

Statistical Frameworks for Seismic Discrimination
DN Anderson KK Anderson DN Hagedorn SR Sain¹ CJ Young²
Pacific Northwest Laboratory
Contract No. DE-AC06-76RLO 1830

Abstract

To verify compliance with a Comprehensive Test Ban Treaty (CTBT), low energy seismic activity must be detected and discriminated. Monitoring small-scale seismic activity will require regional monitoring capabilities (within ≈ 2000 km, U.S. Congress (1988)). The reliable discrimination of small-scale seismic events requires a multi-dimensional representation of the seismic signal. A multi-dimensional characterization might include wave arrival times, magnitudes, and incidence and azimuth angles. These measurements can be used singly or combined to form discriminants, which are then subjected to a set of discrimination rules to categorize the source event. Statistical discrimination methods of this type require a *training sample*, i.e., a set of real or simulated seismic data used to optimize or tune the discrimination algorithm by assigning weights to the various discriminants, or by modifying the structure of the algorithm. Identifying the signatures of various seismic sources also requires a geologic characterization of the shallow structure of the earth in each particular region of interest. The results can be used to construct seismic signals representing nuclear test sources. These synthetic or simulated signals can be combined with empirical signals of earthquakes and mining activities to form a training sample for each region. This paper identifies several statistical issues that must be resolved in order to address the CTBT verification mission. These are all associated with uncertainties in the multidimensional characterization measurements or in the correlations among them. In particular, further research is needed on the statistical properties of:

- wave arrival time estimates, especially for regional wave arrivals, which sometimes tend to be emergent;
- regional velocity tables, i.e., the travel time tables that characterize the regional geology;
- measurements from regional seismic arrays, which can theoretically be combined to provide better estimates of wave arrival times, magnitudes, and direction;
- evasion scenarios (see U.S. Congress, 1988, Chapter 6);
- association, i.e., the agreement among seismic stations or between global and regional seismic networks that a seismic event has occurred and how to classify it;
- training samples, especially how to eliminate bias in the sample;
- robust discrimination algorithms, e.g., algorithms that are less sensitive to data with a poor signal-to-noise ratio;
- Bayesian discrimination algorithms e.g., algorithms that utilize expert opinion to substitute for missing data;
- statistical interdependencies in a regional seismic analysis, e.g., the relationship between the uncertainties in detection, phase identification, event association and discrimination.

Also discussed are several common statistical discrimination methods including linear discrimination, classification and regression trees (CART) and logistic regression.

Key Words: statistical discrimination, CART, logistic regression, nearest neighbor

¹Department of Statistical Sciences, Southern Methodist University

²Sandia National Laboratories

19960624 086

Objective

To meet the exacting demands of monitoring a CTBT, tools that integrate both seismic and statistical technologies are needed. Discriminating small seismic events in a region of interest requires a geological characterization of the crust of the earth in that region. This characterization leads to an understanding of the unique seismic signatures that will be produced by different seismic sources. Characterization of regions of interest to the United States (U.S.) is currently being studied as a part of the U.S. Department of Energy (DOE) CTBT R & D program (DOE, 1994). The findings from this research could conceivably be used to construct synthetic or simulated seismic signals representing the behavior of a nuclear source. These simulated signals can be combined with empirical signals from earthquakes and mining activities to form a *training sample* for the region. Discrimination for low energy seismic events will require a multi-dimensional representation of a seismic signal. Multi-dimensional measurements from a seismic signal (discriminants) can be combined by a variety of statistical techniques to form a unified discrimination method. With a training sample, a statistical discrimination technique is trained to optimally combine discriminants. For example, a discrimination rule might make use of a sum of weighted discriminants with the weights estimated from a regional training sample. Several common statistical discrimination methods are:

- linear discrimination, including Fisher's linear discriminant function
- quadratic discrimination
- nonparametric discrimination, including nonparametric likelihood methods and *k*th nearest-neighbor methods
- classification and regression trees (CART)
- logistic regression.

