Appl. Math. and Comp. Sci., 1998, Vol.8, No.4, 841–864

LAKE BED CLASSIFICATION USING ACOUSTIC DATA[†]

K. KAREN YIN[*], XING LI[*], JOHN BONDE[**]
CARL RICHARDS[**], GARY CHOLWEK[***]

As part of our effort to identify the lake bed surficial substrates using remote sensing data, this work designs pattern classifiers by multivariate statistical methods. Probability distribution of the preprocessed acoustic signal is analyzed first. A confidence region approach is then adopted to improve the design of the existing classifier. A technique for further isolation is proposed which minimizes the expected loss from misclassification. The devices constructed are applicable for real-time lake bed categorization. A minimax approach is suggested to treat more general cases where the *a priori* probability distribution of the substrate types is unknown. Comparison of the suggested methods with the traditional likelihood ratio tests is discussed.

1. Introduction

In Lake Superior, lake trout was an important fishery sustaining a large annual commercial yield, which made it a valuable species. As in the other Great Lakes, the population of lake trout was severely reduced in the 1950's from sea lamprey predation and heavy fishing. The rehabilitation of the species did not begin until the control of sea lamprey was initiated and the closure of commercial lake trout fisheries became effective. With the continued enforcement of sea lamprey control and the limitation of commercial and sport fisheries, the restoration of lake trout appears hopeful. Since the restoration means the creation and conservation of favorable habitats for its subsistence and reproduction, the identification and quantification of its

[†] This work is the result of research funded by the Minnesota Department of Natural Resources through Contract No. 64 of Agreement No. M8110, and by the Minnesota Sea Grant College Program supported by the NOAA Office of Sea Grant, United States Department of Commerce, under grant No. NA46–RG0101. The U.S. Government is authorized to reproduce and distribute reprints for government purposes, notwithstanding any copyright notation that may appear hereon. This paper is Journal Reprint No. 426 of the Minnesota Sea Grant College Program, Publication No. 45 of the Natural Resources Geographic Information Systems Laboratory at the Natural Resources Research Institute, and Contribution No. 987 of the USGS Great Lakes Science Center.

[*] Department of Chemical Engineering, University of Minnesota, Duluth, MN 55812, SA, e-mail: kyin@d.umn.edu.

[**] Natural Resources Research Institute, University of Minnesota, Duluth, MN 55812, USA, e-mail: crichard@sage.nrri.umn.edu.

[***] U.S. Geological Survey, Great Lakes Science Center, Lake Superior Biological Station, Ashland, WI 54806, USA, e-mail: gary_cholwek@usgs.gov.

spawning habitat in the Great Lakes become a crucial component of any restoration program. Given the large area of the shoreline, a detailed survey such as done in conventional hydrographic surveying would be impractical due to its high cost. With recent advances in seafloor mapping as well as real-time data storage and processing technologies, it has become feasible to conduct a survey with relatively inexpensive remote sensing techniques. Because of the depth requirement in an underwater survey, and because acoustic waves can propagate far enough in the water to provide sufficient areal coverage in a timely fashion, an acoustic remote-sensing operation using sonar was adopted for this project.

Bathmetry and seafloor acoustic remote sensing have attracted much attention in recent years. Continuing efforts (Alexandrou *et al.*, 1992; Cervenka and Moustier, 1993; Collins and Gregory, 1996; Curran, 1995; Malinverno *et al.*, 1990; Moustier, 1989; and the references therein) have not only designed novel measuring instruments and improved existing ones, but also led to the development of methodologies for terrain characterization at different scales. Lake floor classification requires the characterization of surficial substrate type. Although lakes are typically shallower than seas, the sensing technologies used are similar.

Using underwater sonar method, the time required for signal transmission is determined by the depth of the water; while the returning energy, or the backscatter, is affected by the characteristics of the lake floor and its immediate subsurface. This work employs a single-beam echosounder to provide both depth data and the signal input into a bottom classification sensor. RoxAnn$_{TM}$, an instrument recently developed by Marine Microsystems limited (Scotland), is used for classification. It processes the output signal from the single-beam echosounder to determine the bottom's roughness and hardness, which are used to infer the lake floor type. In general, the classification requires several steps including data acquisition and pre-processing, geometric registration and rectification, pattern recognition, and high-level geographic information extraction. This paper is concerned with the pattern recognition step. More specifically, it aims to improve the design of the surficial substrate type classifier of RoxAnn$_{TM}$ data.

The design of classifiers amounts to the specification of a set of discriminate functions or tools which are used to divide the measurement space. RoxAnn$_{TM}$ supplies three measurements, e_1, e_2 and depth data, to the data acquisition system. The first two are employed for substrate type classification. Independent DC voltage readings from 0 volts to a maximum of 4.09 volts, e_1 indicates the topographical roughness or the 'texture' of the material on the surface of the lake bed while e_2 provides information on the relative hardness or acoustic impedance of the lake bed. It has been shown both theoretically and experimentally that the e_1 and e_2 data contain information on the substrate particle size and type. Ideally, the same floor type should generate the same e_1 and e_2 pair or a unique vector, which appears as a single point on the two-dimensional e_1-e_2 space. Because of the inherent randomness of variation in particle orientation, underwater condition, and noise in the remote sensing equipment, the signal on the e_1-e_2 plane will appear as localized clusters or clouds of points instead of a single point. Nevertheless, since the regions of the clusters corresponding to specific bottom surficial substrates are more or less distinct,

they can be associated to surficial substrate types and therefore serve as a classifier.

The classifiers that have been employed are the so-called RoxAnn squares in the Hypack$_{TM}$ software. They are constructed by a *supervised* classification method which uses training samples to learn the correspondence between e_1-e_2 values and the substrate types. The normal procedure is: First, choose test beds with known floor features for data acquisition. To gather the ground truth information for quantifying the feature of the lake floor, various methods such as sonar grab samples, visual inspection by divers, lake bed photographs, and underwater video recording can be used. Secondly, collect remote sensing data from the test beds chosen. Based on the correspondence between the floor types and the signals obtained, regions containing distinct categories are circumscribed on the e_1-e_2 plane. This procedure will result in a number of rectangles having different shapes and sizes, separated or more or less overlapped, each representing one type of surficial substrate: fine sand and silt, mixed cobble, gravel, etc. The delineated plane so obtained can then be used as classifier for identification of the coverages of other lake floor areas. Figure 1 shows a typical RoxAnn-square chart in Hypack$_{TM}$, which covers twelve different surficial substrate categories. The future data location on the chart determines the substrate type of the lake floor from which those data have been collected. It can be seen that the construction of classifiers such as RoxAnn squares is a crucial step in lake bed identification and quantification.

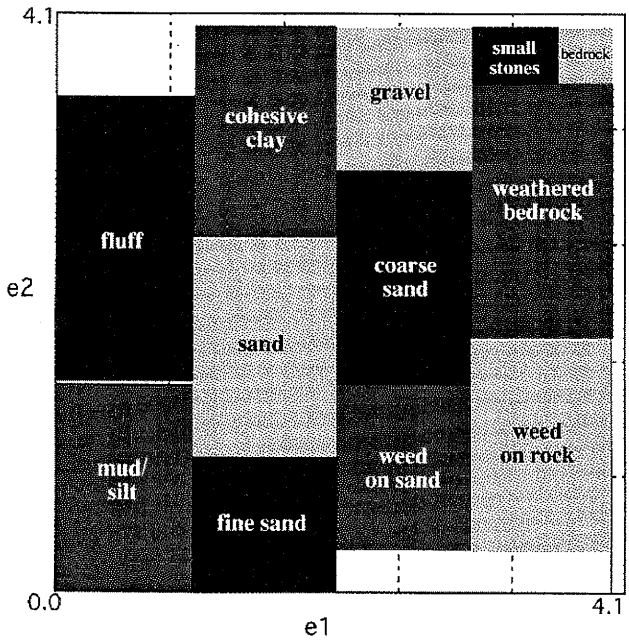


Fig. 1. Typical RoxAnn squares.

