Statistical Pattern Integration for Mineral Exploration*

F.P. Agterberg, G.F. Bonham-Carter, and D.F. Wright
Geological Survey of Canada, Ottawa
Geological Survey of Canada Contribution No. 24088

ABSTRACT

The method of statistical pattern integration used in this paper consists of reducing each set of mineral deposit indicator features on a map to a pattern of relatively few discrete states. In its simplest form the pattern for a feature is binary representing its presence or absence within a small unit cell; for example, with area of 1 km^2 on a 1:250,000 map. The feature of interest need not occur within the unit cell; its "presence" may indicate that the unit cell occurs within a given distance from a linear or curvilinear feature on a geoscience map. By using Bayes' rule, two probabilities can be computed that the unit cell contains a deposit. The log odds of the unit cell's posterior probability is obtained by adding weights W^+ or W^- for presence or absence of the feature to the log odds of the prior probability. If a binary pattern is positively correlated with deposits, W+ is positive and the contrast $C=W^+-W^-$ provides a measure of the strength of this correlation. Weights for patterns with more than two states also can be computed and special consideration can be given to unknown data. Addition of weights from several patterns results in an integrated pattern of posterior probabilities. This final map subdivides the study region into areas of unit cells with different probabilities of containing a mineral deposit. In this paper, statistical pattern integration is applied to occurrence of gold mineralization in Meguma Terrane, eastern mainland Nova Scotia, Canada.

INTRODUCTION

Geoscience maps of different types are to be integrated for target selection in mineral exploration. The geologist stacks these maps and looks for combinations of indicators favorable for occurrence of deposits of different types. The calculations required for mathematical analysis of digitized patterns for points, lines, and areas have been greatly aided by the development of microcomputer based geographic information systems for the treatment of map data (Bonham-Carter, Agterberg, and Wright, 1988). This has led us to develop further a new method for statistical pattern inte- gration simulating the practice of exploration geologists to combine maps for deline- ating favorable areas. This method was proposed initially by Agterberg (in press) for combining geophysical survey data with prior probabilities of occurrence of massive sulfide deposits in the Abitibi area of the Canadian Shield originally obtained in 1971. The prior probability for a massive sulfide deposit being in a small unit cell was assumed to be constant within a larger cell. The frequency of massive sulfide deposits had been estimated by regression analysis from lithological and other variables systematically coded for such larger cells. Other geoscience data for the same area (Bouguer anomalies, aeromagnetic anomalies, and boundaries between tertiary drainage basins) had been quantified later as patterns of two or more mutually exclusive states by Assad and Favini (1980). The patterns for proximity to aeromagnetic anomalies and bouinlaries between drainage basins were binary, needing only two colors (black and white) for representation. It was possible to compute weights W_i^+ and W_i^- representing the states of presence and absence in the unit cell for each binary pattern i.

In Assad and Favini (1980), the pattern for the Bouguer anomaly had five distinct states with different colors. In Agterberg (in press), a weight W_j^+ was computed for each color j of this pattern with more than two states. At any point within the study area, the weights for the geophysical variables were added to the

log odds of the prior probability. This gave the log odds of the posterior probability. Because the patterns combined with one another all consisted of polygons, the final product was also polygonal with different colors for classes defined for the posterior probability per unit cell. The addition of weights W^+ or W^- is permitted only if the patterns being integrated are conditionally independent of occurrence of deposits. In the Abitibi study it was shown that this condition is satisfied approximately for the geophysical variables. In this paper, the method will be applied to gold deposits in Meguma Terrane, Nova Scotia.

An advantage of the statistical pattern integration method with respect to most existing methods in the field of regional resource evaluation (e.g., logistic regression) is that a pattern need be available only for parts of the study region. However, if one or more patterns are missing at a given place, the estimated posterior probability has less certainty than those based on more or all patterns. This type of uncertainty, the result of one or more missing patterns, will be studied later in the paper.

Finally, special attention should be given to verification of the theoretical assumption of conditional independence. The simple addition of weights for different features is permitted only if this assumption is satisfied. In general, the possibility of occurrence of conditional dependence increases with an increasing number of patterns. Failure of the method in this respect would lead to discrepancies between frequencies as predicted by the posterior probability map and the corresponding observed frequencies. If the assumption of conditional independence is not satisfied, the theoretical frequencies would exceed the observed frequencies in the most favorable parts of the region. At the end of the paper, we provide a statistical test for comparing the theoretical and observed frequencies with one another. First, the method of statistical pattern integration will be explained by using a simple artificial example.
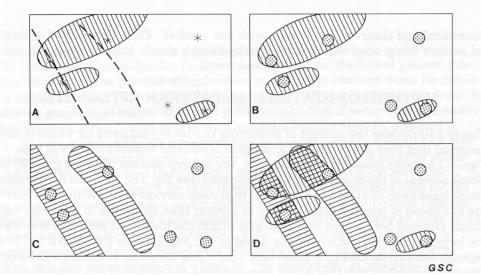

Figure 1. Artificial example to illustrate concept of combining two binary pat- terns related to occurrence of mineral deposits; (A) outcrop pattern of rock type, lineaments, and mineral deposits; (B) rock type and deposits dilatated to unit cells; (C) lineaments dilatated to corridors; (D) super- position of three patterns.