The objective of this research is to identify and resolve the statistical issues associated with monitoring a CTBT and to identify and research various statistical discrimination methods appropriate for regional seismic discrimination.

Preliminary Research Results

Statistical Issues in Seismic Analysis

Several general statistical issues need to be resolved in order to effectively verify a CTBT. Addressing them will undoubtedly uncover more detailed statistical questions. These issues include:

- Statistical properties of wave arrival time estimates—Arrival times of various seismic waves can be estimated in several ways, with associated uncertainties. Hypocenter estimation techniques are based on arrival times of different types of waves from a seismic disturbance. Estimates of depth and epicenter are used as discrimination tools. To ascribe uncertainty to a hypocenter estimate, the uncertainty in a wave arrival time estimate must be resolved. The seismic community is fully aware of the problem of ascribing uncertainty to a wave arrival time estimate in a teleseismic setting. This issue will need to be revisited in a regional setting because regional wave arrivals often exhibit a gradual transition from noise to wave signal, i.e., they are emergent.
- Statistical properties of regional travel time tables—To develop a U.S. CTBT seismic monitoring system, regional travel time tables will need to be developed. At regional distances, small variations in the velocity structure of the earth can have a significant impact on location and source characterization. To construct an uncertainty statement for a regional hypocenter estimate, the uncertainty in a regional travel time table will need to be combined with wave-arrival time uncertainties. Travel time tables provide the single largest source of systematic error in a hypocenter estimate.

- Statistical properties of seismic measurements from arrays—The proposed CTBT monitoring system will use seismic arrays to monitor regions of interest. The statistical properties of teleseismic array data are well understood in the seismic community. The statistical issues associated with using array data in regional seismic analysis will need to be researched. Regional seismic array data will be combined to estimate wave arrival times. The uncertainty in this process will need to be evaluated. Also, calculating wave magnitudes from regional array data will form the foundation for some regional discriminants. The statistical properties of regional array magnitudes and the discriminative power of these magnitudes should be researched.
- Statistical issues of evasion scenarios—The U.S. Congress (1988, Chapter 6) Office of Technology Assessment discusses several viable evasion scenarios that need to be addressed to effectively implement a CTBT. One of these scenarios involves masking or decoupling the energy release from a nuclear weapon test by performing the test in an open underground cavity. The parameters (decoupling factors) associated with an evasive decoupled weapon test need to be estimated. As noted in the congressional report, decoupling factors are needed to establish CTBT monitoring thresholds. The statistical properties of decoupling factors for various geologies and test cavity configurations should be researched in order to effectively establish CTBT monitoring thresholds.
- Statistical properties of association—Errors in the association process often lead to the cataloging of spurious events, or the degradation of the accuracy of seismic event locations. A quantitative measure of the strength of association is necessary. The creation of such a measure is a statistical challenge because it would have to combine measures of the similarity of the associated waveforms, the agreement of the back azimuths and slowness, and the agreement of the phase arrival times, all weighted by some type of observed signal-to-noise ratio. The measure must include penalties for missing data; for example, an analyst may wonder why a high quality, low noise station, which usually detects events from the area in question, did not detect it. If the station is operational, such non-data is strong evidence against the event being a true event. The problem of misassociation of individual events aside, another statistical issue is the estimation of the overall rate of misassociations reported in seismic bulletins and catalogs of events. Perhaps region-by-region rates of misassociation can be estimated by comparison of global seismic catalogs with the seismic catalogs from regional networks. Such regional networks may provide more complete catalogs of the region's events; i.e., the region's near-ground-truth.
- Statistical issues in constructing a training sample—Any regional discrimination process will need a training sample to build individual discriminant weights and possibly the structure of the discrimination algorithm. A proper regional training sample will be similar in every respect to data that would be analyzed in an operational setting. A training sample must not be biased by the elimination of information. Using only "clean" events in a training sample will seriously misrepresent misclassification rates and uncertainties in a discrimination algorithm. A training sample can further misrepresent these misclassification rates and uncertainties if the size of an event is confounded with the source of the event; e.g., if all large magnitude events in a training sample are earthquakes (or conversely if all are explosions). Finally, a training sample may include designed nuclear weapon tests or mining explosions (calibration events) in regions of interest. Statistical experimental design techniques can contribute to an optimally designed calibration event.
- Robustness of statistical discrimination algorithms—The ability of a statistical discrimination method to accurately perform under a less than optimal operational setting (its robustness) should be researched. The statistical discrimination methods discussed below should be considered when developing a U.S. CTBT seismic monitoring system because the robustness properties of these methods can be readily studied. These methods can be synergistically used as evidence when identifying the source of a seismic event. Complementary to research on the robustness of statistical discrimination methods, there should be research on statistical methods to address missing data. It is conceivable that it may not be possible to construct an appropriate training sample for a future region of interest. Expert opinion may be required to construct an initial discrimination algorithm for such a region. In the