Similar to other texture quantification process, the classification of the lake bed is a pattern recognition procedure. Statistical methods are appropriate for pattern recognition for several reasons. They are suitable for dealing with random variations in the remote sensing data. They can be used to construct the confidence regions for classification. And they also allow us to specify overlapped regions and to distinguish between them with certain probabilities. Our work incorporated statistical methods into the existing RoxAnn$_{TM}$ technique. Based on a multivariate statistical approach, confidence RoxAnn ellipses were constructed and were shown to be better classifiers than the squared ones. For further classification, an approach which minimizes the expected loss from misclassification is proposed and used to map several types of lake bed coverage onto the e_1-e_2 plane. To handle more general cases, a minimax procedure is suggested which minimizes the maximum expected loss from misclassification. The relationship between these approaches and the traditional likelihood ratio tests is discussed.

This paper is organized as follows: A brief description of the sonar system and the data acquisition process is given in the next section. The confidence region approach is employed to design the pattern classifiers in Section 3. Section 4 studies several isolation methods and provides a figure to facilitate direct lake bed categorization. Summary and further discussion are provided in Section 5.

2. System Description, Data Acquisition and Preprocessing

Our ultimate goal is to characterize the lake floor according to various surficial substrate types along the entire 350 km Minnesota shoreline of Lake Superior thereby identifying those regions favorable for lake trout spawning. This requires a survey of an area of 120 square kilometers for which there is no data available on bottom-type characteristic for any sizable part. At the initial phase of the project, a number of regions having a total of 8 square kilometers were chosen, which were believed to be areas of interest from the historical information on lake trout habitat. These regions were then prioritized according to their accessibility. In six cruises during the past two summers, eight regions were surveyed to collect data.

The geometric locations of acquired data were registered by using real-time differential GPS, which also provided meter-level position accuracy for the survey vessel in real time, a very important piece of information for navigating transects.

The two types information needed are depth and bottom category. Given the large area to be covered, the sensors have to be relatively efficient. A single-beam echosounder, the Innerspace 448$_{TM}$, provides both depth data and the signal input into the RoxAnn$_{TM}$ bottom classifier. RoxAnn$_{TM}$ processes the output signal from the Innerspace 448$_{TM}$ to determine the bottom's roughness and hardness, which, in turn, are used to infer the lake floor type.

Data integration, reconciliation and initial processing are accomplished with Costal Oceanographics Hypack$_{TM}$ software; later analysis was performed with Arc/Info (Environmental Systems Research Institute). Similar to other remote sensing technologies, the first step in data analysis is to assess data quality and to remove

the bad data. From our experience, bad data happen occasionally with all echo sounders. Therefore it is necessary to filter data in Hypack$_{TM}$ first before further analysis. Two types of reference data are chosen to support the analysis of RoxAnn data: video recording and grab sampling.

3. Confidence Region Approach

Due to the random effects from a number of sources, the remote sensing data acquired are random variables. Using statistical method requires the knowledge of their probability distribution, which is unknown and must be estimated from the data collected. In this section, the probability distribution of the remote sensing data is analyzed first, which is the basis of the next step, building the confidence RoxAnn squares with predefined significance levels. Such an approach implies the assumption of the independence of e_1-e_2 data, which is not always true. Therefore classifiers are designed for the more realistic conditions where e_1 and e_2 are correlated. Elliptical regions are constructed and compared with squared regions for classification.

To proceed, a few words about the notation are in order. In the sequel, z' will be used to denote the transpose of z; and $\hat{z}$ stands for the estimate of z.

3.1. The Probability Distribution of the Remote Sensing Signal

Supplied by RoxAnn$_{TM}$, each of the two parameters e_1 and e_2 consists of two components: deterministic and random. In this work we make the following assumptions.

(A1) The noise is additive. Let $e_i(x)$ be the signal taken from an arbitrary point of the lake floor x, where $x = (x_1, x_2)'$ is the geometric coordinate of the point being surveyed. Then

$$e_i(x) = e_{i,d}(x) + e_{i,r}(x), \quad i = 1, 2 \tag{1}$$

where $e_{i,d}(x)$ and $e_{i,r}(x)$ denote the deterministic and random components of $e_i(x)$. The former is the 'true' value while the latter reflects a combined random effects originated from various sources.

(A2) The noise is spatially stationary and follows a normal distribution with mean

$$E\{e_{i,r}(x)\} = 0$$

and covariance

$$E\Big\{e_{i,r}(x)e'_{i,r}(x + \Delta x)\Big\} = \begin{cases} 0 & \text{if} \quad \Delta x \neq 0 \\ \sigma_i^2 & \text{if} \quad \Delta x = 0 \end{cases} \tag{2}$$

where $i = 1, 2$; $\Delta x = (\Delta x_1, \Delta x_2)'$; and E is the expectation operator.

(A3) For the same type of ground coverage, $e_{i,d}(x)$ is a constant regardless of its geometric location, i.e.,

$$e_{i,d}(x) = \mu_i, \quad i = 1, 2$$

Based on assumptions (A1)–(A3), the $e_1(x)$ and $e_2(x)$ signals have normal distribution with means $\mu_1 = e_{1,d}(x)$, $\mu_2 = e_{2,d}(x)$ and variance σ_1^2 and σ_2^2, the same as the variances of the random components $e_{1,r}(x)$ and $e_{2,r}(x)$.

Remark. Recall that a spatially stationary process is one with constant mean and covariance. Assumption (A2) indicates that the process is uncorrelated in the sense of (2).

In view of the assumptions above, the process $e_i(x)$ is roughly 'independent' of the spatial variable x. In reality, $e_i(\cdot)$ clearly should be x-dependent. However, in our data acquisition process, we let the survey boat drift over different bottom types with the sensor collecting data while an underwater video camera recorded the bottom features directly below it. Data acquired from areas of different substrate types were stored in separate files. In other words, each area from which data were taken is essentially 'homogeneous.' Therefore the assumption of spatial stationality is valid.

Although (A1)–(A3) are imposed, the means μ_i and the variances σ_i^2 are unknown. In the subsequent development, we will first obtain their estimates and perform related data analysis, which are the necessary steps in the determination of the pattern classifiers. The analysis to follow belongs to the off-line category. The estimation requires both point estimation method and interval estimation procedure.

Because $e_i(x)$ are assumed to be spatially stationary, a natural approach is to use the sample mean $\bar{e}_i$ and the sample variance S_i^2 as estimates of their corresponding population parameters, μ_i and σ_i^2. For $i = 1, 2$, denote the sample collected by e_i^j $(j = 1, \ldots, n)$. Define

$$\bar{e}_i = \frac{1}{n} \sum_{j=1}^{n} e_i^j \tag{3}$$

$$S_i^2 = \frac{1}{n} \sum_{j=1}^{n} \left(e_i^j - \bar{e}_i\right)^2 \tag{4}$$

Since this is a large sample case, i.e., $n \gg 1$, $1/n$ can be used in the calculation of S_i^2 to replace $1/(n-1)$ without introducing bias. Different from the description in the preceding paragraphs where the signals are registered according to their spatial or geometrical location, e_i^j $(i = 1, 2;\ j = 1, \ldots, n)$ are arranged in a sequential order of its collection as presented in Fig. 2. We then compared these sample distributions with normal distributions having mean μ_i and variance σ_i^2 estimated from $\bar{e}_i$ and S_i^2. For each of the eight categories prescribed we displayed the empirical distribution of the collected data. A typical comparison is presented in Fig. 3, which suggests a good fit to normal distribution.

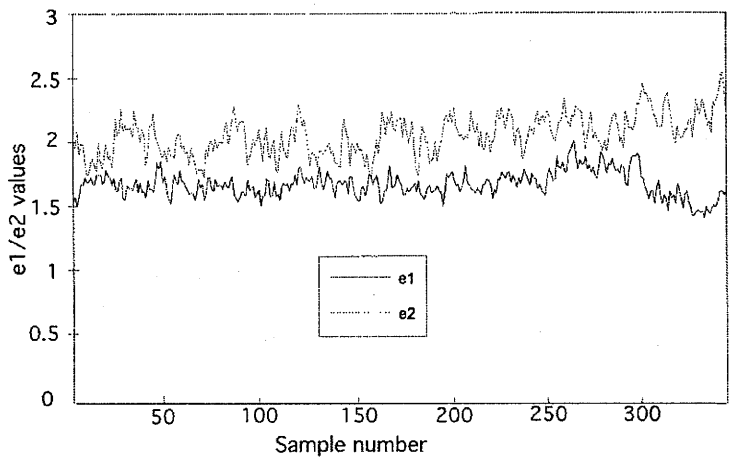


Fig. 2. An example of observed series of e_1 and e_2.

