From Davis, J. C. and Herzfeld, U.C. (eds.), 1993, *Computers in Geology, 25 Years of Progress*, Oxford University Press, Oxford, pp. 13 - 32.

WEIGHTS OF EVIDENCE MODELING AND WEIGHTED LOGISTIC REGRESSION FOR MINERAL POTENTIAL MAPPING*

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

During the past few years, we have developed a method of weights of evidence modeling for mineral potential mapping (cf. Agterberg, 1989; Bonham-Carter et al., 1990). In this paper, weights of evidence modeling and logistic regression are applied to occurrences of hydrothermal vents on the ocean floor, East Pacific Rise near 21 N. For comparison, logistic regression is also applied to occurrences of gold deposits in Meguma Terrane, Nova Scotia.

The volcanic, tectonic, and hydrothermal processes along the central axis of the East Pacific Rise at 21 N were originally studied by Ballard et al. (1981). Their maps were previously taken as the starting point for a pilot project on estimation of the probability of occurrence of polymetallic massive sulfide deposits on the ocean floor (Agterberg and Franklin, 1987). In the earlier work, presence or absence of deposits in relatively large square cells was related to explanatory variables quantified for small square cells (pixels) by means of stepwise multiple regression and logistic regression. In this paper, weights of evidence modeling and weighted logistic regression are applied to the same maps but a geographic information system (Intera-TYDAC SPANS, 1991) was used to create polygons for combinations of maps. These polygons can be classified taking the different classes from each map. Probabilities estimated for the resulting "unique conditions" can be classified and displayed. The vents are correlated with only a few patterns and it is relatively easy to interpret the weights and final probability maps in terms of the underlying volcanic, tectonic, and hydrothermal processes. The vents are situated along the central axis of the rise together with the youngest voicanies. They occur at approximately the same depth below sea level, tend to be associated with pillow flows rather than sheet flows, and with absence of fissures which are more prominent in older volcanics.

Contrary to weights of evidence modeling, weighted logistic regression (cf. Agterberg, 1992, for discussion of algorithm) can he applied when the explanatory variables are not conditionally independent. This method was previously applied by Reddy et al. (1991) to volcanogenic massive suifide deposits in the Snow Lake area of Manitoba.. The assumptions underlying these methods will be evaluated in detail for the seafloor example.

The gold deposits in Meguma Terrane, Nova Scotia, were previously used for weights of evidence modeling (Bonham-Carter et al., 1988; Wright, 1988; Agterberg et al., 1990; Bonham-Carter et al., 1990). It will be shown in this paper that similar results are obtained when weighted logistic regression is used. The degree of fit of the different statistical models is evaluated for all applications in this paper. A difference between the two examples of application is that the number of hydrothermal vents on the seafloor is small in comparison to the number of Nova Scotia gold deposits. For this reason, the posterior probabilities have greater relative precision in the Nova Scotia example.

Hydrothermal Vents on the Ocean Floor

The maps used for this example are shown in Plate 1 (color illustrations are grouped together; Plates 2-5 follow Plate 1). The volcanos on the East Pacific Rise are of two types. (1) pillow flows, and (2) sheet flows. There are three age classes. Ballard *et al.* (1981) determined relative age by measuring relative amounts of sediments deposited on top of the voleanies. The resulting six litho-age units are shown in Plate la together with the occurrences of 13 hydrothermal vents. The other map patterns of Plate 1 are for

topography (depth to sea bottom, Pl. 1b), distance to contact between youngest pillow flows and youngest sheet flows (Pl. le), and distance to fissures (Pl. 1d). The first step in weights of evidence modeling consisted of constructing binary patterns which are relatively strongly correlated with the vents. Five binary patterns for the seafloor example are shown in Plate 2. Each pattern has positive weight W^+ for presence of a feature and negative weight W^- for its absence. The contrast $C = W^+ - W^-$ is a measure of the strength of correlation between the vents and a binary pattern. A binary pattern can be constructed by maximizing the contrast. For example, from Plate lb for topography it can he seen that nearly all (12 of 13) vents belong to a single 20-m topographic interval and this prompted the choice of the binary pattern of Plate 2b. A table of contrast versus corridor width can be used for deciding on the binary pattern

Plates

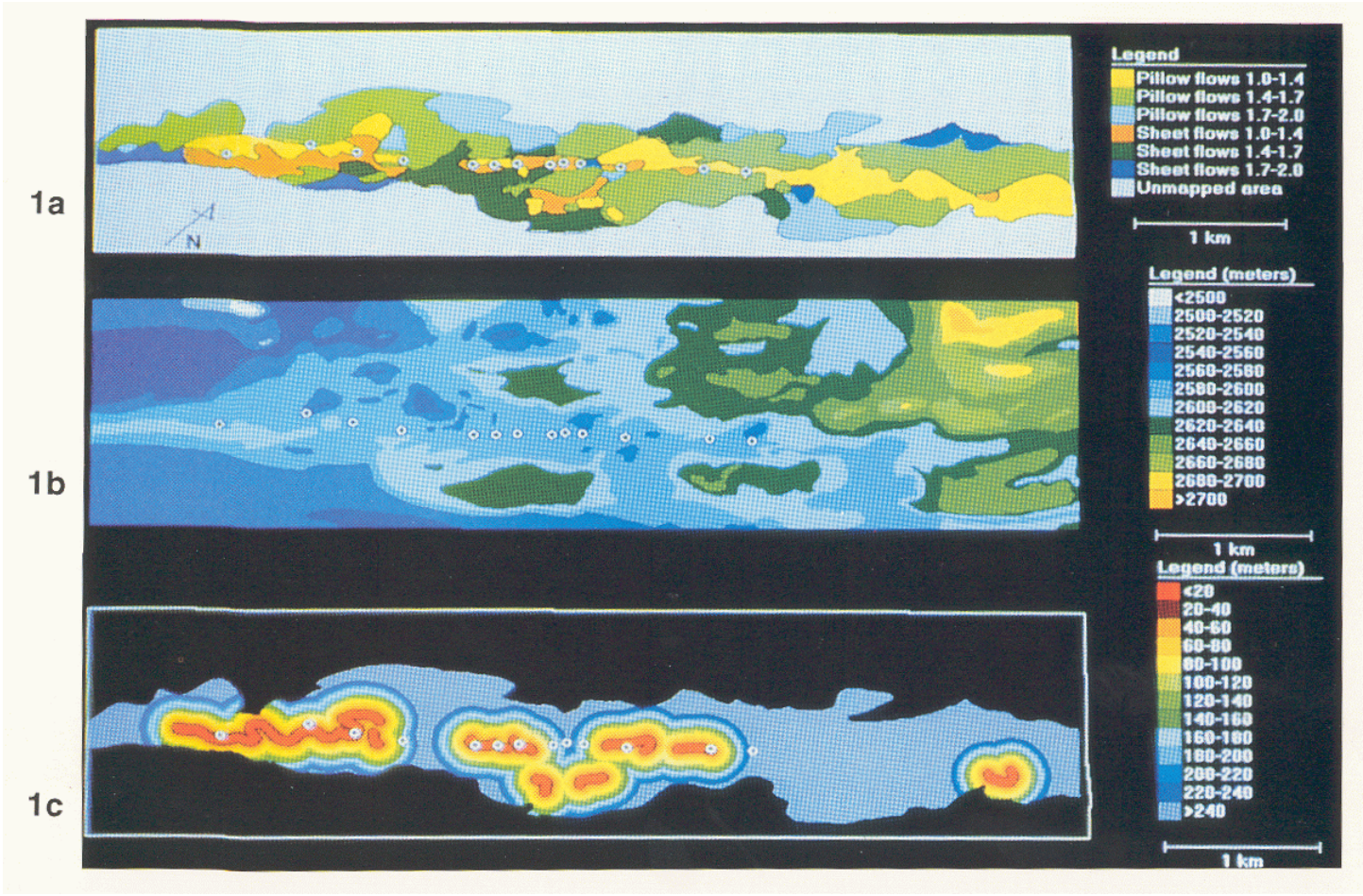
Plate 1: Patterns used for example of occurrence of hydrothermal vents on the seafloor (East Pacific Rise, 21 N; based on Fig. 5 of Ballard *et al.*, 1981). (a) litho-age units and hydrothermal vents (dots); relative age classes are 1.0-1.4 (youngest), 1.4-1.7 (intermediate), and **1.7-2.0** (oldest); (b) topography (depth below sea level); (c) corridors around contact between youngest pillow and sheet flows; (d) corridors around fissures.

