Statistical Classification Methods Applied to Seismic Discrimination

D. N. Anderson   N. E. Miller
K. K. Anderson   T. Redgate
D. N. Hagedorn   A. C. Rohay
K. T. Higbee

June 11, 1996

Battelle

MASTER

Statistical Classification Methods Applied to Seismic Discrimination

D.N. Anderson
K.K. Anderson
D.N. Hagedorn
K.T. Higbee
N.E. Miller
T. Redgate
A.C. Rohay
Technical Editor: F.M. Ryan

June 11, 1996

DISCLAIMER
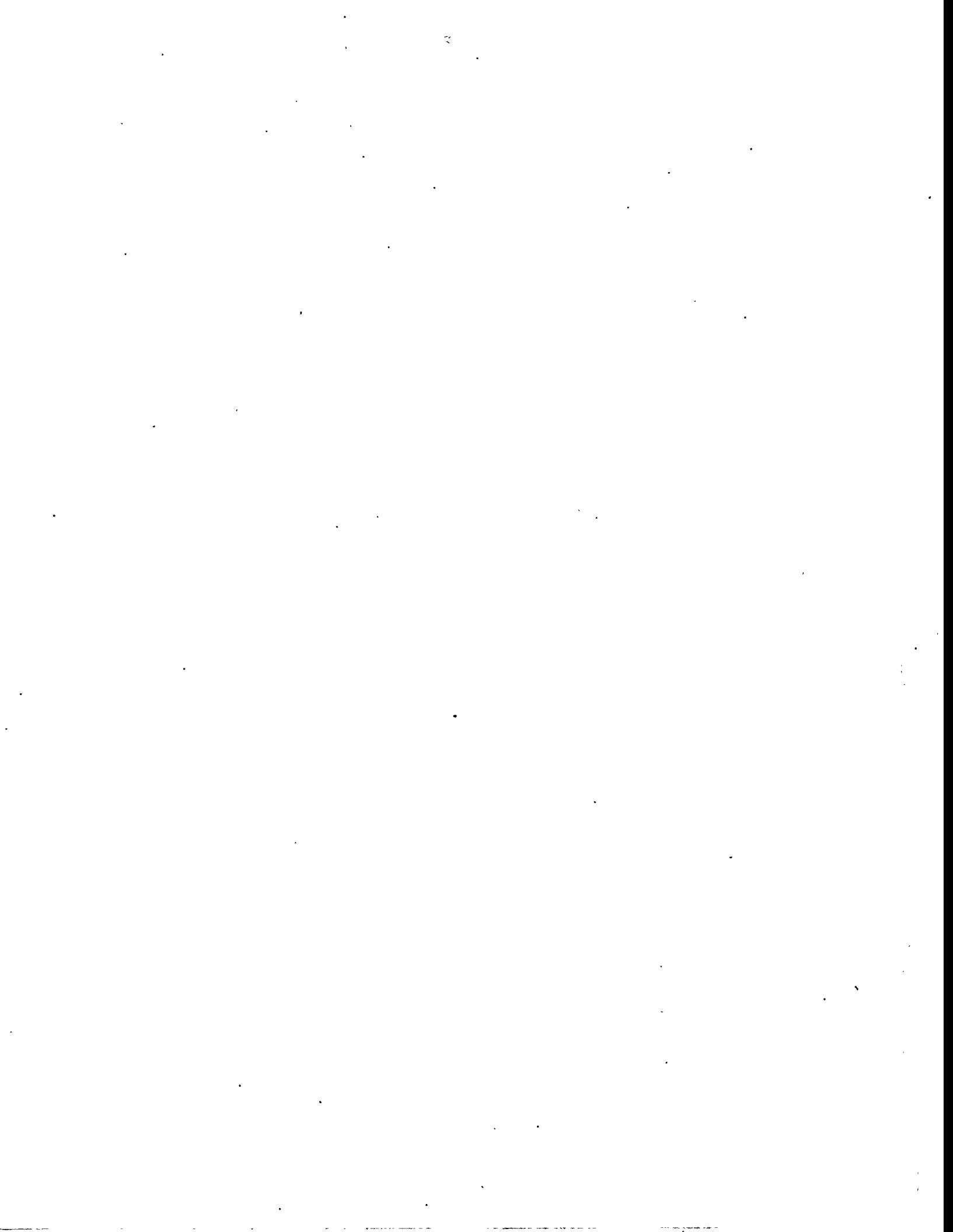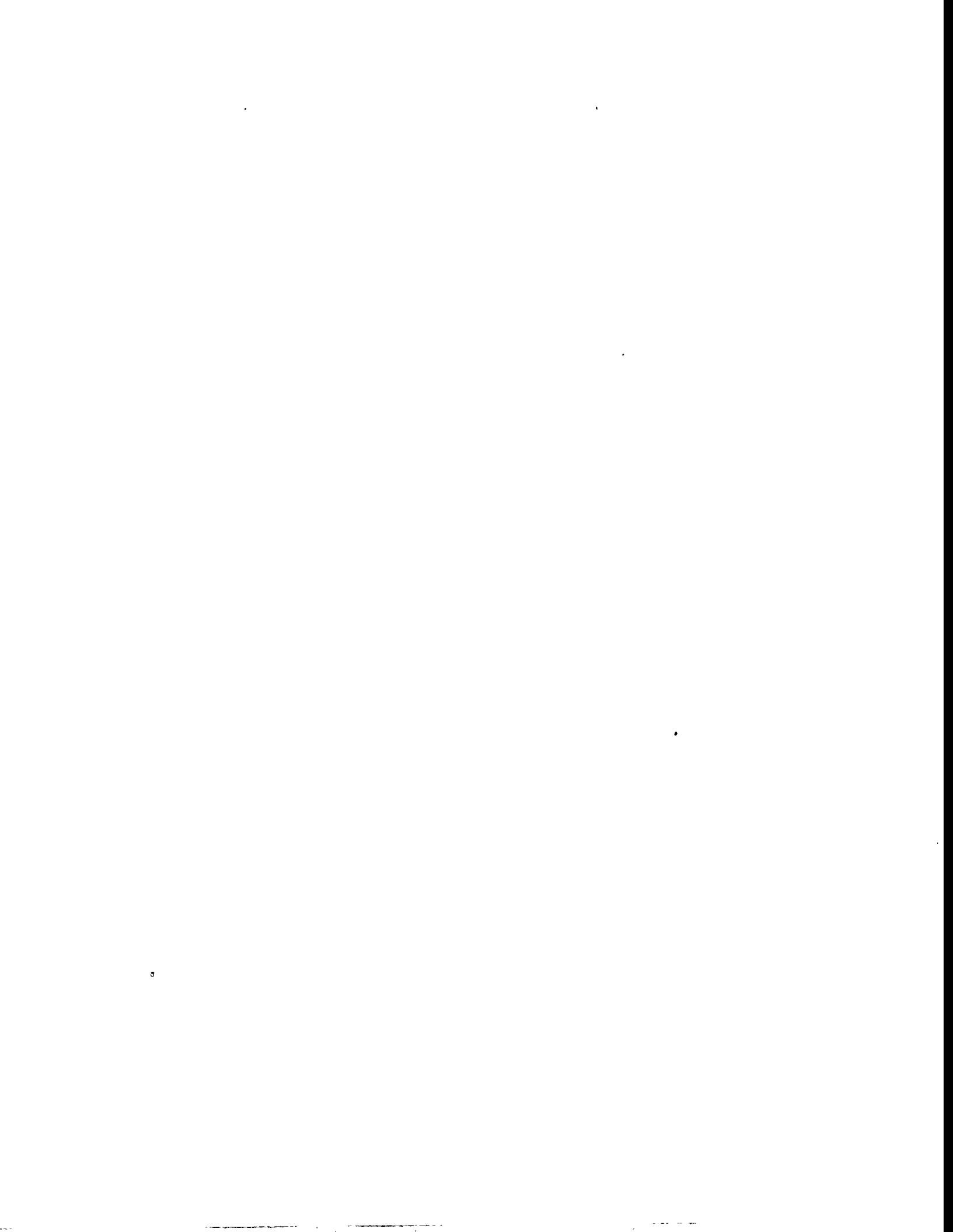
Executive Summary

