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

G.F. Bonham-Carter¹, F.P. Agterberg¹ and D.F. Wright¹


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 variance 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 $\sigma(C)$. 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 source 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 variations 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 $C$ (C). Les facteurs de pondération sont calculés 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 territoire 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 montrait des concentrations d’or anomales dans des sapis baux-miers, donnant 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 montreront 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 cumulativée, on peut voir que les emplacements des plus grands gisements aurifères sont associés 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 volcanicogenic 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 balds 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 (Agerberg et al., 1990; Bonham-Carter and Agerberg, 1990) and are briefly reviewed here.

If the study area is broken down into unit cells with a fixed area, \( u \text{ km}^2 \), and the total area is \( T \text{ km}^2 \), 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 \( \text{km}^2 \); the area where the pattern is present is \( B_j \) which equals \( T - B \) 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 \( B_j \cap D, B_j \cap \bar{D}, B_j \cap D \) and \( B_j \cap \bar{D} \). The conditional probability of choosing a cell with an occurrence, given that the cell contains pattern \( B_j \) is

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

Similarly, three more conditional probabilities can be defined:

\[
P(D | \bar{B}_j) = \frac{B_j \cap \bar{D}}{\bar{D}},
\]

\[
P(D | B_j) = \frac{B_j \cap D}{B_j}, \text{ and}
\]

\[
P(D | \bar{B}_j) = \frac{B_j \cap \bar{D}}{\bar{B}_j}.
\]

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

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

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

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

\[
P(DB_1B_2) = P(DB_2 | DB_1) P(B_1 | D) P(D).
\]

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:

\[
\log_e O(D | B_1B_2) = W_1^+ + W_2^+ + \log_e O(D),
\]

\[
\log_e O(D | B_1B_2) = W_1^- + W_2^- + \log_e O(D),
\]

\[
\log_e O(D | \bar{B}_1\bar{B}_2) = W_1^- + W_2^- + \log_e O(D), \text{ and}
\]

\[
\log_e O(D | B_1\bar{B}_2) = W_1^- + W_2^- + \log_e O(D).
\]
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:

$$\log O(D | B_1 \cap B_2 \cap \ldots \cap B_n) = \sum_{j=1}^{n} W^*_j + \log O(D)$$

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

$$W^*_j = \begin{cases} 
W_j & \text{for } j\text{-th pattern present} \\
W_j & \text{for } j\text{-th pattern absent} \\
0 & \text{for } j\text{-th pattern present.}
\end{cases}$$

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 $W^-$. 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 $\sigma^2$ (weights) have the same trends as the map of $P_{post}$. However, if $P_{post}$ is studentized by forming the ratio $P_{post} / \sigma$, 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.

<table>
<thead>
<tr>
<th>$B_1 \cap B_2$</th>
<th>$B_1 \cap \bar{B}_2$</th>
<th>$B_1 \cap B_2$</th>
<th>$B_1 \cap \bar{B}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D$</td>
<td>$N(DB_1B_2)$</td>
<td>$N(DB_1\bar{B}_2)$</td>
<td>$N(DB_2B_2)$</td>
</tr>
<tr>
<td>$D$</td>
<td>$N(DB_1\bar{B}_2)$</td>
<td>$N(DB_2B_2)$</td>
<td>$N(DB_2\bar{B}_2)$</td>
</tr>
</tbody>
</table>

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

$$\hat{m} = N(B_1B_2) \cdot P(D | B_1B_2) \cdot T$$
where \( P(D | B_i B_j) \) is calculated from the weights of evidence model multiplied by total area \( T \) to give area, and \( N(B_i, B_j) \) is measured from the overlap of \( B_i \) and \( B_j \) in unit cells. Then

\[
G^2 = -2 \sum_{i=1}^{s} x_i \log \left( \frac{\hat{m}_i}{x_i} \right)
\]

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

Where the modelled areas, \( \hat{m}_i \), differ strongly from the observed areas, \( x_i \), the value of \( G^2 \) will be large and the hypothesis of conditional independence of \( B_i \) and \( B_j \) 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 \( B_i \) and \( B_j \)) 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 \( B_i \) and \( B_j \), and the test fails because \( N(B_i, B_j, D) \) is too large, then the three states could be chosen as \( B_i \cap B_j \), \( B_i \cap B_j \) and \( B_i \cup B_j \) 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 antiformal 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 greenstone 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 antiformal folds, 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

![Geology and Location Map](image)

**Figure 1.** Location map (inset) and geological map of the study area, showing known gold occurrences.
Table 1. Weights, contrasts and their standard deviations for predictor maps. The last column is the 'studentized' value of C, for testing the hypothesis that C = 0. Values greater than 1.96 indicate that the hypothesis can be rejected at $\alpha = 0.05$. Note that this hypothesis cannot be rejected for the granite contact and NW lineaments.