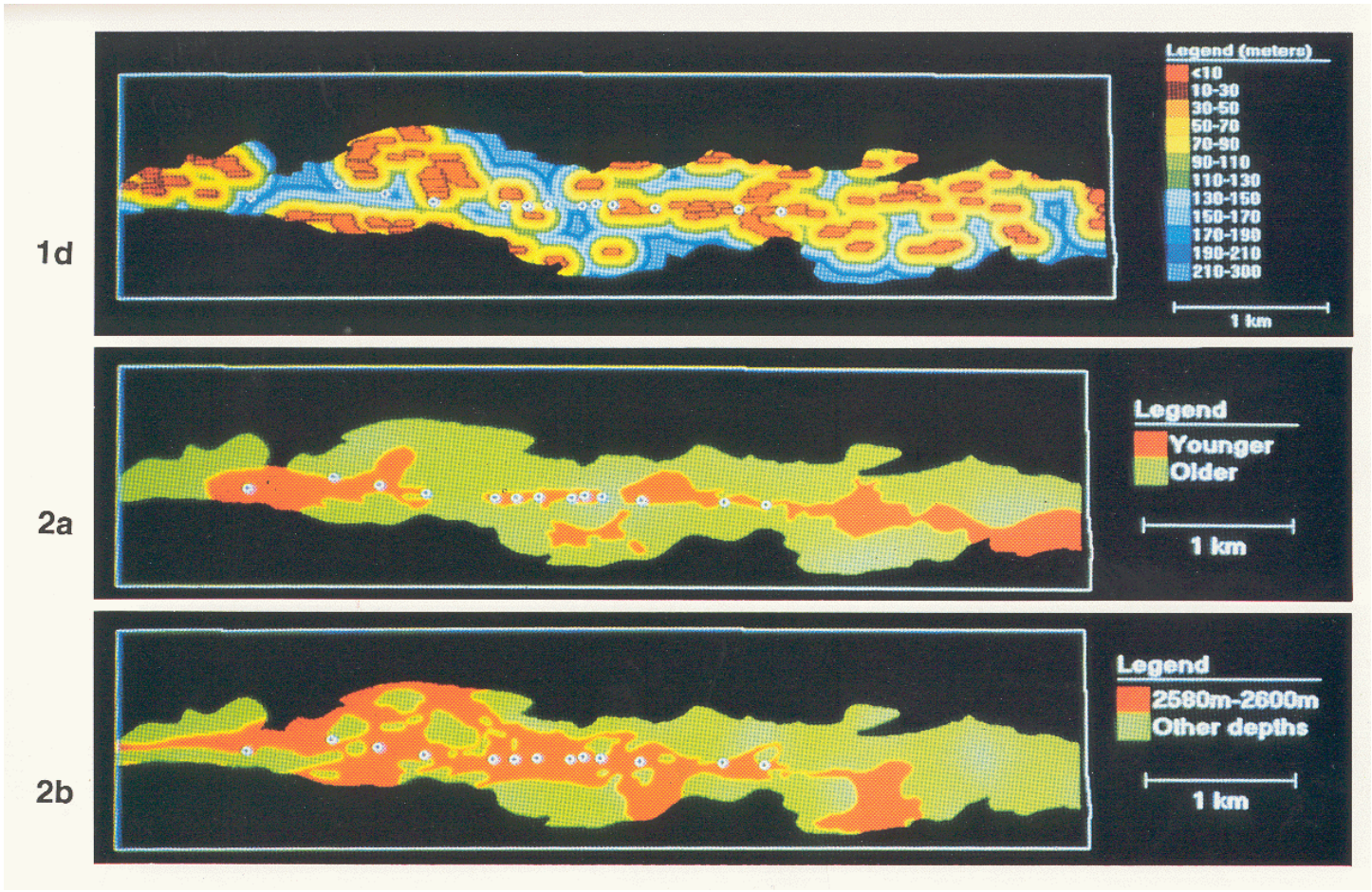
Plate 2: Binary patterns derived from Plate 1 (study area as in Pl. la). Weights for presence or absence of features were calculated for 0.01 km^2 unit cell size. (a) age, $W^+ = 2.251$ for presence of youngest volcanos, $W^- = -1.231$; (b) topography, $W^+ = 2.037$ for presence within zone with water depth between 2580 m and 2600 m, $W^- = -0.811$; (c) rock type, $W^+ = 0.632$ for presence of pillow flows, $W^- = -0.338$ for presence of sheet flows; (d) proximity to contact between youngest volcanics, $W^+ = 2.259$ for points within 20 m, $W^- = -0.570$; (c) absence/presence of fissures, $W^+ = 0.178$ for points at distances greater than 110 m, $W^- = -0.097$ for points within 110 m.

Plate 3: Seafloor example. (a) Posterior probability map with eight unique conditions for the overlap of first three binary patterns of Plate 2, unit cell size = 0.01 km^2; (b) t-value map corresponding to Plate 3a (t-value is ratio of posterior probability and its standard deviation); (c) posterior probability map with 31 unique conditions for the five binary patterns of Plate 2, unit cell size = 0.001 km .