statistical community, methods of integrating expert opinion into a statistical methodology are known as Bayesian methods. Determining if a statistical discrimination technique is amenable to Bayesian methods should be an important CTBT research task.

- Statistical interdependencies in a seismic analysis—All aspects of a regional seismic analysis are related or interdependent. These interdependencies will most likely produce statistical correlations that must be addressed. For example, a hypocenter estimate is used in forming some seismic discriminants. Understanding the correlations between various seismic measurements and sub-analyses is necessary to develop general uncertainty statements.

Statistical Techniques for Seismic Discrimination

In a seismological setting, statistical discrimination is the process of classifying a candidate seismic event as an earthquake, a chemical explosion, or a nuclear detonation using information from seismic discriminants (variables containing information derived from a seismic waveform). For a lucid discussion of potential regional seismic discriminants see Blandford (1995). The goal of discriminant analysis is not only to identify important or relevant discriminants but also to design a procedure incorporating these discriminants that accurately classifies the source of a seismic disturbance. In this section, some basic statistical multivariate discrimination methods are reviewed. Examples that illustrate each of these statistical discrimination methods are included. The data used in these examples were collected by Walter, Mayeda, and Patton (1994). *It is important to remember that these examples are not intended to be an authoritative seismic analysis of these data. Rather, the goal is to use data with seismic characteristics to illustrate the features of statistical discrimination methods. When presented with these examples, the reader should focus on the potential utility of the statistical discrimination methods and not the specific inferences from this small data set.*

For a seismic event, a vector of p discriminants, $\mathbf{x} = (x_1, \dots, x_p)'$, is measured or derived from a seismic waveform. The vector \mathbf{x} might include wave arrival times, magnitudes, incidence and azimuth angles and other potential discriminants. Note that these discriminants, x_i , $i = 1, \dots, p$, can take on any value, real (e.g., focal depth) or categorical (e.g., polarity of first motion). A classifier or discrimination rule is defined as a function $d(\mathbf{x})$ that mathematically combines the discriminants in \mathbf{x} . The value of the function $d(\mathbf{x})$ indicates the most likely source of a seismic event. An alternative formulation of the discrimination rule is to consider the vector of discriminants, \mathbf{x} , as a point in a p -dimensional space. The discrimination rule $d(\mathbf{x})$ can then be thought of as partitioning or dividing this p -dimensional space into sections. Each section would then be associated with a particular seismic source. For example, if a single real discriminant is considered, then $d(\mathbf{x})$ represents a "cut" on that discriminant, dividing the real line into "right" and "left" sections. All candidate events with values of this discriminant to the left of the cut could be labeled as explosions, while all values to the right could be labeled as earthquakes. If two discriminants are considered, then $d(\mathbf{x})$ is a line dividing the plane of real numbers into "left" and "right" areas. If three discriminants are considered, then $d(\mathbf{x})$ is a plane slicing through three-dimensional space. Note that more complex rules are not restricted to a single partition, nor are the partition boundaries always straight lines.

