Computer Programs for Mineral Exploration

FREDERIK P. AGTERBERG

The search for undiscovered orebodies often begins by targeting areas on two-dimensional surface maps that are favorable for more detailed exploration, such as three-dimensional geophysical surveys and drilling to obtain large numbers of core samples for geochemical analysis. Target selections are based on diverse geological information on rock types, mineralogy, stratigraphic relations, and structure. Geological maps for a study region usually are augmented by remote-sensing data and information on regional geochemistry and geophysics. Theoretical models of the genesis and occurrence of orebodies are used for integrating the many different types of data.

During the past few years there has been an upsurge in the use of computers for finding orebodies. Geoscientists have either written their own programs, mostly in Fortran, LISP, C, Pascal, or BASIC, or developed their techniques within the framework of commercially available software packages that include intelligent geographic information systems (GISs) and expert system shells.

Computer-based techniques for problem-solving in mineral exploration data integration are discussed in this review, with an emphasis on the models on which the software is based. The output of most of these programs is a prognostic map that indicates where hidden orebodies may occur. Such maps may highlight combinations of features which, in places that are relatively well explored, are spatially correlated with known orebodies, or display estimated probabilities for the occurrence of mineral deposits. How the uncertainties of facts and the relations between facts are determined and propagated through the system is of critical importance in these programs.

Historical Perspective

The history of computer applications in mineral exploration is illustrated in Fig. 1. The earliest computer programs were for regression, discriminant analysis, and other multivariate techniques (*1*). For example, total amount of ore per unit area was regressed on explanatory variables for the type of rock, the geological age, and tectonic environment. Similar problems can now be solved rapidly with commercially available statistical software packages such as SAS (2), which also provide procedures for obtaining various regression diagnostics.

The diagnostic problems in mineral exploration are similar to those in medicine. As early as 1556, Agricola in "De Re Metallica" proposed methods for reading signs on the surface of the earth in order to find mineral deposits. At about the same time, Paracelsus (1493–1541) created a theory of similarity for treating disease. It can be argued (*3*) that probability concepts grew out of such early rules for reading signs. The development of PROSPECTOR (*4*), an expert system for mineral exploration, began in about 1975. It was preceded by MYCIN (*5*), one of the first expert systems in medicine for handling subjective and heuristic knowledge of expert physicians to diagnose infectious diseases and to provide antimicrobial therapy. In PROSPECTOR, statements on evidence and hypotheses are linked through a hierarchical network. Each link between statements is characterized by numbers that are not based on direct observation but are subjective guesses by experienced scientists.

The logic for propagating uncertainty through the networks of expert systems generally is not in agreement with basic principles of probability theory. For example, in both MYCIN and PROSPECTOR, "fuzzy" logic may be used to combine uncertain facts at the base of the network (*6*). If these facts have different probabilities of being true, then the probability that they are all true simply is set equal to the smallest of the different probabilities. Such a rule may not give good results in practice. Mathematical statisticians (7) have proposed the use of networks that are logically coherent, of which GLADYS (*8*) is an example.

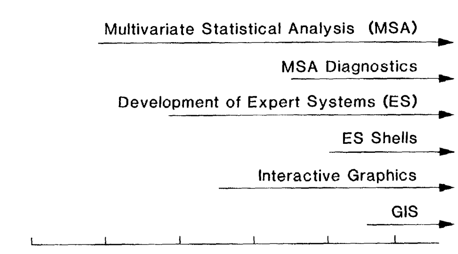


Fig. 1. Time-line representation of the applications of computer-based techniques in mineral exploration.

The way in which MYCIN made inferences was found suitable for other medical diagnosis problems and led to the development of EMYCIN (or empty MYCIN), one of the first expert system "shells" (*9*). A shell does not contain a knowledge base, but consists only of internal operative mechanisms for knowledge representation and probability updating. Expert system shells currently used by geoscientists include GEOMYCIN (*10*) and The Deciding Factor (*11*), the latter being derived from PROSPECTOR. Commercially available expert shells generally are useful for decision-making in complex deterministic situations with many "if-then" rules.

The advent of interactive graphics in the late 1970s facilitated the development of spatially based systems in the earth sciences that allowed rapid comparison of different map patterns. The extraction and display of salient features from large digital databases is widely used in mineral exploration. Early interactive graphic systems included SIMSAG (*12*) for multivariate statistical analysis of data in mineral resource evaluation, GIAPP (*13*) for geological image analysis, and NCHARAN (*14*) for characteristic analysis, this latter being a simplified form of principal component analysis. A follow-up study on target areas outlined by NCHARAN in the Grong area of central Norway resulted in the identification of a previously unknown sulfide vein system (*14*). During the 1980s most mineral exploration programs, including interactive graphic systems, could be run on microcomputers.

Recent GISs such as MAPS (*15*) resemble expert system shells in their knowledge representation as well as logical operations. For mineral exploration, intelligent GISs have the advantage of being based on maps. For example, SPANS (*16*) allows the rapid processing of map images because it has a hierarchical quadtree data structure (that is, the region containing the data, or the cell, is initially divided into squares by a gridding procedure; a process is then started such that each square containing data points is subdivided into four squares).

Examples of computer programs recently developed by geoscientists are muPROSPECTOR (*17*) and PROSPECTOR III (*18*), which are microcomputer and symbolic workstation-based versions of PROS-

Geological Survey of Canada, Ottawa, Ontario, Canada K1A 0E8.