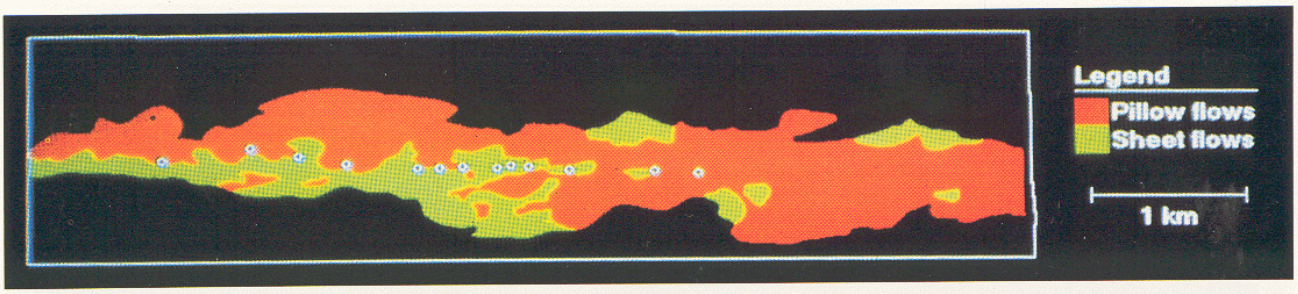
Plate 4: Weighted logistic regression applied to seafloor example. (a) Posterior probability map with 31 unique conditions for the five binary patterns of Plate 2, unit cell size = 0.001 km^2 *(cf.* Pl. 3c); (b) t-value map corresponding to Plate 4a; (c) Posterior probability map with 196 unique conditions for modified logistic model. See text and caption of Table 3 for further explanation.

Plate 5: Weighted logistic regression applied to gold deposits (circles) in Meguma Terrane, Nova Scotia. (a) Posterior probability map with 91 unique, conditions for seven binary patterns without missing data, unit cell size = 1 km^2, (b) t-value map corresponding to Plate 5a.

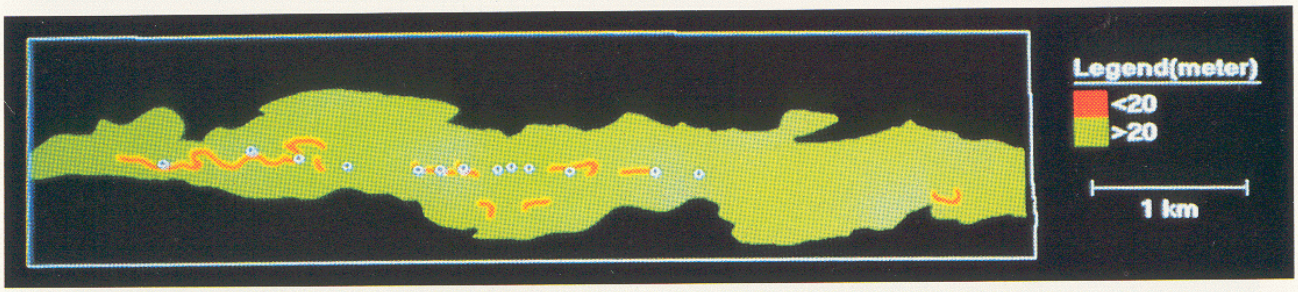

1a
1b
1c
N
Legend
Pillow flows 1.0-1.4
Pillow flows 1.4-1.7
Pillow flows 1.7-2.0
Sheet flows 1.0-1.4
Sheet flows 1.4-1.7
Sheet flows 1.7-2.0
Unmapped area
1 km
Legend (meters)
<2500
2500-2520
2520-2540
2540-2560
2560-2580
2580-2600
2600-2620
2620-2640
2640-2660
2660-2680
2680-2700
>2700
1 km
Legend (meters)
<20
20-40
40-60
60-80
80-100
100-120
120-140
140-160
160-180
180-200
200-220
220-240
>240
1 km


1d
Legend (meters)
<10
10-30
30-50
50-70
70-90
90-110
110-130
130-150
150-170
170-190
190-210
210-300
1 km
2a
Legend
Younger
Older
1 km
2b
Legend
2580m-2600m
Other depths
1 km