The error involved in a classification scheme is governed by the rule that partitions the relevant multi-variable space. Some insight into the sources and behavior of the misclassification error can be found by studying how a classification rule partitions the space of possible discriminant values. Let the discriminants \mathbf{x} for a particular seismic source be generated from a probability model (distribution) that is distinct from the probability model describing another seismic source. A classification rule divides the variable space into sections, with each section representing a seismic source. The probability of misclassification is simply the probability of making an incorrect classification. The total misclassification probability is a sum of the individual source misclassification probabilities.

Discrimination rules are constructed based on past experience in the form of a training sample. A training sample is a set of discriminant vectors \mathbf{x} with known classification that is representative of the distribution of the seismic sources. This set of data is used to "train" the discrimination rule. A training sample is used to build a discrimination rule and to test its performance or accuracy with cross-validation methods. The

performance of a discrimination rule is generally ascertained through some measure of misclassification cost. For example, Taylor et.al. (1989) and Glaser et.al. (1986) discuss the cost function

$$C(x|q)\pi_q P(x|q) + C(q|x)\pi_x P(q|x), \quad (1)$$

where π_q and π_x are the prior probability of an earthquake and explosion, $C(x|q)$ and $C(q|x)$ are the costs or penalties associated with mislabeling an explosion as an earthquake or an earthquake as an explosion, and $P(x|q)$ and $P(q|x)$ are the misclassification probabilities.

Such a criterion allows considerable flexibility to account for a variety of situations. For example, in a CTBT setting, the relative frequency of nuclear explosions versus earthquakes should be quite small. Hence, π_x should be set quite small relative to π_q in a CTBT setting. Perhaps more important is the cost associated with the different types of error. A false alarm (labeling an event as an explosion when in fact it is an earthquake) may be thought of as less serious than the failure to detect a violation of the CTBT (labeling an event as an earthquake when in fact it is an explosion). Hence, the cost associated with failure to detect a violation, $C(q|x)$, would be set higher than the cost of a false alarm, $C(x|q)$. The probabilities in Equation (1) are generally unknown quantities, but they can easily be estimated with cross-validation methods.

Linear Discrimination

One of the most conceptually simple rules, linear discrimination, is based on the assumption that the sources exhibit Gaussian distributions with identical covariance structure (i.e., differing only in location). A linear discrimination rule assigns a candidate event to the source with centroid closest to the position of the \mathbf{x} in the sample space. Many distance metrics are possible, but the most natural is the Mahalanobis distance, using the pooled within-group sample variances (an unbiased estimator of the common covariance matrix of the groups). For example, consider two sources, earthquake and nuclear detonation (NUDET). Then the estimated covariance is written as $\mathbf{s} = (n_x \mathbf{s}_x + n_q \mathbf{s}_q) / (n_x + n_q - 2)$, where \mathbf{s}_x and \mathbf{s}_q are the sample covariances estimated from the training sample and n_x and n_q are the sample sizes of the training sample for the two sources. A new observation \mathbf{x} is labeled as a NUDET if $(\bar{\mathbf{x}}_x - \bar{\mathbf{x}}_q) \mathbf{s}^{-1} (\mathbf{x} - \frac{1}{2}(\bar{\mathbf{x}}_x + \bar{\mathbf{x}}_q)) > 0$, where $\bar{\mathbf{x}}_x$ and $\bar{\mathbf{x}}_q$ are the means of the training sample for the two sources. A quadratic discrimination rule is possible if the covariances are not equal.