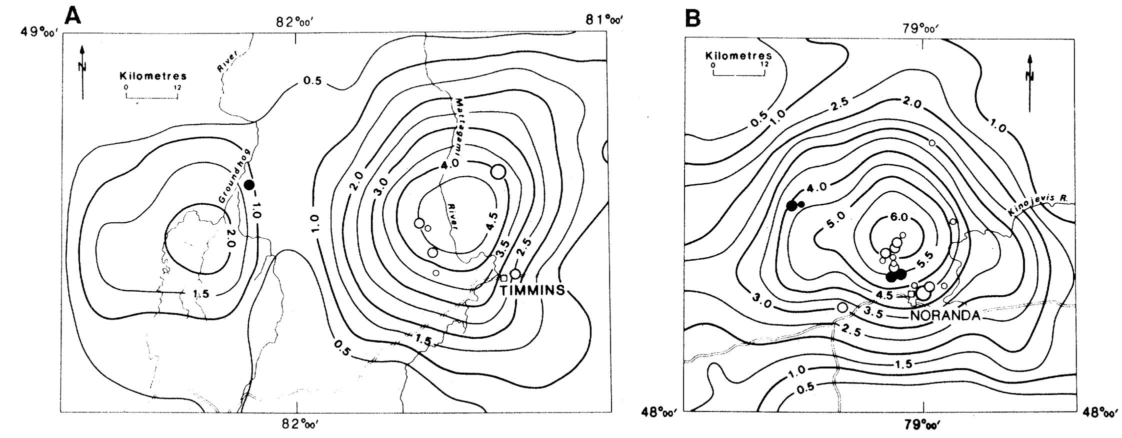


Fig. 2. Prognostic contours and occurrences of copper deposits in favorable areas near (**A**) Timmins, Ontario, and (**B**) Noranda, Quebec. Open circles denote known deposits used in 1972 to construct contours for predicted number of cells containing one or more copper deposits per unit area. Solid circles denote later discovered copper deposits.

PECTOR, respectively; GEOVALUATOR (*19*) for regional mineral resource appraisal (based on muPROSPECTOR); EXPLOR (*20*), which is primarily designed to analyze contour maps derived from wells to find oil deposits; and FINDER (*21*), which can be used to identify massive sulfide deposits through geochemical alteration data. These programs are usually used by small teams of scientists for research; information on obtaining and implementing these programs is given in the references cited.

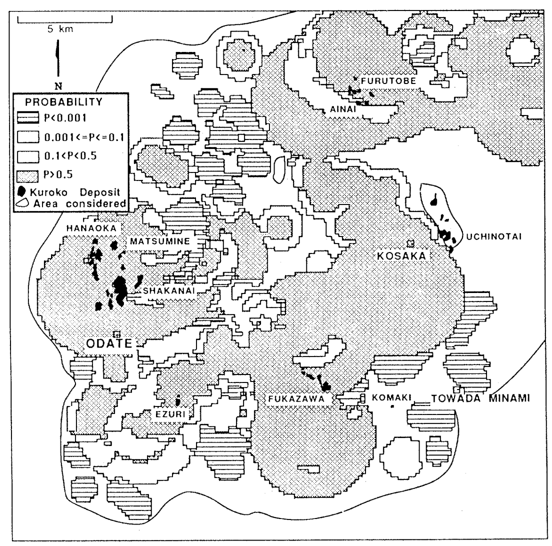


Fig. 3. Prognostic probability map generated by FINDER for Kuroko deposits in Hokuroku Mining District, Japan, based on sodium depletion and sericite and gypsum plus anhydrite (*21*).

Examples of Multivariate Statistical Analysis

In about 1970 it became possible to perform multivariate statistical analysis of large regional data sets that were systematically coded for small cells belonging to grids superimposed in maps. Probabilities estimated either by stepwise regression or the logistic model can be used to construct contour maps of the expected frequency of undiscovered mineral deposits per unit area. In Fig. 2 the locations of later discoveries are compared with frequency contours for polymetallic massive sulfide deposits in the Abitibi Volcanic Belt of the Canadian Shield (*22*) on a mineral potential map derived by multivariate analysis in 1972. The dependent variable for the presence or absence of sulfide deposits was regressed on explanatory variables obtained by coding rock types and regional geophysical maps for square cells 10 km on a side. The estimated probabilities for such cells were added to obtain frequencies for unit areas 40 km by 40 km and contoured. The validity of such prognostic contours could be assessed statistically (*23*). However, in most multivariate statistical analysis methods it is difficult to interpret the estimated coefficients. Although diagnostics now can be used to evaluate the effect of individual observations on estimated probabilities and coefficients, the latter generally cannot be used as weights for extrapolation to other regions.

In applications of multivariate analysis, the study region should contain sufficient information on the type of deposit considered and its relations with the explanatory variables quantifying the geological framework. These conditions are rarely fulfilled. Normally, facts for similar mineral deposits elsewhere should be included. Polymetallic massive sulfide deposits, which were originally formed on the seafloor, are exceptional in that they permit multivariate analysis on a regional basis.

The program FINDER (*21*) was used recently for targeting the Kuroko deposits in the Hokuroku District in Japan, which are polymetallic massive sulfide seafloor deposits of Miocene age. FINDER combines concepts of spatial statistics with frequency distribution information. It is assumed that every sulfide deposit occurs at the center of an ellipse representing its alteration zone. Inside the ellipse, chemical composition of the volcanic rocks hosting the deposit is assumed to differ from the regional background. The study area is divided into a number of small cells. All of these are assigned a prior probability of containing a deposit. Revised (posterior) probabilities are calculated for each cell on the basis of the spatial arrangement of the samples and the size of the elliptically shaped target. The latter can be regarded as a template that is successively centered over each of the small cells. All samples under the template are used to revise the prior probability. The underlying statistical model is that altered rock and background each have a multivariate normal distribution for all chemical elements considered. An example of FINDER's output is shown in Fig. 3. Most of the known Kuroko deposits occur in areas with relatively high posterior probability. The map can be used for targeting unknown Kuroko deposits in the same district.