To verify compliance with the CTBT, seismic measures must be incorporated in an appropriate multivariate statistical classification framework. Various issues to consider in selecting a classification method are outlined. The issues or selection criteria are grouped into the categories of simplicity, robustness, applicability, and performance. Some of the most notable issues for the CTBT setting include:

- Physical interpretability of the method and the classification decision boundaries.

- Proper handling of missing measurements.

- Incorporation of both discrete and continuous measurements.

- Classification accuracy.

The data from Taylor et al. (1989), augmented with 98 additional explosions, is used to illustrate the need for an appropriate multivariate approach that incorporates both seismic knowledge and statistical techniques in determining the appropriate seismic features to use in classifying seismic events. This data is also used to show the importance of proper handling of missing observations. A classification method must either directly handle missing observations or appropriately compensate for their absence. The method of multiple imputation for incorporating the uncertainty due to missing observations is shown to help avoid an overly optimistic estimation of the classification error rate. Apparent error rates (those obtained by training and then testing on the same data) are shown to be poor estimators of the true classification error rate, as compared to cross-validated (hold one out) error rates.

Eight statistical classification methods, representing a wide range of possible approaches, are explained and discussed in relation to the selection criteria outlined in this report. The eight methods are:

- Linear discriminant analysis (LDA)

- Quadratic discriminant analysis (QDA)

- Variably regularized discriminant analysis (VRDA)

- Flexible discriminant analysis (FDA)

- Logistic discrimination

- K-th nearest neighbor (KNN) discrimination

- Kernel discrimination

- Classification and regression trees (CART)

A summary of several important selection criteria for each of the eight methods reveals that no one method is best on all issues. However, the CART classification approach appears to satisfy many important issues with regard to the CTBT setting. Further analysis based

on more realistic operational data, is currently underway to better assess the important issue of classification accuracy for the various methods. After this analysis is completed, a recommendation for an appropriate CTBT classification framework will be made, based on the analysis and the conclusions from this report.

The evaluation of statistical classification methods in this report not only helps in determining an appropriate CTBT operational classification algorithm, but is also helpful in determining which classification methods to include in an "analyst tool box" to be used to understand new regions and to examine seismic events of interest. Examining the results from several different classification algorithms and understanding the strengths and weaknesses of each can help in better understanding and characterizing events of interest.

Acknowledgments

Contents

List of Tables

List of Figures

List of Statistical Abbreviations

CART   Classification and Regression Tree
EM   Expectation Maximization
FDA   Flexible Discriminant Analysis
KNN   K-th Nearest Neighbor
LDA   Linear Discriminant Analysis
MARS   Multivariate Adaptive Regression Splines
QDA   Quadratic Discriminant Analysis
RDA   Regularized Discriminant Analysis
VRDA   Variably Regularized Discriminant Analysis

1 Introduction

To verify compliance with a Comprehensive Test Ban Treaty (CTBT), low energy seismic activity must be detected and discriminated. Monitoring small-scale activity will require regional (within $\approx$ 2000 km) monitoring capabilities (U.S. Congress, 1988). This report provides background information on various statistical classification methods and discusses the relevance of each method in the CTBT seismic discrimination setting. Criteria for classification method selection are explained and examples are given to illustrate several key issues. This report describes in more detail the issues and analyses that were initially outlined in a poster presentation at a recent American Geophysical Union (AGU) meeting (Anderson et al., 1995).

Section 2 of this report describes both the CTBT seismic discrimination setting and the general statistical classification approach to this setting. Seismic data examples illustrate the importance of synergistically using multivariate data as well as the difficulties due to missing observations. Classification method selection criteria are presented and discussed in Section 3. These criteria are grouped into the broad classes of simplicity, robustness, applicability, and performance. Section 4 follows with a description of several statistical classification methods: linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), variably regularized discriminant analysis (VRDA), flexible discriminant analysis (FDA), logistic discriminant analysis, K-th Nearest Neighbor discrimination (KNN), kernel discrimination, and classification and regression tree discrimination (CART). The advantages and disadvantages of these methods are summarized in Section 5.

2 Background

In a seismological setting, statistical discrimination is the process of classifying a candidate seismic event as an earthquake, a single point explosion (either chemical or nuclear), or a quarry blast (ripple-fired event) using information from seismic waveform features (seismic discriminants). Blandford (1995) provides a lucid discussion of potential regional seismic discriminants. 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 and geophysical characterization of the earth's crust in that region. This knowledge will be used to accurately model the earth's effects on the seismic signal, and enable compensation for these effects, which often obscure the true nature of the source. Characterization of regions of interest to the United States is currently being pursued as part of the U.S. Department of Energy (DOE) CTBT R&D program (DOE, 1994). The findings from this research can assist in constructing the probability structure of regional discriminants from a nuclear source. The result of a classification algorithm is a prediction of the source of a seismic event and an estimate of the uncertainty in the prediction.

Sections 2.1 through 2.3 are provided to give further background into the seismic discrimination problem and motivation for the work. A general description of the statistical classification approach is outlined in Section 2.1. Section 2.2 demonstrates the importance of approaching seismic discrimination from a multivariate perspective. The crucial issue of

Table 1: Definition of Data Set Measurements. See Taylor et al. (1989).

Measurement	Definition
S_n	Ratio of the maximum amplitude of S_n to L_g.
m_b	Body-wave magnitude measured from the maximum amplitude of P_n at frequency near 1 Hz.
$m_b - M_s$	Body-wave magnitude relative to regional surface-wave magnitude, the latter measured from maximum amplitude of regional Rayleigh-wave amplitude at periods near 12s.
$\log(SH)$	Log ratio of the mean-square energy after SH arrival to energy before SH arrival.
$\log(L_g/P_g)$	Log ratio of the maximum amplitudes of L_g to P_g.
MURA L_g	Ratio of $0.5 - 1.0$ Hz average spectral amplitudes to 2.0 to 4.0 Hz average spectral amplitudes, with a correction for attenuation with distance.
$\log(S_t) - 2m_b$	Log energy of long-period (7-30s) Love wave (transverse horizontal) amplitudes relative to short period energy ($2m_b$).

accounting for missing observations in seismic discrimination is illustrated in Section 2.3.