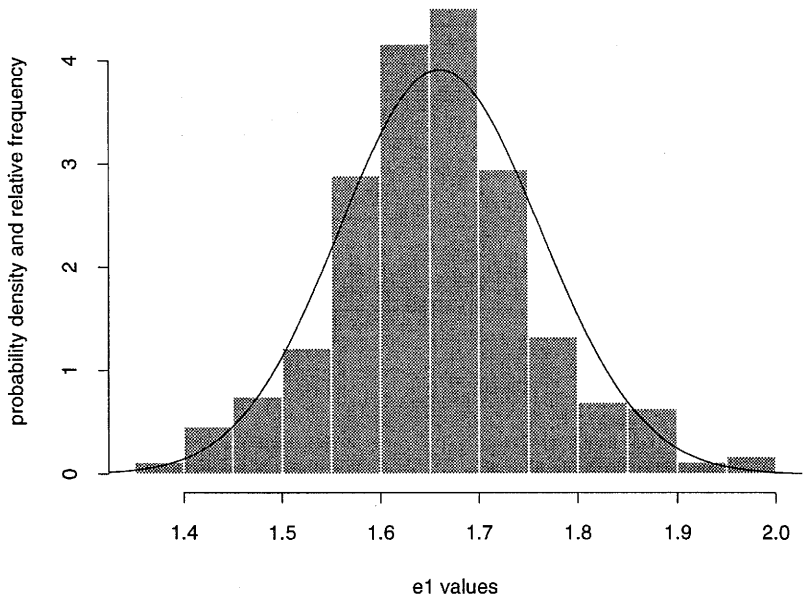


Fig. 3. A comparison of the empirical and its corresponding normal distributions.

To verify this quantitatively, a Kolmogorov-Smirnov goodness of fit test was conducted. It requires a hypothesis testing of

$$H_0 : f_i(\cdot) \sim \mathcal{N}(\mu_i, \sigma_i^2) \qquad \text{vs.} \qquad H_a : f_i(\cdot) \not\sim \mathcal{N}(\mu_i, \sigma_i^2) \tag{5}$$

where $f_i(\cdot)$ is the probability density function of e_i^j. Consider the sample cumulative distribution function T_{e_i} given by

$$T_{e_i}(u) = \frac{k}{n} \qquad \text{for} \qquad \tilde{e}_i^{(k)} \leq u < \tilde{e}_i^{(k+1)}$$

for $k = 1, \ldots, n$ where $\tilde{e}_i^{(1)} < \cdots < \tilde{e}_i^{(n)}$ are the order statistics of the sample e_i^j which take values in (0, 4.09). Choose the test statistics to be the least-upper-bound of the absolute difference between the empirical and hypothetical distribution functions and denote it as D_u, i.e.,

$$D_u = \sup_{0<u\leq 4.09} \left| T_{e_i}(u) - \Phi_i(u) \right| \tag{6}$$

where $\Phi(u)$ is the cumulative function of $\mathcal{N}(\mu_i, \sigma_i^2)$. For a designated confidence level $(1-\alpha) \times 100\%$, the D_u value is compared with the Kolmogorov-Smirnov acceptance limit $\Delta = KS_{\text{limit}}(1-\alpha)$. If $D_u \leq \Delta$ we accept the null hypothesis and conclude that the probability of e_i^j having normal distribution is $100(1-\alpha)\%$. Table 1 lists one of the typical test results. Using sample size of 345, it suggests, at a 95% confidence level, that both e_1^j and e_2^j have normal distribution.

Table 1. Kolmogorov-Smirnov test for fit to $\mathcal{N}(\mu_i, \sigma_i^2)$.

Du(e_1)	Du(e_2)	$\Delta = K\,S$ limit (95%)
0.0527	0.0381	0.0657

It should be pointed out that if the sample mean rather than the sample itself is used in the statistical method, according to the Central Limit Theorem, the sample mean follows a normal distribution $\mathcal{N}(\mu_i, \sigma_i^2/n)$ when the sample size n is large enough. Consequently, normal distribution can be adopted directly without further verification.

3.2. Confidence RoxAnn Squares

Training sample data were collected from forty-five areas during our three cruises. Each area was believed to have the same ground coverage. Those data were then used to train the pattern classifier—to circumscribe the ranges of e_1 and e_2 for each individual substrate type. The pattern classifiers are defined based upon the correspondence between the values of e_1-e_2 pair and the type of surficial substrates. In the statistical approach each pattern is characterized by its probability distribution. Since

it has been concluded that each class can be approximated by a normal distribution, to characterize a pattern only requires to estimate the parameters μ_i $(i = 1, 2)$ and to define their ranges at a certain confidence level from the training data collected.

On the bases of assumptions (A1)–(A3), for a specified positive number $0 \leq \alpha \leq 1$, $P(|e_i - \mu_i|/\sigma_i \leq z_{\alpha/2}) = 1-\alpha$. Equivalently, we have $100(1-\alpha)\%$ confidence that μ_i is within the region

$$l_i = (e_i - z_{\alpha/2}\,\sigma_i, e_i + z_{\alpha/2}\,\sigma_i) \quad \text{for} \quad i = 1, 2 \tag{7}$$

Equation (7) provides confidence intervals for the means μ_1 and μ_2. Using the data from a surveyed area with certain type of coverage, a confidence square centering at $(\bar{e}_1, \bar{e}_2)$ and having an area $l_1 \times l_2$ can be built on the two-dimensional e_1-e_2 plane. This square can serve as a classifier of the same type of lake floor for data taken from other areas without ground truth information. For instance, if the square was obtained from the lake floor covered with gravel under $\alpha = 0.05$, for any data acquired falling on this square, we conclude with 95% confidence that the ground coverage of the area from which the data was taken is gravel. Eight such squares were constructed under 90% confidence level, each corresponding to one type of floor type as shown in Fig. 4(a)–(b). The two-dimensional e_1-e_2 plane consisting of these confidence squares can then be used for future classification. Those areas without enough data support are labeled as 'unknown', or might be types of surficial substrates that do not occur in the areas sampled or are mixtures that require more data to sort them out from other categories.

3.3. Confidence Ellipse

The confidence RoxAnn squares were constructed based on individual confidence intervals for μ_1 and μ_2, which implies the independence between the e_1 and e_2 data. A typical set of e_1-e_2 data output from RoxAnn$_{\text{TM}}$ is displayed in Fig. 5. The pattern suggests that these two parameters are not independent. A positive correlation can be easily recognized. Data analysis showed that for different types of substrates the correlation coefficients between e_1 and e_2 range from 0 to 0.8. To illustrate this phenomenon properly, we replace assumptions (A1) and (A2) with (A4) and (A5):

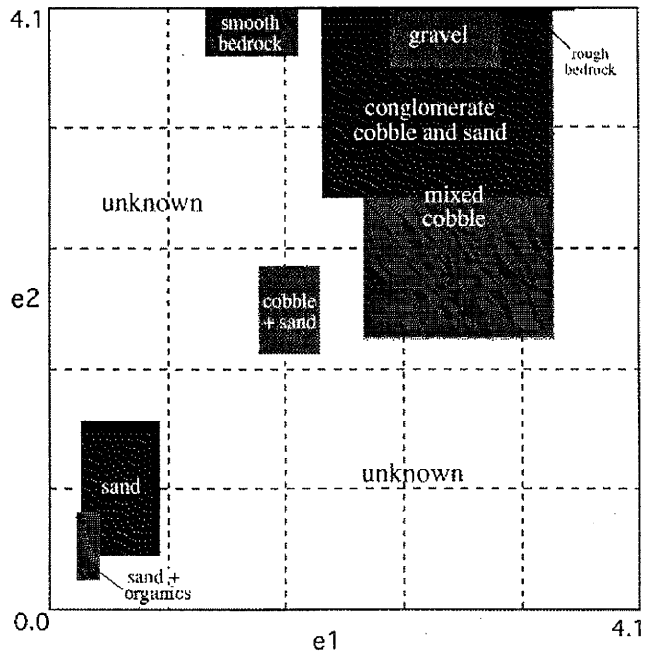
(A4) The noise is additive. Write $e(x) = (e_1(x)\ e_2(x))'$,

