicates the relationship between the two maps. Note that the interquartile ranges are fairly distinct for the 11%, 60% and 100% decision tree results but the 0% interquartile range overlaps the 60% and 100%. As discussed above, the 0% class was defined by only 1 case in the DTA (Rule 9 in Figure 5.17). Removing this node (FLT400 split on SWF into 0.17 to 0.63 and 0.64 to 1.0) would result in this sample (*i.e.* this group of SWF from 0.17 to 0.64) being assigned a probability of 93.3% instead of 0%.

Even though DTA and LRA modelling methods are conceptually different, generally the agreement between the two methods is good, where both produced a classification accuracy of approximately 79%.

6.3 Discussion of Reliability and Favourability

The reliability of the DTA and LRA mineral potential maps indicates a primary difference between the two models. The reliability for the DTA mineral potential map was based on mapping individual predictor reliability on mutually exclusive areas, whereas the reliability for the LRA mineral potential map was based on the joint probability model. Due to the multiplication of predictor reliability in the joint probability model, the LRA reliability map had much lower values (maximum of 67% versus 90% for the DTA reliability map). Therefore, when comparing the final favourability maps, the top 6 percentile was chosen, rather than choosing specific favourability values which would bias the favourability due to the DTA reliability.

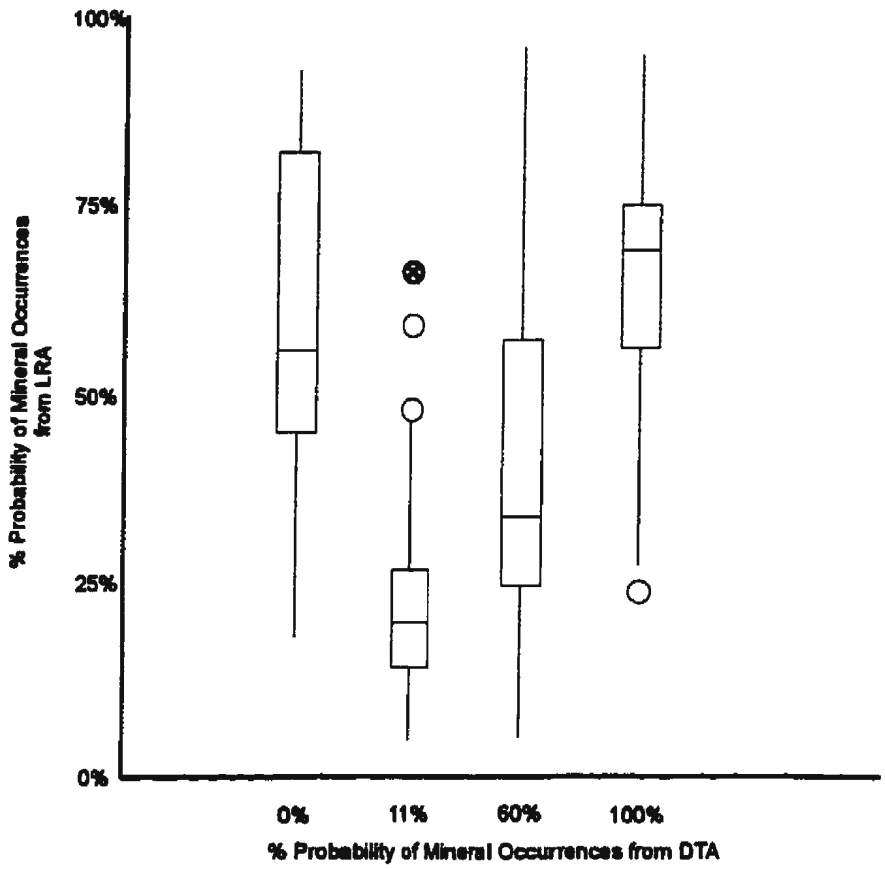

Figure 6.1 : Box-and-whisker summary of the continuous logistic regression analysis results in comparison to the 4 unique decision tree analysis results.

The final favourability map (Figure 5.26), combining the top 6% favourability from both the DTA and LRA models, highlighted those areas of felsic volcanic rock type in close proximity to faults and other mineral occurrences. Sixty percent of the mineral occurrences were within this very favourable mineral potential area. Comparing the favourable areas with the latest published mineral claims map (Sept. 30, 2000) indicates most of the favourable areas have already been staked by exploration companies. A few small areas (Figure 5.26) remain available.