Figure 1 illustrates the concept of combining two binary patterns for which it may be assumed that they are related to occurrence of mineral deposits of a given type. Figure 1A shows locations of six deposits, the outcrop pattern of a rock type (B) with which several of the deposits may be associated (see Fig. 1B), and two lineaments which have been dilated in Figure 1C to provide corridors (C). Within the corridors the likelihood of locating deposits may be greater than elsewhere in the study region. Points situated both on the rock type and within the lineament corridors may have the largest probability of containing deposits (see Fig. 1D). In Figures 1B to 1D, the deposits are surrounded by a small unit area. This allows us to estimate the unconditional probability p(d) that a unit area contains one deposit if it is located randomly

within the study area, and the conditional probabilities p(d | b), p(d | c), and p(d | be) for occurrences on rock type, corridors and overlap of rock type and corridors, respectively. These probabilities are estimated by counting how many deposits occur within the areas occupied by the polygons of their patterns. The relationships between the two patterns B and C, and the deposits, D, can be represented by Venn diagrams as shown schematically in Figure 2.
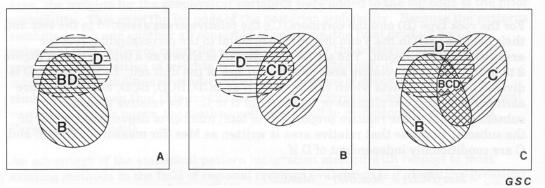

Figure 2. Venn diagrams corresponding to areas of binary patterns in Figure 1; (A) is for Figure 1 B; (B) is for Figure 1 C; (C) is for Figure 1 D.

METHOD OF STATISTICAL PATTERN INTEGRATION

For the rock type (B) and the corridors (C), the relative areas assigned to the sets and their overlap (BC) in the Venn diagrams are equal to the corresponding relative areas in the study region. The set for deposits (D) is shown as a broken line in Figure 2 to indicate that its relative area depends on size of the unit cell. In Figure 2C, D is divided into four subsets which can be written as BCD, BCD, BCD, and BCD where each bar indicates complement or "absence" of B or C. The relative areas of the subsets is equal to the relative proportions of total number of deposits belonging to the subsets. Suppose that relative area is written as Mes (for measure). Then B and C are conditionally independent of D if

$$p(bc|d) = \frac{Mes(BCD)}{Mes(D)} = \frac{Mes(BD)}{Mes(D)} \times \frac{Mes(CD)}{Mes(D)} = p(b|d) \times p(c|d) \qquad (1)$$

This is equivalent to assuming either

$$p(b|cd) = \frac{Mes(BCD)}{Mes(CD)} = \frac{Mes(BC)}{Mes(D)} = \frac{p(b|d)}{p(d)} \quad or \quad p(c|bd) = \frac{Mes(BCD)}{Mes(BD)} = \frac{Mes(CD)}{Mes(D)} = \frac{p(c|d)}{p(d)} \quad (2)$$

The latter two expressions can be readily visualized by comparing Figure 2A to Figure 2C and Figure 2B to Figure 2C, respectively. Obviously, B and C are not necessarily conditionally independent of D. However, this assumption is considerably weaker than assuming that two patterns are statistically independent. For example, if B were independent of C, we would have exactly:

$$p(bc) = Mes(BC) = Mes(B) \times Mes(C) = p(b) \times p(c) \qquad (3)$$

In our type of application, Mes(BC) is greater than Mes(B) x Mes(C) because both B and C are positively correlated with D.

The relationship between B, C, and D can be expressed by the following (2x2x2) table of probabilities:

$$
\begin{array}{llll}
& \multicolumn{2}{c}{D=d} \\
& c & \bar{c} \\
b & p(bcd) & p(b\bar{c}d) \\
\bar{b} & p(\bar{b}cd) & p(\bar{b}\bar{c}d)
\end{array}
\qquad
\begin{array}{llll}
& \multicolumn{2}{c}{D=\bar{d}} \\
& c & \bar{c} \\
b & p(bc\bar{d}) & p(b\bar{c}\bar{d}) \\
\bar{b} & p(\bar{b}c\bar{d}) & p(\bar{b}\bar{c}\bar{d})
\end{array}
$$

Here B, C, and D are regarded as random variables which are either present or absent in a unit cell. Absence is indicated by a bar. The eight probabilities in this table add up to one. If the assumption of conditional independence of B and C holds true, the eight probabilities in the table also are mutually related by:

$$
p(bcd) = p(b|d)\ p(c|d)\ p(d) \quad \text{and}
$$
$$
p(bc\bar{d}) = p(b|\bar{d})\ p(c|\bar{d})\ p(\bar{d}) \tag{4}
$$

This result follows from combining the first part of Equation (2) with the identities:

$$
p(bcd) = p(b|cd)\ p(c|d)\ p(d) \quad \text{and}
$$
$$
p(bc\bar{d}) = p(b|c\bar{d})\ p(c|\bar{d})\ p(\bar{d}). \tag{5}
$$

Equation (4) implies that all eight probabilities in the table can be determined from only five individual probabilities or functions of probabilities. In our approach we
will use for these five constants, the prior probability p(d) and the weights W_b^+, W_b^-, W_c^+ and W_c^- defined as:

$$
W/\ p(b|\bar{d})\};\quad W_b^- = \log_e\{\ p(\bar{b}|d)\ /\ p(\bar{b}|\bar{d})\};
$$
$$
W_c^+ = \log_e\{p(c|d)\ /\ p(c|\bar{d})\};\quad W_c^- = \log_e\{p(\bar{c}|d)\ /\ p(\bar{c}|\bar{d})\}. \tag{6}
$$

Weights of evidence W^+ and W^- were previously used by Spiegelhalter (1986).

Two binary patterns, B and C, give four posterior probabilities for D=d. These are $p(d|\bar{b}c)$, $p(d|b\bar{c})$, $p(d|\bar{b}c)$ and $p(d|\bar{b}\bar{c})$. It is convenient to work with odds (O) instead of probabilities with O=p/(1-p) and p=O/(1+O). Then:

$$
\log_e O(d|bc) = W_b^+ + W_c^+ + \log_e O(d),
$$
$$
\log_e O(d|\bar{b}c) = W_b^- + W_c^+ + \log_e O(d),
$$
$$
\log_e O(d|b\bar{c}) = W_b^+ + W_c^- + \log_e O(d),\ \text{and}
$$
$$
\log_e O(d|\bar{b}\bar{c}) = W_b^- + W_c^- + \log_e O(d), \tag{7}
$$

This is the extension of Bayes' rule which holds true only if B and C are conditionally independent with

$$p(bc|d_i) = p(b|d_i) \times p(c|d_i) \tag{8}$$

with di=d (for i=l) or d_i=d (for i=O).