The data examples in Sections 2.2 and 2.3 come from seismic measurements of earthquakes and nuclear explosions provided by Dr. Ron Glaser of Lawrence Livermore National Laboratory (LLNL) and Dr. Steve Taylor of Los Alamos National Laboratory (LANL), see Taylor et al. (1989). The seven variables shown in Table 1 were used in the examples of this section. The original data from (Taylor et al., 1989) is derived from 233 explosions and 130 earthquakes in the magnitude range 2.5 to 6.5, recorded at four broadband seismic stations operated by LLNL. The data provided to us covers an additional 98 explosions, bringing the total to 331 explosions and 130 earthquakes. The stations surround the Nevada Test Site (NTS) at distances of about 200 to 400 km. *It is important to remember that our analysis of this data is not intended to be an authoritative seismic analysis. Rather, it illustrates statistical concepts and issues in seismic discrimination. When presented with the real data examples in Sections 2.2 and 2.3, the reader should focus on the statistical concept being presented and not on the specific seismic inferences drawn from this data set.*

2.1   General Discrimination Setting

For a seismic event, a vector of p waveform features (discriminants), $\mathbf{x} = (x_1, \ldots, x_p)'$, is measured from a seismic waveform. The vector $\mathbf{x}$ might include wave arrival times, magnitudes, signal-to-noise measures, calculated location and depth, and other potential waveform measurements. Note that individual discriminants, x_i, $i = 1, \ldots, p$, can be real (depth) or discrete (polarity of first motion). As applied to the CTBT problem, statistical classification involves two basic activities. First, we identify seismic discriminants that evidentially predict the source of a seismic disturbance, i.e., that have strong discriminative ability. This aspect of statistical classification is accomplished with statistical methods and seismic theory. Secondly, we integrate these discriminants into an appropriate statistical classification framework. Statistical classification necessarily involves these two activities, since it is con-

ceivable that strong discriminants can be integrated into an improper framework, which would likely result in less than optimal decision error rates. As an example, if the covariance structure of x is different for earthquakes and explosions, then linear statistical classification methods would give a less than optimal performance.

From a decision theory approach, a classification framework for a CTBT setting is composed of three basic components — the probability structures (parametric or non-parametric) for each seismic source, the prior-to-event probabilities that an event will be generated by a particular seismic source, and the costs associated with incorrectly predicting the source of an event. We illustrate, with a fabricated example, how these components contribute to the construction of a classification framework. For a discussion of the mathematical details of the decision theory approach to discrimination, see McLachlan (1992). In Figure 1(a), two normal (Gaussian) probability densities are given. To link these fabricated models to the CTBT problem, we may think of the "red" and "green" densities as the probability models for seismic discriminants generated from an explosion and earthquake, respectively. A classification region can be formed by determining the values of the discriminant for which the explosion density is larger than the earthquake density. The classification region is shown in Figure 1(b). If the prior probability for an explosion is small, the region changes as shown in Figure 1(c). Intuitively, in this case, a discriminant must exhibit overwhelming evidence that an event is an explosion, since it is very unlikely for an explosion to occur. If we assign an extremely high cost to erroneously declaring an event to be an earthquake, then the region in Figure 1(c) is adjusted again, as in Figure 1(d). This region minimizes the expected cost of making an incorrect source prediction.

Statistical classification frameworks are tuned, based on past experience, with a training or calibration sample. A training sample is a set of discriminant vectors x_i generated by known seismic sources. Ideally, a training sample should be sufficiently large to characterize the multivariate probability structure of the discriminants for each seismic source. Otherwise, the probability structure of x must be conjectured for each seismic source. A training sample is used to estimate the statistical parameters (such as covariance matrices) in a classification framework and to test the framework performance or accuracy. With decision theory methods, the classification framework parameters are tuned to minimize the expected cost of making an erroneous source prediction. The accuracy of a classification framework is assessed by estimating the probabilities associated with incorrectly predicting the source of an event. Estimating these probabilities is accomplished with cross-validation methods.

2.2 Multivariate Discrimination

Several discriminants in combination provide greater classification power than the same discriminants applied individually. For instance, a classification rule may change with the size of an event. The data provided by Dr. Ron Glaser of LLNL and Dr. Steve Taylor of LANL (Taylor et al., 1989) illustrates the need for a multivariate classification process. Figure 2 illustrates the discriminative power of combining $m_b - M_s$ with Murphy's L_g spectral ratio. Additionally, the co-plots of Figure 2 shows how these two waveform features ($m_b - M_s$ and Murphy's adjusted L_g) vary with a third feature (m_b). Note that in these co-plots a
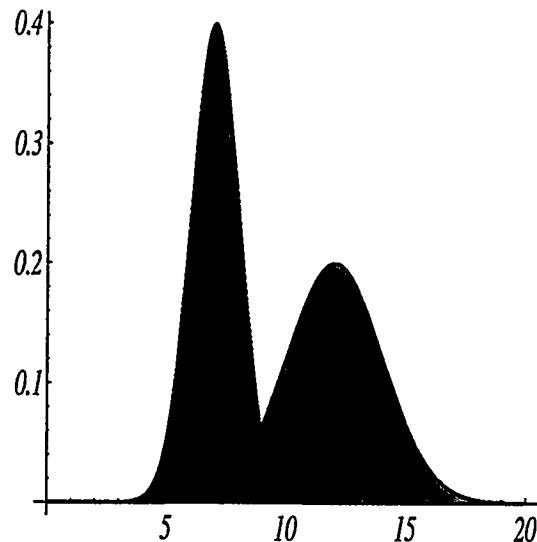
(a) Probability Models of a Single Discriminant for Two Types of Events. Distribution of "Red" on the left and "Green" on the right.
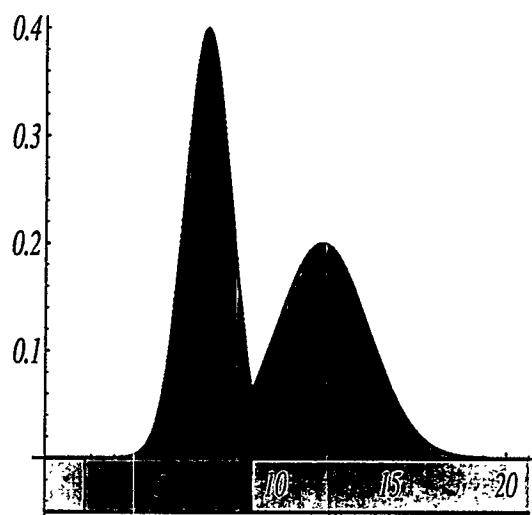
(b) Classification Region Constructed from the Ratio of the Probability Models (Likelihood Ratio).
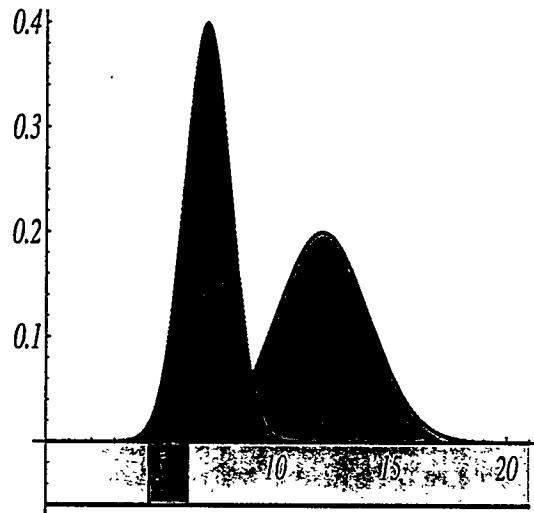
(c) Classification Region Constructed From the Combination of the Likelihood Ratio and Prior Probabilities.
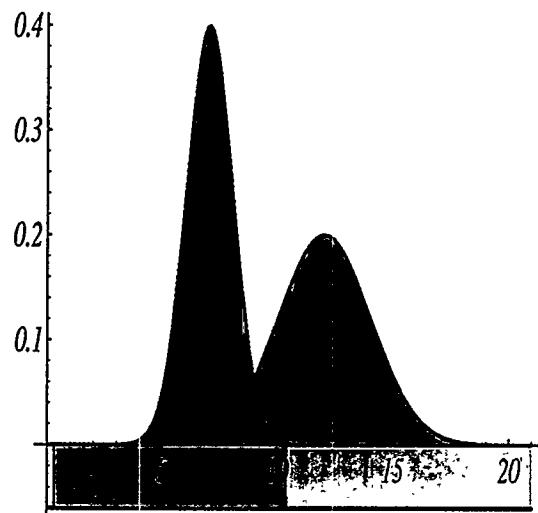
(d) Classification Region Constructed From the Combination of the Likelihood Ratio, Prior Probabilities, and Misclassification Costs.