6.4 Conclusions

The relationship between mineral occurrences and spatial, geological and geochemical factors were assessed using the quantitative modelling techniques of decision tree analysis and logistic regression analysis. These two modelling techniques indicated the significant predictors in both models were the spatial weighting function (a measure of the proximity of mineral occurrences), volcanic felsic rock type and proximity to faults. In addition, the LRA determined that copper in till samples was somewhat significant as a predictor. The felsic volcanics and fault proximity predictors support the VMS deposit model, and the spatial weighting function supports the exploration geologists general "rule-of-thumb" that deposits occur in neighbourhoods.

Till and lake sediment geochemistry have traditionally been used in mineral exploration to indicate areas of potential interest. The results of the quantitative modelling did not indicate a clear spatial association between the geochemistry and mineral occurrences. For the lake sediments, this may have been due to the sparseness of the dataset (*i.e.* only 54 catchment basins contained both mineral/nonmineral occurrences and lake sediment samples). Only copper values in till had a somewhat significant association with mineral occurrences. To better determine the association between the mineral occurrences and the till geochemistry, infill sampling to provide a better distribution over the study area would be beneficial, as well as lithogeochemical studies to better trace the source rocks. With this detailed information in the database, the glacial striation database may be useful for helping to trace till geochemistry back to its souce area.

The till and lake sediment PCA results did not provide any significant information for the determination of mineral potential. But it was determined that the time-consuming screening procedure can be avoided prior to running PCA by using Spearman's rank correlation coefficient matrix of raw till values instead of the default Pearson correlation data.

Advantages of the DTA over the LRA was the inclusion of missing samples in the DTA. Also, the DTA method of splitting data into smaller groups provided insights into the nature of the subgroups, whereas the LRA analyzed the data as a whole. Both methods had the advantage, over other quantitative modelling techniques, of handling continuous and categorical data and of not being strict regarding the distribution of the data.

One of the problems with the study area was the sparseness of the mineral occurrence dataset. This resulted in the lack of an independent dataset to test the reliability of the results. Therefore, the results of the two models (LRA and DTA) were compared with one another to provide an indication of the agreement, if not the accuracy, of the quantitative results. The resulting agreement of 84.3% between the 96 sites of the DTA and LRA mineral potentials was quite good. This was substantiated by Yule's α=0.54, which indicated a good agreement between the two mineral potential maps. Overall, both models correctly classified approximately 79% of the 96 mineral/nonmineral occurrences.

In addition to the agreement, the reliability of the models was assessed. Due to the nature of the reliability calculations (*i.e.* rule-based conditions for the DTA map versus

joint probability model for the LRA map), the DTA model appeared to have the better reliability with a maximum reliability of 90% versus only 67% for the LRA model. Using a weighted technique as suggested by Berry (1993), instead of the joint probability model, to calculate the reliability for the LRA model may produce more similar reliability values to the DTA.

The final favourability map (Figure 5.26), based on high mineral potential and high reliability from both the DTA and LRA models, indicates areas of interest for exploration for VMS deposit types. As of September 2000, only a few of the favourable areas identified in this study were not staked by exploration companies. The two areas of primary interest are located in the Tally Pond volcanics and the Tulks Hill volcanics. Applying the results of the DTA and LRA modelling to the whole of the Victoria Lake Group of volcanic rocks will help to refine the modelling technique, as well as defining other favourable areas for exploration of VMS deposits.

References

Angoss Software (1993) *KnowledgeSEEKER, Version V3.00 Beta*. Toronto, Canada.

Anonymous (3 August 2000; date viewed) *Logistic Regression*. http://www.spss.com/tech/stat/Algorithms.htm/logregre.pfd.

Ash, J.S. and Colman-Sadd, S.P. (1997) *Digital geological maps of insular Newfoundland*. In: Current Research, Newfoundland Department of Mines and Energy, Geological Survey, Report 97-1, pp. 141-149.

Batterson, M.J. (1989) *Quaternary geology and glacial dispersal in the Strange Lake area, Labrador*. Newfoundland Department of Mines and Energy, Geological Survey Branch, Report 89-3, 51 pp.

Batterson, M.J., Taylor, D.M., and Davenport, P.H. (1998) *Till geochemistry of the Grand Falls – Mount Peyton area (NTS 2D/13, 2D/14 and part of 2E/03)*. Newfoundland Department of Mines and Energy, Geological Survey, Open file NFLD/2664.

Berry, J.K. (1993) *Beyond mapping: concepts, algorithms and issues in GIS*. GIS World Books, Fort Collins, Colorado, USA, 246 pp.