$$e(x) = e_d(x) + e_r(x) \tag{8}$$

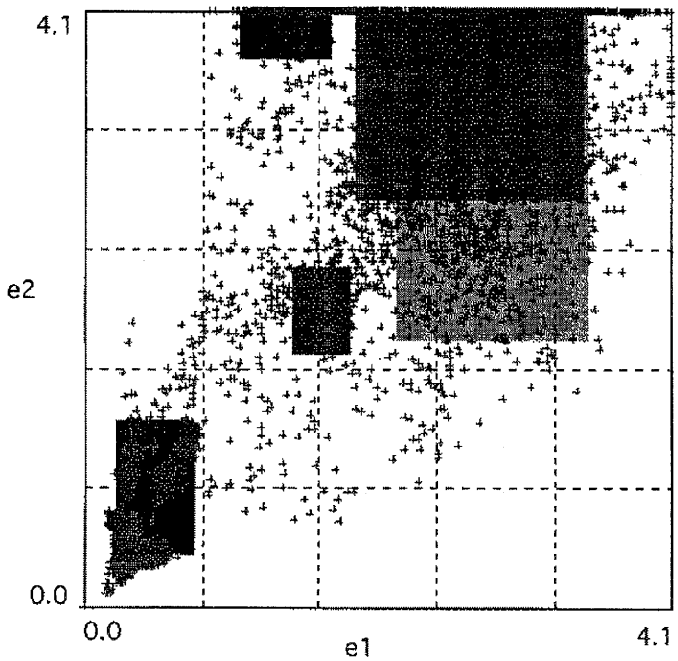
where $e_d(x)$ and $e_r(x)$ are its deterministic and random components, respectively.

(A5) The random component $e_r(x)$ is normally distributed and

$$E\{e_r(x)\} = 0, \quad E\{e_r(x)e_r'(x + \Delta x)\} = \begin{cases} 0, & \text{if} \quad \Delta x \neq 0 \\ \Sigma, & \text{if} \quad \Delta x = 0 \end{cases}$$



(a)



(b)

Fig. 4. (a) Confidence RoxAnn squares, (b) Collected e_1-e_2 data on confidence RoxAnn squares.

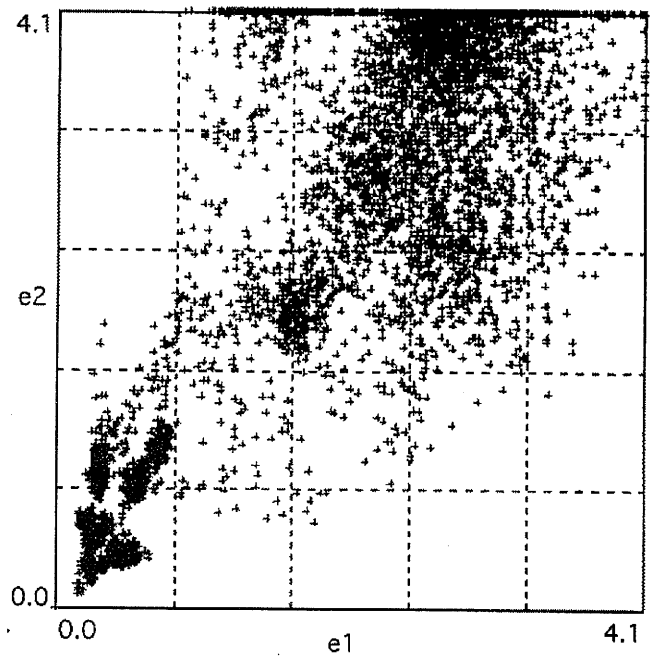


Fig. 5. A typical set of e_1-e_2 data.

From assumptions (A4) and (A5), the random vector $e(x) = (e_1(x)\ e_2(x))'$ follows a bivariate normal distribution with probability density function

$$f(e) = (2\pi)^{-1}|\Sigma|^{-1/2} \exp\left[-\frac{1}{2}(e-\mu)'\Sigma^{-1}(e-\mu)\right] \tag{9}$$

having mean

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} e_{1,d} \\ e_{2,d} \end{pmatrix}$$

and the covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 \end{pmatrix}$$

Being a measure of the association between e_1 and e_2, the correlation coefficient ρ_{12} is defined as

$$\rho_{12} = \frac{\sigma_{12}}{\sigma_1\sigma_2}$$

Consequently, eqns. (3) and (4) should be replaced by

$$\bar{e} = \frac{1}{n}\sum_{j=1}^{n} e^j \tag{10}$$

$$S^2 = \frac{1}{n}\sum_{j=1}^{n} (e^j - \bar{e})(e^j - \bar{e})' \tag{11}$$

where $\bar{e}$ is a vector having density $\mathcal{N}(\mu, \Sigma/n)$ and S^2 is a matrix. Since random vector e follows $\mathcal{N}(\mu, \Sigma)$, $e'\Sigma^{-1}e$ is distributed according to a non-central χ^2-distribution with two degrees of freedom and noncentrality parameter $\mu'\Sigma^{-1}\mu$. Equivalently, $n(\bar{e} - \mu)'\Sigma^{-1}(\bar{e} - \mu)$ has a central χ^2-distribution with two degrees of freedom. On the basis of this fundamental fact, we can establish confidence regions for μ. Let $0 < \alpha < 1$, e.g., $\alpha = 0.05$, and $\chi_2^2(\alpha)$ be the number such that $\mathrm{P}\{\chi_2^2 > \chi_2^2(\alpha)\} = \alpha$. Hence

$$\mathrm{P}\Big\{n(\bar{e} - \mu)'\Sigma^{-1}(\bar{e} - \mu) > \chi_2^2(\alpha)\Big\} = \alpha$$

To test the hypothesis of the mean being $\mu_0 = (\mu_{1,0}\ \mu_{2,0})'$, we may use the critical region defined by

$$n(\bar{e} - \mu_0)'\Sigma^{-1}(\bar{e} - \mu_0) > \chi_2^2(\alpha) \tag{12}$$

In the $\mathbb{R}^2$ space, inequality (12) defines the contour and the exterior of an ellipse centered at μ_0. The shape of the ellipse depends on Σ^{-1}. Its major axis makes an angle $\theta = 1/2[\tan^{-1}[2\rho\sigma_1\sigma_2/(\sigma_1^2 - \sigma_2^2)]$ with the e_1-axis. Its size is determined by $\chi_2^2(\alpha)$ for a given Σ^{-1}. If the true mean of e, μ, is very much different from μ_0, the point $\bar{e}$ in the two dimensional e_1-e_2 space will be far from the point of μ_0 and will fall near the edge or even outside the ellipse. Therefore if (12) is satisfied we conclude that the mean of the signal is not μ_0. Equivalently, the type of the bottom coverage is not as initially suggested.

To decide whether the signal came from a particular type of surficial substrate requires a test of the population mean. From the data collected for each category, we calculated the estimates of the parameters $\hat{\mu} = \bar{e}$, $\hat{\Sigma} = S^2$. For designated significance levels (10% and 5%), confidence ellipses are constructed from (12) with the parameters replaced by their estimates. Figure 6 presents the RoxAnn confidence ellipses obtained from the survey data at a 90% confidence level, which, similary to the RoxAnn confidence squares, can be used as a substrate type classifier.

3.4. Comparison of Two Approaches

To compare the two classifiers discussed above, the data collected from known types of lake beds were spotted on the e_1-e_2 plane. The percentages of their correct coverages by both confidence ellipses and confidence squares were computed. Table 2 and 3 are comparisons between these coverages under significance levels 0.1 and 0.05, respectively. They show clearly that the elliptical-region approach is superior to the

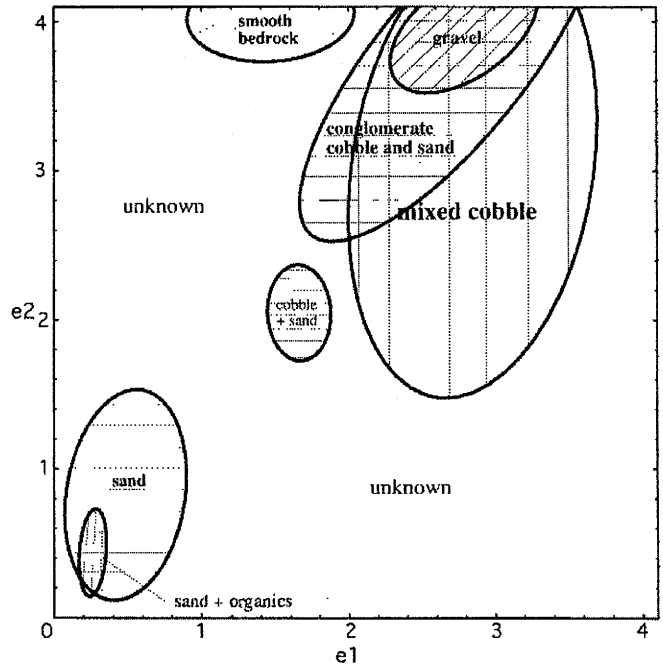


Fig. 6. Confidence ellipses from survey data, confidence level: 90%.

