Weights of evidence modelling:
a new approach to mapping mineral potential

G.F. Bonham-Carter[1], F.P. Agterberg[1] and D.F. Wright[1]

Abstract

Seven maps have been combined using a weights of evidence model to predict gold potential in the Meguma terrane of eastern shore Nova Scotia. The model uses the spatial distribution of known mineral occurrences to calculate a multi-map signature for gold mineralization, which is then employed to map gold potential.

In weights of evidence modelling, the log of the posterior odds of a mineral occurrence lying within a unit area is determined by adding a weight for each input map to the log of the prior odds; the ultimate product is a map of posterior probability, or mineral potential. When the input maps are binary, the weight added is either W^+ (binary pattern present) on W^- (binary pattern absent). The variances of the weights permit the calculation of an uncertainty map, which is augmented further in areas where one or more of the input maps is missing. The strength of association between the input map and the known mineral occurrence points is expressed as a contrast $C=W^+-W^-$, and the significance of C can be tested by estimating σ (C). The weights are calculated as log ratios of conditional probabilities. The model assumes that the input maps are conditionally independent from one another, with respect to the known occurrence points. This assumption is tested a) by pairwise tests, and b) by an overall comparison of model predictions with observations, using all input maps.

Previous work demonstrated that in order of contrast, the relative importance of the input maps for predicting known gold in eastern shore Meguma terrane is 1) presence of the Goldenville Formation, 2) proximity of anticlinal axes, 3) presence of a multi-element lake-sediment anomaly, 4) proximity to the Goldenville-Halifax contact, 5) proximity to the granite contact, and 6) proximity to NW structural lineaments. The new map showing Au in balsam fir anomalies is found to be strongly predictive of the known gold occurrences, with a value of C just larger than that for the lake-sediment signature. Pairwise tests show that the new map is conditionally independent of the other input maps with respect to the gold occurrences. By plotting past production from known gold occurrences on a graph of posterior probability versus cumulative area, it is shown that the larger gold producers are associated with map areas having a higher posterior probability than those points with no known production.

Three new areas predicted by the model are proposed as good prospects for gold mineralization. One is about 6 km south of Goldenville, one is between the Sherbrooke pluton and Seal Harbour, and a smaller one is about 2 km north of the Sherbrooke pluton.

Résumé

On a combiné sept cartes en utilisant un modèle à pondération de données pour prévoir les possibilités de minéralisation en or du terrane de Meguma sur la côte est de la Nouvelle-Écosse. À partir de la répartition spatiale de venues minérales connues, le modèle calcule une signature multicarte pour la minéralisation en or; cette signature sert ensuite à cartographier le potentiel aurifère.

[1] Geological Survey of Canada, 601 Booth St., Ottawa, Ontario K1A 0E8

Dans la modélisation à pondération de données, on détermine le logarithme de la probabilité postérieure qu'une venue minérale soit située dans une aire unitaire en additionnant un facteur de pondération pour chaque carte d'entrée au logarithme de la probabilité antérieure; le produit final est une carte de la probabilité postérieure ou du potentiel minéral. Lorsque les cartes d'entrée sont binaires, le facteur de pondération ajouté est soit W^+ (présence de configuration binaire) soit W^- (absence de configuration binaire). Les variances des facteurs de pondération permettent de calculer une carte d'incertitude; cette incertitude est d'ailleurs plus grande dans les régions où manquent une ou plusieurs cartes d'entrée. La force de la relation entre la carte d'entrée et les venues minérales connues est exprimée sous la forme d'un contraste $C = W^+ - W^-$, et on peut vérifier la signification de C en évaluant $\sigma (C)$. Les facteurs de pondération sont calculées sous forme de rapports logarithmiques de probabilités conditionnelles. Il est supposé, dans le modèle, que les cartes d'entrée sont indépendantes conditionnellement les unes des autres en ce qui a trait aux venues connues. On vérifie cette hypothèse a) en effectuant des essais par couples, et b) en effectuant une comparaison générale des prévisions du modèle avec des observations, à l'aide de toutes les cartes d'entrée.

Des travaux antérieurs ont montré que par ordre de contraste, l'importance relative des cartes d'entrée pour la prévision de minéralisations en or connues de la côte est du terrane de Meguma est: 1) la présence de la formation de Goldenville, 2) la proximité d'axes anticlinaux, 3) la présence d'une anomalie due à des sédiments lacustres renfermant plusieurs éléments, 4) la proximité du contact de Goldenville et Halifax, 5) la proximité du contact granitique et 6) la proximité de linéaments structuraux NO. On a trouvé que la nouvelle carte montrant des concentrations d'or anormales dans des sapins baumiers, donnait d'excellentes prévisions des venues d'or connues, la valeur de C étant dans ce cas légèrement plus élevée que dans le cas de la signature des sédiments lacustres. Des essais par couples montrent que la nouvelle carte est indépendante conditionnellement des autres cartes d'entrée en ce qui concerne les venues d'or. En reportant la production du passé provenant de venues d'or connues sur un graphique de probabilité postérieure en fonction d'une surface cumulative, on peut voir que les emplacements des plus grands gisements aurifères sont asociés à des zones cartographiques ayant une probabilité postérieure plus grande que celle des points ne présentant aucune production connue.

On propose trois nouvelles régions prévues par le modèle comme présentant de bonnes possibilités de minéralisation en or: la première se trouve à environ 6 km au sud de Goldenville; la deuxième, entre le pluton de Sherbrooke et Seal Harbour; et la troisième, plus petite, à environ 2 km au nord du pluton de Sherbrooke.

INTRODUCTION

This paper describes the application of a new statistical approach for making a regional map of mineral potential by combining evidence from geological (including structural), geophysical and geochemical surveys. The method is called weights of evidence modelling, and is based on a statistical method developed for medical diagnosis, (Spiegelhalter, 1986; Spiegelhalter and Knill-Jones, 1984). It has been extended to deal with spatial prediction, "diagnosing" mineral deposits using the "symptoms" of geological, geophysical and geochemical signatures. The method has been applied to the prediction of volcanogenic massive sulphides in the Abitibi region of Quebec (Agterberg, 1989), to gold in the Meguma terrane of Nova Scotia (Agterberg et al., 1990; Bonham-Carter et al., 1988; and Bonham-Carter and Agterberg, 1990), and to gold in New Brunswick by Watson et al. (1989).