Berry, J.K. (1995) *Spatial reasoning for effective GIS*. GIS World Books, Fort Collins, Colorado, USA, 208 pp.

Biggs, D., de Ville, B. and Suen, E. (1991) *A method of choosing multiway partitions for classification and decision trees*. Journal of Applied Statistics, Vol. 18, No. 1, pp. 49-62.

Bishop, M.M., Fienberg, S.E., and Holland, P.W. (1975) *Discrete multivariate analysis: theory and practice*. MIT Press, Cambridge Massachusetts, 587 pp.

Bonham-Carter, G.F. (1994) *Geographic information systems for geoscientists: modelling with GIS*. Computer Methods in the Geosciences, D.F. Merriam, series editor, Elsevier Science Inc., Tarrytown, New York, 398 pp.

Bonham-Carter, G.F., and Chung, C.F. (1983) *Integration of mineral resource data for Kasmere Lake area, Northwest Manitoba, with emphasis on uranium*. Computers and Geosciences, Vol. 15, pp. 25-45.

Bonham-Carter, G.F. and Goodfellow, W.D. (1984) *Autocorrelation structure of stream-sediment geochemical data: interpretations of zinc and lead anomalies, Nahanni*

River area, Yukon-Northwest Territories, Canada. In: G. Verly *et al.* (Ed.): Geostatistics for natural resources characterization. Reidel Dordrecht, pt. 2, pp. 817-829.

Bonham-Carter, G.F. and Goodfellow, W.D. (1986) *Background corrections to stream geochemical data using digitized drainage and geological maps: application to Selwyn Basin, Yukon and Northwest Territories.* Journal of Geochemical Exploration, Vol. 25, pp. 139-155.

Bonham-Carter, G.F., Rogers, P.J. and Ellwood, D.J. (1987) *Catchment basin analysis applied to surficial geochemical data, Cobequid Highlands, Nova Scotia.* Journal of Geochemical Exploration, Vol. 29, pp. 259-278.

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (1984) *Classification and regression trees.* Wadsworth, Inc., California, USA, 354 pp.

Brower, J.C., and Merriam, D.F. (1990) *Geological map analysis and comparison by several multivariate algorithms.* In: Agterberg, F.P. and Bonham-Carter, G.F. (Ed.): Statistical applications in the earth sciences. Geological Survey of Canada, Paper 89-9, pp. 123-134.

Chou, Y., Minnich, R.A., Salazar, L.A., Power, J.D., and Dezzani, R.J. (1990) *Spatial autocorrelation of wildfire distribution in the Idyllwild quadrangle, San Jacinto Mountain, California.* Photogrammetric Engineering & Remote Sensing, Vol. 56, No. 11, pp. 1507-1513.

Cliff, A.D. and Ord, J.K. (1973) *Spatial Autocorrelation.* Pion Ltd., London, 178 pp.

Cliff, A.D. and Ord, J.K. (1981) *Spatial Processes: Models and Applications.* Pion Ltd., London, 266 pp.

Coker, W.B., Hornbrook, E.H.W., and Cameron, E.M. (1979) *Lake sediment geochemistry applied to mineral exploration.* In: Hood, P.J. (Ed.): Geophysics and geochemistry in the search for metallic ores. Geological Survey of Canada, Economic Geology Report, 31, pp. 435-477.

Colman-Sadd, S. (compiler; 2000) *Unpublished GeoLegend digital data for Map 94-223, Open File 012A/10/0686.* Government of Newfoundland and Labrador, Department of Mines and Energy, Geological Survey.

Cook, S.J., Levson, V.M., Giles, T.R. and Jackaman, W. (1995) *A comparison of regional lake sediment and till geochemistry surveys: A case study from the Fawnie*

Creek area, central British Columbia. Exploration and Mining Geology, Vol. 4, No. 2, pp. 93-110.

Daultry, S. (1976) *Principal components analysis.* Concepts and techniques in modern geography, No. 8, Geo. Abstracts, University of East Anglia Press, Norwich, England, 51 pp.

Davenport, P.H. (1990) *A comparison of regional geochemical data from lakes and streams in northern Labrador; implication for mixed-media geochemical mapping.* Journal of Geochemical Exploration, Vol. 39, pp. 117-151.

Davenport, P.H., Christopher, T.K., Vardy, S. and Nolan, L.W. (1993) *Geochemical mapping in Newfoundland and Labrador: its role in establishing geochemical baselines for the measurement of environmental change.* Journal of Geochemical Exploration, Vol. 49, pp. 177-200.