Table 2. Percentages of correct coverage of the training data.

Significance Level: 0.1

Category	1	2	3	4	5	6	7	8
RoxAnn squares	87.00%	89.45%	86.09%	91.95%	79.25%	82.52%	80.00%	80.16%
Ellipses	90.77%	94.58%	92.20%	91.95%	99.74%	89.92%	89.28%	90.42%

Table 3. Percentages of correct coverage of the training data.

Significance Level: 0.05

Category	1	2	3	4	5	6	7	8
RoxAnn squares	92.60%	94.58%	93.09%	94.46%	99.04%	87.33%	88.41%	91.72%
Ellipses	94.79%	96.43%	96.54%	94.46%	100.00%	92.69%	94.49%	95.49%

squared-region one. The former provides a higher percentage (up to 20.49%) of correct coverage. An 'amplified' comparison is given in Fig. 7, where the advantage of using ellipse can be recognized.

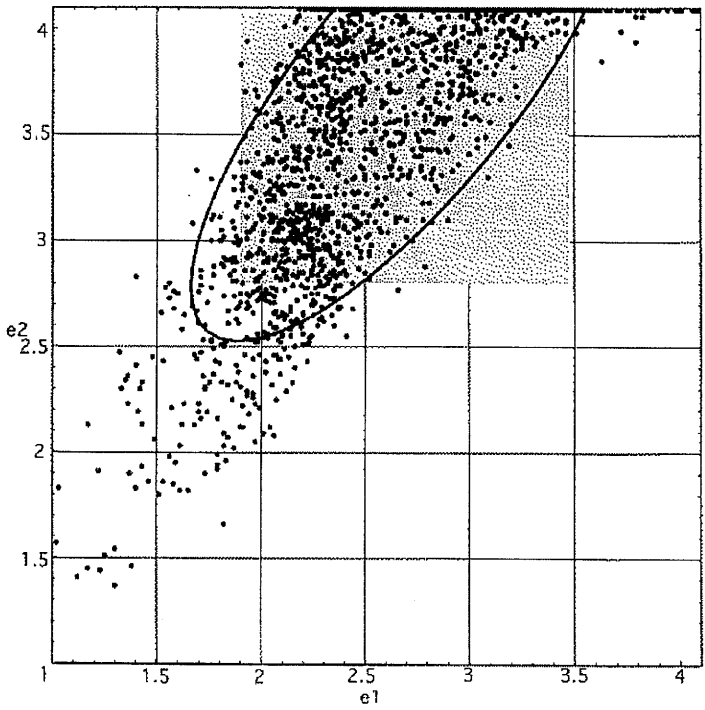


Fig. 7. Comparison of two approaches: confidence RoxAnn squares and confidence ellipses, confidence level: 90%.

This method relies on the assumption that the signals e_1 and e_2 have bivariate normal distribution. It is necessary to ensure that such an assumption is valid and to detect those cases where the assumption is seriously violated, such as the cases where the probability distribution is multimodal, and to develop methods to separate the offending classes into approximately normal subclasses. The areas to be analyzed are large and heterogeneous, which makes it very difficult to collect sufficient training samples for all types of lake surficial substrates. Moreover, it is almost impossible to locate large homogeneous areas with single coverage for collecting 'pure' samples. Therefore, a purely supervised analysis is not sufficient, and hence the adoption of a hybrid approach by incorporating the unsupervised method into supervised ones. Specifically, by displaying and analyzing the signals collected, we applied a clustering method to separate the data into unimodal classes.

4. The Classification Problem

The confidence regions, squares or ellipses alike, are centered at the means of the e_1 and e_2 signals, $\bar{e} = (\bar{e}_1\ \bar{e}_2)'$. Their shapes are determined by Σ^{-1}, and their sizes are decided by the prescribed significance level for a given Σ^{-1}. For lake floor having

m possible types of surficial substrates, the design of the pattern classifier requires circumscribing the regions of $\mathbb{R}_1^2, \mathbb{R}_2^2, \ldots, \mathbb{R}_m^2$ in the entire two-dimensional e_1-e_2 space. In a future classification, the location of the observed signals (or their mean) is checked. If it falls onto $\mathbb{R}_j^2$, it is classified as coming from type j. Such underwater pattern classification using remote sensing data remains to be a challenging problem partly because the physics of the interaction of acoustic waves with the lake bed is not completely understood; it has been observed that different substrates can have similar acoustic signatures. Furthermore, due to the inherent randomness of the data and the various shapes and/or angles of the confidence regions, it is not uncommon that there are areas overlapped by territories belonging to different floor coverages (see Figs. 4(a)–(b) and 6). Consequently, decisions are frequently required to categorize an individual measurement falling onto the overlapping area into one of the two or more types—a classification (isolation) problem.

4.1. The Cost, the Loss and the Risk in Classification

In developing a procedure for isolation, it is desired to minimize the loss from misclassification. Consider the simplest case—an area overlapped by only two types of surficial substrates, e.g., sand and cobble, hence there are only two kinds of errors involved, i.e., a signal actually from sand (category 1) being classified as from cobble (category 2) or *vice versa*. We want to minimize the probability of these errors weighted with their 'undesirability.' This undesirability is usually characterized by the average or expected loss being involved in this procedure.

Let $f_j(e)$ be the probability density of measurements resulting from type j coverage; the region $\mathbb{R}_j^2$ is specified on the basis of $f_j(e)$. In the two-category case the probability of correct classification of a measurement actually coming from type 1 is

$$P(1|1,d) = \int_{\mathbb{R}_1^2} f_1(e)\,\mathrm{d}e \tag{13}$$

and an incorrect decision which classifies an observation actually resulting from type 1 as from type 2 has the probability

$$P(2|1,d) = \int_{\mathbb{R}_2^2} f_1(e)\,\mathrm{d}e \tag{14}$$

where d is the decision rule for isolation. If the cost of this misclassification is $C(2|1)$, then the loss associated with the above misclassification $L(1,d)$ is

$$L(1,d) = C(2|1)P(2|1,d) \tag{15}$$

Assume the probabilities (proportion) of the coverages by these two types are known *a priori* as q_1 and q_2, respectively. Then the expected loss from misclassification of a signal which actually belongs to type 1 as from type 2 will be

$$\mathcal{L}(1,d) = q_1 C(2|1)P(2|1,d)$$

Hence the total expected loss (the risk) from costs of misclassification in the two-category case is

$$\mathcal{L}(d) = q_1 C(2|1) P(2|1,d) + q_2 C(1|2) P(1|2,d) \tag{16}$$

It is this expected loss that we wish to minimize. The above procedure can be easily generalized to multi-type cases. For an area overlapped by m possible coverages, the total probability of incorrect classifications of measurements from type i as from other types is

$$\sum_{j=1,j\neq i}^{m} P(j|i,d) = \sum_{j=1,j\neq i}^{m} \int_{\mathbb{R}_j^2} f_i(e)\,\mathrm{d}e \tag{17}$$

The loss associated with these misclassifications can be computed from

$$L(i,d) = \sum_{j=1,j\neq i}^{m} C(j|i) P(j|i,d) \tag{18}$$

Then the total *expected* loss we wish to minimize is

$$\mathcal{L}(d) = \sum_{i=1}^{m} q_i L(i,d) = \sum_{i=1}^{m} q_i \left\{ \sum_{j=1,j\neq i}^{m} C(j|i) P(j|i,d) \right\} \tag{19}$$

where q_i is the *a priori* probability of the proportion of the coverage by category i. A function $d(\cdot)$ which minimizes the risk (19) is the decision rule selected for classification.