Past work on making maps of mineral potential using statistical methods has predominantly used regression techniques (e.g. Agterberg et al., 1981, Harris 1984). The discovered mineral occurrences in a region are used to develop a multivariate signature for mineralization, expressed as a vector of coefficients for the predictor variables. The coefficients are calculated using least squares regression; the resulting equation is used to generate regression scores whose magnitude reflect mineral potential.

Weights of evidence modelling also uses the locations of known mineral occurrences to determine coefficients for each predictor map. However, there are two coefficients, or weights, W^+ and W^-, for each predictor map; predictor maps are usually binary, and W^+ and W^- refer to those areas where the binary pattern is either present, or not present, respectively. A weight of 0 is used where the pattern is unknown or missing. Weights are calculated using measurements of the area of binary pattern, the total study area, the number of mineral occurrences within the binary pattern, and the total number of occurrences in the study area. Evidence of mineralization is combined from several predictive maps, using a formulation of Bayes Rule. Starting with a prior probability of a mineral deposit occurring in a unit area, a posterior probability is calculated, which may be larger or smaller than the prior probability, depending on the overlap combination of predictor maps and their weights.

In comparison with the regression method, weights of evidence are easy to interpret, simple to program, missing data can be accommodated, and patterns with complex spatial geometry can be modelled with the same computational effort as those with simple geometry. On the other hand, the assumption that the predictor patterns are conditionally independent with respect to the points, implicit in weights of evidence modelling, must be tested and satisfied; in regression modelling, no such assumption is required.

In this paper, the weights of evidence method is outlined, and its application to gold prediction in Meguma terrane is discussed. In the earlier papers describing this application, such as Bonham-Carter et al. (1988), the maps used to predict gold were lithology, lake sediment geochemistry, proximity to anticlinal axes, distance to two types of contact, and distance to NW structures. Since then, an interesting new data set, the biogeochemistry of balsam fir twigs (Dunn et al., 1989) has come available. The effect of adding this new information to the prediction of gold potential is discussed, and the rather simple computational steps required to add a new predictor map are illustrated.

WEIGHTS OF EVIDENCE METHOD

Assume that for a particular region, a series of binary maps are known, and are to be used as predictors of mineral potential of a particular type. Further, assume that the locations of a number of mineral deposits, or occurrences, are known. The occurrences are treated as points. The binary predictor maps can be thought of as input maps; the desired end-products are output maps showing probability of occurrence and the associated uncertainty of the probability estimates.

The weights of evidence calculations involve several steps: 1) the estimation of a prior probability, i.e. the probability of mineral occurrence in a unit area, given no further information; 2) the calculation of positive and negative weights for each binary predictor map, using conditional probability ratios; 3) the application of a test for conditional independence of each pair of input maps with respect to the mineral occurrence points, possibly leading to the rejection or amalgamation of some input maps; 4) the calculation of posterior probability and uncertainty for each unique overlap combination of the binary predictor maps; and 5) the application of a goodness-of-fit test for testing the overall conditional independence assumption. These operations have been described previously (Agterberg et al., 1990; Bonham-Carter and Agterberg, 1990) and are briefly reviewed here.

If the study area is broken down into unit cells with a fixed area, u km$_2$, and the total area is t km$_2$, then $T = t/u$ is the total number of unit cells in the study area. If there are D unit cells containing an occurrence, equal to the number of occurrences if u is small enough (i.e. one occurrence per cell), then the prior probability that a unit cell chosen at random will contain an occurrence is $P(D) = D/T$, expressed as odds by

$$O(D) = \frac{P(D)}{1 - P(D)} = \frac{D}{T - D}.$$

For the j-th binary predictor map, $j = 1, 2, \ldots, n$, the area of pattern present in terms of unit cells is $B_j = b_j/u$, b_j is the area in km^2; the area where the pattern is not present is $\bar{B}_j$ which equals $T - B_j$ unless some of the region is unknown with respect to the j-th map. The areas of overlap between known occurrences and the j-th binary pattern are $B_j \cap D$, $\bar{B}_j \cap D$, $B_j \cap \bar{D}$ and $\bar{B}_j \cap \bar{D}$. The conditional probability of choosing a cell with an occurrence, given that the cell contains pattern B_j is

$$P(D|B_j) = \frac{B_j \cap D}{B_j}.$$

Similarly, three more conditional probabilities can be defined:

$$P(\bar{D}|B_j) = \frac{B_j \cap \bar{D}}{B_j},$$

$$P(D|\bar{B}_j) = \frac{\bar{B}_j \cap D}{\bar{B}_j}, \text{ and}$$

$$P(\bar{D}|\bar{B}_j) = \frac{\bar{B}_j \cap \bar{D}}{\bar{B}_j}.$$

But according to Bayes' rule

$$P(D|B_j) = \frac{P(B_j|D) \, P(D)}{P(B_j)}, \text{ and}$$

$$P(D|\bar{B}_j) = \frac{P(\bar{B}_j|D) \, P(D)}{P(\bar{B}_j)}.$$

So if the weights for pattern j are defined as

$$W_j^+ = log_e \frac{P(B_j|D)}{P(B_j|\bar{D})}, \text{ and}$$

$$W_j^- = log_e \frac{P(\bar{B}_j|D)}{P(\bar{B}_j|\bar{D})},$$

it can be shown that:

$$log_e O(D|B_j) = W_j^+ + log_e O(D), \text{ and}$$

$$log_e O(D|\bar{B}_j) = W_j^- + log_e O(D).$$

Suppose there are two binary predictor patterns, B_j, $j = 1, 2$. From probability theory

$$P(DB_1B_2) = P(B_2|DB_1) \, P(B_1|D) \, P(D).$$

If B_1 and B_2 are conditionally independent with respect to the mineral occurrence points, then:

$$P(B_2|DB_1) = P(B_2|D), \text{ thus}$$

$$P(DB_1B_2) = P(B_1|D) \, P(B_2|D) \, P(D).$$

It can then readily be shown that:

$$log_e O(D|B_1B_2) = W_1^+ + W_2^+ + log_e O(D),$$

$$log_e O(D|B_1\bar{B}_2) = W_1^+ + W_2^- + log_e O(D),$$

$$log_e O(D|\bar{B}_1B_2) = W_1^- + W_2^+ + log_e O(D), \text{ and}$$

$$log_e O(D|\bar{B}_1\bar{B}_2) = W_1^- + W_2^- + log_e O(D).$$

Similarly, if more binary predictor maps are used, they can be added provided that they are also conditionally independent with respect to the mineral occurrence points. In general, with B_j, $j = 1, 2, \ldots, n$ binary predictor maps, the log posterior odds are:

$$log_e O(D \mid B_1^k \cap B_2^k \cap B_3^k \ldots B_n^k) = \sum_{j=1}^{n} W_j^k + log_e O(D)$$

where the superscript k refers to the presence or absence of the binary pattern, and

$$W_j^k = \begin{cases} W_j^+ & \text{for } j\text{—th pattern present} \\ W_j^- & \text{for } j\text{—th pattern absent} \\ 0 & \text{for } j\text{—th pattern present.} \end{cases}$$

The posterior probability is then calculated using $P = O/(1+O)$. For each predicted map, the contrast $C = W^+ - W^-$ gives a useful measure of correlation with the mineral occurrence points. The weights W^+ and W^- have opposite signs, except that both become zero, and C becomes zero, where a map pattern has a distribution spatially independent of the points. For a positive spatial association, C will have positive values, usually in the range 0-2; C would take on negative values in a similar range for a negative association. Except in the special case of $C = 0$, the sign of W^+ will always be opposite that of W^-.

Two components of uncertainty of the posterior probability can be estimated: the uncertainty due to the variances of the weights, and the uncertainty due to one or more of the predictor maps being incomplete (partially known or missing). For each map, the variances of the weights can be calculated as:

$$\sigma^2(W_j^+) = \frac{1}{B_j \cap D} + \frac{1}{B_j \cap \bar{D}}, \text{ and}$$

$$\sigma^2(W_j^-) = \frac{1}{\bar{B}_j \cap D} + \frac{1}{\bar{B}_j \cap \bar{D}}.$$

These expressions use an asymptotic result (Bishop et al., 1975) which assumes that the number of occurrences is large. Assuming that the prior probability of an occurrence is D/T, it can be shown that the variance of the prior odds, $O(D)$, is $1/D$. The summed effect of uncertainty due to the weights for each unique overlap condition of the predictor maps is then

$$\sigma^2(P_{post}) = \left[\frac{1}{D} + \sum_{j=1}^{n} \sigma^2(W_j^k) \right] \cdot P_{post}^2.$$

where the superscript k is $+$ for presence, $-$ for absence, as before.

In order to estimate the uncertainty in posterior probability due to incomplete or missing data in the j-th binary predictor map, the following variance component can be calculated

$$\sigma_j^2(P_{post}) = \{P(D \mid B_j) - P(D)\}^2 P(B_j) + \{P(D \mid \bar{B}_j) - P(D)\}^2 P(\bar{B}_j)$$

For any unique overlap condition, $P(D)$ is here the posterior probability calculated from the non-missing binary maps. The terms $(P(D \mid B_j)$ and $P(D \mid \bar{B}_j)$ are the updated posterior probabilities assuming that the j-th binary pattern is present and absent, respectively. Thus the equation expresses the average squared changes in posterior probability if the pattern were known, weighted by the areal proportion $P(B_j)$ or $P(\bar{B}_j)$, respectively. For areas where two patterns are unknown, two variance components can be calculated, and so on. The total uncertainty for a unique overlap condition is then

$$\sigma^2(total) = \sigma^2(weights) + \sum_{j=1}^{n} \sigma_j^2(missing).$$

The uncertainty due to the weights, which includes the uncertainty of the prior probability, is in general correlated to the posterior probability, and maps of σ^2 (weights) have the same trends as the map of P_{post}. However, if P_{post} is studentized by forming the ratio P_{post}/σ, in effect applying a test to determine whether P_{post} is significantly greater than zero, the relative uncertainty of P_{post} is revealed, and areas with a ratio less than some cutoff, such as 1.96, can be masked out as being too uncertain.

Two tests to determine whether the assumption of conditional independence is satisfied can be applied. First, every possible pair of the binary predictor maps can be tested, and if a test fails, one of the maps can be rejected unless the other one is missing. Second, an overall test of goodness-of-fit can be applied, using a Kolmogorov-Smirnov statistic.

The pairwise test involves the calculation of observed and expected frequencies of unit cells for each possible pair of the input maps. If B_1 and B_2 are the two binary maps and D represents deposit points, then there are eight overlap possibilities between the points and the map patterns. This is shown in the following table of areas, where N indicates number of unit cells.

	$B_1 \cap B_2$	$B_1 \cap \bar{B}_2$	$\bar{B}_1 \cap B_2$	$\bar{B}_1 \cap \bar{B}_2$
D	$N(DB_1B_2)$	$N(DB_1\bar{B}_2)$	$N(D\bar{B}_1B_2)$	$N(D\bar{B}_1\bar{B}_2)$
$\bar{D}$	$N(\bar{D}B_1B_2)$	$N(\bar{D}B_1\bar{B}_2)$	$N(\bar{D}\bar{B}_1B_2)$	$N(\bar{D}\bar{B}_1\bar{B}_2)$

For each of these eight overlap possibilities, the observed area (in unit cells), x, is measured directly. For example, the x in the first row, first column of the table is obtained by counting the number of deposits (unit cells) occurring where both B_1 and B_2 are present. The predicted area, $\hat{m}$, is given by

$$\hat{m} = N(B_1B_2) \, P(D \mid B_1B_2) \, T$$

where $P(D|B_1B_2)$ is calculated from the weights of evidence model multiplied by total area T to give area, and $N(B_1B_2)$ is measured from the overlap of B_1 and B_2 in unit cells. Then

$$G^2 = -2 \sum_{i=l}^{8} x_i \log_e \frac{\hat{m}_i}{x_i}$$

is distributed as χ^2 with 2 degrees of freedom (Bishop et al., 1975).