Previous applications of the assumption of conditional independence in mineral exploration include those by Duda and others (1977) and Singer and Kouda (1988). Even if this assumption is not satisfied, we always have:

$$\log_e O(d|b) = W_b{}^+ + \log_e O(d) \tag{9}$$

and equivalent expressions for O(d | b), O(d | c), and O(d | c). The latter are formulations of Bayes' rule which has had many previous geological applications (cf. Harbaugh, Doveton, and Davis, 1977). Extensions of Equation (7) to more than two patterns are readily made. For example, if A is conditionally independent of B and C, then:

$$\log_e O(d|abc) = W_a{}^+ + W_b{}^+ + W_c{}^+ + \log_e O(d) \tag{10}$$

with seven equivalent expressions.

Part of the usefulness of this approach for mineral exploration results from the fact that it can be assumed that weights such as W_b+ are independent of the prior probability p(d). For example, if there would be as many undiscovered deposits in the region as there are known deposits, then the prior probability p(d) becomes twice as large. However, weights such as W_b+ =$\log_e$ p(b | d)/p(b | d) remain the same even if p(d) is changed provided that the proportion of new deposits associated with B=b would not change during exploration in future.

APPLICATION TO GOLD DEPOSITS IN MEGUMA TERRANE

Wright, Bonham-Carter, and Rogers, (1988) have used regression analysis to determine the multielement lake-sediment geochemical signature that best predicts the catchment basins containing gold occurrences in Meguma Terrane, eastern mainland Nova Scotia (see Fig. 3). Their geochemical signature was reduced to a ternary pattern (Fig. 4) for this study. Bonham-Carter and others (1988) have coregistered and analyzed a variety of regional geoscience data sets for this same study area using a geographic information system. A number of these data sets also are used in this paper.

Bonham-Carter and others (1988) have pointed out that the mechanism of gold mineralization in the study area is not well understood. Various authors have proposed different genetic models, emphasizing stratigraphic control, structural control, or importance of the intrusive granites as a source of mineral-rich hydrothermal fluids. Different processes have played a role in the formation of some or all of the gold deposits. By the method given in this paper, the spatial relationships to gold mineralization of patterns based on different genetic models can be compared and integrated with one another.
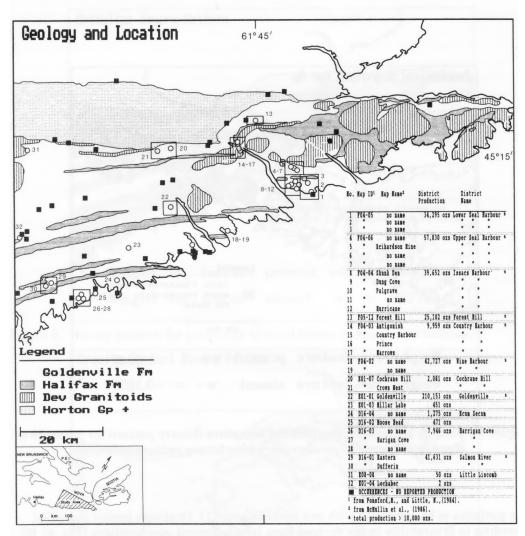
Figure 3. Location of study area with gold deposits in Meguma
Terrane, eastern mainland Nova Scotia.

Geology and Location   61°45′

45°15′

Legend

Goldenville Fm
Halifax Fm
Dev Granitoids
Horton Gp +

20 km

No.	Map ID[1]	Map Name[2]	District Production	District Name
1	P04-05	no name	34,295 ozs	Lower Seal Harbour [a]
2	"	no name	"	" "
3	"	no name	"	" "
4	P04-06	no name	57,830 ozs	Upper Seal Harbour [a]
5	"	Richardson Mine	"	" " "
6	"	no name	"	" "
7	"	no name	"	" "
8	P04-04	Skunk Den	39,652 ozs	Issacs Harbour [a]
9	"	Dung Cove		" "
10	"	Palgrave		" "
11	"	no name		" "
12	"	Hurricane		" "
13	P05-12	Forest Hill	25,102 ozs	Forest Hill [a]
14	P04-03	Antigonish	9,959 ozs	Country Harbour [a]
15	"	Country Harbour		" "
16	"	Prince		" "
17	"	Narrows		" "
18	P04-02	no name	42,727 ozs	Vine Harbour [a]
19	"	no name		
20	K01-07	Cochrane Hill	2,081 ozs	Cochrane Hill
21	"	Crows Nest		" "
22	E01-01	Goldenville	210,153 ozs	Goldenville [a]
23	K01-03	Millar Lake	451 ozs	
24	D16-04	no name	1,275 ozs	Ecum Secum
25	D16-02	Moose Head	471 ozs	
26	D16-03	no name	7,946 ozs	Harrigan Cove
27	"	Harrigan Cove		"
28	"	no name		"
29	D16-01	Eastern	41,631 ozs	Salmon River [a]
30	"	Dufferin		" "
31	E08-08	no name	50 ozs	Little Liscomb
32	K01-04	Lochaber	2 ozs	

■ OCCURRENCES - NO REPORTED PRODUCTION
[1] from Ponsford,N., and Little, W.,(1984).
[2] from McMullin et al., (1986).
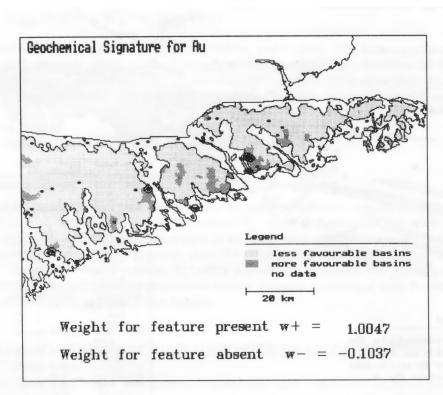[a] total production > 10,000 ozs.