PROSPECTOR

In the original PROSPECTOR, the need was recognized to integrate, on a worldwide basis, all information that may be relevant to the possible occurrence of a type of mineral deposit. For each deposit type, a network of statements was created by an expert economic geologist assisted by a knowledge engineer (*24*). PROSPECTOR has been credited with the recognition of a hidden molybdenum deposit near Mount Tolman in Washington (*25*). Some of the rules used in this application schematically are shown in Fig. 4A (*26*). The boxes represent statements of evidence or hypotheses, whose interrelation is represented by connecting arrows. Most of these arrows are accompanied by two numbers (LS and LN) that represent the likelihood measures for "sufficiency" and "necessity," respectively. The prior probability of a statement at the end of an arrow is either increased or decreased by LS or LN, depending on the presence or absence of the supporting evidence. The prior probabilities and likelihood measures were subjectively estimated.

The rules of PROSPECTOR are a mixture of statistical procedures and heuristic reasoning, that is, an approximate version of Bayes' rule is used. At the base of the particular network shown in Fig. 4A, which uses information from soil geochemistry, the statements deal with the situation that the molybdenum deposit of interest is likely to occur at a positive molybdenum anomaly but away from a gold anomaly. Suppose that the first of these two statements of evidence (X near peak Mo) is labeled E. If E is true, the prior probability $P(H) = 0.17$ of H, which represents the subsequent hypothesis (suggestive Au/Mo at X) being true, is changed by the factor $LS = 20$ as follows. The probability P is replaced by its corresponding odds $O = P/(1 - P)$ $[O(H) = 0.2048]$. The prior probability of E ($= 0.1$ in Fig. 4A) is ignored because E is true. The conditional odds $O(H|E)$ of H being true given the evidence satisfy $O(H|E) = LS \times O(H) = 4.096$, so that the posterior probability $P(H|E)$ becomes 0.804. However, if $\tilde{E}$ is true (that is, E is false), $O(H)$ would be multiplied by $LN = 1$, so that $P(H)$ would remain unchanged. The subjective numbers of Fig. 4A are not logically coherent. If Bayes' rule was satisfied, $LN = 1$ would imply $LS = 1$ (see below).

If, in addition to E being true, the other evidence statement (X "away from" Au) is true, then $O(H|E) = 4.096$ is multiplied by 3 (= LS for other evidence) to yield a higher posterior probability for H (suggestive Au/Mo at X) equal to 0.925. The logical AND statement ($\cap$) in the box for the hypothesis "favorable Au/Mo at X" in Fig. 4A means that the previous hypothesis "suggestive Au/Mo at X" may be replaced by this stronger statement if there is "200 to 400 ppm at X."

PROSPECTOR assumes that if "favorable Au/Mo at X" is at its prior probability $P(E) = 0.1$, the prior probability of "favorable soil geochemistry" (0.17) would be changed into the posterior probability $P(H|E) = 0.381$ through $O(H|E) = 0.6145$. The previously computed probability (= 0.925) of "suggestive Au/Mo at X" then is propagated by the following rule. The posterior probability of "favorable soil geochemistry" should be the fraction $(0.925 - 0.1)/(1 - 0.1) = 0.9167$ of the way between 0.17 and 0.6145, or $0.17 + 0.9167 \times (0.6145 - 0.17) = 0.577$. The probability of the evidence being false is not considered in this rule.

Weights of Evidence

The preceding rules in PROSPECTOR are based only in part on probability theory (*27*). The two main obstacles to logical coherence in the network of Fig. 4A are that:

1) Each rule for proceeding from evidence (E) to hypothesis (H) in the system is overspecified because four parameters, $P(E)$, $P(H)$, LS, and LN, were defined, whereas only three parameters are necessary statistically; and

2) If LS and LN were likelihood ratios in a statistical sense, then $LN = 1$ would imply $LS = 1$, and vice versa. Relatively minor changes in the numbers of the network of Fig. 4A would make them logically coherent, as is shown in Fig. 4, B and C, for two portions of the network, at its top and base, respectively. The four probabilities for intersections of E, H, and their complements can be represented as

1	$P(E)$	$P(\tilde{E})$
$P(H)$	$P(H \cap E)$	$P(H \cap \tilde{E})$
$P(\tilde{H})$	$P(\tilde{H} \cap E)$	$P(\tilde{H} \cap \tilde{E})$

Any three probabilities in this table that are not in the same row or column completely define the system because the sum of the four probabilities for the logical intersections equals 1. If H and E are statistically independent, $P(H \cap E) = P(H) \times P(E)$ and $P(H) = P(H|E) = P(H|\tilde{E})$.