4.2. Decision Rules for Classification

Given q_i and $f_i(e)$, we need to choose decision functions to minimize the expected loss (16) or (19). This can be accomplished under the framework designed in (Anderson, 1984). In the two-category case, the regions of classification, $\mathbb{R}_1^2$ and $\mathbb{R}_2^2$, should be chosen according to the following two inequalities:

$$\begin{aligned} \mathbb{R}_1^2 &: \ \frac{f_1(e)}{f_2(e)} \geq \frac{C(1|2)q_2}{C(2|1)q_1} \\ \mathbb{R}_2^2 &: \ \frac{f_1(e)}{f_2(e)} < \frac{C(1|2)q_2}{C(2|1)q_1} \end{aligned} \tag{20}$$

Whenever the first inequality is true, we conclude that the lake bed is covered by category 1; otherwise, category 2.

The above procedure can be generalized to the multi-category cases, in which the total expected loss is given in (19). The signal e comes from type k coverage if the following inequality holds:

$$\sum_{i=1,i\neq k}^{m} q_i C(k|i) f_i(e) < \sum_{i=1,i\neq j}^{m} q_i C(j|i) f_i(e), \quad j = 1,\ldots,m, \quad j \neq k \tag{21}$$

This can be further simplified if the costs $C(j|i)$ $(i = 1, \ldots, m;\ j = 1, \ldots, m;\ i \neq j)$ are all the same. Then

$$\mathbb{R}_k^2 : \quad \frac{f_k(e)}{f_j(e)} \geq \frac{q_j}{q_k}, \quad j = 1, \ldots, m,\ j \neq k \tag{22}$$

If the inequality (22) is satisfied, it is concluded that the lake floor should be categorized as type k.

4.3. Classification with Known *A Priori* Probabilities

In an area having two possible types of coverage, classification requires to designate the measurements as from one of the two distributions $\mathcal{N}(\mu_1, \Sigma_1)$ and $\mathcal{N}(\mu_2, \Sigma_2)$. The classification can be accomplished by testing the hypothesis H_0: type 1 coverage, against H_1: type 2 coverage, based on the available observation e. This is equivalent to testing:

$$H_0 : f_{e|H_0}(e) = \frac{1}{(2\pi)|\Sigma_1|^{1/2}} \exp\Big[-\frac{1}{2}(e-\mu_1)'{\Sigma_1}^{-1}(e-\mu_1)\Big]$$

vs.

$$H_1 : f_{e|H_1}(e) = \frac{1}{(2\pi)|\Sigma_2|^{1/2}} \exp\Big[-\frac{1}{2}(e-\mu_2)'{\Sigma_2}^{-1}(e-\mu_2)\Big]$$

The ratio of densities in (20) is

$$\frac{f_1(e)}{f_2(e)} = \left\{\frac{|\Sigma_2|}{|\Sigma_1|}\right\}^{1/2} \exp\left\{-\frac{1}{2}\Big[(e-\mu_1)'{\Sigma_1}^{-1}(e-\mu_1) - (e-\mu_2)'{\Sigma_2}^{-1}(e-\mu_2)\Big]\right\} \tag{23}$$

with $\tilde{\gamma}_{12}$ defined as

$$\tilde{\gamma}_{12} = \frac{C(1|2)q_2}{C(2|1)q_1}$$

Let

$$\gamma_{12} = \tilde{\gamma}_{12}\left\{\frac{|\Sigma_1|}{|\Sigma_2|}\right\}^{1/2}$$

The first inequality in (20), the region for $\mathbb{R}_1^2$ or the classification into type one, is obtained as the signal e satisfying

$$\frac{f_1(e)}{f_2(e)} \geq \tilde{\gamma}_{12}$$

or

$$-\frac{1}{2}\Big[(e-\mu_1)'{\Sigma_1}^{-1}(e-\mu_1) - (e-\mu_2)'{\Sigma_2}^{-1}(e-\mu_2)\Big] \geq \ln \gamma_{12} \tag{24}$$

The same arguments can be made for the classification into category two. Therefore with known *a priori* probability, let $c = -2\ln\gamma_{12}$. The decision rule is

$$\begin{aligned} \mathbb{R}_1^2 &: \ (e-\mu_1)'\Sigma_1{}^{-1}(e-\mu_1) - (e-\mu_2)'\Sigma_2{}^{-1}(e-\mu_2) < c \\ \mathbb{R}_2^2 &: \ (e-\mu_1)'\Sigma_1{}^{-1}(e-\mu_1) - (e-\mu_2)'\Sigma_2{}^{-1}(e-\mu_2) \geq c \end{aligned} \tag{25}$$

The decision procedure (25) minimizes the expected loss from misclassification. To carry out this procedure, the required population means μ_i and variances Σ_i need to be replaced by their estimates, e.g., their maximum likelihood estimates $\bar{e}$ and S^2 can be used.

For a lake floor with m substrate types, assuming that the costs of misclassification $C(j|i)$ are the same for all i and j and that the *a priori* probabilities q_i are known, define $\tilde{\gamma}_{jk}$ as

$$\tilde{\gamma}_{jk} = \frac{q_k}{q_j}$$

for $j = 1, 2, \ldots, m$ and $j \neq k$, the region $\mathbb{R}_j^2$ is defined by those e satisfying

$$\frac{f_j(e)}{f_k(e)} = \left\{\frac{|\Sigma_k|}{|\Sigma_j|}\right\}^{1/2} \exp\left\{-\frac{1}{2}\left[(e-\mu_j)'\Sigma_j{}^{-1}(e-\mu_j) - (e-\mu_k)'\Sigma_k{}^{-1}(e-\mu_k)\right]\right\} \geq \tilde{\gamma}_{jk}$$

$$k = 1, \ldots, m, \quad k \neq j \tag{26}$$

4.4. Classification with Unknown *A Priori* Probabilities—the Minimax Procedure

For a given decision rule $d(\cdot)$, the risk function $\mathcal{L}(\cdot)$ is a function of the parameter $q = (q_1, q_2, \ldots, q_m)'$. In the above discussion, q has been taken as a constant vector. If the parameter $q \in Q$ is a random variable distributed according to the *a priori* probability density function (or mass function in the discrete case) $\pi(q)$, then the average loss from the use of $d(\cdot)$ is

$$\mathcal{B}(\pi(q), d) = E_\pi \mathcal{L}(q, d) = E_\pi E_q L(q, d)$$

The risk $\mathcal{B}(\pi(q), d)$ is the so-called Bayes risk and the decision function $d^*(\cdot)$ that minimize this risk is the Bayes procedure (rule), namely,

$$\mathcal{B}(\pi(q), d^*) = \inf_d \mathcal{B}(\pi(q), d)$$

We are interested in the Bayes procedure because it is admissible in the sense that there are no other procedures better (incurring less risk) than this. In addition, the Bayes procedure can be used to facilitate the determination of the minimax method which will be used for classification in the later sections.

An ideal solution for making decisions would be the choice of $d(\cdot)$ such that the risk function $\mathcal{L}(q, d)$ attains a minimum for all $q \in Q$. Unfortunately this is usually

impossible, i.e., we cannot find a decision rule that minimizes $\mathcal{L}(q,d)$ uniformly in Q. Since it is not clear *a priori* which q is the best choice, to balance between optimality and robustness, a minimax procedure

$$\min_d \max_q \mathcal{L}(q,d)$$

is often employed. This procedure is designed to choose $d^\star(\cdot) \in \mathcal{D}$ so that the maximum expected loss is a minimum, i.e.,

$$\max_q \mathcal{L}\big(q,d^\star(\cdot)\big) \leq \max_q \mathcal{L}\big(q,d(\cdot)\big)$$

The idea here is to consider the worst case scenario and to minimize the largest expected loss, i.e., the largest risk.

In lake bed classification, the *a priori* probability q_j $(j=1,\ldots,m)$ is corresponding to $\pi(q)$ in the above discussion. Considering the fact that in general Bayes procedures are admissible and admissible procedures are Bayes or limit of Bayes procedures and further assuming the costs $C(j|i)=1$ $(j\neq i)$, Anderson (1984) has given the minimax solutions for classification of random variables into two and multiple populations. He showed that in the bivariate case, the minimax procedure is the Bayes procedure for *a priori* q_1^* (and q_2^*) such that

$$P(2|1,q_i^*) = P(1|2,q_i^*) \quad (i=1,2)$$

and in the multivariate case, the minimax procedure is the Bayes procedure for *a priori* q_i^* such that

$$P(i|j,q_i^*) = P(j|i,q_i^*), \quad (j=1,\ldots,m, j\neq i)$$

If $C(j|i)$ are the same for all i,

$$L(i,d) = \sum_{j=1,j\neq i}^{m} C(j|i)P\big(j|i,d\big) = \sum_{j=1,j\neq i}^{m} P\big(j|i,d\big)$$

In general, therefore, the minimax procedure is the Bayes procedure for which the risks $\mathcal{L}(i,d^*)$ are equal.