Figure 4. Ternary pattern for geochemical signature (binary pattern for presence or absence of favorable geochemistry plus binary pattern unknown).

The patterns to be combined with one another are: (1) Drainage basins classified according to favorability index derived from lake sediment geochemistry (Fig. 4); (2) Bedrock geology (see Fig. 3); (3) Proximity to axial traces of Acadian anticlines (see Fig. 5); (4) Proximity to NW-trending lineaments; (5) Proximity to Devonian granites (see Fig. 6); and (6) Proximity (within the Goldenville Formation) to the contact between Goldenville and Halifax Formations. The weights estimated for these six patterns are shown in Table 1. The final map (Fig. 7) obtained by adding the computed weights to the log prior odds delineates subareas where most or all favorable conditions exist and can be used in gold exploration.

Four of the six patterns integrated with one another are for proximity to linear or curvilinear features. Binary patterns (e.g., Figs. 5 and 6) were selected in each of these situations after studying how size of neighborhood influences the contrast $C = W^+ - W^-$ which provides a measure of the strength of correlation between a point
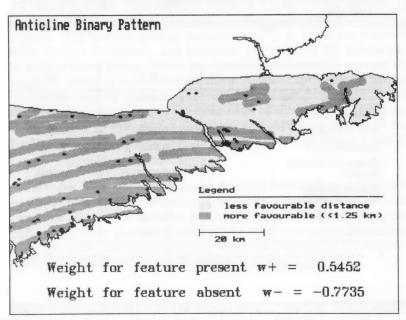
Figure 5. Binary pattern for proximity to axial traces of Acadian anticlines.
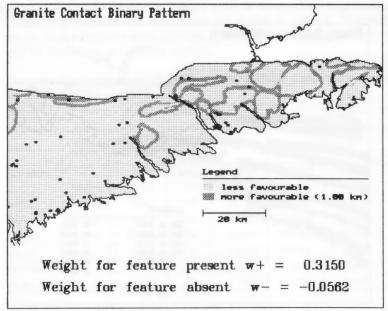

Figure 6. Binary pattern for proximity to Devonian granite contact within Goldenville Formation.

Table l. Weights for modeling posterior probability of a gold deposit occuring in a 1 km^2 area.

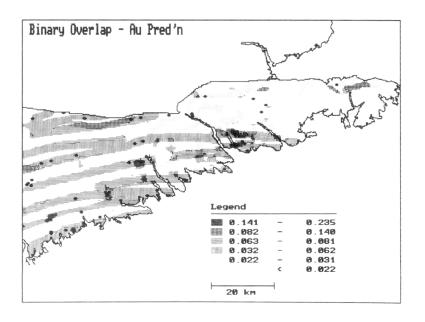| Map Pattern	W$^+$	W$^-$
Geochemical Signature	1.0047	-0.1037
Anticline Axes	0.5452	-0.7735
N.W. Lineaments	-0.0185	0.00062
Granite Contact	0.3150	-0.0562
Goldenville-Halifax Contact	0.3682	-0.2685
Bedrock Geology*		
Halifax Formation	-1.2406	
Goldenville Formation	0.3085	
Granite	-1.7360	
* A ternary pattern where units are mutually exclusive, and weights W$^-$ for absence are not used.		

Figure 7. Final map of posterior probabilities using all weights shown in Table 1.

pattern and a binary pattern. The expected value of C is zero if the deposits are randomly distributed with respect to the pattern. The properties of C will be studied later in this paper. The bedrock geology pattern (see Fig. 3) has three states (Goldenville Formation, Halifax Formation, and Devonian granite) and W_j^+ values were computed for each of these states. When the feature is unknown in parts of the study region, no weight is added or subtracted for the unit cells. The feature then has a ternary pattern with discrete states for presence, absence and unknown, respectively. The geochemical favorability index for lake drainage basins was quantified as a ter- nary pattern (see Figure 4) with W^+ for most favorable signature basins, W^- for less favorable basins, and zero weight ($W^o=0$) for parts of the region without lake drainage basins.

Weights for presences or absences of features obtained from different map patterns can be added if the theoretical assumption of conditional independence is satisfied. Although it may not be possible to verify this assumption for all pairs of patterns combined with one another, statistical tests can be used to compare theoretical frequencies of deposits with their corresponding observed frequencies for subareas with the same posterior probabilities on the integrated pattern.

PRACTICAL EXAMPLE OF ESTIMATION OF WEIGHTS

The weights for individual patterns in Table 1 used to obtain the integrated pattern of Figure 7 were obtained from the data which are shown in Table 2. The widths of the corridors for linear features were selected by studying contrasts for different widths as will be explained in the next section.

Table 2 Data used to compute weights W^+ and W^- of Table 1 and their standard deviations $s(W^+)$ and $s(W^-)$

Map Pattern	Corridor width	Area (in km²)	Gold occ.	W+	s(W+)	W-	s(W-)
Geochemical Signature		164.9	10	1.0047	0.3263	-0.1037	0.13278
Anitcline Axes	2.5 km	1276.4	.50	0.5452	0.1443	-0.7735	0.2370
N.W. Lineaments	1.0 km	749.7	17	-0.0185	0.2453	0.0062	0.1417
Granite Contact	1.0 km	3825.5	12	0.3150	0.2932	-0.0563	0.1351
Goldenville/ Halifax	2.0 km	1029.4	34	0.3682	0.1744	-0.2685	0.1730
Halifax Formation		441.9	3	-1.2406	0.5793	0.1204	0.1257
Goldenville Formation		2020.9	63	0.3085	0.1280	-1.4690	0.4484
Devonian Granite		482.2	2	-1.7360	0.7086	0.1528	0.1248