Where the modelled areas, $\hat{m}$, differ strongly from the observed areas, x, the value of G^2 will be large and the hypothesis of conditional independence of B_1 and B_2 with respect to the points will be rejected. For example, if more deposit points occur in the region where both binary patterns are present than are predicted, and this statistical test fails, the final posterior probabilities (using multiple input maps including this pair B_1 and B_2) will be too large in some areas of the map. To avoid the problem, one of the patterns can be omitted from the final combined model. Alternatively, the two binary maps might be combined as a ternary map with three states. For example, if the two maps are B_1 and B_2, and the test fails because $N(B_1B_2D)$ is too large, then the three states could be chosen as $B_1 \cap B_2$, $\bar{B}_1 \cap B_2$ and $B_1 \cup B_2$ where $\cup$ is the exclusive or.

The overall goodness-of-fit test is applied after the final posterior probability map has been calculated. As with the pairwise test, each unique overlap condition of the input maps is determined. The actual number of unit cells occupied by deposit points occurring in each unique condition region of the map is measured, and compared to the number predicted from the model. Either a chi-squared test (Agterberg et al., 1990) or a Kolmogorov-Smirnov test (Bonham-Carter and Agterberg, 1990) can be used, the latter having some advantage because it avoids the requirement of binning the data. The results of the Kolmogorov-Smirnov test can be illustrated graphically. If the observed curve stays within a confidence envelope surrounding the predicted curve (Fig. 3), the hypothesis of conditional independence is not rejected, and the assumptions of the method are satisfied.

APPLICATION TO GOLD POTENTIAL, EASTERN NOVA SCOTIA

Study area and geological background

The "eastern shore" portion of the Meguma terrane in Nova Scotia (Fig. 1), is underlain by lower Paleozoic turbidites (Goldenville and Halifax formations) intruded by Devonian granites. Gold occurs in quartz veins, usually confined to the upper part of the Goldenville Formation. Within the study area, 68 occurrences have been documented (McMullin et al., 1986) of which 33 have recorded production. Several of the occurrences occur fairly close together, and they may be grouped into districts, e.g. Upper Seal Harbour (Fig. 1).

The arrays of gold-bearing veins occur predominantly on domes, flanks or plunges of regional anticlines, (Henderson, 1983; Keppie, 1976). Besides the strong relationship

to the anticline structures, various other regional controls on the distribution of gold districts have been discussed in the literature.

For example, the majority of gold-bearing veins are located within or on the upper margins of incompetent, impermeable slate horizons in the Goldenville Formation (Smith and Kontak, 1986). Most of the districts occur within greenschist facies rocks though some of them (Cochrane Hill, Forest Hill) are within rocks of the amphibolite facies (Taylor and Schiller, 1966). Mawer (1986) has suggested a positive correlation between gold occurrences and horizontal distance from the Goldenville-Halifax Formation transition zone, Devono-Carboniferous granitic intrusions and the chlorite-biotite isograd. Further, the transition between the Goldenville and Halifax formations appears to be a control for other metal concentrations besides gold (Graves and Zentilli, 1988). High levels of arsenic, tungsten and antimony are associated with much of the gold mineralization (Kontak and Smith, 1987). Finally, a study of the relationship of gold occurrences to lineaments in the Meguma rocks of the Halifax-Windsor area (Bonham-Carter et al., 1985), indicated that lineaments with a NNW-NW orientation had a spatial association with gold occurrences.

No consensus on the origin of the deposits has been achieved, but proposals include a) synsedimentary deposition on the seafloor, b) deposition early in the geological history of the area from metamorphic fluids, and multi-cyclic remobilization of components during deformation, and c) deposition late in the orogenic history either from granitic magmas or other deep crustal sources.

The maps used as predictor patterns were chosen to reflect as far as possible some of the current ideas about gold genesis, subject to the constraint that each map must provide either universal coverage or coverage of the majority of the area. For example, instead of using proximity to anticlinal axes, it would have been desirable to use a map showing the fold curvature index greater than 600 degrees (Keppie, 1976), but such a map was not available for the study area. It would also have been desirable to use a map of the greenschist facies, but the greenschist/amphibolite isograd is only patchily mapped.

The datasets used for the study consisted of 1) a geological map (Keppie, pers. comm., 1985), digitized by raster scanning; 2) a lake-sediment geochemical survey (Bingley and Richardson, 1978); 3) a biogeochemical survey using balsam fir twigs (Dunn et al., 1989) and 4) lineaments derived from a combination of mapped faults and features identified on satellite and vertical gradient magnetic images.

In an initial phase of the study, Wright (1988) and Wright et al. (1988) made digital images of the lake-sediment catchment basins, and derived a geochemical signature (combination of Au, Sb, As and W in lakes) that best predicted the known Au mineral occurrences. Maps showing distance 1) to NW lineaments, 2) to anticlinal fold axes, 3) to the Goldenville-Halifax contact, and 4) to the Devonian granite contact were generated by successively dilating these linear features.

Results

The map operations and data integration were carried out using the SPANS geographic information system (TYDAC, 1989). The GIS was useful not only for building a co-registered database, but also for allowing the combination of diverse data types (point, line, polygon, raster), and for carrying out the weights of evidence measurements and calculations.

The thresholding of the six original predictor maps to binary form (Wright, 1988; Agterberg et al., 1989) was optimized to maximize the contrast C, as summarized in Table 1. The balsam fir twig data for Au, not available for the earlier studies, was converted from point into map form by a weighted moving average technique (POTMAP), available in SPANS. An indicator variogram was first calculated, showing a range of about 3 km, for values thresholded at the 90th percentile (137 ppm), (George et al., in press). For the SPANS algorithm, a circular zone of influence, radius 3 km, with an arbitrary exponential decay function was used. Figure 2b shows the resulting map, using a percentile classification. Note that regions farther than 3 km from a sample point are masked out, and classified as missing.