Davenport, P.H., Honarvar, P., Hogan, A., Kilfoil, G.J., King, D., Nolan, L.W., Ash, J.S., Colman-Sadd, S.P., Hayes, J.P., Liverman, D.G.E., Kerr, A., and Evans, D.T.W. (1996) *Digital geoscience atlas of the Buchans-Robert's Arm belt.* Government of Newfoundland and Labrador, Department of Mines and Energy, Geological Survey, Open File NFLD/2611, version 1.0.

Davenport, P.H., Hornbrook, E.H.W., and Butler, A.J. (1974) *Regional lake sediment geochemical survey for zinc mineralization in western Newfoundland.* In: Elliot, I.L. and Fletcher, W.K. (Ed.): Geochemical Exploration, Amsterdam, Elsvier, pp. 555-578.

Davenport, P.H., and Nolan, L.W. (1991) *Definition of large-scale zones of hydrothermal alteration by geochemical mapping using organic lake sediment.* Transactions of the Institute of Mining and Metallogeny (section B: Applied Earth Sciences), Vol. 100, pp. B111-B121.

Davenport, P.H., Nolan, L.W. and Honarvar, P. (1994) *The digital geochemical atlas of Newfoundland.* In: Current Research, Newfoundland Department of Mines and Energy, Geological Survey Branch, Report 94-1, pp. 279-299.

Davenport, P.H., Nolan, L.W., Honarvar, P. and Hogan, A.P. (1990) *Gold and associated elements in lake sediment from regional geochemical surveys in the Red Indian Lake map area (NTS 12A).* Newfoundland Department of Mines and Energy, Geological Survey, Open File 12A/561, 34 pp.

Davis, J.C. (1986) *Statistics and data analysis in geology, second edition.* John Wiley & Sons, Toronto, 646 pp.

Eastman, J.R. (1993) *IDRISI, version 4.1, Update Manual.* Clark University Publishing, Worcester, Massacusetts, USA, 209 pp.

Environmental Systems Research Institute Inc. (1996) *ArcView Spatial Analyst.* Redlands California, USA, 148 pp.

Evans, D.T.W., and Kean, B.F. (1987) *Gold and massive sulphide mineralization in the Tulks Hill volcanics, Victoria Lake Group, central Newfoundland.* In: Current Research, Newfoundland Department of Mines and Energy, Mineral Development Division, Report 87-1, pp. 103-111.

Evans, D.T.W., Kean, B.F., and Dunning, G.R. (1990) *Geological studies, Victoria Lake Group, central Newfoundland.* In: Current Research, Newfoundland Department of Mines and Energy, Geological Survey Branch, Report 90-1, pp. 131-144.

Evans, D.T.W., Kean, B.F., and Mercer, N.L. (1994) *Geology and Mineral Occurrences of Lake Ambrose, Map 94-223.* Government of Newfoundland and Labrador, Department of Mines and Energy, Geological Survey Branch, Open File 12A/10/0686.

Franklin, J.M (1993) *Volcanic-associated massive sulphide deposits.* In: Kirkham, R.V. Sinclair, W.D., Thorpe, R.I. and Duke, J.M. (Ed.): Mineral deposit modelling. Geological Association of Canada, Special Paper 40, pp. 315-334.

Foody, G.M. (1992) *On the compensation for chance agreement in image classification accuracy assessment.* Photogrammetric Engineering and Remote Sensing, Vol. 58, No. 10, pp. 1459-1460.

George, H. and Bonham-Carter, G.F. (1989) *Spatial modelling of geological data for exploration, Star Lake area, Saskatchewan.* In: Agterberg, F.P. and Bonham-Carter, G.F. (Ed.): Statistical applications in the earth sciences. Geological Survey of Canada, Paper 89-9, pp. 157-169.

Grant, D.R. (1974) *Prospecting in Newfoundland and the theory of multiple shrinking ice caps.* Geological Survey of Canada, Paper 74-1, Part B, pp. 215-216.

Grant, D.R. (1975) *Surficial geology of the Red Indian Lake map area, Newfoundland – a preliminary interpretation.* In: Report of Activities, Part B, Geological Survey of Canada, Paper 75-1, pp. 111-112.