Nonparametric Discrimination

The classical approach to statistical discrimination involves assuming a parametric form for the probability distribution of each group and using a training sample to estimate the relevant parameters. A candidate event is then classified to the group with the largest likelihood. For example, one might choose the Gaussian distribution to model the earthquakes. A training sample would then be used to estimate the mean and covariance. The distribution for NUDETs would be handled similarly. A candidate event would then be classified as an earthquake if $\hat{f}_q(\mathbf{x}; \bar{\mathbf{x}}_q, \mathbf{s}_q^2)$ is greater than $\hat{f}_x(\mathbf{x}; \bar{\mathbf{x}}_x, \mathbf{s}_x^2)$ or as a NUDET otherwise. Here, \hat{f}_q and \hat{f}_x are the earthquake and NUDET distributions with parameters $\bar{\mathbf{x}}_q$, $\bar{\mathbf{x}}_x$, \mathbf{s}_q^2 , and \mathbf{s}_x^2 estimated from training data.

Hand (1981) and Silverman (1986) have studied the use of nonparametric methods for use in classification problems, replacing the parametric probability models in the classical procedure with nonparametric density estimates. Examples of nonparametric density estimates include the histogram and the kernel estimator. Recent advances in multivariate probability density estimation (see Scott (1992)) have led to further work in nonparametric methods for discrimination. Hall and Wand (1988) study the use of nonparametric methods with probability model differences as a discrimination tool. Holmstrom and Sain (1993) successfully apply a ratio of nonparametric probability models to applications in particle physics.

An example of nonparametric discrimination is shown in Figure 1. The plot on the right shows decision boundaries based on a bivariate product kernel estimator of the distributions of the earthquakes and the NUDETs. New events are classified according to which model yields a higher density value at the new point,

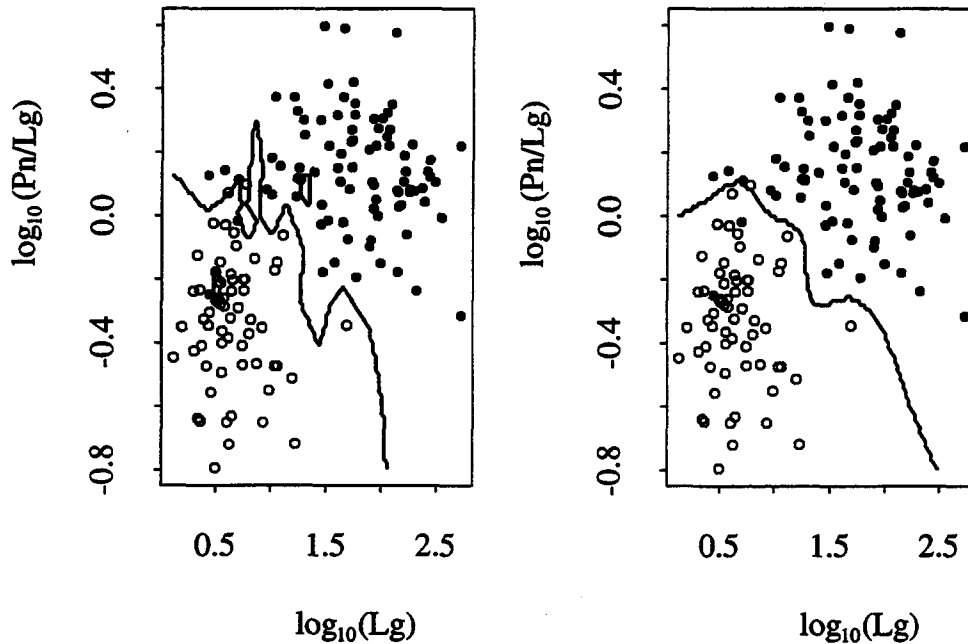


Figure 1: Example of nearest neighbor (left plot) and kernel (right plot) discrimination.

x. Note that the boundaries are highly nonlinear, offering greater flexibility. This flexibility can become increasingly important for situations with more complex structure or as the dimensionality increases.