The area of, and number of gold occurrences in each class on the resulting Au in balsam fir map were measured using SPANS. These data were fed into a short FORTRAN program outside SPANS to calculate the weights, contrasts and standard deviations for a series of Au levels (Table 2). Although the maximum value of W^+ occurs by thresholding at the 90th percentile, the maximum contrast, C, occurs for the 80th percentile. At this level, 24 out of 68 occurrences fall within the balsam fir Au anomaly, which occupies 435 km^2 out of the total area of 2591 km^2. Note
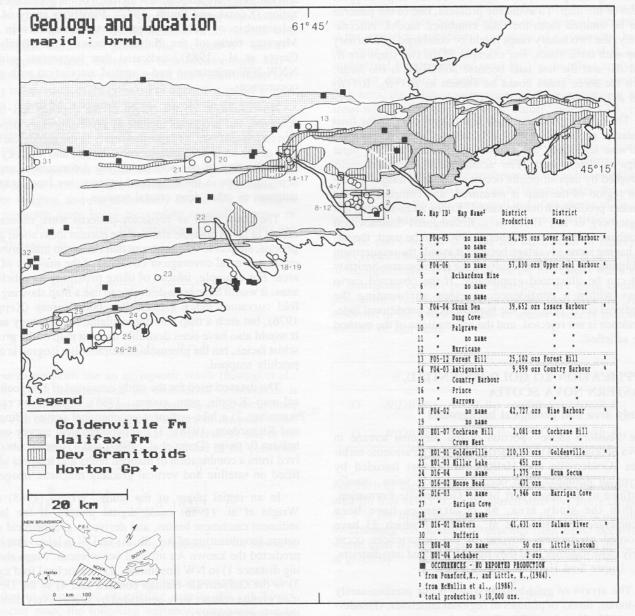

No.	Map ID[1]	Map Name[2]	District Production	District Name
1	F04-05	no name	34,295 ozs	Lower Seal Harbour *
2	"	no name	"	"
3	"	no name	"	"
4	F04-06	no name	57,830 ozs	Upper Seal Harbour *
5	"	Richardson Mine	"	"
6	"	no name	"	"
7	"	no name	"	"
8	F04-04	Skunk Den	39,652 ozs	Issacs Harbour *
9	"	Dung Cove	"	"
10	"	Palgrave	"	"
11	"	no name	"	"
12	"	Hurricane	"	"
13	F05-12	Forest Hill	25,102 ozs	Forest Hill *
14	F04-03	Antigonish	9,959 ozs	Country Harbour
15	"	Country Harbour	"	"
16	"	Prince	"	"
17	"	Harrows	"	"
18	F04-02	no name	42,727 ozs	Wine Harbour *
19	"	no name	"	"
20	E01-07	Cochrane Hill	2,081 ozs	Cochrane Hill
21	"	Crows Nest	"	"
22	E01-01	Goldenville	210,153 ozs	Goldenville *
23	E01-03	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	"	Harigan Cove	"	
28	"	no name	"	
29	D16-01	Eastern	41,631 ozs	Salmon River *
30	"	Dufferin	"	
31	E08-08	no name	50 ozs	Little Liscomb
32	E01-04	Lochaber	2 ozs	

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

Figure 1. Location map (inset) and geological map of the study area, showing known gold occurences.

	W^+	$\sigma(W^+)$	W^-	$\sigma(W^-)$	C	$\sigma(C)$	$C/\sigma(c)$
Goldenville Fm	0.3085	0.1280	-1.4689	0.4484	1.7774	0.4663	3.8117
Anticline axes	0.5452	0.1443	-0.7735	0.2370	1.3187	0.2775	4.7521
Au, biogeochem.	0.9045	0.2100	-0.2812	0.1521	1.1856	0.2593	4.5725
Lake sed. signature	1.0047	0.3263	-0.1037	0.1327	1.1084	0.3523	3.1462
Golden-Hal contact	0.3683	0.1744	-0.2685	0.1730	0.6368	0.2457	2.5918
Granite contact	0.3419	0.2932	-0.0562	0.1351	0.3981	0.3228	1.2332
NW lineaments	-0.0185	0.2453	0.0062	0.1417	-0.0247	0.2833	0.0872
Halifax Fm.	-1.2406	0.5793	0.1204	0.1257	-1.4610	0.5928	2.4646
Devonian Granite	-1.7360	0.7086	0.1528	0.1248	-1.8888	0.7195	2.6253

Cut off		Cumulative		W^+	$\sigma(W^+)$	W^-	$\sigma(W^-)$	C	$\sigma(C)$	$C/\sigma(C)$
%ile	ppb	area, km^2	occurrences#							
98	137	42	0	—	—	—	—	—	—	—
95	24	93	3	0.3438	0.5869	-0.0133	0.1254	0.3571	0.6002	0.2090
90	16	227	13	0.9439	0.2856	-0.1349	0.1362	1.0788	0.3164	3.4095
80	12	435	24	0.9045	0.2100	-0.2812	0.1521	1.1856	0.2593	4.5724
70	10	848	31	0.4733	0.1830	-0.2745	0.1659	0.7479	0.2470	3.0279
60	8	1070	35	0.3582	0.1719	-0.2771	0.1756	0.6353	0.2457	2.5853
50	7	1360	45	0.3701	0.1516	-0.4732	0.2100	0.8433	0.2590	3.2560
<50	3-6	2591	64	0.0695	0.1266	-0.7295	0.5028	0.7900	0.5185	1.5237
outside		2945	68	—	—	—	—	—	—	—

	$B_1 \cap B_2$	$B_1 \cap \bar{B_2}$	$\bar{B_1} \cap B_2$	$\bar{B_1} \cap \bar{B_2}$
D	3.0 (2.3)	7.0 (7.7)	2.0 (2.7)	10.0 (9.3)
$\bar{D}$	27.2 (21.8)	114.2 (119.6)	205.1 (210.5)	1159.4 (1154.0)