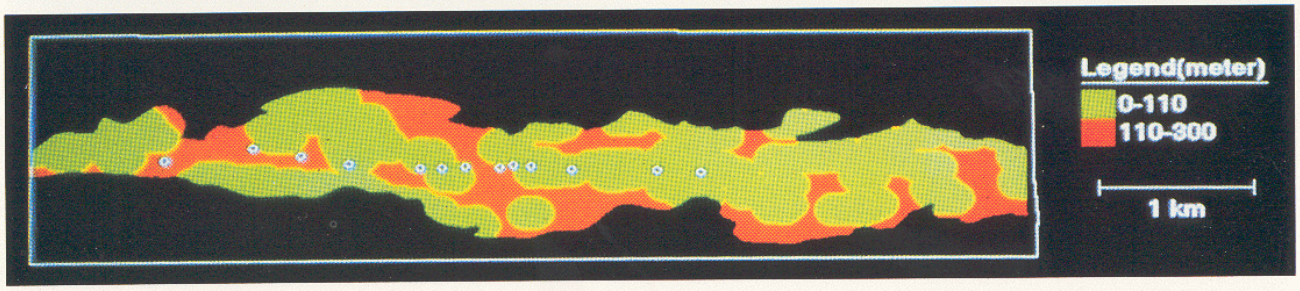
2c


Legend
Pillow flows
Sheet flows
1 km

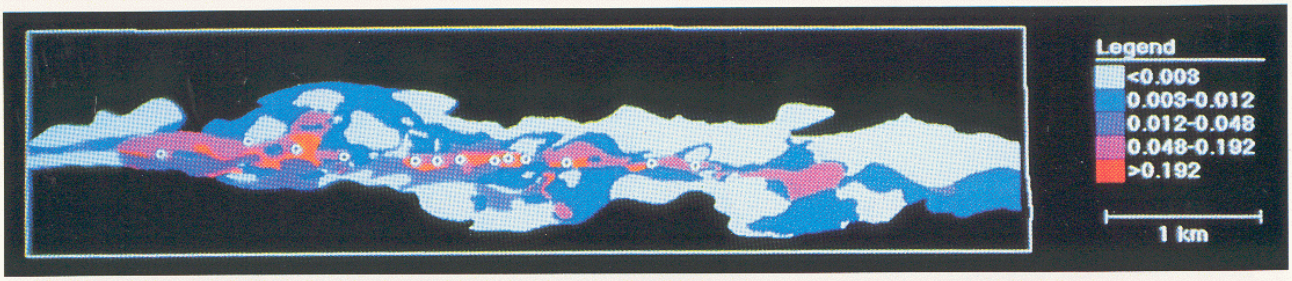
2d


Legend(meter)
<20
>20
1 km

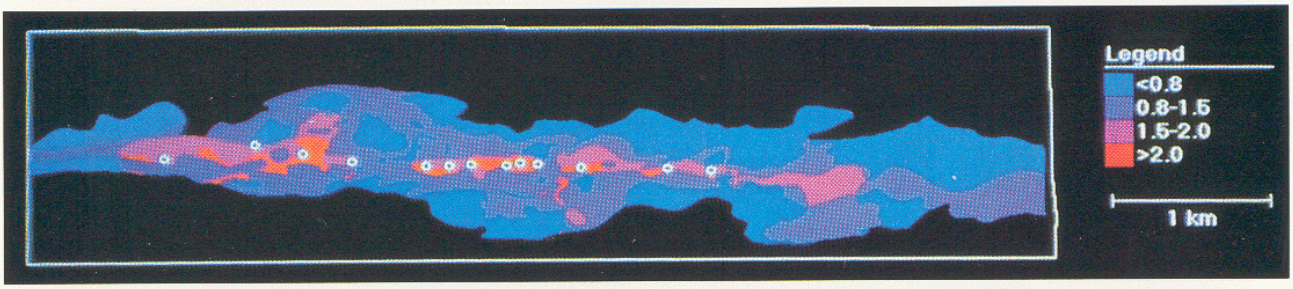
2e


Legend(meter)
0-110
110-300
1 km

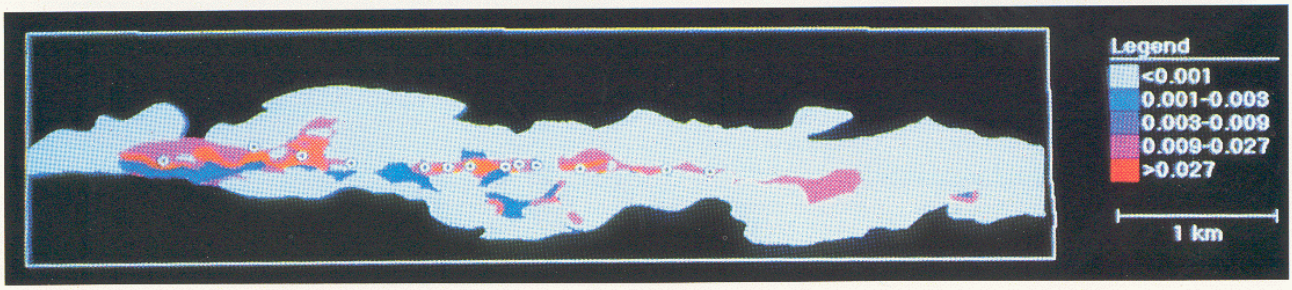
3a

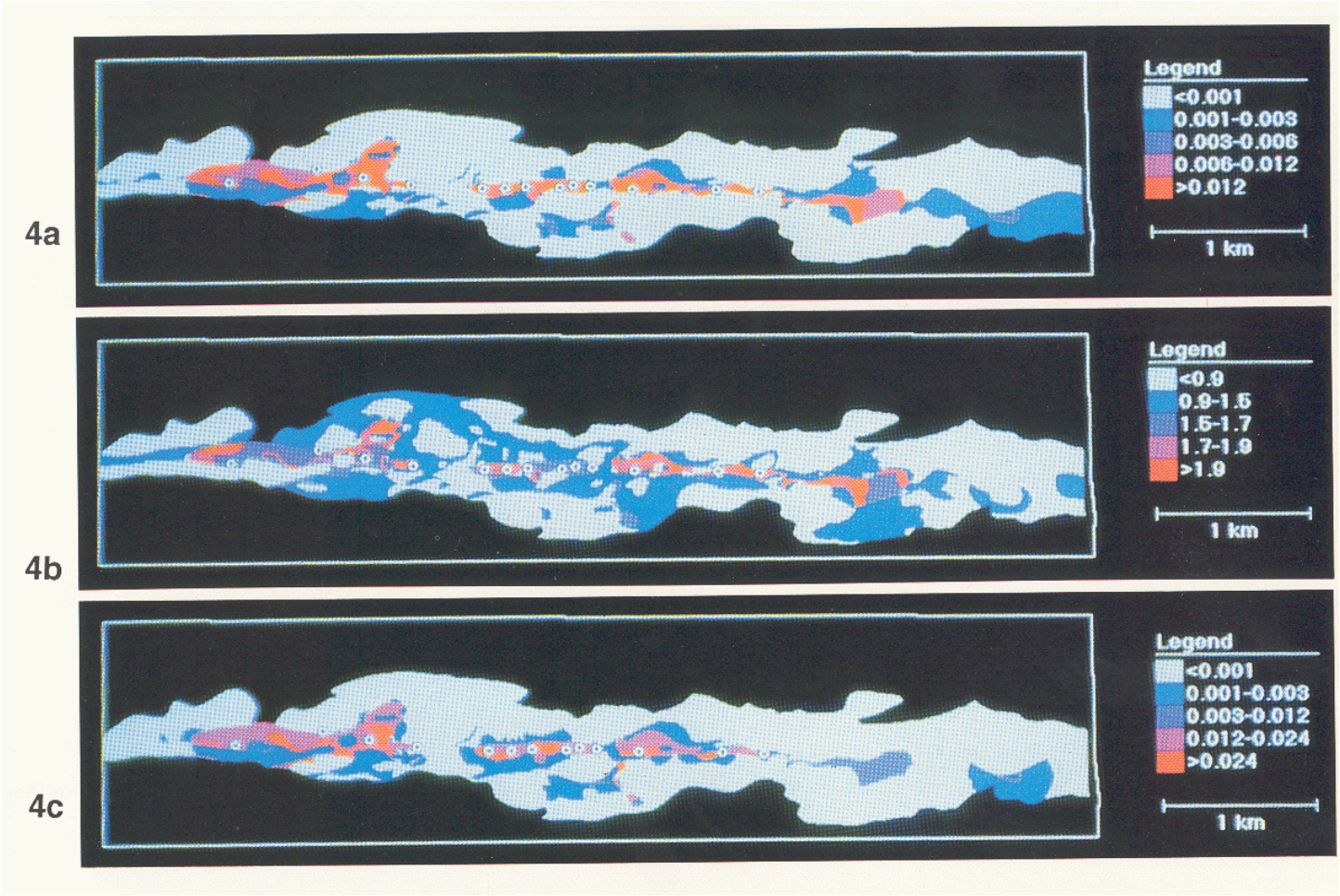

Legend
<0.003
0.003-0.012
0.012-0.048
0.048-0.192
>0.192
1 km

3b

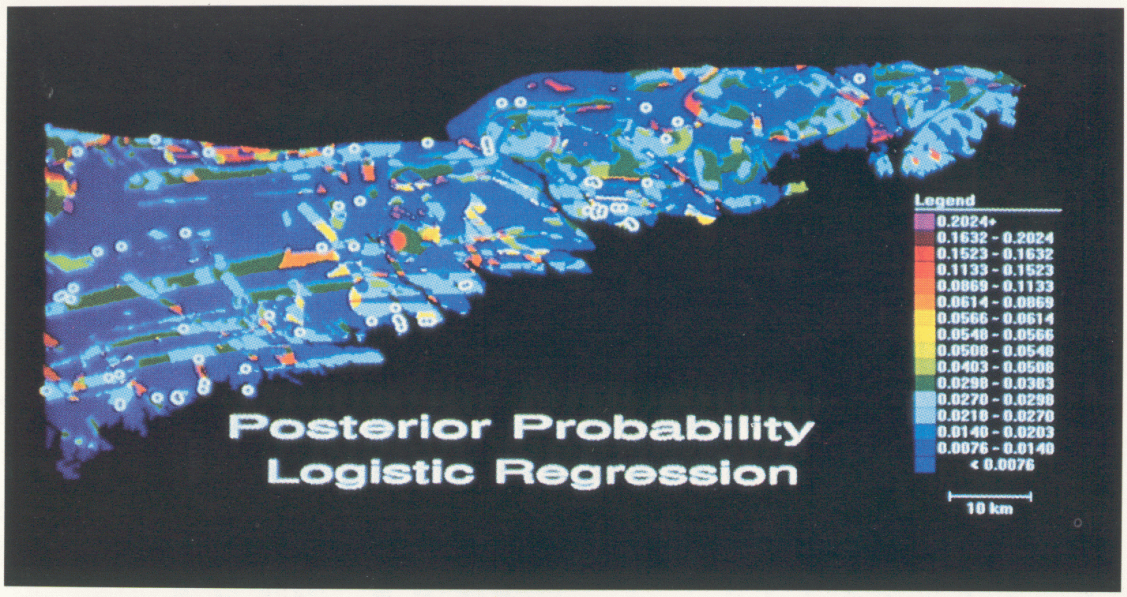

Legend
<0.8
0.8-1.5
1.5-2.0
>2.0
1 km

3c


Legend
<0.001
0.001-0.003
0.003-0.009
0.009-0.027
>0.027
1 km


4a
Legend
<0.001
0.001-0.003
0.003-0.006
0.006-0.012
>0.012
1 km
4b
Legend
<0.9
0.9-1.5
1.5-1.7
1.7-1.9
>1.9
1 km
4c
Legend
<0.001
0.001-0.003
0.003-0.012
0.012-0.024
>0.024
1 km

5a


Posterior Probability
Logistic Regression
Legend
0.2024+
0.1632 - 0.2024
0.1523 - 0.1632
0.1133 - 0.1523
0.0869 - 0.1133
0.0614 - 0.0869
0.0566 - 0.0614
0.0548 - 0.0566
0.0508 - 0.0548
0.0403 - 0.0508
0.0298 - 0.0303
0.0270 - 0.0298
0.0210 - 0.0270
0.0140 - 0.0203
0.0076 - 0.0140
< 0.0076
10 km

5b

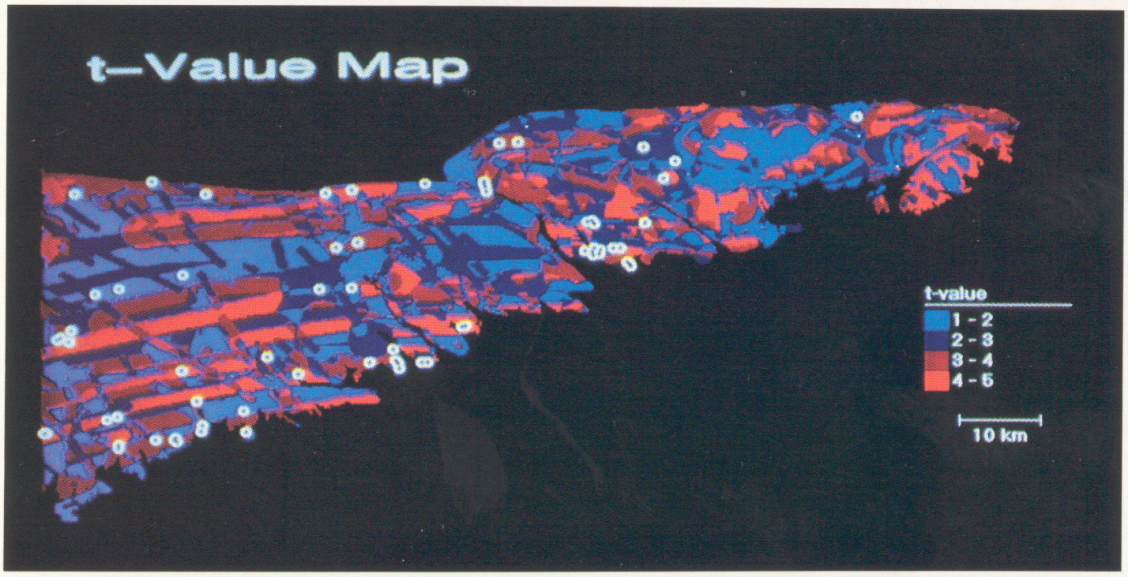

t–Value Map
t-value
1 - 2
2 - 3
3 - 4
4 - 5
10 km

Table 1: Positive Weights (W^+), Contrasts (C), and Standard Deviations (s) for Corridors around Contact between Youngest Volcanics
(Area measured in km2. Last column shows *t*-value of C[*t* = C/s(C)]. The 13 corridors are displayed in Plate 1c.)

No.	Width	Area	Vents	W^+	s(W^+)	C	s(C)	C/s(C)
1	20m	0.197	6	2.259	0.416	2.829	0.561	5.041
2	40m	0.338	7	1.863	0.382	2.549	0.559	4.557
3	60m	0.532	8	1.538	0.356	2.352	0.572	4.112
4	80m	0.721	8	1.232	0.356	1.989	0.572	3.481
5	100m	0.931	8	0.973	0.355	1.664	0.571	2.912
6	120m	1.097	8	0.808	0.355	1.443	0.571	2.526
7	140m	1.281	9	0.771	0.335	1.564	0.602	2.598
8	160m	1.480	11	0.827	0.303	2.237	0.769	2.908
9	180m	1.660	11	0.712	0.303	2.047	0.769	2.661
10	200m	1.808	11	0.626	0.302	1.896	0.769	2.464
11	220m	2.000	11	0.525	0.302	1.702	0.769	2.212
12	240m	2.164	12	0.533	0.290	2.317	1.041	2.225
13	> 240m	3.984	13					