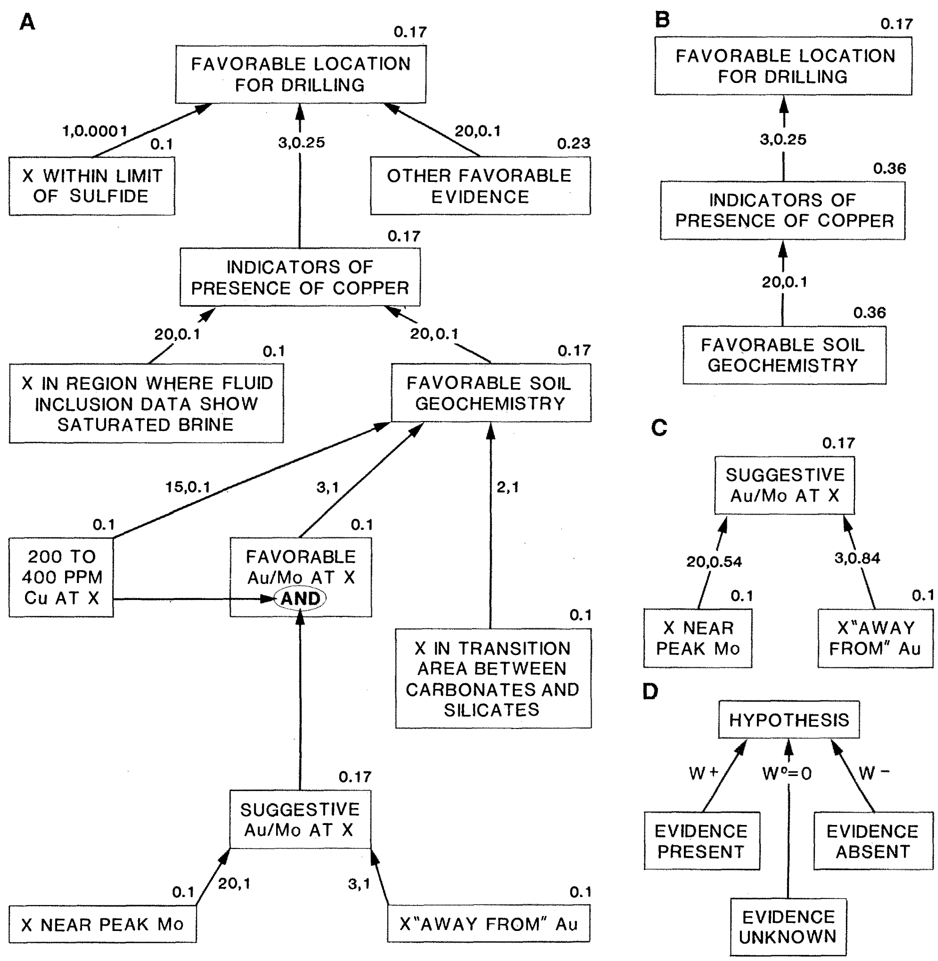


Fig. 4. (**A**) Part of the PROSPECTOR network for porphyry molybdenum deposits (*26*). (**B**) Example of two consecutive links in network of uncertain statements that are logically coherent. (**C**) Example of two parallel links in network of uncertain statements that are logically coherent. (**D**) Connection between "evidence" and "hypothesis" in the weights of evidence method.

If two likelihood ratios $L_1 = P(E|H)/P(E|\tilde{H})$ and $L_2 = P(\tilde{E}|H)/P(\tilde{E}|\tilde{H})$ are defined, then $O(H|E) = L_1 \times O(H)$ and $O(H|\tilde{E}) = L_2 \times O(H)$. Although it would seem that L_1 and L_2 are identical to LS and LN, it also follows that $P(E) = [P(H) - P(H|\tilde{E})]/[P(H|E) - P(H|\tilde{E})]$. In Fig. 4A, the prior probability of "favorable location for drilling" is $P(H) = 0.17$. Setting $L_1 = 3$ and $L_2 = 0.25$ gives $P(H|E) = 0.381$ and $P(H|\tilde{E}) = 0.049$, respectively, so that $P(E) = 0.364$ (Fig. 4B) instead of 0.17 as in Fig. 4A. Downward propagation of 0.364 by the same method yields a prior probability of "favorable soil geochemistry," which is equal to 0.358 (Fig. 4B) instead of 0.17 (Fig. 4A). A logically coherent network in which all three prior probabilities remain equal to 0.17 (as in Fig. 4A) could be constructed, but then at least one of the numbers in each of the two pairs (LS, LN) must be changed. The other difference between the pairs (L_1, L_2) and (LS, LN) is that L_1 and L_2 either both equal 1 or differ from 1 [the latter occurs if $P(H) = P(H|E) = P(H|\tilde{E})$]. The partial network of Fig. 4C was made logically coherent by changing the values of LN that were equal to 1 in Fig. 4A.

The previous discussion was given in detail to help preface the weights of evidence method illustrated in Fig. 4D (*27*). A distinction is now made between evidence that is either present, absent, or unknown. The weights of evidence corresponding to these three possibilities are $W^+ = \ln L_1$, $W^- = \ln L_2$, and $W^0 = 0$. The third possibility ($W^0 = 0$) represents the situation that $P(H|E) = P(H|\tilde{E}) = P(H)$, implying that the prior probability of the hypothesis remains unchanged when there is no evidence. This situation, in which $L_1 = L_2 = 1$, cannot be distinguished from presence (or absence) of "evidence" that is statistically independent of the hypothesis.

Unless both are equal to zero, W^+ and W^- must have opposite signs. Normally, the existence of evidence (such as a "positive" anomaly) implies that the corresponding hypothesis is strengthened, so that normally $W^+ > 0$ and $W^- < 0$. However, it may be useful to define evidence with $W^+ < 0$ and $W^- > 0$, such as when the weights of different features are compared with one another.