According to the above discussion, classification of signals having normal distribution with unknown *a priori* probability can be carried out. Define a random variable y as

$$y = (e-\mu_1)'{\Sigma_1}^{-1}(e-\mu_1) - (e-\mu_2)'{\Sigma_2}^{-1}(e-\mu_2) \tag{27}$$

Comparing (25) and (27), it can be seen that when $y<c$ or $y\in(-\infty,c)$, $e\in\mathbb{R}_1^2$; i.e., the substrate is type one; while when $y\geq c$ or $y\in[c,+\infty)$, $e\in\mathbb{R}_2^2$; i.e., the substrate belongs to type two. If q_1 and q_2 are unknown, $c=-2\ln\gamma_{12}$ is unknown. To find the value of c becomes the key issue in isolation. We resort to the minimax procedure for solution. Since the minimax procedure is the Bayes procedure $d^*(x)$ for q_1^* such that it makes $C(2|1)P(2|1,q_1^*) = C(1|2)P(1|2,q_1^*)$, if the signals have normal distribution, and further assume that $\Sigma_1=\Sigma_2=\Sigma$, the minimax solution can be obtained by using essentially the same derivation as in (Anderson, 1984).

Write

$$\eta^2 = (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)$$

It can be seen that the random variable y distributes according to $\mathcal{N}(\frac{1}{2}\eta^2, \eta^2)$ or $\mathcal{N}(-\frac{1}{2}\eta^2, \eta^2)$ if the signals e have distribution $\mathcal{N}(\mu_1, \Sigma)$ or $\mathcal{N}(\mu_2, \Sigma)$. Therefore, the misclassification of e from type one as from type two is equivalent to the misclassification of y into $\mathcal{N}(-\frac{1}{2}\eta^2, \eta^2)$. Introducing a new variable y here serves to simplify the calculation. Observe that

$$\begin{aligned} P(2|1) &= \int_{\mathbb{R}_2^2} f_1(x)\,\mathrm{d}x = \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi}\eta} e^{-\frac{1}{2}(y-\frac{1}{2}\eta^2)^2/\eta^2}\,\mathrm{d}y \\ &= \int_{-\infty}^{(c-\frac{1}{2}\eta^2)/\eta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\,\mathrm{d}z \end{aligned}$$

Use this change of variables to the calculation of $P(1|2)$ again. With given $C(1|2)$ and $C(2|1)$, seeking a minimax procedure which makes $C(2|1)P(2|1, q_1^*) = C(1|2)P(1|2, q_1^*)$ is equivalent to finding the value of c which satisfies the following equation:

$$C(1|2)\int_{(c+\frac{1}{2}\eta^2)/\eta}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\,\mathrm{d}z = C(2|1)\int_{-\infty}^{(c-\frac{1}{2}\eta^2)/\eta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\,\mathrm{d}z \quad (28)$$

To solve (28) for c, a trial and error procedure is often needed. For the cases where $\Sigma_1 \neq \Sigma_2$, let

$$\eta_1^2 = 2\mu_2'\Sigma_2\mu_1 - \mu_1'\Sigma_2\mu_1 - \mu_2'\Sigma_2\mu_2$$

$$\eta_2^2 = -2\mu_1'\Sigma_1\mu_2 + \mu_1'\Sigma_1\mu_1 - \mu_2'\Sigma_1\mu_2$$

The normal random variable y has mean η_1^2 or η_2^2 if the mean of the signals e is μ_1 or μ_2. A procedure similar to the above derivation gives the covariance of y. However it has a much more complicated format than the equal covariances case.

In summary, if the *a priori* distribution is known, then $c = -2\ln\gamma_{12}$ can be calculated, and use (25) for classification. In the case of unknown *a priori* probabilities, using a minimax approach leads to (28). The solution of (28) provides the required value of c. With c available, (25) can be used for isolation. It should be mentioned that the cost from misclassification, $C(i|j)$, is determined from the importance of the category concerned. For instance, it is known that the floor covered by mixed cobbles is most favorable to lake trout spawning. Therefore it is reasonable to put more weight on the cost of misclassification of this category. The above method can be generalized to a higher dimension case easily.

4.5. The Likelihood Ratio Tests

Hypothesis testing can also be constructed based on the likelihood ratios. Let Θ be the set of all possible parameters. For the two subsets Θ_1 and Θ_2, $\Theta_1 \subset \Theta$, $\Theta_2 \subset \Theta$ and $\Theta_1 \cap \Theta_2 = \emptyset$, an empty set. In our case $\theta = (\mu, \Sigma)$, $\theta_1 = (\mu_1, \Sigma_1)$ and $\theta_2 = (\mu_2, \Sigma_2)$. Given a sample of size n, $e = (e^1, e^2, \ldots, e^n)'$, to test H_0: $e \sim f_\theta, \theta \in \Theta_1$ against H_1: $e \not\sim f_\theta, \theta \in \Theta_1$, the likelihood ratio test is one having the form: For certain constant c_1, reject H_0 if and only if the likelihood ratio $\lambda_1(e) < c_1$, where

$$\lambda_1(e) = \frac{\sup_{\theta\in\Theta_1} f_\theta(e^1, e^2, \ldots, e^n)}{\sup_{\theta\in\Theta} f_\theta(e^1, e^2, \ldots, e^n)} \tag{29}$$

The numerator is the best explanation of e in the sense of maximum likelihood that the null hypothesis H_0 can provide, and the denominator is the best possible explanation of e. Owing to the fact that $\Theta_1 \subset \Theta$, $0 \leq \lambda_1 \leq 1$. For a predefined level $0 \leq \alpha \leq 1$, the constant c_1 is determined from

$$\sup_{\theta\in\Theta_1} P_\theta\{e : \lambda_1(e) < c_1\} = \alpha$$

For a normal distribution with known Σ, let $\Theta_1 = \{\mu_1\}$. It is easy to show that to test $\lambda_1(e) < c_1$ is equivalent to testing

$$z(\bar{e}) = \frac{\sqrt{n}(\bar{e} - \mu_1)'\Sigma(\bar{e} - \mu_1)}{|\Sigma|^{1/2}} > c_2 \tag{30}$$

Under H_0 and large n, the statistic $z(\bar{e})$ has a normal distribution with zero mean. By choosing $c_2 = z_{\alpha/2}$, the hypothesis testing can be carried out.

In the classification into one of two categories, we can use the likelihood ratio test described above for categories 1 and 2, respectively, then compare their respective λ_1 values. As an alternative, we may define the likelihood ratio as $\lambda_2'(e)$

$$\lambda_2'(e) = \frac{\sup_{\theta\in\Theta_1} f_\theta(e^1, e^2, \ldots, e^n)}{\sup_{\theta\in\Theta_2} f_\theta(e^1, e^2, \ldots, e^n)}$$

$$= \left\{\frac{|\Sigma_2|}{|\Sigma_1|}\right\}^{n/2} \exp\left\{-\frac{1}{2}\sum_n \left[(e-\mu_1)'{\Sigma_1}^{-1}(e-\mu_1) - (e-\mu_2)'{\Sigma_2}^{-1}(e-\mu_2)\right]\right\} \tag{31}$$

and the decision is made by comparing $\lambda_2'(e)$ with a constant c_3'. Again, the numerator is the best explanation of e in the sense of maximum likelihood that the null hypothesis H_0 ($\theta \in \theta_1$) can provide, and the denominator is the best possible explanation H_1 ($\theta \in \theta_2$) can provide. The null hypothesis (type 1 coverage) is rejected when a better explanation of e is provided by H_1 (type 2 coverage); or when $\lambda_2'(e) < c_3'$. Note that different from $\lambda_1(e)$, $\lambda_2'(e)$ can take any nonnegative value. For a predefined level $0 \leq \alpha \leq 1$, the constant c_3' is determined from

$$\sup_{\theta\in\Theta_1} P_\theta\{e : \lambda_2'(e) < c_3'\} = \alpha$$

Let

$$c_3 = c_3'\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right)^{n/2} \quad \text{and} \quad \lambda_2(e) = \lambda_2'(e)\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right)^{n/2}$$

The decision is made by comparing $\ln \lambda_2(e)$ with the threshold value c_3, i.e.,

$$\ln \lambda_2(e) = -\frac{1}{2}\sum \left[(e-\mu_1)'{\Sigma_1}^{-1}(e-\mu_1) - (e-\mu_2)'{\Sigma_2}^{-1}(e-\mu_2)\right] \geq c_3$$

It can be seen that the result is essentially the same as the one presented in (25). The classification problem now comes down to the choice of c_3.