Figure 1: Decision Theoretic Classification: An illustration of the effect that prior probabilities and misclassification costs have on a simple likelihood ratio classification region. The prior probability of the "Red" population is very small —1(c), and the misclassification cost of incorrectly predicting "Red" as "Green" is large — 1(d).

(a) m_b Between 1 and 3

(b) m_b Between 3 and 3.5

(c) m_b Between 3.5 and 4

(d) m_b Between 4 and 4.5
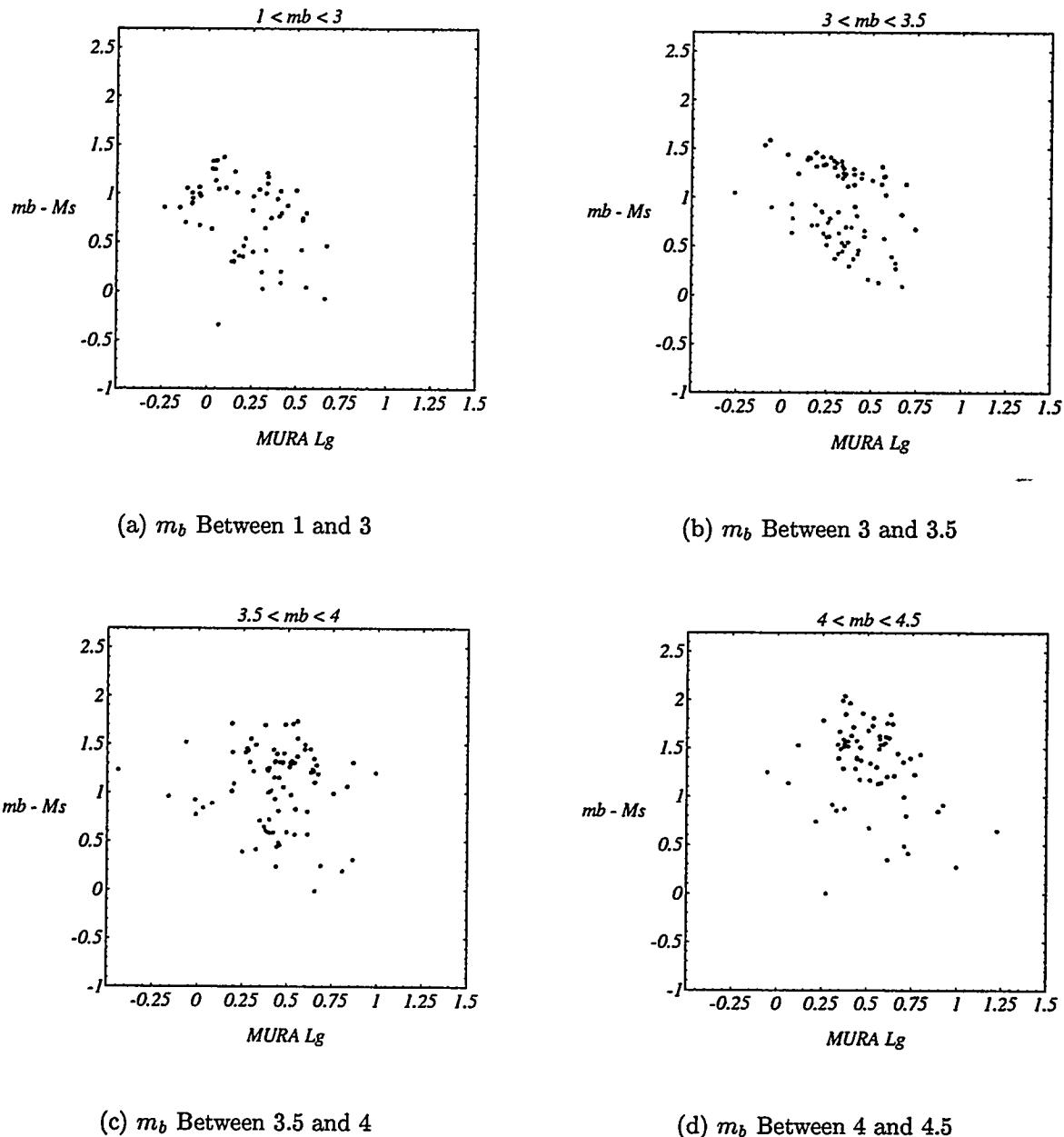
Figure 2: MURA L_g Versus $(m_b - M_s)$ for Various Values of m_b; Earthquakes = green, Nuclear Detonations = red

distinct separation emerges for events where $3 < m_b < 3.5$ (see Figure 2(b)).

Using the seven variables: S_n, m_b, $m_b - M_s$, $Log(SH)$, $\log(L_g/P_g)$, MURA L_g, and $\log(S_t) - 2m_b$ (see Table 1), both linear and quadratic discriminant analysis (LDA and QDA — see Section 4.1) were applied to illustrate the improved discriminative performance obtained as additional features are incorporated. Figure 3 shows box plots of the cross-validated misclassification rates for all possible two, three, four, and five variable combinations of the seven variables using LDA (Figure 3(a)) and QDA (Figure 3(b)). Figure 3 shows that the misclassification rates decrease as the number of terms in the model increases. The increase in discriminative ability as more features are added becomes less pronounced after the first few. In general, it has been shown that as the number of features increases, the discriminative ability generally also increases. However, after a certain point, adding more features can actually harm performance. This peaking phenomenon (not shown in Figure 3) is due to the classification algorithm fitting to random noise rather than to seismic structure, because of the finite number of training observations compared to the number of features incorporated. If the number of training observations were infinite, there would be no performance loss when adding additional features.

2.3   Missing Data in Discrimination

In the CTBT discrimination setting, missing data can cause formidable problems. Missing data can occur in three main ways as illustrated by the following three example cases. First, a station may be off line. Second, the energy signal may be clipped by the seismometer or recording system. Third, the missing data may be unmeasurable, e.g., surface wave seismic signals for extremely deep earthquakes. Seismic measurements can be missing from the regional training samples used to build and test the discrimination rules, as well as at the operational stage when unknown events are classified. In both cases, consideration must be given to how best to proceed. In this section, we discuss the impact that missing data has on the training and on discrimination performance evaluation.