An example of calculation of the positive weights W^+ and the negative weights of one of the features W^- is as follows: Fifty gold occurrences are situated on the corridors of the anticline axes. The combined area of these corridors is 1276.4 km². The total study area contains n(d)=68 gold occurrences and measures 2945.0 km². The total number of unit cells can be set equal to n=2945. It follows that n(i)= n-n(d)=2877. Our calculations may he based on frequencies (= p×n) instead of on probabilities p. Then:

$$W^+ = \log_e \frac{n(bd)}{n(d)} \bigg/ \frac{n(b\bar{d})}{n(\bar{d})} \; ; \; W^- = \log_e \frac{n(\bar{b}d)}{n(d)} \bigg/ \frac{n(\bar{b}\bar{d})}{n(\bar{d})} \qquad (11)$$

From n(bd) = 50 and $n(\bar{b}d)$ = 18, it follows that $n(b\bar{d})$ = 1226 and $n(\bar{b}\bar{d})$ = 1651. Consequently,

$$W^+ = \log_e \frac{50}{68} \Big/ \frac{1226}{2877} = 0.5455;$$

$$W^- = \log_e \frac{18}{68} \Big/ \frac{1651}{2877} = -0.7738$$

The weights reported in Table 2 differ slightly from these numbers, because they were based on slightly more precise estimates of areas.

$$s^2(W^+) = \frac{1}{n(bd)} + \frac{1}{n(b\,\bar{d})};$$

$$s^2(W^-) = \frac{1}{n(\bar{b}\,d)} + \frac{1}{n(\bar{b}\,\bar{d})} \tag{12}$$

Table 2 also shows estimates of standard deviations of W^+ and W^-. These were obtained from the variances:

These formulae are consistent with the asymptotic expression for the contrast to the discussed in the next section. Spiegelhalter and Knill-Jones (1984) have used similar formula to obtain standard errors of the weights. The only difference between their formulae and ours is that Spiegelhalter and Knill-Jones (1984) applied a correction based on the theory of binary data analysis to help remove bias from their estimated weights as well as from the corresponding variances.

Eight of the sixteen weights in Table 2 are more than twice as large, in absolute value, as their standard deviation. These eight weights probably are different from zero, because the 95 percent confidence interval for hypothetical zero weight is ap- proximately equal to ± 2s.

We have used asymptotic maximum likelihood expressions (cf. Bishop, Fienberg, and Holland, 1975, chapter 14) for s. Such expressions are valid only if a number of conditions are satisfied including the condition that the probabilities in the (2x2) table are neither large (= close to one) nor small (= close to zero). The latter condition may have been violated during estimation of the relatively large standard deviations of negative weights for rock types in the lower part of Table 2, because these are based on relatively few deposits. For example, only two gold occurrences on Devonian granite contribute 0.5 to the variance of their weight (= -1.7360), and therefore, account for most of the value of s(W+) = 0.7086 (bottom line of Table 2) which is probably too large.

The standard deviation of a posterior probability can he estimated as follows. The variance $s^2(p)$ of the prior probability p satisfies approximately p/n. For p = 68/2945 = 0.0231, this yields the standard deviation s(p) = 0.0028. The corresponding standard deviation of $\log_e(p/(1-p))$ = -3.7450 is approximately equal to s/p = 0.1213. This follows from the approximate identity for any variable x with mean x:

$$\frac{s(\log_e x)}{s(x)} \simeq \frac{d(\log_e x)}{dx}\Big|_{x=\bar{x}} = \frac{1}{\bar{x}} \tag{13}$$

Suppose that a unit cell has the following features: Its geochemical signature is unknown; it occurs in the Goldenville Formation not near a granite contact, and in the proximities of an anticline axis, NW lineament, and Goldenville/Halifax contact. Then the log posterior odds is -2.598 as can be seen when the appropriate weights are added. The variance of the log posterior odds is derived by adding variances of weights to the variance of the log prior odds. It follows that the standard deviation of the log posterior odds amounts to 0.401. The posterior probability of the unit cell containing a deposit becomes 0.069 with approximate standard deviation equal to $0.069 \times 0.401 = 0.028$. In this way, a standard deviation can be estimated for each of the posterior probabilities on a final integrated pattern. However, it will be shown later that if one or more patterns are missing, the standard deviation of the posterior probability should be increased due to the lack of knowledge. Because no information on geochemical signature is available for the unit cell in preceding example, the final standard deviation becomes 0.042 instead of 0.028 (see later). Although this final value (=0.042) is greater than the standard deviation (=0.028) computed from the uncertainties associated with the prior probability and the weights of Table 2, it is less than the standard deviation (=0.087) of the posterior probability (=0.169) arising when the unit cell considered for example in this section would have favorable geochemical signature.

CORRELATION BETWEEN PATTERN AND DEPOSITS

The contrast $C_b = W_b^+ - W_b^-$ for a pattern B provides a convenient measure of the strength of correlation between B and the pattern of deposits. The (2x2) table of probabilities with marginal totals for B and D is:

	d	d̄	
b	p(bd)	p(bd̄)	p(b)
b̄	p(b̄d)	p(b̄d̄)	p(b̄)
	p(d)	p(d̄)	1.00

If the deposits are randomly distributed within a study region, without preference for b or b̄, this table becomes

	d	d̄
b	$p(b) \times p(d)$	$p(b) \times p(d̄)$
b̄	$p(b̄) \times p(d)$	$p(b̄) \times p(d̄)$

By using the previous definitions of W_b^+ and W_b^-, it then is readily shown that $W_b^+ = W_j$ and $C = W_b^+ - W_b^- = 0$.

Table 3 for proximity of gold occurrences to anticline axes in Meguma Terrane, Nova Scotia, shows the contrast $C(x)$ as a function of distance x by which these linear features were dilated (in both directions) to define the binary pattern previously shown as Figure 5. Thus x is equal to one-half the width of the corridors. Inspection of $C(x)$ as a function of x provides a useful tool for deciding on a good value of x. It should be kept in mind, that $C(x)$ will be less precise for smaller values of x. This is because the number of deposits n(bd) from which p(bd) is estimated then may be small and subject to considerable uncertainty. If, as before, total number of unit cells is written as n, we have p(bd)=n(bd)/n with equivalent expressions for the other elements of the (2x2) table.