To carry out hypothesis testing using any statistics requires its probability distribution. In this case, the probability density function of $\ln \lambda_2(e)$ is needed. Observe that a function of the likelihood ratio, $-2\ln \lambda_2(e)$, is the same as the statistic y defined in (27). As analyzed above, $y \sim \mathcal{N}(\frac{1}{2}\eta^2, \eta^2)$ if $e \sim \mathcal{N}(\mu_1, \Sigma)$, and $y \sim \mathcal{N}(-\frac{1}{2}\eta^2, \eta^2)$ if $e \sim \mathcal{N}(\mu_2, \Sigma)$. With the probability density function of y available, a decision can be made upon choosing an appropriate threshold value c_3.

Using the regular confidence region approach, designate a confidence level α, the constant c_3, i.e., the critical region is determined from

$$\sup_{\theta \in \Theta_1} P_\theta\{x : \ln \lambda_2(x) < c_3\} = \sup_{\theta \in \Theta_1} P_\theta\{y : -\frac{1}{2}y < c_3\} = \alpha$$

As an alternative, c_3 can be chosen according to the minimax criterion. Following the same derivation presented in the previous section, the same result as given in (28) is obtained with c replaced by c_3.

In short, the classical procedure for constructing tests based on likelihood ratios can also be used for classification. These tests have some intuitive appeal and frequently lead to uniformly most powerful procedure. The likelihood ratio (29) can be used as the test statistic. It has normal distribution when n is large. If the test statistic as defined in (31) is chosen, either a test resulting from the traditional confidence approach, or a test derived from the minimax procedure can be applied, which provides the same result as the one obtained in the previous sections.

4.6. Classifier Capable of Direct Classification

Figure 6 circumscribes $e1$-$e2$ ranges for various categories on the basis of the 90% confidence level. For signals falling onto those overlapped areas, inequalities (25) and (26) can be used for real-time classification. It is conceivable that a higher confidence level will result in a confidence ellipse diagram with much larger overlapped areas, which will certainly increase the work load in classification. In addition, it is desirable to have more efficient tools capable of bypassing the computation and comparison steps required by (25) or (26) and providing the result directly. To facilitate future classification, by choosing the *a priori* probabilities $q_1, q_2, \ldots, q_m$ according to past experience and by applying (26) to each point on the $e1$-$e2$ space ($0 < e_1 < 4.1$, $0 < e_2 < 4.1$), we have constructed Fig. 8, where the eight delineated regions serve as classifiers of

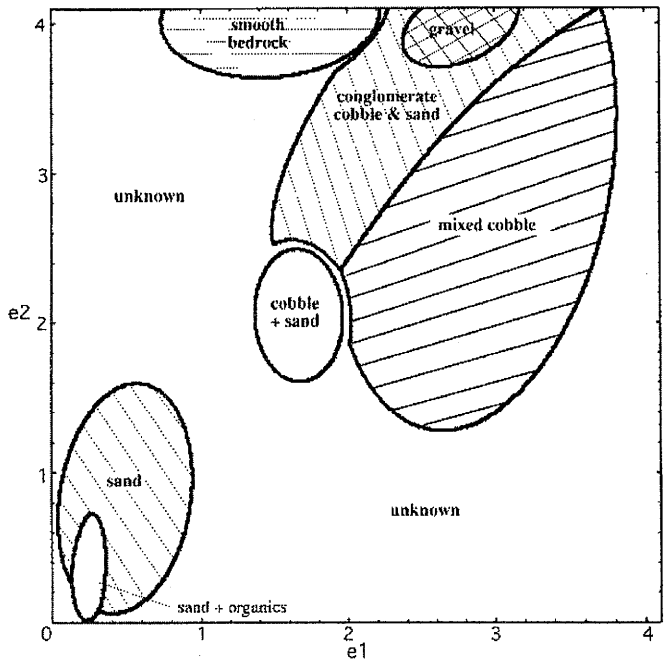


Fig. 8. Classifiers of lake bed surficial substrates.

the corresponding eight different lake floor categories. These categories occupy most of the lake bed of Lake Superior and include those which are most favorable to lake trout spawning.

5. Summary and Discussion

This work is concerned with the design of pattern classifiers for the lake bed terrain using the multivariate statistical method. Although the methods utilized belong to off-line procedures, the classifiers obtained are applicable to on-line isolation. Terrain characterization can be carried out with the confidence RoxAnn ellipses provided, or conducted using the categorization figure (Fig. 8) if a further isolation is necessary. Embedding these figures into the existing RoxAnn$_{TM}$ technique will enable an automated classification of lake bottom surficial substrates.

Although the devices are specifically developed for RoxAnn$_{TM}$, a relatively new system, the methodology and rationale are applicable in other systems. For remote sensing systems which use the backscatter of the acoustic signal directly, other approaches are needed to provide meaningful interpretation of the reverberation. It has been found that characterization is more difficult for those small-scale features in otherwise homogeneous regions. Due to the complicity of the underlying problem, simple interpolation and/or extrapolation are usually not sufficient. There have been

numerous reports of different approaches (e.g., Jhung and Swain, 1996; Masson and Pieczynski, 1993; Paola and Schowengerdt, 1995) such as developing a model for fusion of optical images, radar images and Geographic Information Systems data using Bayesian estimation techniques, designing backpropagation neutral network as classifiers, and employing a random field model for Bayesian image segmentation of satellite remote sensing problems. In order to design more universal classifiers and to achieve more accurate results, our effort has been directed to developing new classification algorithms using stochastic approximation methods (Yin and Yin, 1993; 1994; 1996), as well as designing a new procedure based on a discrete Markov Random Field.

References

Alexandrou D., Moustier C. and Haralabus G. (1992): *Evaluation and veriation of bottom acoustic reverberation statistics predicted by the point scattering model.* — J. Acoust. Soc. Am., Vol.91, No.3, pp.1403–1413.

Anderson T.W. (1984): *An Introduction to Multivariate Statistical Analysis.* — New York: Wiley.

Cervenka P. and Moustier C. (1993): *Sidescan sonar imagine processing techniques.* — IEEE J. Oceanic Eng., Vol.18, No.2, pp.108–122.

Collins W. and Gregory R. (1996): *A digital approach to seabed classification.* — Sea Technol., No.72, pp.83–87.

Curran T. (Ed.) (1995): *Remote sensing techniques for subtidal classification.* — Workshop Proceedings, Institute of Ocean Science, Sidney, British Columbia, Canada.

Jhung Y. and Swain P. (1996): *Bayesian contexual classification based on modified M-estimates and Markov random field.* — IEEE Trans. Geoscience and Remote Sensing, Vol.34, No.1, pp.67–75.

Malinverno A., Edwards M.H. and Ryan W.B.F. (1990): *Processing of SeaMARC swath sonar data.* — IEEE J. Oceanic Eng., Vol.15, No.1, pp.14–23.

Masson P. and Pieczynski W. (1993): *SEM algorithm and unsupervised statistical segmentation of satellite images.* — IEEE Trans. Geoscience and Remote Sensing, Vol.31, No.3, pp.618–633.

Moustier C. (1989): *Guest editorial and overview.* — IEEE J. Oceanic Eng., Vol.14, No.1, pp.286–288.

Paola J. and Schowengerdt A. (1995): *A detailed comparison of backpropagation neutral network and maximum-likelihood classifiers for urban land use classification.* — IEEE Trans. Geoscience and Remote Sensing, Vol.33, No.4, pp.981–996.

Yin G. and Yin K. (1994): *Asymptotically optimal rate of convergence of smoothed stochastic recursive algorithms.* — Stochastics and Statistics Reports, Vol.47, pp.21–46.

Yin G. and Yin K. (1996): *Passive stochastic approximation with constant step size and window width.* — IEEE Tran. Automat. Contr., Vol.41, No.1, pp.90–106.

Yin K. and Yin G. (1993): *Numerical studies of stochastic optimization algorithms with averaging.* — Proc. IFAC World Congress, Sydney, Australia, pp.427–432.