Some discrimination algorithms can be trained even if the regional training samples have missing data (Anderson, 1995). Other discrimination algorithms require complete data and hence the missing data must be filled in (as in the Taylor et al. (1989) study). Filling in missing data is complicated by the necessity to take into account the inherent uncertainty in the data. Failing to do so can lead to over-optimistic evaluation of the discriminative ability of the regional monitoring system. One way to take account of the uncertainty in the missing data is through the use of multiple imputation methods (Rubin and Schenker, 1986).

Other complications arise when the data are not "missing at random." Consider the NTS discrimination data introduced in Section 2.2, which is the data from Taylor et al. (1989), augmented by data from 98 additional explosions. The numbers in Table 2 indicate that the missing data are not randomly missing. The diagonal elements of Table 2 represent the observed percentages of missing measurements on an individual measurement basis. The numbers in red (above the diagonal) represent the observed percentages of missing measurements on a joint (pairwise) basis. The green elements (below the diagonal) are the estimated percentages of measurements missing, assuming pair-wise independence. If the
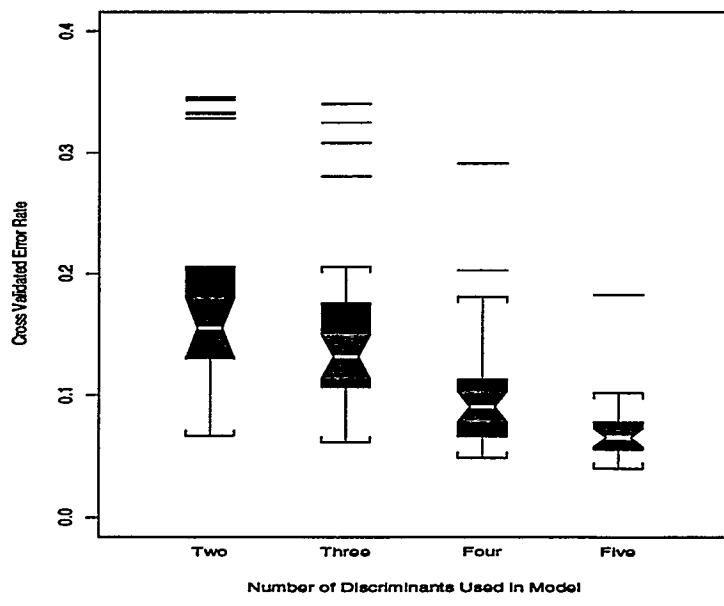
(a) LDA
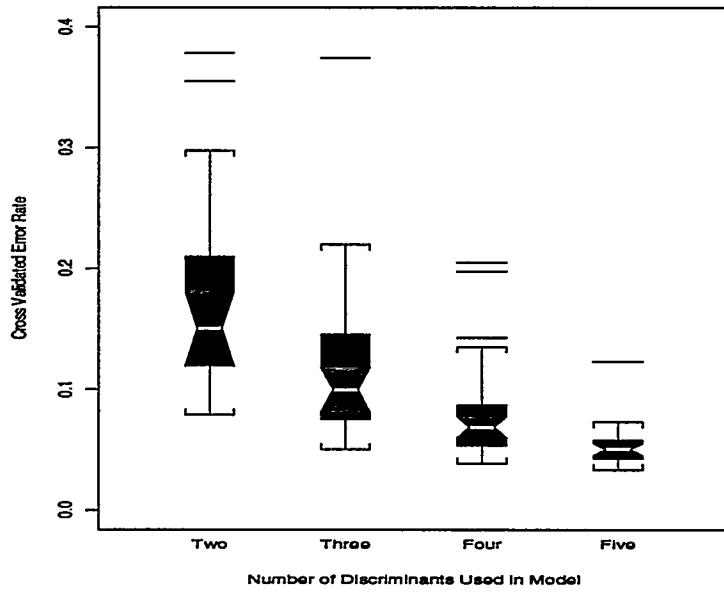
(b) QDA

Figure 3: Misclassification Rates For All Possible Combinations of 7 Discriminants Taken 2, 3, 4, and 5 at a Time Using LDA and QDA

Table 2: Missing at Random?

	S_n	m_b	$m_b - M_s$	$\log(SH)$	$\log(L_g/P_g)$	MURA L_g	$\log(S_t) - 2m_b$
S_n	**2.39**	2.39	2.39	1.08	2.39	2.39	2.39
m_b	0.06	**2.39**	2.39	1.08	2.39	2.39	2.39
$m_b - M_s$	0.94	0.94	**39.48**	15.18	2.39	6.51	32.54
$\log(SH)$	0.96	0.96	15.93	**40.35**	1.08	3.04	40.35
$\log(L_g/P_g)$	0.06	0.06	0.94	0.96	**2.39**	2.39	2.39
MURA L_g	0.16	0.16	2.66	2.71	0.16	**6.72**	6.51
$\log(S_t) - 2m_b$	1.50	1.50	24.75	25.29	1.50	4.22	**62.69**

Observed % missing individually (diagonal)
Observed % missing jointly (upper triangle)
Estimated % missing jointly assuming pairwise
independence (lower triangle), e.g. $P_{ij} = P_{ii}P_{jj}/100$

data were truly missing at random, corresponding off-diagonal (red and green) percentages would be close in value. It is easy to understand why the pattern of missing observations shows dependence on the particular seismic measurements, since many of the measurements are related (if one is difficult to obtain then a related measurement will also be difficult to obtain).

Because the data may not be missing at random, discrimination results which require filling in of the missing data may not be valid. Reliable analysis requires an understanding of why the data is missing. A full discussion of the reasons for missing data in the seismic setting is beyond the scope of this report. However, the available data (with missing observations) is used here to demonstrate that failing to account for the uncertainty in filling in missing data leads to over-optimistic evaluation of discriminative ability. The well-known fact that apparent misclassification rates (obtained by assessing the discrimination algorithm on the training samples) are over-optimistic is also shown.

Two simulation studies were performed, which provide a worst-case and best-case scenario for filling in the missing data. The discrimination algorithm selected for the simulation studies is CART (see Section 4.5), but any other discrimination algorithm could have been used.

In the first simulation, illustrated in Figure 4, the data are randomly partitioned into 80% and 20% subsets by source. The 80% group is the training set and the 20% is the holdout prediction set. A robust Expectation Maximization (EM) algorithm (Anderson, 1995) is used separately on the earthquake and explosion training subsets to produce estimated means and covariance matrices for each source. These means and covariance matrices are used to calculate imputed values, which are then used to fill in the missing data in the 80% training set (Note that the EM algorithm uses source information, since the estimation procedure is

performed separately for earthquakes and explosions). One hundred iterations of the EM algorithm are used for each source subset to ensure convergence. CART is then used to build a classification tree based on the 80% EM-filled subset. We do not fill in the missing data in the 20% holdout group with imputed values from the EM algorithm. Since in a real-time setting one will not know the source for a new event, we "naively" fill in the missing values in the holdout 20% with marginal means calculated without regard to source. We next use the CART "EM-filled" tree to predict the source for the 20% holdout set. This entire process is repeated 100 times. Thus, we calculate 100 apparent misclassification rates for the training sets and 100 corresponding misclassification rates for the holdout sets.