for distance from a linear feature. For example, in Table 1 the contact between youngest pillows and sheets (Pl. le) has the largest contrast for corridor no. 1 which was selected for the binary pattern of Plate 2d. The fissure binary pattern (Pl. 2c) is also for the corridor with the largest contrast. Weights and contrasts for the binary pat- terns of Plate 2 are summarized in Table 2. From the standard deviations it can be seen that the correlation between fissures and vents is probably not significant. The binary pattern of Plate 2e is only weakly correlated with the vents. In weights of evidence modeling, the binary patterns

Table 2: Weights and Contrast (with Standard Deviations) for Five Binary Patterns of Plate 2

Pattern	W^+	s(W^+)	W^-	s(W^-)	C	s(C)	C/s(C)
Age	2.251	1.000	-1.231	0.290	3.481	1.041	3.343
Topography	2.037	1.000	-0.811	0.290	2.848	1.041	2.735
Contact	2.259	0.415	-0.570	0.378	2.829	0.561	5.041
Rock type	0.632	0.410	-0.338	0.378	0.970	0.558	1.740
Fissures	0.178	0.448	-0.097	0.354	0.275	0.571	0.481

are combined with one another by the addition of weights for very small unit cells where the features are either present or absent. From a statistical point of view, this addition is only permitted if the binary patterns are conditionally independent of the vent pattern. Chi-squared statistics for conditional independence testing (Agterberg, 1992) cannot be used here, because the required frequencies of points are too small. It is likely that the binary patterns are not conditionally independent. For example, the age (Pl. 2a), rock type (Pl. 2c) and contact corridor (Pl. 2d) patterns were constructed from the litho-age units of Plate la. This assumption is corroborated by performing the following pattern correlation analysis. Yule's measure of association Q for binary variables resembles the product-moment correlation coefficient

in that it is equal to zero if there is no correlation and cannot exceed one (for exact linear relationship) in absolute value. It is close to zero for the three pairs: age-topography (0.19), age-rock type (0.18), and topography-rock type (-0.02). In absolute value it is relatively large for con- tact corridor correlated with age (0.95), topography (0.53), and rock type (0.53), respectively. These results suggest that the contact corridor pattern may be redundant, as will be demonstrated later by statistical tests.