Combining Weights of Evidence

The weights of different types of evidence for the same hypothesis can be added provided that the features considered are conditionally independent of the hypothesis. The validity of the assumption of conditional independence can only be tested if probabilities are available for the hypothesis being true or false under all possible combinations of evidence. Suppose that "X near peak Mo" in Fig. 4C is written as E_1, "X away from Au" as E_2, and "suggestive Au/Mo at X" as H. The eight different outcomes imply seven degrees of freedom. Acceptance of the seven numbers specified for the two links in Fig. 4C would mean that two parameters remain to be specified. (Separate links between boxes have three parameters, so that two links have only five parameters if one box is shared.) Conditional independence would imply that $P(E_1 \cap E_2|H) = P(E_1|H) \times P(E_2|H)$ and that $P(E_1 \cap E_2|\tilde{H}) = P(E_1|\tilde{H}) \times P(E_2|\tilde{H})$, which is equivalent to $P(E_1 \cap E_2 \cap H) = P(E_1|H) \times P(E_2|H) \times P(H)$ and $P(E_1 \cap E_2 \cap \tilde{H}) = P(E_1|\tilde{H}) \times P(E_2|\tilde{H}) \times P(\tilde{H})$. With the seven numbers shown in Fig. 4C, these two additional conditions result in a fully specified system. If the weights for E_1 and E_2 are written as W_1 and W_2, then $\ln O(H|E_1 \cap E_2) = W_1^+ + W_2^+ + \ln O(H)$, and similar rules apply for combining evidence including $\tilde{E}_1$ or $\tilde{E}_2$. Thus if E_1 and $\tilde{E}_2$ are true, H has a posterior probability $P(H|E_1 \cap \tilde{E}_2) = 0.77$, as follows from $\ln O(H|E_1 \cap \tilde{E}_2) = \ln(20) + \ln(0.84) + \ln(0.17/0.83) = 1.236$. In the situation of Fig. 4B, it could be assumed that H and E_1 are conditionally independent of E_2, implying that $P(H|E_1 \cap E_2) = P(H|E_2)$ and $P(\tilde{H}|E_1 \cap E_2) = P(\tilde{H}|E_2)$. If E_2 is true, the probability H would increase from 0.17 to 0.381, independent of whether E_1 is true.

Suppose that, as is common practice in backward chaining expert systems, the user is asked questions regarding evidence in sequence, moving down from the top toward the base of the network. Then, if evidence near the top is available, there may be no need to proceed further downward through the network. Obviously, this strategy is in accordance with the preceding assumption of conditional independence of H and E_1 with respect to E_2. Suppose that E_2 is missing but that E_1 is true. One could proceed one step downward into the network and calculate $P(H|E_1) = P(E_1 \cap H)/P(E_1) = [P(E_1 \cap E_2 \cap H) + P(E_1 \cap \tilde{E}_2 \cap H)]/P(E_1)$. From the numbers of Fig. 4B, $P(H|E_1) = 0.354$, which is only slightly less than $P(H|E_2) = 0.381$.

By continuing to make assumptions of conditional independence, large, logically coherent networks can be constructed. GLADYS (*8*) provides an example of an expert system based on the assumption $P(H|E_1 \cap E_2 \cap ... E_n) = P(H|E_1) \times P(H|E_2) \times ... P(H|E_n)$. The probabilities of n symptoms in GLADYS were determined by means of sampling experiments and the assumption of conditional independence is tested statistically. A similar approach discussed below can be used in regional mineral exploration when the probabilities are determined from areas measured by using a geographic information system.

Integration of Datasets for Gold Exploration in Nova Scotia

In this example, the weights of evidence method was combined with a GIS called SPANS (*16*) to determine probabilities for map patterns. SPANS uses a raster data structure with a variable pixel size. Raster images up to a maximum resolution of 2^{15} by 2^{15} pixels can be handled, although normally most SPANS applications deal with maps with a quad level of 10 to 12, that is, with a size between 2^{10} and 2^{12} (1024 by 1024 and 4096 by 4096 pixels). Typical hardware needed to run SPANS is an 80386 DOS-based personal computer with a 70-Mbyte hard drive. The system accepts a variety of vector and raster data inputs, allows forward and backward transformations from about 20 cartographic projections to geographic coordinates and provides a powerful set of analytical tools for analyzing multiple maps. Because SPANS permits the user to move readily to DOS, other DOS-compatible software (such as editors, statistical packages, and locally de-

Table 1. Meguma Terrane, central Nova Scotia. Binary map patterns were used to estimate posterior probabilities of a gold deposit occurring in a 1-km² unit. Weights W^+ for presence or W^- for absence were added to the logarithms of odds of prior probability. Errors are standard deviations.

Map pattern	Corridor width (km)	Area (km²)	Gold deposits	W^+	W^-
Geochemical signature		165	10	1.00 ± 0.33	−0.10 ± 0.13
Anticline axes	2.5	1276	50	0.55 ± 0.14	−0.77 ± 0.24
Northwest lineaments	1.0	750	17	−0.02 ± 0.25	0.01 ± 0.14
Granite contact	1.0	383	12	0.32 ± 0.29	−0.06 ± 0.14
Goldenville-Halifax contact	2.0	1030	34	0.37 ± 0.17	−0.27 ± 0.17
Halifax Formation		442	3	−1.24 ± 0.58	0.12 ± 0.13
Goldenville Formation		2021	63	0.31 ± 0.13	−1.47 ± 0.45
Devonian granite		482	2	−1.74 ± 0.71	0.16 ± 0.12

veloped programs) can be executed on mutually shared data files.

In addition to presence or absence of rock types, five potentially useful exploration indicator patterns for occurrence of gold deposits were digitized for Meguma Terrane in central Nova Scotia (*28*) (Table 1). The mechanism of gold mineralization is not well understood, and each of the five presence-absence indicator patterns reflects a separate prospecting philosophy for the favorable occurrence of gold deposits (*29*). Positive lake sediment geochemical anomalies provided a polygonal pattern for parts of the region only. The other four patterns consisted of corridors representing proximity to linear or curvilinear features. The half-width of these corridors was optimized by maximizing the contrast $C = |W^+| + |W^-|$, which measures the strength of correlation between a corridor pattern and the point pattern for known gold deposits. For example, two corridor patterns are shown in Fig. 5, A and B. The contrast C as a function of half-corridor width for proximity to anticline axes is shown in Table 2.