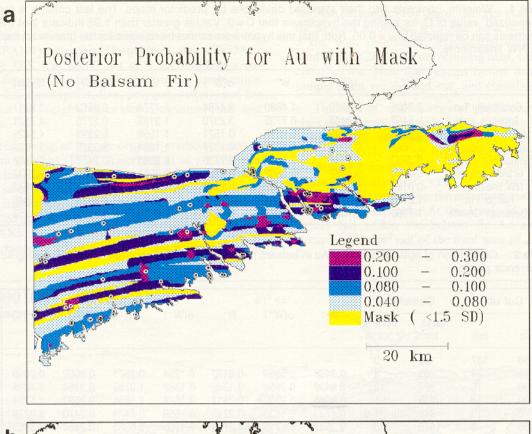Test statistic G^2 = 2.195, distributed as χ^2 with 2 df.

that the studentized value of C indicates that C is significantly greater than zero, and that when all the predictor maps are ranked by the magnitude of C the balsam fir Au map is among the most significant predictors of the occurrences. As expected the values of $\sigma(W^+)$ vary inversely with cumulative area, whereas those for $\sigma(W^-)$ vary in proportion to the cumulative area. Thus the value of W^+ at the 95% cutoff is not significantly different from zero (0.3438 ± 0.5869), but at the 80% cutoff, W^+ is much less uncertain (0.9045 ± 0.2100).

The relative size of W^+ as compared to W^- for each input map varies markedly (Table 1). For anticlinal axis corridors, there is about the same contribution to the contrast

from W^+ and W^-. On the other hand, in the balsam fir Au and lake sediment signature maps, the presence of the anomalous pattern has much more influence than the absence of the pattern, i.e. regions with anomalous geochemical patterns score very strongly, but regions that are not anomalous are not greatly downweighted. However, notice that the absence of Goldenville Formation causes a large downweighting, being strongest for areas of Devonian granite, but also being strong for areas of Halifax formation.

The degree of conditional independence of the lake-sediment and biogeochemical maps with respect to the mineral occurrences was tested (Table 3). By assuming conditional independence, the model predicts that 2.3 occurrences should be present in areas where both patterns overlap, compared with an observed number of 3, and in each of the eight overlap categories, there is good agreement between observed and predicted area, as shown by the test statistic G^2 = 2.195, which is much smaller than tabled chi-squared values for 2 d.f. and α=0.01. Note that these calculations can only be carried out where both maps are known, so the test is applied only to a subset of the total area.

As might be expected, the effect of adding the Au in balsam fir data is quite pronounced, as shown by the posterior probability maps without (Fig. 2a) and with (Fig. 2c) the extra predictor map. Because the balsam fir data is a good predictor, the effect of missing data (i.e. areas farther than
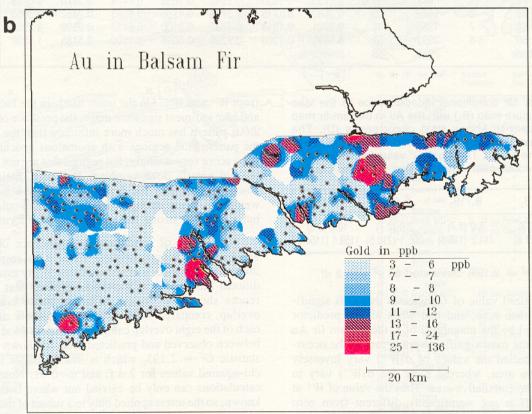
Figure 2. **a)** Posterior probability without balsam fir data. **b)** Au in balsam fir map. **c)** Posterior probability with balsam fir data.

3 km from the closest biogeochemical sample) is to raise the combined uncertainty ($\sigma_{weights}$ and $\sigma_{missing}$) in the missing areas to a level where $P_{post} < 1.5\sigma_{total}$. The areas masked out in Figures 2a, and 2c have been eliminated because the 'studentized' posterior probability (P_{post}/σ_{total}) is less than 1.5, indicating a relatively large uncertainty. These masked areas include the outcrop regions of Halifax Formation and granite (because almost no occurrences are known on these formations in the study area); they also include regions with missing balsam fir data in Figure 2b.

To illustrate the calculation of posterior probability, three examples are shown (Table 4). In the area of the Goldenville deposit, the four most important predictor patterns are present; the three least important are absent. Note that the posterior probability of 0.2132 ± 0.1109 is significantly greater than the prior probability of 0.023. In the case of Forest Hill before knowing the biogeochemical results, the posterior probability was only 0.0180 ± 0.0071, i.e. less than the prior probability. But with the biogeochemical data, the posterior probability increased to 0.0433 ± 0.0193, i.e. double the prior probability.

The overall conditional independence test (Fig. 3) indicates that the hypothesis of conditional independence is satisfied, because nowhere does the observed curve break through the confidence envelope surrounding the predicted curve. An interesting result is that if the gold occurrences with known production are plotted on a graph of posterior probability versus cumulative area, there appears to be a positive correlation between production and posterior probability (Fig. 4). In other words, the larger gold districts are associated with higher predictions of gold potential. Although this makes sense geologically, the mineral occurrence points are not numerically weighted by production or deposit size, and there is no statistical reason that the posterior probability would automatically correlate with gold production. However, the effect may be caused by spatial clustering of points in the more important gold districts, and this has yet to be tested.