The plot on the left illustrates a discrimination method in which new events are classified according to a k th nearest-neighbor rule (Fix and Hodges, 1951) with $k = 1$. Here a candidate event is classified to the group in which the nearest point to that event belongs. The boundaries in this case are highly irregular due to the lack of “smoothing”. This single nearest-neighbor method represents an extreme case, when no smoothing is performed. The decision boundary for the k th nearest-neighbor rule will “smooth” as the value of k increases.

Tree-Based Methods

Binary tree methods represent an important improvement over some of the basic methods of statistical discrimination and in the use of standard linear and additive models for classification problems. First and foremost, binary tree methods can incorporate both numeric and categorical discriminants. Complicated discriminant behavior can also be modeled easily. Furthermore, binary tree methods are conceptually simple and yield a nice graphical representation of the final decision tree and the resulting classifications. This is especially valuable when dealing multi-dimensional data. For an overview of the theory and methodology, see Breiman et.al. (1984). Artificial Neural Networks (ANN) also have these features, however binary tree methods have advantages over ANN in seismic discrimination applications. For a comparison of binary trees and neural networks see Blough and Anderson(1994).

Binary tree methods are based on the notion of recursive partitioning. To illustrate, consider again the notion of a vector of discriminants, \mathbf{x} , as a distinct point lying in a multivariate space. To build a binary decision tree, the discrimination algorithm recursively divides this multivariate space into smaller and smaller subregions. This dividing process is based on the training sample, and continues until each subregion is homogeneous with respect to one of the sources.

This tree-growing process leads to a large number of regions and can overfit by becoming overly representative of the training sample. To prevent overfitting and extreme complexity, a tree is grown and then pruned

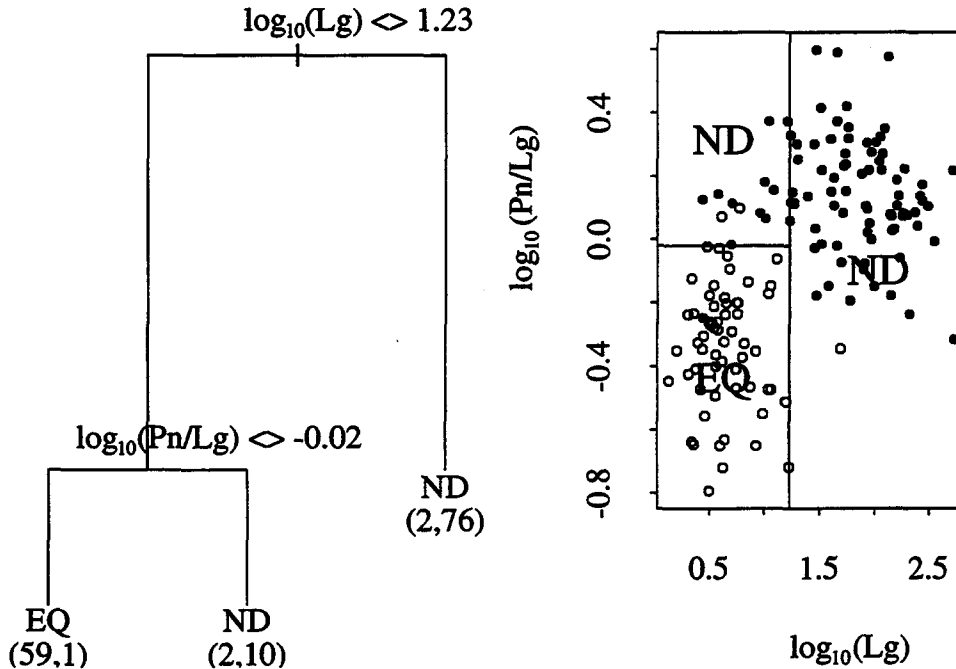


Figure 2: Binary tree example.

or shrunk to a more manageable size. This pruning is generally done by removing the least important splits, based on a cost-complexity measure that is designed to balance the homogeneity of the final regions and the complexity of the tree. This pruning can be thought of as combining adjacent regions of the multivariate space that are very much alike.