Table 3. Weights and contrast for anticline binary patterns as function of one-half-width of corridor. Total area sampled = 2945 km^2; total number of gold occurrences = 68; * denotes maximum contrast.

CORRIDOR HALF-WIDTH (in km)	CORRIDOR AREA (in km²)	GODL OCC. ON CORRIDOR	W⁺	W⁻	CONTRAST C = W⁺ - W⁻	STANDARD DEV. OF C
0.25	257	16	1.033	-0.181	1.213	0.294
0.50	614	31	0.811	-0.382	1.193	0.248
0.75	809	37	0.707	-0.473	1.180	0.247
1.00	995	43	0.648	-0.599	1.246	0.255
0.25	1276	50	0.545	-0.774	1.319*	0.278
0.50	1488	51	0.408	-0.694	1.101	0.283
1.75	1641	54	0.364	-0.778	1.142	0.302
2.00	1838	57	0.303	-0.857	1.160	0.332
2.25	2007	59	0.248	-0.892	1.140	0.360
2.50	2128	60	0.205	-0.872	1.077	0.379
2.75	2226	61	0.176	-0.878	1.053	0.401
3.00	2341	61	0.124	-0.701	0.824	0.402

Writing $\alpha = e^c$, the following asymptotic result for large n (see Bishop, Fienberg, and Holland, 1975, **p.** 377) can be used:

$$\hat{\sigma}^2_{\infty}(\hat{a}) = \frac{\hat{a}^2}{n} \left\{ \frac{1}{p(bd)} + \frac{1}{p(b\bar{d})} + \frac{1}{p(\bar{b}d)} + \frac{1}{p(\bar{b}\bar{d})} \right\} \tag{14}$$

If $\hat{\sigma}_{\infty}(\hat{a})$ is small compared to α, it follows from Equation (13) that the standard deviation of C is approximately equal to

$$\hat{\sigma}_{\infty}(C) \simeq \left\{ \frac{1}{n(bd)} + \frac{1}{n(b\bar{d})} + \frac{1}{n(\bar{b}d)} + \frac{1}{n(\bar{b}\bar{d})} \right\}^{\frac{1}{2}} \tag{15}$$

In the last column of Table 3, it is shown how this asymptotic standard deviation initially decreases as a function of distance. Once the one-half-width exceeds 0.75 km, the standard deviation continually increases. An approximate 95 percent confidence interval for C is provided by $\pm 2\hat{\sigma}_{\infty}(C)$. From this it may be concluded that the values of C shown in Table 3 are significantly greater than zero. Table 4 provides another example of C(x) as a function of x. Both positive and negative values of C occur in Table 4 which is for proximity to Devonian granites. The standard deviation of C now continues to decrease for wider corridors and it is likely that none of the values of C are significantly different from zero. The maximum value of C corresponds to a proximity of 1 km and this binary pattern was selected for use (cf., Tables 1

Table 4. Weights and contrast for granite contact corridors as function of corridor width. Total area samples = 2945 km²; total number of gold occurrences = 68; * denotes maximum contrast.

CORRIDOR WIDTH (in km)	CORRIDOR AREA (in km)	GOLD OCC. ON CORRIDOR	W^+	W^-	CONTRAST $C = W^+ - W^-$	STANDARD DEV. OF C
0.25	121	3	0.074	-0.003	0.077	0.598
0.50	247	6	0.052	-0.005	0.056	0.433
0.75	319	7	-0.051	0.006	-0.057	0.404
1.00	383	12	0.315	-0.056	0.371*	0.323
1.25	478	13	0.167	-0.036	0.203	0.313
1.50	528	13	0.065	-0.015	0.080	0.312
1.75	582	14	0.043	-0.012	0.054	0.304
2.00	670	14	-0.102	0.028	-0.130	0.303
2.25	715	14	-0.168	0.049	-0.217	0.303
2.50	756	14	-0.226	0.068	-0.299	0.303
2.75	799	15	-0.211	0.069	-0.280	0.293
3.00	865	17	-0.170	0.061	-0.226	0.283

and 2). The corresponding weights for proximity to Devonian granite (cf. Fig. 6) are relatively small and had relatively little effect on the final map (Fig. 7).

In mathematical statistics, various functions of α have been proposed to express correlation between two binary variables. Yule's "measure of association" $Q=(\alpha-1)/(\alpha+1)$ (see Bishop, Fienberg, and Holland, 1976, p. 3 78) is comparable to the ordinary product-moment correlation coefficient for two continuous variables in that it is confined to the interval [-1, 1] with $E(Q)=0$ for uncorrelated binary patterns. It is readily shown that

$$\frac{dQ}{d\alpha} = 2/(\alpha + 1)^2 \tag{16}$$

which is always positive. Consequently, $\hat{Q}(x)$ as a function of x would reach its maximum at the same value of x as $C(x)$ (cf. Table 3). It may be concluded that the contrast $C=W+ -W-$ provides a convenient measure of strength of correlation between the patterns B and D.

UNCERTAINTY BECAUSE OF ONE OR MORE MISSING PATTERNS

In the Introduction, it was pointed out that posterior probabilities do not all have the same precision if some of them are based on fewer patterns than others. This situation arises when data for a pattern are missing in parts of the study region. For example, the geochemical signature based on lake drainage basins is only available for parts of our study area (Meguma Terrane, Nova Scotia). Spiegelhalter (1986, p. 37) has proposed to regard any prior probability $p(d)$ as the expectation of the possible final probabilities $p(d \mid x)$ that may be obtained on observing data x

$$p(d) = E_X[p(d|X)] = \int p(d|x)p(x)dx \tag{17}$$

In general,
For the relationship between B, C, and D:

$$p(d) = \sum_{ij} p(d|b_i c_j)p(b_i c_j) = p(d|bc)p(bc) + p(d|b\bar{c})p(b\bar{c}) + p(d|\bar{b}c)p(\bar{b}c) + p(d|\bar{b}\bar{c})p(\bar{b}\bar{c}) \tag{18}$$