<table>
<thead>
<tr>
<th></th>
<th>W⁺</th>
<th>σ(W⁺)</th>
<th>W⁻</th>
<th>σ(W⁻)</th>
<th>C</th>
<th>σ(C)</th>
<th>C/σ(C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goldenville Fm.</td>
<td>0.3085</td>
<td>0.1280</td>
<td>-1.4689</td>
<td>0.4484</td>
<td>1.7774</td>
<td>0.4663</td>
<td>3.8117</td>
</tr>
<tr>
<td>Anticline axes</td>
<td>0.5452</td>
<td>0.1443</td>
<td>-0.7735</td>
<td>0.2370</td>
<td>1.3187</td>
<td>0.2775</td>
<td>4.7251</td>
</tr>
<tr>
<td>Au, biogeochem.</td>
<td>-0.9045</td>
<td>0.2100</td>
<td>-0.2812</td>
<td>0.1521</td>
<td>1.1856</td>
<td>0.2593</td>
<td>4.7252</td>
</tr>
<tr>
<td>Lake sed. signature</td>
<td>1.0047</td>
<td>0.3263</td>
<td>-1.0373</td>
<td>0.1327</td>
<td>1.1084</td>
<td>0.3524</td>
<td>3.1462</td>
</tr>
<tr>
<td>Golden-Hal contact</td>
<td>0.3683</td>
<td>0.1744</td>
<td>-0.2685</td>
<td>0.1730</td>
<td>0.6368</td>
<td>0.2457</td>
<td>2.5918</td>
</tr>
<tr>
<td>Granite contact</td>
<td>0.3419</td>
<td>0.2932</td>
<td>-0.0662</td>
<td>0.1351</td>
<td>0.3981</td>
<td>0.3228</td>
<td>1.3232</td>
</tr>
<tr>
<td>NW lineaments</td>
<td>-0.0185</td>
<td>0.2453</td>
<td>0.0062</td>
<td>0.1417</td>
<td>-0.0247</td>
<td>0.2833</td>
<td>0.0872</td>
</tr>
<tr>
<td>Halifax Fm.</td>
<td>-1.2406</td>
<td>0.5793</td>
<td>0.1204</td>
<td>0.1257</td>
<td>-1.4641</td>
<td>0.5928</td>
<td>2.4646</td>
</tr>
<tr>
<td>Devonian Granite</td>
<td>-1.7360</td>
<td>0.7086</td>
<td>0.1528</td>
<td>0.1248</td>
<td>-1.8888</td>
<td>0.7195</td>
<td>2.6253</td>
</tr>
</tbody>
</table>

Table 2. Calculation of optimal cut off of Au in balsam fir, to maximize the contrast, C, with known gold occurrence points.

<table>
<thead>
<tr>
<th>Cut off</th>
<th>Cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>%ile</td>
<td>ppb</td>
</tr>
<tr>
<td>98</td>
<td>137</td>
</tr>
<tr>
<td>95</td>
<td>24</td>
</tr>
<tr>
<td>90</td>
<td>16</td>
</tr>
<tr>
<td>80</td>
<td>12</td>
</tr>
<tr>
<td>70</td>
<td>10</td>
</tr>
<tr>
<td>60</td>
<td>8</td>
</tr>
<tr>
<td>50</td>
<td>7</td>
</tr>
<tr>
<td>&lt; 50</td>
<td>3-6</td>
</tr>
<tr>
<td>outside</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Test for conditional independence of the lake sediment signature map (B₁) with the Au in balsam fir map (B₂) with respect to known gold occurrences (D). The observed areas are shown first, followed by the areas predicted by the model. The small G² value indicates that a hypothesis of conditional independence is not rejected.