The weights in Table 1 suggest that positive geochemical anomalies are the best indicators of gold mineralization, followed by proximity to anticline axes. Relatively few gold deposits occur outside the corridors in Fig. 5A, and this is reflected by the absolute value of $W^-(= -0.7735)$ exceeding $W^+(= 0.5452)$ for this pattern. The maximum contrast for proximity to Goldenville-Halifax contact ($= 0.3683 + 0.2685 = 0.6368$) is less than one-half of that for the anticline axes ($= 1.3187$) (compare Fig. 9A with 9B).

In this application, the probability that a small unit area, measuring 1 km on a side, contains a gold deposit was set equal to total number of known gold deposits divided by total area measured in square kilometers. The weights for presence or absence of each of the patterns available for each unit area were added to the logarithm of the prior odds in order to obtain the logarithm of the posterior odds. The corresponding posterior probabilities are shown for part of the study area in Fig. 5C. Weights of evidence and prior probability have their own uncertainties, which can be estimated and added to obtain standard deviations for the estimated posterior probabilities. Missing patterns provide additional uncertainty which also was considered. Only posterior probabilities with t test values greater than 2 ($P < 0.05$) are shown in Fig. 5C.

The prior probability $P(H)$ of occurrence of a gold deposit in a unit cell was based on known deposits only. Suppose that it would be increased by assuming that there are undiscovered deposits in the region. New posterior probabilities can be computed with the weights kept constant, which would be equivalent to assuming that the undiscovered deposits are associated with the indicator patterns in the same way as the known deposits.

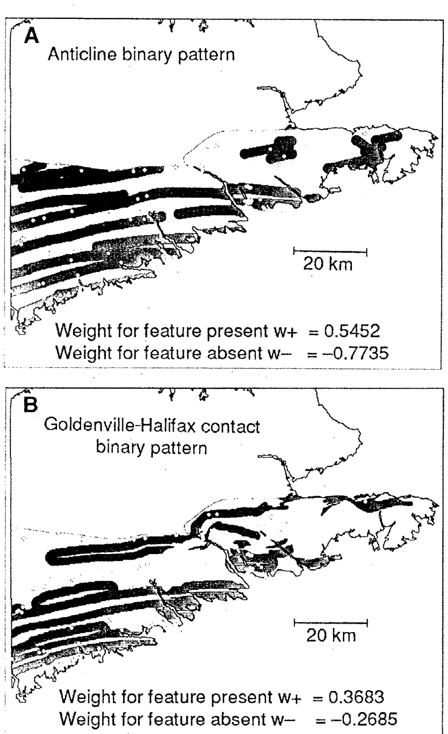
As explained in the previous section, weights may only be added if their patterns are conditionally independent. This assumption was tested (i) for all possible pairs of indicator patterns by using procedures from discrete multivariate analysis (*30*) and (ii) by using the pattern of posterior probabilities to obtain expected frequencies for occurrence of gold deposits for comparison with observed frequencies in unit cells with the same posterior probabilities. Each chi-square test for conditional independence of a pair of two patterns has two degrees of freedom. The estimated chi-square value exceeded the 5% significance level but not the 1% level for only one pair of patterns (proximity to anticline axes and proximity to contact between Halifax-Goldenville Formations), which indicates that the assumption of conditional independence of the patterns listed in Table 1 is approximately satisfied.

A comparison of frequency distributions in a Kolomogorov-Smirnov test using the posterior probabilities of all unit cells to provide the expected frequencies is shown in Fig. 6. If two or more patterns would be conditionally independent, the expected frequencies in this second test would overestimate the observed frequencies when the

Table 2. Weights and contrast ($W^+ - W^-$) of binary patterns for proximity to anticlinal axes as a function of corridor half-width; total area, 2945 km^2; total number of known gold deposits, 68. Errors are standard deviations.

Corridor half-width (km)	Corridor area (km^2)	Gold deposits	W^+	W^-	C
0.25	257	16	1.03	−0.18	1.21 ± 0.29
0.50	614	31	0.81	−0.38	1.19 ± 0.25
0.75	809	37	0.71	−0.47	1.18 ± 0.25
1.00	995	43	0.65	−0.60	1.25 ± 0.26
1.25	1276	50	0.55	−0.77	1.32 ± 0.28*
1.50	1488	51	0.41	−0.69	1.10 ± 0.28
1.75	1641	54	0.36	−0.79	1.14 ± 0.30
2.00	1838	57	0.30	−0.86	1.16 ± 0.33
2.25	2007	59	0.25	−0.89	1.14 ± 0.36
2.50	2128	60	0.21	−0.87	1.08 ± 0.38
2.75	2226	61	0.18	−0.88	1.05 ± 0.40
3.00	2341	61	0.12	−0.70	0.82 ± 0.40

*Maximum contrast for corridor pattern selected.



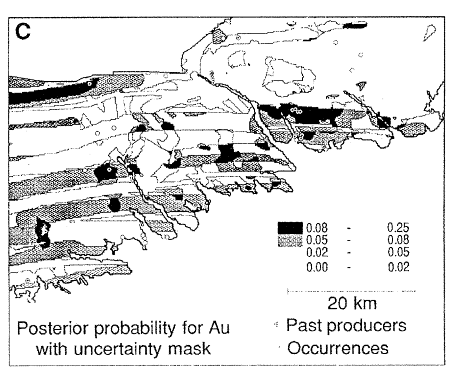


Fig. 5. Corridor patterns for (**A**) anticline axes and (**B**) contact between Goldenville and Halifax Formation, central Nova Scotia. These are examples of binary patterns obtained with SPANS for "corridors" around linear or curvilinear geological features believed to be favorable for occurrence of gold deposits. (**C**) Posterior probability of gold occurrence per unit area (1 km by 1 km) in part of Meguma Terrane, central Nova Scotia. A blank pattern denotes an uncertainty mask where the estimated posterior probability was less than twice its standard deviation.