The corresponding variance is:

$$\sigma_2^2[p(d)] = \sum_{ij} \left\{ p(d|b_i c_j) - p(d) \right\}^2 p(b_i c_j) \tag{19}$$

If only B is unknown, the information on C can be added to the prior probability in order to obtain updated prior probabilities Pb(d) with variance:

$$\sigma_1^2[p_b(d)] = \left\{ p(d|b) - p(d) \right\}^2 p(b) + \left\{ p(d|\bar{b}) - p(d) \right\}^2 p(\bar{b}) \tag{20}$$

follows from:

$$\sum_{j} p(d|b_i c_j)p(b_i c_j) = p(d|b_i)p(b_i) \tag{21}$$

The expressions for the variances $\sigma 1^2$ (one pattern missing) and $\sigma 2^2$ (two patterns ,missing) are independent of any other patterns for which data were available and ised to change the prior probability. Extensions to situations with three or more missing patterns are readily made. In our example, only one pattern is incomplete: geochemical signature for gold deposits in Meguma Terrane. The ternary pattern representing geochemical signature (Fig. 4) shows those parts of the area where this feature could not be determined. In these places, the probability $p_b(d)$ on the final map (Fig. 7) has partial uncertainty that can be expressed by the standard deviation $\hat{\sigma} 1[P_b(d)]$. This uncertainty is partial because it becomes zero in places where all patterns including the geochemical signature are available, although the posterior probabilities in these places have their own uncertainties which can be estimated by using the standard deviations of the weights (see before). The latter type of uncertainty of the posterior probability increases when the pattern for geochemical signature is added. Of course, the uncertainty because of a missing pattern decreases when information on the pattern is added.

The weights $W^+=1.0047$ and $W^-=-0.1037$ for the geochemical signature (cf., Table 1) were determined from likelihood ratios for the entire area. For example, $W^+=\log_e p(b|d)/p(b|\bar{d})=1.0047$ was based on (1) $p(b|d)=p(bd)/p(d)=n(bd)/n(d)$ with $n(bd)=10$ and $n(d)=68$; and (2) $P(b|\bar{d})=p(b\bar{d})/p(\bar{d})=n(b\bar{d})/n(\bar{d})$ with $n(b\bar{d})=164.9 - 10=154.9$ and $n(\bar{d})=2945.0-68=2877.0$. As discussed before, the weight W^+ can be regarded as independent of the prior probability. For this example, approximately the same value of W^+ is obtained when (1) the calculation is based on the subarea (=1765.8 km') with known geochemistry; and (2) the prior probability within the area with known geochemistry is equal to that for the total study area (=2945.0 km^2) . The second condition implies that there would be about 41 deposits within the area with known data.

In reality, this subarea contains only 24 gold occurrences. A revised weight based on the subarea only would amount to 1.5444 which is greater than W^+=1.0047, because the subarea contains a larger proportion (=10124) of the deposits. The lesser weight (W^+=1.0047) was used in Figure 7 and now will be employed for estimating $\sigma\ [p_b(d)]$.

For example, the modified prior probability p(d), which is based on all patterns except geochemical signature, will be set equal to 0.05 and 0.10 within the area without definable lake drainage basins. The log odds of these values are -2.9444 and -2.1972, respectively. Addition of W^+ and W^- provides the required estimates of p(d | b)and p(d | $\bar{b}$). For p(d)=0.05, these conditional probabilities are equal to 0.1257 and 0.0453, respectively. For p(b) which also is needed to determine σ1 the ratio of favorable area (=164.9 km^2) to known area (=1765.8 km^2) can be used. This gives p(b)=0.0934 and p(b)=1-p(b)=0.9066. Consequently, σ1 (0.05)=0.024. By the same method, it follows that σ1 (0.10)=0.042. Previously, it was pointed out that if a unit cell in the Goldenville Formation with unknown geochemical signature is in the proximity of all linear features except granite contact, then its posterior probability is 0.070 with standard deviation equal to 0.028. Addition of the uncertainty because of the missing pattern results in the larger standard deviation of 0.042.

For both revised prior probabilities p(d)(=0.05 and 0.10), the standard deviation expressing uncertainty is the result of missing information is about one-half of p(d), or σ1[p(d)]≈0.5 p(d). This indicates that outside the lake drainage basins where geochemical information is not available, the posterior probabilities on the final map (Fig. 5) are less precise than would follow from the uncertainties associated with the prior probability and the weights (Table 2). It is convenient to express uncertainty due to ignorance by a single statistical parameter (standard deviation a, in this section). It should be kept in mind, however, that this parameter is estimated from a discrete probability distribution approximating an unknown continuous frequency distribution.

TEST FOR GOODNESS-OF-FIT

As pointed out in the Introduction, the final posterior probability map (Fig. 7) provides expected frequencies that can be compared to observed frequencies for the known occurrences. Suppose that p, represents the posterior probability after classification. For example, p, may be set equal to the midpoints of the classes of probabilities used for constructing the map on which Figure 7 is based. Suppose that, in total, there are n deposits (n=68 in Fig. 7). For each pi, the expected frequency amounts to

$$f_{ei} = \frac{A_i p_i n}{\Sigma A_i p_i} \qquad (22)$$

where A_i is the joint area of all polygons with posterior probability p_i. The corresponding observed frequency foi is obtained by counting how many deposits actually occur in the polygons with posterior probability p_i. Table 5 shows that expected and observed frequencies are nearly equal to one another for the pattern of Figure 7. It is possible to apply the chi-square test with

$$\hat{\chi}^2 = \sum_i \frac{(f_{oi} - f_{ei})^2}{f_{ei}} = 9.786 \qquad (23)$$

The number of degrees of freedom for the corresponding theoretical $\chi^2(\nu)$ is not known. Setting ν equal to number of classes -1 would give $\chi^2$2 0.05(5)=11.1 for level of significance α=0.05. The estimated value $\chi^2$2 (=9.8) is less than 11.1 suggesting a good fit of the model.
In this type of application, the theoretical frequencies were determined by assuming conditional

independence of all patterns. The test for goodness-of-fit used in this section would suggest that this hypothesis is approximately satisfied. Care has to be taken, however, in interpreting these results, because an upper bound for the number of degrees of freedom (v) was used. Comparison of observed and expected frequencies in Table 5 suggests that observed values tend to exceed expected values in the upper part of the table where p_i is relatively large and that the reverse holds that in the lower part of Table 5. This might indicate a minor violation of the assumption

Table 5. Comparison of observed and theoretical frequencies for final integrated pattern of Figure 7.