Gold prospects

Figure 5 shows an enlargement of the region roughly centred on the Sherbrooke pluton. The masked areas are where $P_{post}/\sigma < 1.5$, i.e. where P_{post} is not significantly greater than zero. The geological contacts are superimposed in black, and the masked areas are either granite, or Halifax Formation, or where the biogeochemical Au map is uncertain. The rectangular areas A to E are of interest, with $P_{post} > 0.3$, i.e. the probability of a gold occurrence within a 1 km² area is about 1 in 3. Two of the areas, B (Goldenville) and E (Seal Harbour), are known gold districts. A, C and D, on the other hand, have no reported occurrences, yet they contain essentially the same signatures as B and E. No follow-up work has yet been undertaken to investigate these prospects.
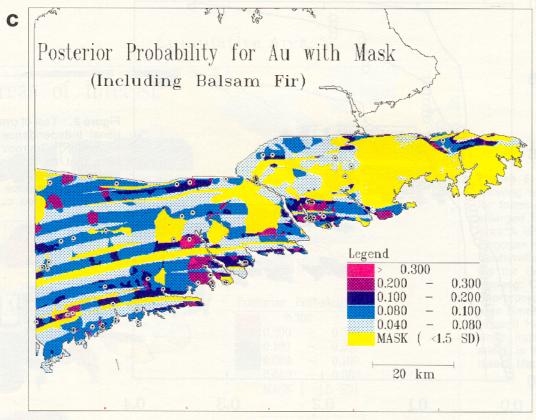

Figure 2. Continued

Table 4. Sample calculations of P_{post} and $\sigma(P_{post})$ for three cases A: Goldenville (including balsam fir data), B: Forest Hill (including balsam fir data), and C: Forest Hill (without balsam fir data). Note that the P_{post} for case B is about twice the prior probability (0.023), but P_{post} for case C is less than the prior probability.

	Case A: Goldenville			Case B: Forest Hill +			Case C: Forest Hill −	
	Status	Weight	St. dev.	Status	Weight	St. dev.	Weight	St. dev.
Log prior odds		-3.7500	0.1213		-3.7500	0.1213	-3.7500	0.1213
Goldenville Fm	+	0.3085	0.1280	+	0.3085	0.1280	0.3085	0.1280
Anticline axes	+	0.5452	0.1443	−	-0.7735	0.2370	-0.7735	0.2370
Au, biogeochem	+	0.9045	0.2100	+	0.9045	0.2100	−	
Lake sed. signature	+	1.0047	0.3263	−	-0.1037	0.1327	-0.1037	0.1327
Golden-Halifax contact	−	-0.2685	0.1730	+	0.3683	0.1744	0.3683	0.1744
Granite contact	−	-0.0562	0.1351	−	-0.0562	0.1351	-0.0562	0.1351
NW lineaments	−	0.0062	0.1417	−	0.0062	0.1417	0.0062	0.1417
Log posterior odds		-1.3056	—		-3.0960	—	-4.0004	—
Posterior probability, st. dev.		0.2132	0.1109		0.0433	0.0193	0.0180	0.0071
Studentized post prob.		1.922			2.2390		2.5370	

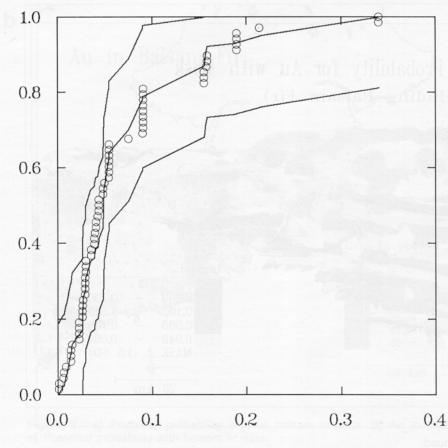+ = pattern present, − = pattern absent

Figure 3. Test of overall conditional independence, using a Kolmogorov-Smirnov statistic. Note that the observed curve (open circles) stays within the confidence envelope surrounding the predicted curve (solid line).
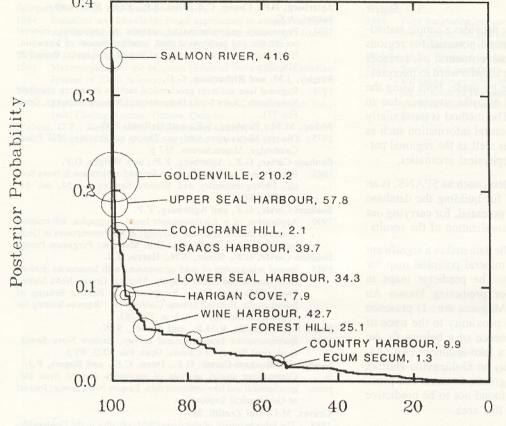
POSTERIOR PROBABILITY

Figure 4. P_{post} plotted against cumulative area, with the producing gold mines shown as circles whose radii reflect magnitude of reported production.
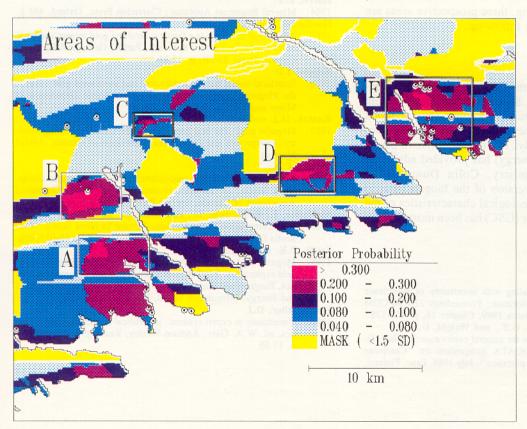
Chart labels (y-axis: Posterior Probability; x-axis: Cumulative Area, %):

- SALMON RIVER, 41.6
- GOLDENVILLE, 210.2
- UPPER SEAL HARBOUR, 57.8
- COCHCRANE HILL, 2.1
- ISAACS HARBOUR, 39.7
- LOWER SEAL HARBOUR, 34.3
- HARIGAN COVE, 7.9
- WINE HARBOUR, 42.7
- FOREST HILL, 25.1
- COUNTRY HARBOUR, 9.9
- ECUM SECUM, 1.3

Areas of Interest

Posterior Probability

>	0.300
0.200	– 0.300
0.100	– 0.200
0.080	– 0.100
0.040	– 0.080
MASK	(<1.5 SD)

10 km

Figure 5. Map of P_{post} showing areas suggested for follow-up exploration, enlarged from Figure 2c. Area A is at the head of Gegogan Harbour; B is the Goldenville district, including the Goldenville mine working; C is north of the Sherbrooke pluton; D is an area almost 6 km north of Holland Harbour, through which Indian River flows; and E is the area around Isaacs Harbour inlet.