Grant, D.R. and Tucker, C.M (1976) *Preliminary results of terrain mapping and base metal analysis of till in the Red Indian Lake and Gander Lake map areas of central*

Newfoundland. In: Report of Activities, Part A, Geological Survey of Canada, Paper 76-1A, pp. 283-285.

Hammond, R. and McCullagh, P. (1978) *Quantitative techniques in geography: an introduction (2nd edition).* Oxford University Press, Oxford, 364 pp.

Honarvar, P., Hogan, A., Kilfoil, G.J., King, D., Nolan, L.W., Ash, J.S., Colman-Sadd, S.P., Hayes, J.P., Liverman, D.G.E., Davenport, P.H., Kerr, A., and Evans, D.T.W. (1996) *A digital geoscience atlas on CD-ROM of the Buchans-Robert's Arm Belt.* In: Current Research, Newfoundland Department of Natural Resources, Geological Survey, Report 96-1, pp. 253-281.

Hornbrook, E.H.W., Davenport, P.H. and Grant, D.R. (1975) *Regional and detailed geochemical exploration studies in glaciated terrain in Newfoundland.* Department of Mines and Energy, Mineral Development Division, Report 75-2, 116 pp.

Hosmer, W.H, and Lemeshow, S. (1989) *Applied logistic regression.* John Wiley & Sons, Inc., Toronto, Canada, 307 pp.

James, L.D. and Perkins, E.W. (1981) *Glacial dispersion from sulphide mineralization, Buchans area, Newfoundland.* In: Swanson, E.A., Strong, D.F. and Thurlow, J.G. (Ed.): The Buchans orebodies: fifty years of geology and mining. Geological Association of Canada, Special Paper 22, pp. 269-283.

Kass, G.V. (1975) *Significance testing in automatic interaction detection (A.I.D.).* Applied Statistics, Vol. 24, No. 2, pp. 178-189.

Kass, G.V. (1980) *An exploratory technique for investigating large quantities of categorical data.* Applied Statistics, Vol. 29, No. 2, pp. 119-127.

Kean, B.F., and Jayasinghe, N.R. (1980) *Geology of the Lake Ambrose (12A/10) – Noel Paul's Brook (12A/9) map areas, central Newfoundland.* Newfoundland Department of Mines and Energy, Mineral Development Division, Report 80-2, 29 pp.

Keckler, D. (1995) *Surfer's triangulation with linear interpolation method in Surfer for Windows: contouring and 3D surface mapping, User's Guide, version 6.* Golden Software, Inc., Golden, Colorado, USA.

Kirkham, R.V., Sinclair, W.D., Thorpe, R.I. and Duke, J.M., (Ed.) (1993) *Mineral deposit modelling.* Geological Association of Canada Special Paper 40, St. John's, Canada, 798 pp.

Klassen, R.A. (1994) *Till geochemistry and ice flow data, central Newfoundland (NTS 12A/10, 12A/15, 12A/16, 12H/1).* Geological Survey of Canada, Open File 2823, 29 pp.

Klassen, R.A. (1997) *Surficial geology of the Lake Ambrose map area (NTS 12A/10).* Geological Survey of Canada, unpublished map.

Kvamme, K.L. (1990) *Spatial autocorrelation and the classic Maya collapse revisited: refined techniques and new conclusions.* Journal of Archaeological Science, Vol. 17, pp. 197-207.

Levinson, A.A. (1980) *Introduction to Exploration Geochemistry, Second Edition.* Applied Publishing Ltd., Wilmette, Illinois, 924 pp.

Lindqvist, L., Lundholm, I., Nisca, D., Espensen, K., Wold, S. (1987) *Multivariate geochemical modelling and integration with petrophysical data.* Journal of Geochemical Exploration, Vol. 29, pp. 279-294.

Lydon, J.W. (1988) *Volcanogenic massive sulphide deposits, Part 1: A descriptive model.* In: Roberts, R.G. and Sheahan, P.A. (Ed.): Ore deposit models. Geoscience Canada, Reprint Series 3, GSC, Ottawa, Ontario, 194 pp.

Merchant, D.C. (1987) *Spatial accuracy specifications for large scale topographic maps.* Photogrammetric Engineering and Remote Sensing, Vol. 53, No. 7, pp. 958-961.

Mihychuck, M. (1985) *Drift prospecting in the Victoria and Tally Pond areas, central Newfoundland.* In: Current Research, Newfoundland Department of Mines and Energy, Report 85-1, pp. 99-104.

Murray, R.C. (1955) *Directions of glacier ice motion in south-central Newfoundland.* Journal of Geology, Vol. 63, pp. 268-274.