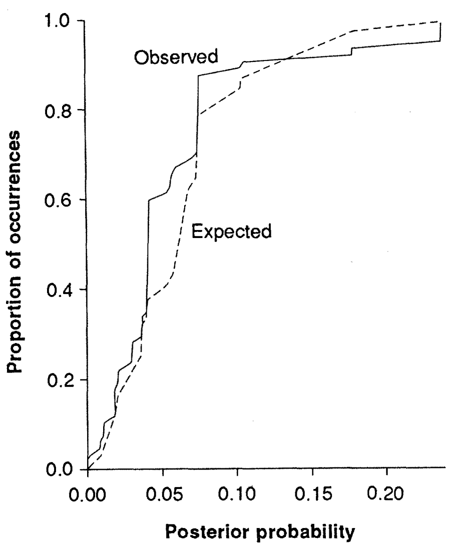


Fig. 6. Frequency distribution curves of observed and expected gold deposits as a function of their posterior probability. The expected frequency would exceed the observed frequency for larger posterior probabilities if one or more pairs of indicator patterns were not conditionally independent of the pattern of gold deposits. Differences between curves are statistically significant at the 5% but not at the 1% significance level in the Kolmogorov-Smirnov test.

posterior probability is relatively large and underestimate them when it is small. Some conditional dependence may indeed exist, but it is as not significant at the 1% level for the Kolmogorov-Smirnov test.

Concluding Remarks

Future mineral exploration expert systems will probably remain map-oriented both in input and output. In such systems the rules for spatial pattern integration should be as simple as possible but logically coherent. They should allow verification of assumptions such as that two or more prognostic patterns are conditionally independent with respect to occurrence of mineral deposits. Estimated probabilities for occurrence of hidden ore deposits should be accompanied by measures of uncertainty, including uncertainty due to missing data.