Graphically, the splitting of the discriminant space into regions can be displayed as a binary tree. An example is shown in Figure 2, using the data from Walter, Mayeda, and Patton (1995). The figure on the left shows a tree that has been grown and then pruned using a cost-complexity function as discussed in the previous paragraph. The final partitioning of the discriminant space is shown in the figure on the right. In this example, 59 earthquakes were correctly identified as earthquakes and 1 nuclear detonation was incorrectly identified as an earthquake. The counts at the “ND” branches of the tree are interpreted similarly.

Logistic Regression

Standard linear regression techniques have also been used extensively for discrimination tasks. The procedure is to model the responses, in this case a dummy variable taking on values of 0 or 1 depending on the candidate event being an earthquake or NUDET, as a linear function of the predictor variables, in this case the discriminants. In the simplest form, the model is written as

$$\ln \frac{p}{1-p} = \beta_0 + \beta_1 x_1, \quad (2)$$

where p is the probability of NUDET as calculated using the training sample and x_1 is a discriminant measured from a waveform. The parameters β_0 and β_1 are estimated using an iterative maximum likelihood procedure. The form of the left hand side of Equation (2) is used to ensure that predicted values for p are bona fide probabilities, i.e., that they lie in the interval $[0, 1]$. Once the parameters are estimated from a training sample, a new candidate event is classified as a NUDET if the estimated probability is greater than 0.5.

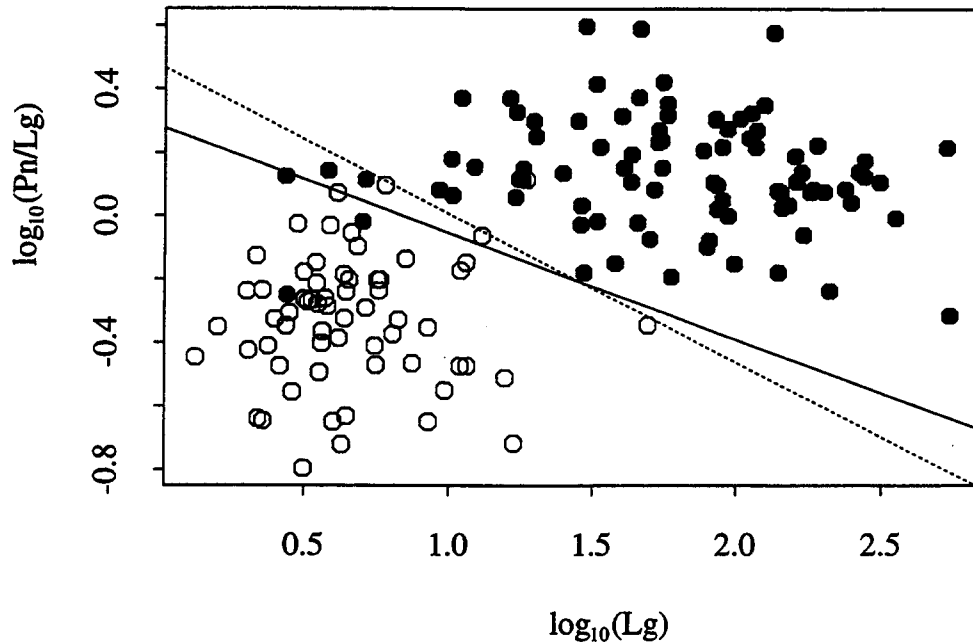


Figure 3: Logistic regression example. Dotted line represents linear discrimination.