Neary, G.N. (1981) *Mining history of the Buchans area.* In: Swanson, E.A, Strong, D.F., and Thurlow, J.G. (Ed.): The Buchans orebodies: fifty years of geology and mining. Geological Association of Canada Special Paper 22, pp. 1-64.

Nolan, L.W. (1990) *UNISTAT: statistical and graphics package for geochemists.* In: Current Research, Newfoundland Department of Mines and Energy. Geological Survey Branch, Report 90-1, pp. 129-130.

Norušis, M.J. (1990) *Logistic Regression analysis.* In: SPSS/PC+ Advanced Statistics 4.0. SPSS Inc., Chicago, USA, pp. B39 – B61.

O'Connor, P.J., Reimann, C. and Kuerzl, H. (1988) *An application of exploratory data analysis (EDA) techniques to stream sediment surveys for gold and associated elements in County Donegal, Ireland.* In: MacDonald, D.R. and Mills, K.A. (Ed.): Prospecting in areas of glaciated terrain - 1988. CIMM Geology Division, pp. 449-467.

O'Donnell, N.D. (1973) *Glacial indicator trains near Gullbridge, Newfoundland.* Unpublished M.Sc. thesis, Faculty of Graduate Studies, University of Western Ontario, 259 pp.

Reddy, R.K.T., and Bonham-Carter, G.F. (1991) *A decision-tree approach to mineral potential mapping in Snow Lake area, Manitoba.* Canadian Journal of Remote Sensing, Vol. 17, No. 2, pp. 191-200.

Robertson, G.P. (1998) *GS$^+$: Geostatistics for the environmental sciences.* Gamma Design Software, Plainwell, Michigan, USA, 65 pp.

Rock, N.M.S. (1988) *Numerical geology*. Springer-Verlag, New York, USA, 427 pp.

Rogers, P.J. (1988) *Gold in lake sediments: implications for precious metal exploration in Nova Scotia.* In: MacDonald, D.R. and Mills, K.A. (Ed.): Prospecting in areas of glaciated terrain - 1988. CIMM Geology Division, pp. 357-381.

Shilts, W.W. (1976) *Glacial till and mineral exploration.* In: Legget., R.F. (Ed.): Glacial till. Royal Society of Canada, Special Publication 12, pp. 205-224.

Shilts, W.W. (1993) *Geological Survey of Canada's contributions to understanding the composition of glacial sediments.* Canadian Journal of Earth Sciences, Vol. 30, pp. 333-353.

Sibley, D. (1991) *Spatial applications of exploratory data analysis.* Concepts and Techniques in Modern Geography, No. 49, Hull University, England, 37 pp.

Sparkes, B.G. (1985) *Quaternary mapping, Central Volcanic Belt.* In: Current Research, Newfoundland Department of Mines and Energy, Geological Survey Branch, Report 85-1, pp. 94-104.

Sparkes, B.G. (1987) *Glacial geology and till geochemistry of the Buchans (12A/15) map area, Newfoundland.* Mineral Development Division, Newfoundland Department of Mines and Energy, Open file 12A/15 (396), 11 pp.

Stapleton, G.J. (1999) *Guidelines and recommendations standard for the entry of data into the MODS database (Version I).* Geological Survey, Newfoundland and Labrador Department of Mines and Energy, pp. 61.

Stapleton, G.J., and Smith, J.L. (1999) *Mineral inventory reports for Newfoundland and Labrador.* Government of Newfoundland and Labrador, Department of Mines and Energy, Geological Survey, Open File, NFLD/2688 (CD).

Stapleton, G., Smith, J.L., Pollock, J.C., and Way, B.C. (2000) *Mineral occurrence data system.* In: Current Research, Newfoundland Department of Mines and Energy, Geological Survey, Report 2000-1, pp. 341-348.

Strobel, M.L., and Faure, G. (1987) *Transport of indicator clasts by ice sheets and transport half-distance: a contribution to prospecting for ore deposits.* Journal of Geology, Vol. 95, pp. 687-697.

Swinden, H.S. (1992) *Metallic mineral deposits in Newfoundland and Labrador: A workbook prepared for The Prospectors Training Course, Stephenville, Newfoundland, June 3, 1992.* Geological Survey Branch, Department of Mines and Energy, 94 pp.

Swinden, H.S., Jenner, G.A., Kean, B.B. and Evans, D.T.W. (1989) *Volcanic rock geochemistry as a guide for massive sulphide exploration in central Newfoundland.* In: Current Research, Newfoundland Department of Mines and Energy, Geological Survey of Newfoundland, Report 89-1, pp. 201-219.