181

CONCLUSIONS

1) Weights of evidence modelling provides a simple statistical method for predicting mineral potential for regions where a number of representative mineral occurrences are known. The weights are straightforward to interpret; an estimate of uncertainty can be made, both using the variances of the weights and also the variance due to missing or incomplete data. The method is particularly well-suited for modelling structural information such as proximity to linear features, as well as the regional patterns of geochemical and geophysical anomalies.

2) A geographic information system, such as SPANS, is an excellent computing platform for building the database required for mapping mineral potential, for carrying out model calculations, and for visualization of the results.

3) The addition of Au in balsam fir data makes a significant contribution to the predicted mineral potential map for the eastern Meguma terrane. The predictor maps in order of their importance for predicting known Au occurrences in eastern shore Meguma are: 1) presence of Goldenville Formation; 2) proximity to the trace of an anticlinal axis; 3) the presence of a balsam fir Au anomaly; 4) the presence of a lake-sediment (Au, As, Sb, W) anomaly; 5) proximity to Goldenville-Halifax contact; and (6) proximity to granite contact. The proximity to NW lineaments was found not to be predictive of the known occurrences in this area.

4) The posterior probability calculated for the Au occurrences shows a positive correlation with production. This gives added strength to the predictions.

5) As a result of the modelling, three prospective areas are suggested for exploration follow-up.

ACKNOWLEDGMENTS

REFERENCES

Agterberg, F.P.
1989: Systematic approach to dealing with uncertainty of geoscience information in mineral exploration; Proceedings 21st APCOM Symposium, Las Vegas, March 1989, Chapter 18, p. 165-178.

Agterberg, F.P., Bonham-Carter, G.F., and Wright, D.F.
1990: Statistical pattern integration for mineral exploration; in Gaal. G, (ed.) Proceedings COGEODATA Symposium on "Computer Applications in Resource Exploration", July 1988, Espo, Finland, Pergamon Press.

Agterberg, F.P., Chung, C.F., Divi, S.R., Eade, K.E., and Fabbri, A.G.
1981: Preliminary geomathematical analysis of geological, mineral occurrence and geophysical data, southern district of Keewatin, Northwest Territories; Geological Survey of Canada, Open File 718, 29 p.

Bingley, J.M. and Richardson, G.G.
1978: Regional lake sediment geochemical surveys in eastern mainland Nova Scotia; Nova Scotia Department of Mines and Energy, Open File 371.

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 geological datasets for gold exploration in Nova Scotia; Photogrammettry and Remote Sensing, v. 54, no. 11, p. 1585-1592.

Bonham-Carter, G.F. and Agterberg, F.P.
1990: Application of a microcomputer-based geographic information system to mineral potential mapping; in Microcomputers in Geology, v. 2, ed. T. Hanley and D.F. Merriam; Pergamon Press.

Bonham-Carter, G.F., Rencz, A.N., Harris, J.R.
1985: Spatial relationship of gold occurrences with lineaments derived from Landsat and Seasat imagery, Meguma Group, Nova Scotia; Proceedings International Symposium on Remote Sensing of Environment, Fourth Thematic Conference: "Remote Sensing for Exploration Geology". v. 2, p. 755-767.

Dunn, C.E., Banville, R.M.P., and Adcock, S.W.
1989: Reconnaissance biogeochemical survey, Eastern Nova Scotia; Geological Survey of Canada, Open File 2002, 95 p.

George, H., Bonham-Carter, G.F., Dunn, C.E., and Rogers, P.J.
in Comparative spatial analysis of anomalies derived from bio-
press: geochemical and lake-sediment data, Eastern Nova Scotia; Journal of Geochemical Exploration.

Graves, M.C. and Zentilli, M.
1988: The lithochemistry of metal-enriched coticules in the Goldenville-Halifax transition zone of the Meguma Group, Nova Scotia; in Current Research, Part B, Geological Survey of Canada, Paper 88-1B, p. 251-261.

Harris, D.P.
1984: Mineral Resources Appraisal; Clarendon Press, Oxford, 445 p.

Henderson, J.R.
1983: Analysis of structure as a factor controlling gold mineralization in Nova Scotia, in Current Research, Part B, Geological Survey of Canada, Paper 83-1B, p. 13-21.

Keppie, J.D.
1976: Structural model for the saddle reef and associated gold veins in the Meguma Group, Nova Scotia; Nova Scotia Department of Mines and Energy, Paper 76-1, 34 p.

Kontak, D.J. and Smith, P.K.
1987: Meguma gold: The best kept secret in the Canadian mining industry; Prospectors and Developers Association of Canada, Annual Meeting.

Mawer, C.K.
1986: The bedding-concordant gold-quartz veins of the Meguma Group, Nova Scotia; in Turbidite-Hosted Gold Deposits; Geological Association of Canada, Special Paper 32, p. 135-148.

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 Files 86-055, 056.

Smith, P.K. and Kontak, D.J.
1986: Meguma gold studies: Advances in geological insight as an aid to gold exploration: Tenth Annual Open House and Review of Activities, Program and Summaries; Nova Scotia Department of Mines and Energy, Information Series, No. 12, p. 105-114.

Spiegelhalter, D.J.
1986: Uncertainty in expert systems; in Artifical Intelligence and Statistics, ed. W.A. Gale; 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; Journal of the Royal Statistical Society, A, Part 1, p. 35-77.

Taylor, F.C. and Schiller, E.A.
1966: Metamorphism of the Meguma Group of Nova Scotia; Canadian Journal of Earth Sciences, v. 3, p. 959-974.

TYDAC
1989: SPANS Users Guide, Version 4.0; TYDAC Technologies Inc., 1600 Carling Avenue, Ottawa, Ontario.

Watson, G.P., Rencz, A.N., and Bonham-Carter, G.F.
1989: Computers assist prospecting; Geos., v. 18, no., 1, p. 8-15.

Wright, D.F.
1988: Data integration and geochemical evaluation of Meguma terrane, Nova Scotia, for gold mineralization; unpublished M.Sc. Thesis, University of Ottawa, 82 p.

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, Nova Scotia; in Prospecting in Areas of Glaciated Terrain — 1988, ed. D.R. MacDonald and K.A. Mills, Canadian Institute of Mining and Metallurgy, p. 501-515.