In the second simulation, illustrated in Figure 5, we assume the most optimistic view of the real-time operational setting, that no missing data were encountered in the 20% holdout set. In this experiment, we calculate the EM algorithm means and covariance matrices (separately for each source), using 100 iterations to ensure convergence. We fill the holes in the entire data matrix using pseudo-random number generation from the appropriate multivariate normal distributions whose mean and covariance matrices were estimated from the source-specific EM algorithm run. Next, a random 80% training set and 20% holdout set are selected for each source. CART is used to build a classification tree on the training set, which is in turn used to predict the source for the holdout set. This process, from the calculation of the randomized imputed values to building a new classification tree and predicting on the 20% subset, is repeated 100 times. Thus, in the second simulation we also have 100 apparent (training set) misclassification rates and 100 corresponding misclassification rates for the holdout sets. We used a more realistic method to fill in missing values, in that we incorporated the variances as well as the means produced by the EM algorithm. By incorporating the variance in filling in the missing data in each simulation run (multiple imputation) we take into account the uncertainty inherent in the missing values.

The first box plot of Figure 6(a) illustrates the apparent misclassification rates obtained for the 100 simulations from the first (worst-case) simulation method. The second box plot of Figure 6(a) illustrates the misclassification rates for the holdout 20% for 100 simulations, also from the first (worst-case) simulation method. The box plots of Figure 6(b) show analogous results for the second (best-case) simulation method. For both simulations, the apparent misclassification rates are lower than the misclassification rates for the holdout sets. It is well-known that the misclassification rates for holdout sets provide more realistic assessment of the discriminative ability for future unknown events than the apparent error rate. The box plots show that the "naive" filling in of the holdout sets for the first simulation method produced increased variation in misclassification rates compared with the second simulation method, although the average misclassification rates were nearly the same.

As seen in Figure 6, the apparent misclassification rates from the first simulation are lower than those from the second simulation. The reason is that the first simulation does not take into account the uncertainty associated with filling in missing values. The holdout misclassification rates should still be viewed with caution, however, because the missing values were not missing randomly.
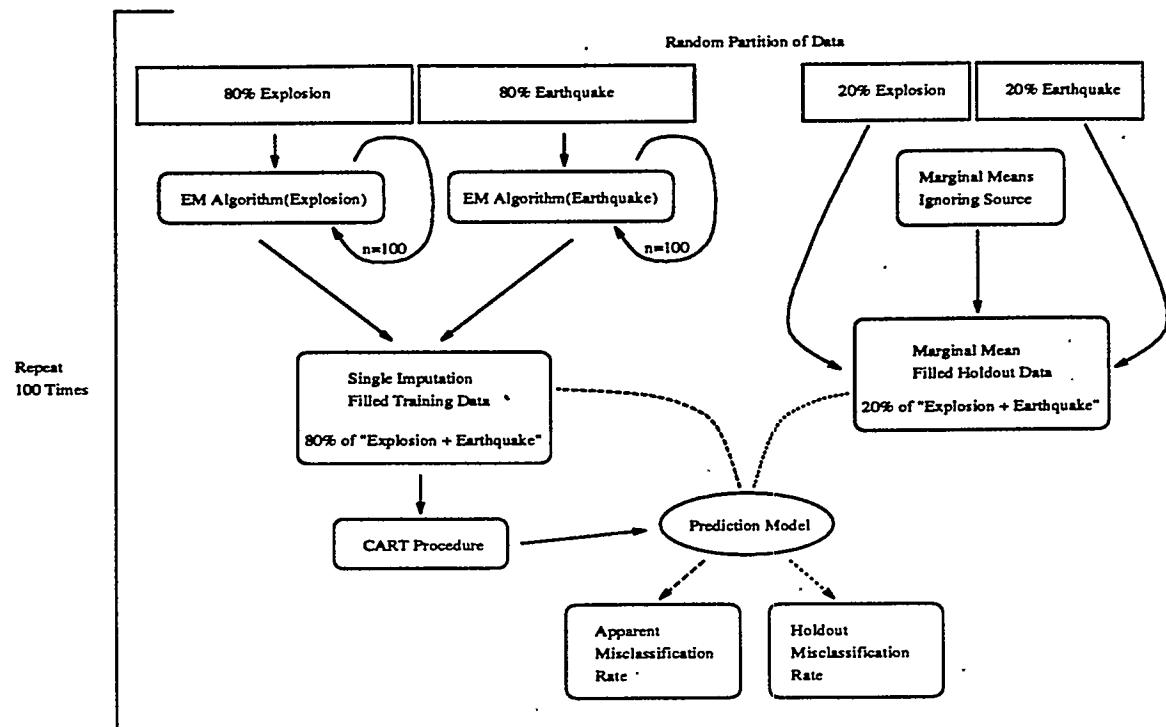
**Figure 4:** Single Imputation Simulation Flow Chart
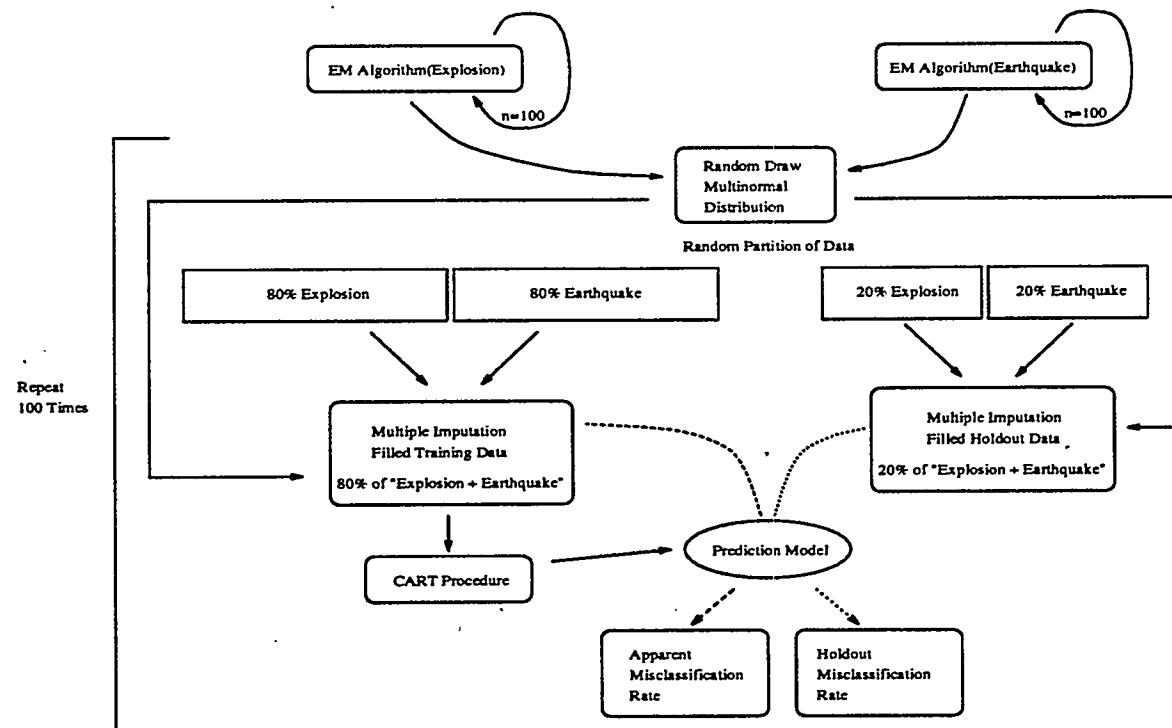
Figure 5: Multiple Imputation Simulation Flow Chart
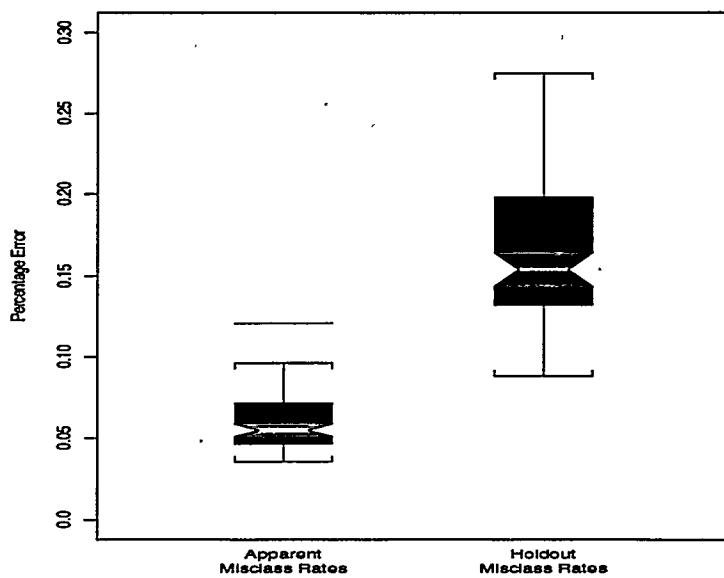
(a) Single imputation for training set. Marginal mean fill for holdout sample. No measurement uncertainty incorporated.
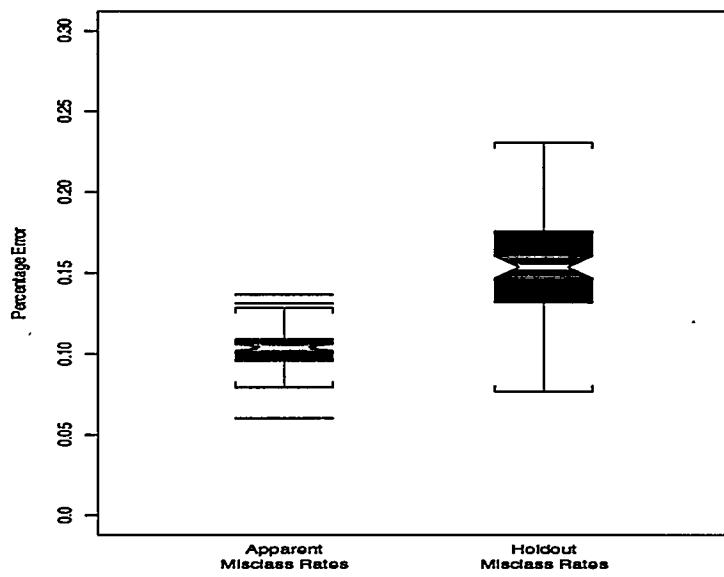
(b) Multiple imputation for training set and holdout sample. Measurement uncertainty incorporated.