<table>
<thead>
<tr>
<th></th>
<th>$B₁ \cap B₂$</th>
<th>$B₁ \cap B₂$</th>
<th>$B₁ \cap B₂$</th>
<th>$B₁ \cap B₂$</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>3.0 (2.3)</td>
<td>7.0 (7.7)</td>
<td>2.0 (2.7)</td>
<td>10.0 (9.3)</td>
</tr>
<tr>
<td>D</td>
<td>27.2 (21.8)</td>
<td>114.2 (119.6)</td>
<td>205.1 (210.5)</td>
<td>1158.4 (1154.0)</td>
</tr>
</tbody>
</table>

Test statistic $G² = 2.195$, distributed as $\chi²$ 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 $σ(W⁺)$ vary inversely with cumulative area, whereas those for $σ(W⁻)$ 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 $G² = 2.195$, 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 (σs and σm) in the month areas to a level where P_post < 1.5σ_total. The areas masked out in Figures 2a, 2b, 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/σ < 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.
Table 4. Sample calculations of $P_{post}$ and $\alpha(P_{post})$ 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.

<table>
<thead>
<tr>
<th></th>
<th>Case A: Goldenville</th>
<th>Case B: Forest Hill +</th>
<th>Case C: Forest Hill -</th>
</tr>
</thead>
<tbody>
<tr>
<td>Status</td>
<td>Weight</td>
<td>St. dev.</td>
<td>Weight</td>
</tr>
<tr>
<td>Log prior odds</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goldenville Fm +</td>
<td>0.3085</td>
<td>0.1260</td>
<td>0.3085</td>
</tr>
<tr>
<td>Anticline axes +</td>
<td>0.5452</td>
<td>0.1443</td>
<td>-0.7735</td>
</tr>
<tr>
<td>Au. biogeochem</td>
<td>+ 0.9045</td>
<td>0.2100</td>
<td>+ 0.9045</td>
</tr>
<tr>
<td>Lake sed. signature</td>
<td>+ 1.0047</td>
<td>0.3263</td>
<td>-0.1037</td>
</tr>
<tr>
<td>Golden-Halifax</td>
<td>-0.2685</td>
<td>0.1730</td>
<td>+ 0.3683</td>
</tr>
<tr>
<td>contact Granite</td>
<td>-0.0562</td>
<td>0.1351</td>
<td>-0.0562</td>
</tr>
<tr>
<td>NW lineaments</td>
<td>0.0062</td>
<td>0.1417</td>
<td>0.0062</td>
</tr>
<tr>
<td>Log posterior odds</td>
<td>-1.3056</td>
<td></td>
<td>-3.0960</td>
</tr>
<tr>
<td>Posterior</td>
<td>0.2132</td>
<td>0.1109</td>
<td>0.0433</td>
</tr>
<tr>
<td>probability,</td>
<td>0.0180</td>
<td>0.0071</td>
<td></td>
</tr>
<tr>
<td>st. dev.</td>
<td>2.3960</td>
<td>2.5370</td>
<td></td>
</tr>
</tbody>
</table>

+ = 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).
Figure 4. $P_{\text{post}}$ plotted against cumulative area, with the producing gold mines shown as circles whose radii reflect magnitude of reported production.

Figure 5. Map of $P_{\text{post}}$ 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.
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 fir 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

This work was supported by the Geological Survey of Canada under the Canada-Nova Scotia Mineral Development Agreement (1984-1989). Duncan Keppie (Nova Scotia Department of Mines and Energy) prepared a special geological basemap for this study. Peter Rogers (Nova Scotia Department of Mines and Energy) has provided advice on the lake-sediment geochemistry. Colin Dunn (GSC) provided a pre-publication version of the biogeochemical data, and discussion of the geological characteristics of the gold deposits with Al Sangster (GSC) has been most helpful.

REFERENCES

Agterberg, F.P., Chung, C.F., Divi, S.R., Eade, K.E., and Fabbrri, A.G.

Bingley, J.M. and Richardson, G.G.

Bishop, M.M., Fienberg, S.E., and Holland, P.W.

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

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., Renz, A.N., Harris, J.R.

Dunn, C.E., Banville, R.L.M., and Adcock, S.W.

George, H., Bonham-Carter, G.F., Dunn, C.E., and Rogers, P.J.

Graves, M.C. and Zentilli, M.

Harris, D.P.

Henderson, J.R.

Keppie, J.D.

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.

McMullin, J., Richardson, G., and Goodwin, T.

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 Artificial Intelligence and Statistics, ed. W.A. Gale; Addison-Wesley, Reading, Massachusetts, p. 17-55.
Spiegelhalter, D.J. and Knill-Jones, R.P.

Taylor, F.C. and Schiller, E.A.

TYDAC

Watson, G.P., Renz, A.N., and Bonham-Carter, G.F.

Wright, D.F.

Wright, D.F., Bonham-Carter, G.F., and Rogers, P.J.