Posterior Probability Maps

Plate 3a is a posterior probability map for a 0.01-km2 unit cell based on only three binary patterns (age, topography, and rock type). The prior probability in this application was set equal to 0.033 for number of vents (= 13) divided by total area (= 398.4 unit cells). The binary patterns of Plate 2 also can be regarded as posterior probability maps. Presence of single features gives posterior probabilities of 0.111 (age), 0.073 (topography), and 0.061 (rock type), respectively. These probabilities are for presence of a vent within a 0.01-km^2 subarea at a particular place where an indicator pattern is present. In general, combining p binary patterns gives 2^P possible combinations for the unique conditions. Plate 3a is based on eight unique conditions with probabilities equal to 0.000, 0.001, 0.006, 0.011, 0.015, 0.030, 0.171, and 0.360. The uncertainties of these probabilities are relatively large, as shown in the corresponding t-value map of Plate 3b where every posterior probability was divided by its standard deviation.

Plate 3e is the posterior probability map for a 0.001-km^2 unit cell using all five binary patterns of Plate 2. Although the patterns of Plates 3a and 3e are similar, a more detailed analysis shows that the results of these two applications of the weights of evidence method are different. Plate 3e is based on 31 unique conditions (one of the possible 32 combinations of five features is not represented), with probabilities ranging from 0.000 to 0.352. The unit cell for Plate 3e is ten times as small as the one used for Plate 3a. Because the posterior probabilities cover approximately the same range of values, this means that the probability of finding a vent per 0.01-km^2 unit cell in the unique conditions with the largest posterior

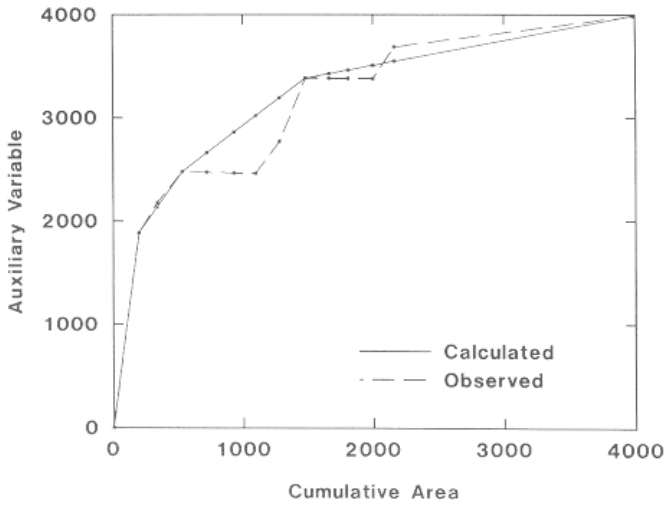


Figure 1: Seafloor example, analysis of relationship between hydrothermal vents and contact between youngest volcanics (cf. Table 1). Auxiliary variable $y = A.\exp(W+)$ is plotted against cumulative area A measured in units of 0.001 km 2. The first derivative dy_c/DA of fitted curve y_c provides estimates of four values of variable weight $W^+(A)$ that depends on distance from the contact. See text for further explanation.

probabilities is about ten times greater in the situation of Plate 3c. We will show later (see sections on weighted logistic regression and goodness-of-fit test) that the model of Plate 3a provides a good fit, whereas the model underlying Plate 3e overestimates the posterior probabilities in the most favorable areas because of lack of conditional independence of the contact corridor binary pattern.

Analysis of Contact Corridor Pattern

The contrast in Table 1 has secondary maxima for corridors 8 and 12. Although the positive weights for these other corridors are less than that for the first corridor used for Plate 2d, their areas are larger. Expected number of vents within a corridor is equal to the product of corridor area and posterior probability. For this reason, a wider corridor (e.g., no. 8) can also be selected as a binary pattern. Another method of modeling the relationship between vents and contact is to estimate weights for the intersections of successive corridors ("classes") shown in Plate lc.

Figure 1 was derived from the data of Table 1 for classes of contact corridors as follows. An auxiliary variable $y = A \exp(W^+)$ is plotted against cumulative area A. Agterberg and Bonham-Carter (1990) have shown that the natural logarithm of the first derivative dy_c/dA of a curve y, fitted to y may provide a good estimate of $W^+(A)$ representing a variable weight that depends on distance from the contact. Suppose m distinct weights are calculated for m classes of distance instead of the two weights corresponding to the two classes of a binary pattern. The observed values of Table 1 (and Fig. 1) are for increasingly wide corridors. Adjoining classes with the smallest difference in y can be combined repeatedly until only m new classes are retained. The result of this iterative process for m=4 is shown in Figure 1 as four straight-line segments approximating y_c. The slopes of the four straight lines can be used to estimate the following four weights: 2.259 (for class 1, as before), 0.566 (for classes 2 and 3), -0.043 (for classes 4 to 8), and -1.431 (for remainder of study area). This pattern suggests an approximately linear decrease in weight with distance from the contact. This, in turn, implies that the probability of finding a vent within a small cell would decrease exponentially with distance. It will be shown next how these results can be incorporated in the modeling.

Weighted Logistic Regression

Weights of evidence modeling and logistic regression with the observations weighted according to their areas of the corresponding unique conditions are different types of application of the loglinear model (cf. Andersen, 1990). In weighted logistic regression, the patterns are not necessarily conditionally independent as in weights of evidence modeling. Plate 4a shows posterior probabilities for a 0.001 km^2 unit cell using the same five binary patterns of Plate 3c. The probabilities of Plate 4a range from 0.000 to 0.054. For the most favorable unique conditions, they are nearly ten times as small as t corresponding values that resulted from applying the weights of evidence method to the five binary patterns. In this respect, the posterior probabilities resulting from weighted logistic regression are close to those obtained by applying the weights of evidence method to three binary patterns only (cf. Pl.. 3a). These results indicate that the large probabilities that arose when the weights of evidence method was used with the five binary variables are, indeed, too large because of lack of conditional independence. The logistic regression coefficients and their standard deviations are shown in Table 3. The t-value map for Plate 4a is shown in Plate 4b.

Weighted logistic regression can also be used in situations where the explanatory variables have many classes or are continuous. In the discussion of Figure 1, it was suggested that probability of occurrence of vents decreases exponentially with distance from contact. In order to incorporate this exponential decrease in the logistic model, a new explanatory variable was created by assigning values decreasing from 13 to 1 to the 13 classes used for Figure 1 *(cf. Pi.* le). Combining this new ordinal variable with the previous four binary variables resulted in an increase in the number of unique conditions (from 31 to 196). Plate 4e shows the posterior probability map for this new model. In general, the pattern of Plate 4e is close to the one of Plate 4a. Although the relationship between vents and contact was modeled in more detail, the overall effect of this refinement becomes small when it is combined with the relationships of the vents with age, elevation, and rock type (cf. Table 3).