Tabachnick, B.G. and Fidell, L.S. (1996) *Using multivariate statistics (3rd edition).* HarperCollins College Publishers, NY, USA, 880 pp.

Taylor, D.M., St. Croix, L. and Vatcher, S.V. (1993) *Newfoundland striation database.* Newfoundland Department of Mines and Energy, Geological Survey Branch, Open File NFLD 2195 (version 2), 142 pp.

Vanderveer, D.G. and Sparkes, B.G. (1979) *Geochemistry of glacial till samples, Lake Ambrose (12A/10)-Noel Paul's Brook (12A/9) map area, Newfoundland.* Newfoundland Department of Mines and Energy, Mineral Development Division, Open File 12A (212), 27 pp.

Vanderveer, D.G. and Sparkes, B.G (1982) *Regional quaternary mapping: an aid to mineral exploration in west-central Newfoundland.* In: Davenport, P.H. (Ed.): Prospecting in areas of glaciated terrain. Canadian Institute of Mining, Geology Division Publication, pp. 284-299.

Williams, H., Colman-Sadd, S.P. and Swinden, H.S. (1988) *Tectonic-stratigraphic subdivisions of central Newfoundland.* In: Current Research, Part B. Geological Survey of Canada, Paper 88-1B, pp. 91-98.

Wright, D.F., Bonham-Carter, G.F., and Rogers, P.J. (1988) *Spatial data integration of lake sediment geochemistry, geology and gold occurrences, Meguma Terrane, eastern Nova Scotia.* In: MacDonald, D.R. and Mills, K.A. (Ed.): Prospecting in areas of glaciated terrain - 1988. CIMM Geology Division, pp. 501-515.

Appendix A: Programs

Program A1: Calculation of 'S2' used in Moran's I Calculations

```
'Program used to calculate S2, used in calculating Moran's I statistical sig-
' nificance test under the randomization assumption using the tables
' as outlined in Kvamme's 1990 paper, pg. 205, for point data
CLS
' A/I pointdistance output file has pnti, pntj, and distance columns
INPUT "Enter ascii distance point file name with extension: "; infile1$

OPEN infile1$ FOR INPUT AS #1

'set sum of weights=0  and total sum of weights=0
sumweight = 0
totalsw = 0

'input 1st pair of points
INPUT #1, pnti1, pntj1, distance
IF distance <> 0 THEN
  sumweight = sumweight + (1 / distance)
END IF

'input subsequent points
DO WHILE NOT EOF(1)
  INPUT #1, pnti2, pntj2, distance
' LOCATE 5, 10: PRINT "Working"

'if first of second pair = first of first pair then keep adding weights
  IF pnti1 = pnti2 THEN
     IF distance <> 0 THEN
       sumweight = sumweight + (1 / distance)
     END IF
  ELSE
     PRINT "For i= "; pnti1; " sum of weights = "; sumweight
     pnti1 = pnti2
     pntj1 = pntj2
     totalsw = totalsw + (sumweight * sumweight)
     IF distance <> 0 THEN
       sumweight = 1 / distance
     ELSE
       sumweight = 0
     END IF
  END IF
LOOP

'print out last sum of weights and final results
   PRINT "For i= "; pnti1; "sum of weights = "; sumweight
   totalsw = totalsw + (sumweight * sumweight)
   PRINT "Total sum of squares for Table A3 = "; totalsw
CLOSE #1
END
```