Figure 6: Single and Multiple Imputation CART Study Using 100 Simulations

3   Classification Method Selection Criteria

No single classification method is best for all applications. According to Hand (1981):

> In any particular application the practical differences rather than the theoretical differences are the determining factors in choice of method, and which practical differences are important depends, unfortunately, on the problem itself. As is often the case, there is no best method.

There are many issues in selecting a classification method for use in the CTBT regional discrimination setting. In addition to the misclassification rate (i.e., the percent of observations incorrectly classified) for the method, there are many other criteria to be considered when selecting a classification method. These criteria can be loosely grouped into the categories of simplicity, robustness, applicability, and performance. Some of the criteria outlined in this section could logically be grouped under more than one of the four categories, but are included only once under the most appropriate category. The full impact of each individual selection criterion can only be seen in context with all of the other issues. Several of the criteria are also related (e.g., for many classification methods, speed and memory use are related — often inversely related). Also, the criteria are not of equal importance to all applications. No single classification method completely satisfies all of the criteria. For each issue, a short description is provided to place the criterion in the proper context.

3.1   Simplicity

Several classification method selection criteria fall under the general heading of simplicity. This applies both to the algorithm itself and to the results of applying the algorithm. On this general subject of comprehensibility (simplicity) of the classification rule the following example is cited in Michie et al. (1994):

> ... It is important also, that human operators believe the system. An oft-quoted example is the Three-Mile Island case, where the automatic devices correctly recommended a shutdown, but this recommendation was not acted upon by the human operators who did not believe that the recommendation was well founded.
> ...

The classification method and resulting classifications in a CTBT setting need to be understood to be accepted. The simplicity criteria are outlined below.

- Easy to Understand — The adopted CTBT classification method should be sophisticated but lucid.

- Easy to Use — The method should be easy to integrate into the CTBT monitoring environment.

- Easy to Interpret Results — The method should give easily interpretable predictions of an unknown event, i.e., probabilities, conditional probabilities, or a class membership score.

- Sequential Decision Rules — The method should not require computation of complex seismic measurements except when necessary for a decision.

- Number of statistical parameters — The method should employ a reasonably small number of statistical parameters.

- Automatic parameter selection — The method should have the ability to automatically estimate or select the values of the model parameters used by the classification method.

- Parsimonious model — The method should not have an excessively complicated structure.

3.2 Robustness

Issues of robustness are important in selecting a classification method. The robustness criteria deal with the appropriate handling of situations that are not ideal. These non-ideal situations are often present in real world applications (such as the CTBT) and the selected classification method must be able to adequately handle them. The robustness criteria are outlined below.

- Physical basis for assumptions — The adopted CTBT classification method should be based on assumptions that agree with physical laws.

- Models structure, not noise (sampling variability) — The method should not over fit a training sample (i.e. it should fit the real structure instead of an apparent structure).

- Handles training outliers — The method should not be excessively influenced by outliers in a training sample.

- Stable for small changes in training data — The classification regions should be reasonably stable with minor additions or deletions to the training data and/or small random changes to the training data.

- Accommodates multi-modal sources — The method should have the ability to accurately differentiate between sources characterized by seismic measurement data with two or more modes (i.e., data with two or more local distribution peaks or density maxima).

- Handles data transformations — The method should not be overly sensitive to mathematical transformations (as may be required for statistical or seismic analysis) of the training data.

- Handles small group sizes — In some regional CTBT settings few training measurements may be available for the explosion group. The classification method should accommodate this.

- Handles small training size — When placed in a regional setting, all groups in the initial training data may be small.

- Handles missing data — In a CTBT setting, it is very probable that not all seismic measurements will be seen by each monitoring station. The method should allow incorporating missing values into the algorithm.

3.3   Applicability

Particular classification methods are more appropriate than others, depending on the special needs of the application. The applicability criterion outlined below detail some of the needs that should be examined for the CTBT discrimination setting.

- Allows prior probabilities — In a CTBT setting, the number of nuclear weapon tests relative to earthquakes and commercial explosions will be very small. The method should have the ability to incorporate this prior information.