Note that the model given by Equation (2) can easily be extended to include multiple discriminants as well as interaction terms. Furthermore, categorical variables can also be included as variables in the model. An example is shown in Figure 3, again using the data collected by Walter, Mayeda, and Patton (1995). The solid line represents the partition of the discriminant space based on the logistic regression approach. The dotted line represents the partition based on linear discrimination, and is shown for comparison. The two methods are quite similar since the fitted logistic model, in this case, included no interaction terms. Logistic regression methods are easily extended to the CTBT problem of discriminating between earthquakes, NUDETs and commercial explosions.

Recommendations and Future Plans

Research to address the statistical issues discussed above is an integral part of the DOE CTBT R & D program (DOE, 1994). In collaboration with AFTAC, PNL plans to research and resolve the statistical issues associated with regional discrimination. This research is comprised of two major issues. First, the operational capabilities of a CTBT monitoring system must be fully understood when monitoring regions of interest. Realistic training samples can be used accurately assess misclassification rates and uncertainties in a discrimination algorithm. Developing appropriate regional training samples is critical to the regional discrimination problem. Second, the statistical discrimination discussed above can be synergistically used by AFTAC as ancillary or corroborative evidence when identifying the source of a seismic event. PNL will continue research on statistical discrimination methods as a collaborative effort with AFTAC and Sandia National Laboratory.

Acknowledgment. The authors thanks Rich Hanlen and Wes Nicholson for useful technical discussions, and Frank M Ryan for editorial assistance. This work was done at the Pacific Northwest Laboratory, a multiprogram laboratory operated for the U.S. Department of Energy under Contract DE-AC06-76RLO 1830.

References

1. Blandford, R. R. (1995), *Regional Seismic Event Discrimination, Monitoring a Comprehensive Test Ban Treaty*, NATO Advanced Studies Institute, Series E, (in press).
2. Blough, D. K., and K. K. Anderson (1994), *A Comparison of Artificial Neural Networks and Statistical Analyses*, Technical Report PNL-9050 UC-705, Pacific Northwest Laboratory.
3. Breiman, L., J. Friedman, R. Olshen, and C. Stone (1984), *Classification and Regression Trees*, Wadsworth International Group.
4. DOE (1994), *Comprehensive Test Ban Treaty Research and Development FY95-96 Program Plan*, United States Department of Energy, Report DOE/NN-0003.
5. Fix, E., and J. L. Hodges (1951), *Nonparametric Discrimination: Consistency Properties*, Project No. 21-49-004, Report No. 4, U.S. Air Force School of Aviation Medicine, Randolph Field, TX.
6. Glaser, R. E., S. R. Taylor, M. D. Denny and E. S. Vergino (1986), *Regional Discrimination of NTS Explosions and Western U.S. Earthquakes: Multivariate Discriminants*, Technical Report UCID-20930, Lawrence Livermore National Laboratory.
7. Hall, P. and M. Wand (1988), *On Nonparametric Discrimination using Density Differences*, *Biometrika*, 75, 541-547.
8. Hand, D. J. (1981) *Discrimination and Classification*, John Wiley & Sons.
9. Holmstrom, L. and S. Sain (1993), *Searching for the Top Quark Using Multivariate Density Estimates*, Technical Report, Department of Statistics, Rice University.
10. Scott, D. W. (1992), *Multivariate Density Estimation*, John Wiley & Sons.
11. Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall.
12. Taylor, S. R., Denny, M. D., Vergino, E. S., and Glaser, R. E. (1989), *Regional Discrimination Between NTS Explosions and Western U.S. Earthquakes*, *Bull. Seism. Soc. Am.*, 79, 1142-1176.
13. U.S. Congress (1988), *Seismic Verification of Nuclear Testing Treaties*, Office of Technology Assessment, OTA-ISC-361.
14. Walter, W. R., K. M. Mayeda and H. J. Patton (1995), *Phase and Spectral Ratio Discrimination between NTS Earthquakes and Explosions, Part I: Empirical Observations*, *Bull. Seism. Soc. Am.*, (in press).