Class No.	Classes of posterior probabilities	Observed frequency (O)		Expected frequency (E)		$\frac{(O-E)^2}{E}$
1	0.171-0.235	4		1.1		
2	0.141-0.171	3	7	4.4	5.5	0.474
3	0.101-0.190	1		2.7		
4	0.082-0.100	1	2	4.3	7.0	3.571
5	0.063-0.081	17		23.9		1.992
6	0.032-0.062	23		16.7		2.377
7	0.022-0.031	5		3.3		0.875
8	0.000-0.021	14		11.6		0.497
				Sum =		9.786

of conditional independence. If two or more patterns are conditionally dependent with positive "partial association" (cf. Bishop, Fienberg, and Holland, 1975, p. 32), the expected frequencies would exceed the observed frequencies when p, is relatively large, whereas they would be smaller when p_i is small.

CONCLUDING REMARKS

The application of statistical pattern integration to gold exploration in Nova Scotia was performed using SPANS-A quadtree-based GIS. SPANS runs on IBM PC compatible under DOS. The work described here was carried out on an 80386 machine with 70 mb hard drive. SPANS accepts a variety of inputs of vector and raster data and permits the user to move readily in and out of DOS, so that other DOS compatible software can be executed on mutually shared data files.

This paper is concerned primarily with three problems: (1) Construction of optimum binary patterns for linear features in order to represent the relationship between these features and occurrence of mineral deposits; (2) Statistical integration of patterns for linear features and polygon patterns for areal features representing geochemistry and rock types; and (3) Development of a measure of uncertainty which is the result of missing information.

In order to resolve the first problem (1), a sequence of increasingly wide corridors around the linear features was constructed using SPANS. The choice of optimum width was made on the basis of the contrast C which measures correlation between a binary pattern and a point pattern. An asymptotic formula was used to estimate the standard deviation of C. Statistical pattern integration (2) was carried out by the addition of weights W^+ or W^- representing presence or absence of features. The addition of weights is based on the assumption of conditional independence of the map patterns with respect to the mineral

deposits. This assumption was tested by comparing the posterior probabilities shown on the final integrated map pattern with observed frequencies of gold deposits. Uncertainty resulting from one or more missing patterns (3) was evaluated by considering that no weights for presence of absence of a feature can be added if it is unknown. A measure of uncertainty was based on differences between posterior probabilities computed without the feature, and posterior probabilities computed using the possible outcomes for the feature if its presence or absence would be known. Contrary to the propagation of uncertainty associated with the weights which increases when more patterns are added, the uncertainty resulting from missing information decreases when patterns are added.

ACKNOWLEDGMENTS

REFERENCES

Agterberg, F.P., in press, Systematic approach to dealing with uncertainty of geoscience information in mineral exploration: Proc. 21st APCOM Symposium (Application of Computers in the Mineral Industry) held in Las Vegas, March 1989; Soc. Mining Engineers, New York.

Assad, R., and Favini, G., 1980, Prévisions de minerai cupro-zincifère dans le nord-ouest Québecois: Ministère de l'Energie et des Ressources, Quebec, DPV-670, 59 P.

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

Bonham-Carter, G.F., Agterberg, F.P., and Wright, D.F., 1988, Integration of mineral exploration data sets using SPANS-A quadtree-based GIS: Application to gold exploration in Nova Scotia; GIS Issue of Photogrammetric Engineering & Remote Sensing, in press.

Bonham-Carter, G.F., Rencz, A.N., George, H., Wright, D.F., Watson, F.G., Dunn, C.E., and Sangster, A.L., 1988, Demonstration of a microcomputer-based spatial data integration system with examples from Nova Scotia, New Brunswick and Saskatchewan: Geol. Survey of Canada, Current Activities Forum Abstract.

Duda, R.0, Hart, P.E., Nilson, N.J., Reboh, R., Slocum, J., and Sutherland, G.L., 1977, Development of a computer-based consultant for mineral exploration: Stanford Research Inst. International, Annual Rept. SRI Projects 5821 and 6915, Menlo Park, California, 202 p.

Harbaugh, J.W., Doveton, J.H., and Davis, J.C., 1977, Probability methods in oil exploration: Wiley Interscience Publ., New York, 269 p.

McMullin, J., Richardson, G., and Goodwin, T., 1986, Gold compilation of the Meguma Terrane in Nova Scotia; Nova Scotia Department of Mines and Energy, Open File 86-055, 056, maps.

Ponsford, M., and Lyttle, N., 1984, Metallic mineral occurrences and map data compilation, central Nova Scotia-map sheets IID, IIE; Nova Scotia Department of Mines Open File Report 599,12 p.

Singer, D.A., and Koada, R., 1988, Integrating spatial and frequency information in the search for Kuroko deposits of the Hokuroku District, Japan: Econ. Geology, v. 83, no. 1, p. 18-29.

Spiegelhalter, D.J., 1986, Uncertainty in expert systems, in W.A. Gale, ed.: Artificial intelligence and statistics: Addison-Wesley, Reading, Massachusetts, p. 17-55.

Spiegelhalter, D.J., and Knill-Jones, R.P., 1984, Statistical and knowledge-based approaches to clinical decision-support systems, with an application in gastroenterology: Jour. Royal Statist. Soc. A, v. 147, pt. 1, p. 35-77.

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: Proc. Can. Inst. Mining Metall. Meeting on "Prospecting in Areas of Glaciated Terrain", held in Halifax, Nova Scotia, September 1988.