Goodness-of-Fit Test

The degree of fit of several models is evaluated in Figure 2. The posterior probability is plotted in the horizontal direction. The product of posterior probability and area per unique condition provides theoretical values for frequency of vents. Corresponding observed frequencies can be obtained by

Table 3: Regression Coefficients for Logistic Model (B) and Modified Logistic Model (B) with Standard Deviations
[The value of x in B contact between youngest volcanics ranges from 13 (corridor no. 1) to 1 (corridor no. 13).]

Pattern	B	s(B)	B	s(B)
Age	2.862	1.076	2.979	1.086
Topography	2.388	1.050	2.458	1.051
Contact	1.114	0.604	0.145	0.579
Rock type	0.2580	0.584	0.420	0.591
Fissures	0.139	0.579	0.062	0.076

counting the number of vents per unique condition. Theoretical and observed frequencies were converted to relative frequencies by dividing by total number of vents (= 13). If a model is good, the predicted total number of vents should be close to 13. This condition is nearly satisfied in Figures 2a (weights of evidence modeling using three binary patterns) and 2b (weighted logistic regression using five binary patterns). In the situation of Figure 2a, the model predicts 14.0 vents which is one too many; the model of Figure 2b predicts 12.6 vents-slightly less than 13. The Kolmogorov-Smirnov (K-S) test can be used to evaluate the largest difference between observed and expected relative frequencies. In Figure 2a, the absolute value of the largest difference is 0.081. In a two-tailed test with eight observations, this value should not exceed 0.454 with a probability of 95%. The corresponding 95% confidence level for Figure 2b with 31 observations is 0.238 which also is greater than the observed value of 0.099 in this diagram. It may be concluded that the models tested in Figures 2a and 2b provide a good fit.

On the other hand, the degree of fit of the models underlying Figures 2e and 2d is poor. Figure 2e corresponds to Plate 3e for which it was already shown that the five binary patterns are not conditionally independent. The predicted total number of vents is 37.6, which is nearly three times too large. Moreover, the absolute value of the largest difference (= 1.892) in Figure 2e exceeds the 95% confidence level (= 0.238) in the K-S test. Figure 2d is for a probability map (not shown) derived from five binary patterns in which the contact pattern was for the wider corridor comprising classes 1 through 8 in Plate le.

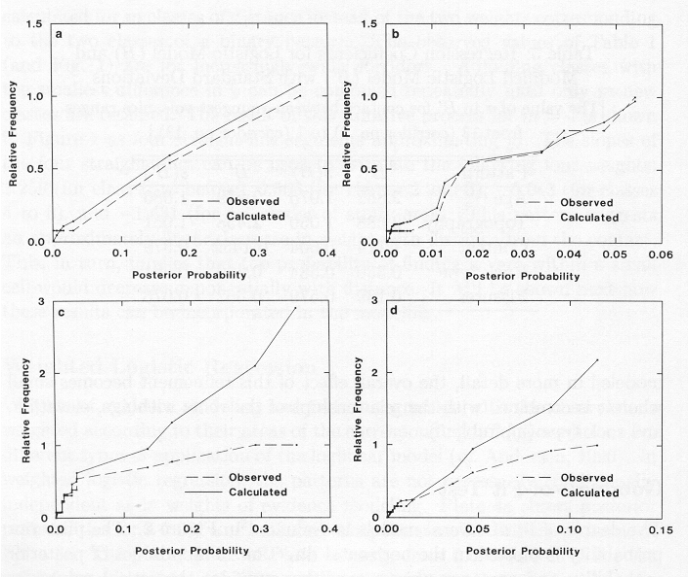


Figure 2: Seafloor example, goodness-of-fit tests. Observed and estimated relative frequencies versus posterior probabilities from (a) Plate 3a, (b) Plate 4a, (c) -Plate 3c, and (d) pattern similar to Plate 3e obtained after using contact corridor no. 8 instead of no. 1 for the contact binary pattern.

The expected total number of vents then is 28.4, which is more than twice the observed total (= 13). The absolute value of the largest difference (= 1.184) in Figure 2d exceeds the 95% confidence level (= 0.238) for a good fit.

The largest posterior probabilities in Figures 2e and 2d are 0.352 and 0.115, respectively. Differences between observed and calculated frequencies do not exceed the 95% confidence level of the K-S test except for the three or four largest posterior probabilities. The models underlying Figures 2e and 2d provide a good fit except in the most favorable unique conditions where the frequencies of vents are overestimated by a wide margin.

The preceding application of the K-S test differs from other applications of this test because in our application the model also predicts total number of discrete events. Normally a non-zero difference between observed and expected frequencies at the largest value does not arise because the observations originate from an infinitely large population. In a strict sense, the Kolmogorov-Smirnov test statistics may only be used when the total number of discrete events is correctly predicted. The approximate K-S test used in this paper loses its validity when the expected relative frequency is not approximately equal to 1.0 at the largest value. Note that in Bonham-Carter et al. (1990) the K-S test was applied, but the theoretical as well as the observed cumulative frequencies were constrained to reach a maximum of 1.0. This had the advantage of satisfying the assumptions for the K-S test, but the disadvantage of failing to recognize theoretical frequencies that are too large.

Also note that possible undiscovered deposits are not considered in the goodness-of-fit test. The reason that results of weights of evidence modeling and logistic regression are useful for mineral potential mapping is that the estimated weights are approximately independent of undiscovered deposits in a study region provided that the known deposits can he regarded as a random subset of all (known + unknown) deposits in the region. Only the prior probability in weights of evidence modeling and the constant term in logistic regression depend strongly on undiscovered deposits (cf. Agterberg, 1992).

Gold Deposits in Central Nova Scotia

In the weighted logistic regression, 68 gold deposits were related to the fol- lowing seven binary patterns (cf. Bonham-Carter et al., 1990). (1) proximity to anticlinal axes, (2) Au in balsam fir, (3) contact between Goldenville and Halifax Formations, (4) Goldenville Formation, (5) Devonian granite contact zone, (6) lake sediment signature, and (7) NW lineaments. The assumption of conditional independence is slightly violated in this application. For example, weights of evidence modeling for a 1-km unit cell on these seven

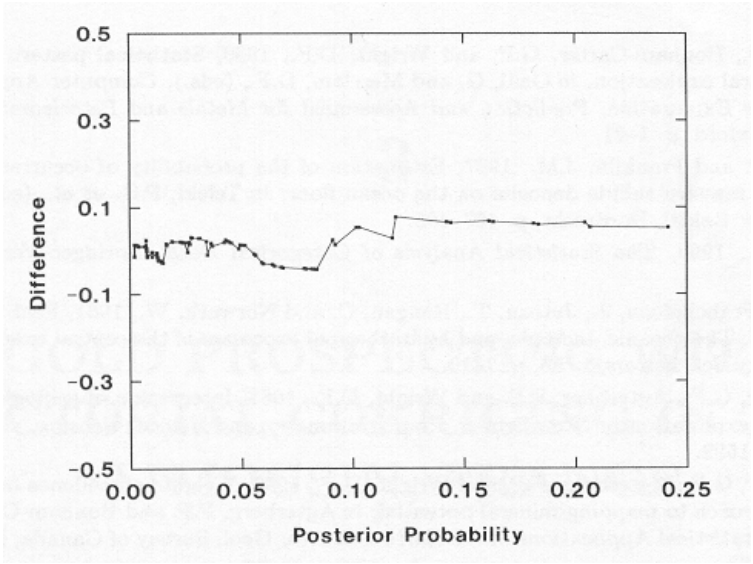


Figure 3: Goodness-of-fit test applied to logistic model for gold deposits, Meguma Terrane (Pl. 5a). The difference between observed and theoretical relative frequencies is plotted against posterior probability. See text for further explanation.