REFERENCES AND NOTES

1. Reviews of multivariate analysis in mineral exploration are given by D. P. Harris, *Mineral Resources Appraisal* (Oxford Univ. Press, New York, 1984), and by T. K. Wignall and J. G. De Geoffroy, *Statistical Models for Optimizing Mineral Exploration* (Plenum, New York, 1987). For other applications, see G. Gaál, Ed., *Exploration Target Selection by Intergration of Geodata Using Statistical Image Processing Techniques* (Geological Survey of Finland, Helsinki, 1988); R. G. Garrett, *Geol. Surv. Canada Pap.* 88-1F (1988); E. C. Grunsky and F. P. Agterberg, *Math. Geol.* **23**, 825 (1988).
2. SAS (Statistical Analysis System), SAS Institute Inc., Box 8000, Cary, NC 27411.
3. I. Hacking, *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference* (Cambridge Univ. Press, London, 1975). The role of mathematical statistics in geoscience is discussed by F. P. Agterberg, *Geol. Min.* **58**, 201 (1979).
4. R. O. Duda, J. G. Gaschnig, P. Hart, in *Expert Systems in the Microelectronic Age*, D. Mitchie, Ed. (Edinburgh Univ. Press, Edinburgh, 1979), pp. 153–167. Also see J. Gaschnig, R. Reboh, J. Reiter, *An Annotated Bibliography About PROSPECTOR* (SRI International, Menlo Park, CA, 1982).
5. B. C. Buchanan and E. H. Shortliffe, *Rule Based Expert Systems—The MYCIN Experiments of the Stanford Heuristic Programming Project* (Addison-Wesley, Reading, MA, 1984). See also R. O. Duda and E. H. Shortliffe, *Science* **220**, 261 (1983).
6. D. J. Spiegelhalter, in *Artificial Intelligence and Statistics*, W. A. Gale, Ed. (Addison-Wesley, Reading, MA, 1986), pp. 17–55.
7. W. A. Gale, Ed., *Artificial Intelligence and Statistics* (Addison-Wesley, Reading, MA, 1986).
8. D. J. Spiegelhalter and R. P. Knill-Jones, *J. R. Stat. Soc. B* **147**, 35 (1984).
9. Reviews of artificial intelligence (AI), expert systems, and expert system shells in geology and mining are given by S. M. Yatabe and A. G. Fabbri, *Sci. Terre. Inf. Geol.* **27** (no. 1), 37 (1988); *Episodes* **12** (no. 1), 10 (1989); K. V. K. Prasad and R. V. Ramani, in *Application of Computers and Operations Research in the Mineral Industry*, A. Weiss, Ed. (Society of Mining Engineers, Inc., Littleton, CO, 1989), pp. 897–910. For a general introduction of AI, see P. H. Winston, *Artificial Intelligence* (Addison-Wesley, Reading, MA, ed. 2, 1984). Limitations of AI and expert systems are discussed by H. Inose and J. R. Pierce, *Information Technology and Civilization* (Freeman, New York, 1984).
10. J. R. Davis and P. M. Nanninga, *J. Environ. Manage.* **211**, 377 (1985). For other strategies, see M. Vanier and R. Woodtli, *Comput. Geosci.* **5**, 369 (1979) and D. Bonnefoy, M. Jebrak, M. C. Rousset, H. Zeegers, *Proceedings of the 12th International Geochemical Exploration Symposium* (Bureau of Geological and Mining Research, Orleans, 1987), p. 33.
11. A. N. Campbell and T. Glover, *The Deciding Factor Expert System Shell* (Channelmark Corp., San Mateo, CA, 1986).
12. C. F. Chung, *Math. Geol.* **15**, 47 (1983).
13. A. G. Fabbri, *Image Processing of Geological Data* (Van Nostrand Reinhold, New York, 1984).
14. R. B. McCammon, J. M. Botbol, R. Sinding-Larssen, R. W. Bowen, *Math. Geol.* **15**, 59 (1983).
15. S. S. Katz, in *Computer-aided Methods and Modelling in Geology and Engineering* (American Association of Peteroleum Geologists, Tulsa, OK, 1987), pp. 175–195. The relation between AI and GISs is discussed by R. N. Coulson, L. J. Folse, D. K. Loh, *Science* **237**, 262 (1987).
16. *SPANS, Spatial Analysis System Reference Guide, Version 4-0* (TYDAC Technologies, Inc., Ottawa, 1989).
17. R. B. McCammon, *U.S. Geol. Surv. Bull. 1967* (1986).
18. ———, *Geol. Surv. Can. Pap.* 89-9, in press.
19. G. S. Koch, Jr., and D. Papacharalampos, in *Quantitative Analysis of Mineral and Energy Resources*, C. F. Chung, A. G. Fabbri, R. Sinding-Larsen, Eds. (Reidel, Dordrecht, 1988), pp. 513–527.
20. R. M. Masly, in *Microcomputer Applications in Geology*, J. T. Hanley and D. F. Merriam, Eds. (Pergamon, Oxford, 1986), pp. 89–103.
21. D. A. Singer and R. Kouda, *Econ. Geol.* **83**, 18 (1988).
22. Original contour map in F. P. Agterberg, C. F. Chung, A. G. Fabbri, A. M. Kelly, J. C. Springer, *Geol. Surv. Can. Pap.* 71-41 (1972); hindsight study by F. P. Agterberg and M. David, in *Computer Methods for the 80's*, A. Weiss, Ed. (Society Mining Engineers, Inc., Littleton, CO, 1979), pp. 90–115.
23. Statistical evaluation and discussion by J. W. Tukey, *Int. Stat. Inst. Bull.* **44**, 597 (1972); *Math Geol.* **16**, 591 (1984). For use of regression diagnostics, see F. P. Agterberg, in *Quantitative Analysis of Mineral and Energy Resources*, C. F. Chung, A. G. Fabbri, R. Sinding-Larsen, Eds. (Reidel, Dordrecht, 1988), pp. 1–28; F. P. Agterberg and J. M. Franklin, in *Marine Minerals: Resource Assessment*, P. Teleki, Ed. (Reidel, Dordrecht, 1987), pp. 467–483.
24. The idea of an expert system developed by an expert and a knowledge engineer is discussed in F. Hayes-Roth, D. A. Waterman, D. B. Lenart, *Building Expert Systems* (Addison-Wesley, Reading, MA, 1983); D. A. Waterman, *A Guide to Expert Systems* (Addison-Wesley, Reading, MA, 1983).
25. A. N. Campbell *et al.*, *Science* **217**, 927 (1982).
26. The original inference network for porphyry molybdenum deposits is discussed by R. O. Duda, *The PROSPECTOR System for Mineral Exploration* (SRI International, Menlo Park, CA, 1980). Figure 4A is an abbreviated and simplified version. This network has also been discussed by S. S. Katz (paper presented at the Fifth National MOSS Users Workshop, New Orleans, April 1988).
27. Axiomatic theory was originally proposed by A. N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Springer, Berlin, 1933; English edition, Chelsea, New York, 1950). For explanation of basic concepts, see, for example, P. A. P. Moran, *An Introduction to Probability Theory* (Oxford Univ. Press, Oxford, 1984). The weights of evidence method used here was originally proposed by Spiegelhalter in (*6*). For a recent application to mineral exploration, see F. P. Agterberg, in *Application of Computers and Operations Research in the Mineral Industry*, A. Weiss, Ed. (Society of Mining Engineers, Inc., Littleton, CO, 1989), pp. 165–178.
28. This example is summarized from the following papers: G. F. Bonham-Carter, F. P. Agterberg, D. F. Wright, *Photogramm. Eng. Remote Sensing* **54**, 1585 (1988); F. P. Agterberg, G. F. Bonham-Carter, D. F. Wright, in *Computer Applications in Resource Exploration*, G. Gaál and D. F. Merriam, Eds. (Pergamon, Oxford, in press); G. F. Bonham-Carter and F. P. Agterberg, in *Microcomputer Applications in Geology II*, J. T. Hanley and D. F. Merriam, Eds. (Pergamon, Oxford, in press).
29. Gold occurs in quartz veins commonly confined to black shale horizons throughout the Goldenville Formation. Discussion relating to the origins of the approximately 70 gold occurrences in the study area is complex, and no consensus has been reached. The indicator patterns of Table 1 are for features believed to be important by M. C. Graves and M. Zentilli, *Can. Inst. Min. Metall. Spec. Pap.*, 233 (1982); J. D. Keppie, in *The Caledonide Orogen*, D. G. Gee and B. Stuart, Eds. (Wiley, New York, 1984); G. F. Bonham-Carter, A. N. Rencz, J. R. Harris, in *Remote Sensing for Exploration Geology*, J. J. Cook and D. R. Morris-Jones, Eds. (Environmental Research Institute of Michigan, Ann Arbor, 1985), vol. 2, pp. 775–768; D. F. Wright, G. F. Bonham-Carter, P. J. Rogers, in *Prospecting in Areas of Glaciated Terrain* (Canadian Institute of Mining and Metallurgy, Montreal, 1988), pp. 501–515.
30. M. M. Bishop *et al.*, *Discrete Multivariate Analysis* (MIT Press, Cambridge, MA, 1975).
31. This software review is Geological Survey of Canada contribution 16989.