Program A2: Calculation of the Local Spatial Weighting Function
(swfindst.bas)

```
'Program used to calculate the local spatial weighting function of mineral
' and nonmineral occurrences.
'The SWF is calculated for each point, using an arcinfo point-distance output file.
'This version calculates the SWF for maximum distances of 2500m radius around
'a mineral occurrence, and uses inverse distance as a weight
'based on concept by Chou et al., 1990
CLS
' A/I pointdistance output file has pnta, pntb, and distance columns
INPUT "Enter ascii distance point file name with extension: "; infile1$
'enter eg dist2500.txt

OPEN infile1$ FOR INPUT AS #1

'output file name is same as input file but with swd extension
n = INSTR(1, infile1$, ".")
filen$ = LEFT$(infile1$, (n - 1))
outfile1$ = filen$ + ".swd"
OPEN outfile1$ FOR OUTPUT AS #2

'  Calculate Spatial Weighting Function
'
'set sum of all weights = 0
sumallw = 0
'set sum of mineral occurrence weights=0
mosumw = 0

'input 1st pair of points
INPUT #1, pnti1, pntj1, distance

IF distance <> 0 THEN

'in this version, weight=(1/d) and sum all weights for min occ and
'divide by weights for min occ and nonmin occ
'  based on Chou's and Kvamme's papers
  weight = (1 / distance)
  sumallw = sumallw + weight
'if j is a mineral occurrence then add weight to summinocc weight
  IF pntj1 < 48 THEN
    mosumw = mosumw + weight
  END IF

END IF
PRINT pnti1; " "; sumallw; " "; mosumw; " "

'input subsequent points
DO WHILE NOT EOF(1)
  INPUT #1, pnti2, pntj2, distance
```

```
' LOCATE 5, 10: PRINT "Working"
'check if pnti2 part of set of pnti1
  IF pnti2 = pnti1 THEN

'check distance
    IF distance <> 0 THEN

      weight = (1 / distance)
      sumallw = sumallw + weight
'check if j is a minocc, if yes then add weight to mosumw
      IF pntj2 < 48 THEN
        mosumw = mosumw + weight
      END IF
    END IF
'otherwise point is part of a new set or not close to a min occ
   ELSE
      IF sumallw > 0 THEN
        swf = mosumw / sumallw
      ELSE
        swf = 0
      END IF
      PRINT #2, USING "### "; pnti1;
      PRINT #2, USING "##.##"; swf

      pnti1 = pnti2
      pntj1 = pntj2
      IF distance <> 0 THEN
        weight = 1 / distance
      ELSE
        weight = 0
      END IF
      sumallw = weight
      IF pntj2 < 48 THEN
        mosumw = weight
      ELSE
        mosumw = 0
      END IF
   END IF
LOOP

'print out last results
   IF sumallw > 0 THEN
     swf = mosumw / sumallw
   ELSE
     swf = 0
   END IF
   PRINT #2, USING "### "; pnti1;
   PRINT #2, USING "##.##"; swf
CLOSE #1
CLOSE #2
END
```

Program A3: Outcrop Point-to-Line Conversion Program

```
'Program to convert outcrop xy point locations to idrisi vec line file
' containing a small line trending in the bedrock direction
'  for use in eliptical buffering to obtain a reliability map
INPUT "Enter outcrop point filename with extension: "; infile$
OPEN infile$ FOR INPUT AS #1
'output is same as input but with vec and dvc ending
n = INSTR(1, infile$, ".")
filen$ = LEFT$(infile$, (n - 1))
outfile1$ = filen$ + "1.vec"
OPEN outfile1$ FOR OUTPUT AS #2
outfile2$ = filen$ + "1.dvc"
OPEN outfile2$ FOR OUTPUT AS #3
CLS
'input angle of trend and lenth of line to buffer
INPUT "Enter angle (in degrees) of geological trend: "; trend
INPUT "Enter length (in m) of line to buffer: "; length
'determine min and max utmeast and utmnorth of a line
' centred on outcrop location of length length and angle trend
radian = trend * 3.141593 / 180
WHILE NOT EOF(1)
'read input file utmeasting, northing and pointype 1=oc, 2=dh
INPUT #1, oceast, ocnorth, pointype
maxeast = oceast + (.5 * length * COS(radian))
maxnorth = ocnorth + (.5 * length * SIN(radian))
mineast = oceast - (.5 * length * COS(radian))
minnorth = ocnorth - (.5 * length * SIN(radian))
PRINT #2, pointype; 2
PRINT #2, USING "###### "; mineast;
PRINT #2, USING "####### "; minnorth
PRINT #2, USING "###### "; maxeast;
PRINT #2, USING "####### "; maxnorth
WEND
PRINT #2, "0 0"
PRINT #3, "file title  : oc buffer lines"
PRINT #3, "id type     : integer"
PRINT #3, "file type   : ascii"
PRINT #3, "object type : line"
PRINT #3, "ref. system : plane"
PRINT #3, "ref. units  : m"
PRINT #3, "unit dist.  : 1"
PRINT #3, "min. x      : 500000"
PRINT #3, "max. x      : 537000"
PRINT #3, "min. y      : 5371600"
PRINT #3, "max. y      : 5399600"
PRINT #3, "pos'n error : unknown"
PRINT #3, "resolution  : 25"
CLOSE #1
CLOSE #2
CLOSE #3, END
```