binary patterns results in a predicted total number of deposits equal to 75.2, which exceeds the observed total by nearly 10%. It is noted that patterns (2) and (6) are missing in parts of the area. In weights of evidence modeling, the weights can be estimated for patterns with missing data by omitting areas that are unknown from the weight calculations. In logistic regression, this procedure would result in significant loss of information because coefficients for all patterns are estimated simultaneously; thus, omitting areas with missing data would eliminate these regions from estimation entirely. For this reason patterns (2) and (6) were modified so that, in regions where the patterns are missing, they were treated as being "not present." Logistic regression on the resulting revised data set predicts 64.3 gold deposits-slightly less than 68.

The weights, their standard deviations, and contrasts of the weights of evidence modeling are compared to the estimated logistic regression coefficients in Table 4. Plate 5a shows the logistic posterior probability map which is similar to weights of evidence modeling results previously shown in Bonham-Carter et al. (1990). Plate 5b shows the posterior probabilities divided by their standard deviations (t-value map). A significant difference between Plate 5b and the t-value maps for the seafloor example (Pls. 3b and 4b) is that the values in Plate 5b are relatively large. In an approximate significance test based on the normal distribution in standard form, a t-value greater than 1.645 indicates that the corresponding posterior

Table 4: Weights and Contrasts (with Standard Deviations) for Seven Binary Patterns Related to Gold Deposits in Meguma Terrane, Nova Scotia

[Regression coefficients for logistic model (B) and their standard deviations, are shown in last two columns. First row (pattern no. 0) is for constant term in weighted logistic regression.]

Pattern No.	W+	s(W+)	W-	s(W-)	C	s(C)	B	s(B)
0							-6.172	0.501
1	0.563	0.143	-0.829	0.244	1.392	0.283	1.260	0.301
2	0.836	0.210	-0.293	0.160	1.129	0.264	1.322	0.267
3	0.367	0.174	-0.268	0.173	0.635	0.246	0.288	0.266
4	0.311	0.128	-1.474	0.448	1.784	0.466	1.290	0.505
5	0.223	0.306	-0.038	0.134	0.261	0.334	0.505	0.343
6	1.423	0.343	-0.375	0.259	1.798	0.430	0.652	0.383
7	0.041	0.271	-0.010	0.138	0.051	0.304	0.015	0.309

probability is greater than 0 with a probability of 95%. This greater degree of precision is due to the larger number of occurrences for the Nova Scotia example.

Finally, Figure 3 is for evaluation of the goodness of fit of the logistic model of Plate 5. The absolute value of the largest difference between expected and observed relative frequencies is 0.0775. This is less than the Kolmogorov-Smirnov statistic (= 0.1426; 95% two-tailed test) and it may be concluded that the fit of the logistic model is good.

Concluding Remarks

Care should be taken in weights of evidence modeling to avoid bias caused by predictive patterns that are mutually interrelated, because violations of the conditional independence assumption usually lead to overestimation of the largest posterior probabilities. The problem of bias is avoided when weighted logistic regression is used. In general, the drawbacks of regression are that it cannot be applied without making assumptions about missing values unless all explanatory patterns are fully known for a study area. Moreover, the standard deviations of regression coefficients can be unreasonably large if there is multicollinearity. The latter problems are of minor significance in this paper where the logistic model

produced satisfactory results in all applications.

It is suggested in this paper that both weights of evidence and logistic regression solutions be routinely compared. The weights of evidence method yields readily interpreted positive and negative weights and is a straight- forward method for determining optimal cutoffs for the creation of binary patterns and for handling missing data. On the other hand, logistic regression provides a cheek on the effects of lack of conditional independence, in addition to the X 2- and K-S tests suggested for the weights of evidence method.

References

Agterberg, F.P., 1989, Computer programs for mineral exploration: Science, v. 245, p. 76-81. Agterberg, F.P., 1992, Combining indicator patterns in weights of evidence modeling for resource evaluation: *Nonrenewable Resources, v.* 1, no. 1, p. 35--50.

Agterberg, F.P. and Bonham-Carter, G.F., 1990, Deriving weights of evidence from geoscience contour maps for the prediction of discrete events, in TUB-Dokumentation Kongresse und Tagungen No. 51, *Proceedings 22nd APCOM Symposium,* Berlin, September 1990: Tech. Univ. Berlin, v. 2, p. 381-396.

Agterberg, F.P., Bonham-Carter, G.F. and Wright, D.F., 1990, Statistical pattern integration for mineral exploration, in Gaal, G. and Merriam, D.F., (eds.), *Computer Applications in Resource Exploration, Prediction and Assessment for Metals and Petroleum:* Pergamon Press, Oxford, p. 1-21.

Agterberg, F.P. and Franklin, J.M., 1987, Estimation of the probability of occurrence Of poly-metallic massive sulfide deposits on the ocean floor, in Teleki, P.G. et al., (eds.), *Marine Minerals*: Reidel, Dordrecht, p. 467-483.

Andersen, E.B., 1990, *The Statistical Analysis of Categorical Data*: Springer-Verlag, Berilli, 523 pp.

Ballard, R.D., Francheteau, J., Juteau, T., Rangan, C. and Norwark, W., 1981, East Pacific Rise at 21 N: The oceanic, tectonic, and hydrothermal processes of the central axis: *Earth and Planetary Sel. Letters, v.* 55, p. 1-10.

Bonham-Carter, G.F., Agterberg, F.P. and Wright, D.F., 1988, Integration of geological datasets for gold exploration in Nova Scotia: *Photogrammetry and Remote Sensing,* v. 54, no. 11, p. 1585-1592.

Bonham-Carter, G.F., Agterberg, F.P. and Wright, D.F., 1990, Weights of evidence modelling: A new approach to mapping mineral potential, in Agterberg, F.P. and Bonham-Carter, G.F., (eds.), *Statistical Applications in the Earth Sciences:* Geol. Survey of Canada, Paper 89-9, p. 171-183.

Reddy, R.K.T., Agterberg, F.P. and Bonham-Carter, G.F., 1991, *Application of GIS-based Logistic Models to Base-metal Potential Mapping In Snow Lake Area, Manitoba:* Proceedings, 3rd Canadian Conference on GIS, Ottawa, Canada, March 18-21, 1991, p. 607A19.

Intera-TYDAC, 1991, *SPANS Users Guide, Version 5.2:* Intera-TYDAC Technologies Inc., Ottawa, Canada, 4 volumes.

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 pp.