- Accounts for misclassification costs — The *cost* of erroneously classifying a nuclear weapon test as an earthquake may be considerably different than the *cost* of classifying an earthquake as a nuclear weapon test. The method should be able to incorporate the costs of errors into a decision process.

- Can handle many discriminants — The method should be able to synergistically combine several seismic measurements into a decision rule.

- Select or weight discriminants — The method should be able to identify or weight the discriminants that best separate the groups.

- Discrete and continuous discriminants in combination — Polarity of first motion is discrete and $m_b - M_s$ is continuous. The classification method should be applicable to both types of measurement.

3.4   Performance

While the misclassification rate is a very important performance issue, there are several other performance issues that should be considered in selecting a classification method. In many real world settings several classification methods will have similar misclassification rates, but will differ in other important ways. Performance criteria to be considered when selecting a classification method for CTBT discrimination are outlined below.

- Results stable over long term — The classification regions should not change drastically as new data is incorporated in the training sample.

- Results accurate over long term — The accuracy of the statistical classification method should not degrade as new data is incorporated in a training sample.

- Provides probability or ranking of source membership — The method should provide a numerical likelihood that an event was generated by a particular source. Such a likelihood can also serve as a tool to rank or prioritize events for further in-depth analysis.

- Amenable to uncertainty analysis — The uncertainties in the classification regions should be calculable (e.g., when the probability that an event was generated by a particular source is estimated to be 0.85, can a confidence interval on this probability be calculated?).

- Detects outliers in new data — The method should flag events that are not similar to any group. Most statistical classification methods are capable of outlier detection (Note: One possible classification approach involves characterizing the probability structure of seismic measurements from only earthquakes. Events that are inconsistent with an earthquake probability model are declared to be outliers. If the classification scheme is based completely on detecting outliers in this manner, then the probability of incorrectly declaring an explosion to be an earthquake cannot be assessed).

- Speed requirements — The method should perform classifications in near real time. In an operational setting, algorithm speed is very important. With regard to this issue, emphasis should be placed on algorithm speed rather than on machine speed.

- Memory requirements — The classification method should not require excessive memory. Some statistical classification methods are memory intensive.

4 Statistical Classification Methods

Eight statistical classification methods are outlined in this section. The eight methods have been chosen to represent a wide range of the numerous methods that are available in the statistics community. Section 2.1 described the three components (prior probability of group membership, cost of misclassification, and likelihood/probability estimation) that are the key ingredients for classification decision making. The discussion in this section is focused on likelihood/probability estimation. The incorporation of prior probabilities and costs can easily be incorporated with all eight methods. Section 4.1 outlines two popular normal distribution based discrimination methods: linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA). Two recent variations of the standard LDA and QDA are presented in Section 4.2. These two methods are variably regularized discriminant analysis (VRDA) and flexible discriminant analysis (FDA). Another popular discrimination method, logistic discrimination is described in Section 4.3. Two non-parametric classification methods are presented in Section 4.4. The two methods are kernel discrimination and K-th nearest neighbor discrimination (KNN). A statistical tree based approach to classification is presented in Section 4.5. The description of each of these methods is provided as a review rather than a full technical exposition. References are given which provide a more detailed and technical explanation. The evaluation and comparison of these statistical methods in light of the classification method selection criteria presented in Section 3 is provided in Section 5.

The data presented in Figures 7 through 15 of this section are fictitious and consist of three groups in two dimensions. This data set is not based on any seismic data, but is helpful in understanding the kind of classification boundaries that are obtained with the different methods. Figures 7 through 15 act as visual icons to reinforce the basic concepts underlying the statistical classification methods. In Figures 7 through 15 the data for the

three groups are represented by circles, stars, and triangles. The classification regions are correspondingly represented by the blue, yellow, and pink shaded areas.

4.1   Linear and Quadratic Discrimination

One of the best known classification rules, linear discriminant analysis (LDA), is based on the assumption that the data are multivariate normal (Gaussian) with equal group covariance structures (McLachlan, 1992). The linear discriminant rule assigns a candidate event to the source whose mean is closest to the candidate event, using what is called the Mahalanobis distance measure. The squared Mahalanobis distance between a point $\mathbf{x}$ and $\bar{\mathbf{x}}_i$, the mean of group i, is

$$D_i^2 = (\mathbf{x} - \bar{\mathbf{x}}_i)' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_i) \tag{1}$$

where $\mathbf{S}$ is the covariance matrix pooled (appropriately averaged) across the groups. If there are two groups, explosions and earthquakes, then $i = x$ or $i = q$ respectively. If $S = I$ (the identity matrix) then Mahalanobis distance simplifies to simple Euclidean distance. LDA can also provide for estimation of the probabilities of group membership. Figure 7 illustrates the linear regions formed by applying a linear discrimination rule.
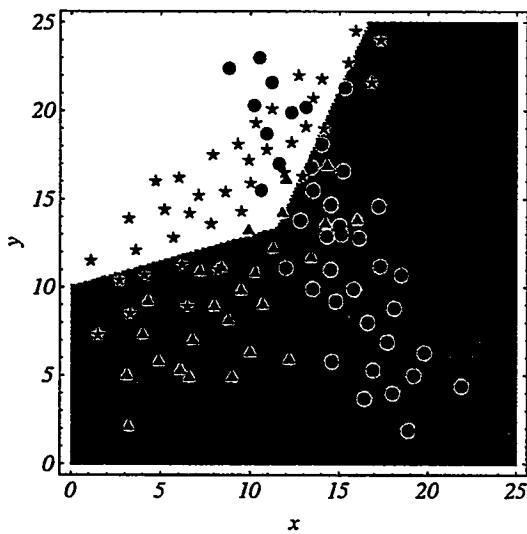
| Figure 7: LDA Regions | Figure 8: QDA Regions |
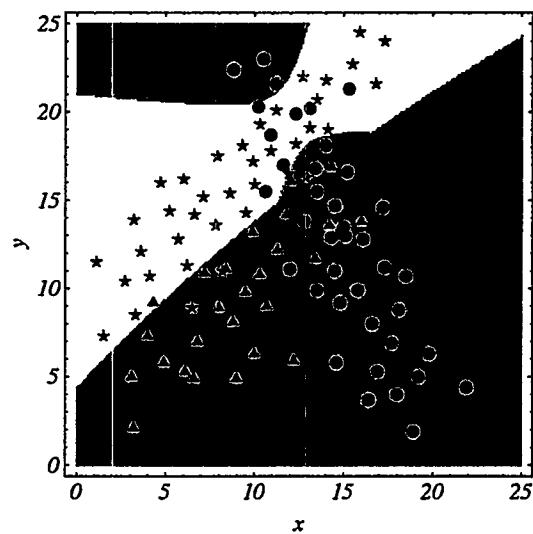
Quadratic discriminant analysis (QDA) is also based on the assumption that the data are multivariate normal. However, each group is allowed to have different covariance structures (McLachlan, 1992). A more general form of the Mahalanobis distance is used as the basis for QDA. The $\mathbf{S}$ used in LDA is replaced with $\mathbf{S}_i$, the covariance matrix for group i.

$$D_i^2 = (\mathbf{x} - \bar{\mathbf{x}}_i)' \mathbf{S}_i^{-1} (\mathbf{x} - \bar{\mathbf{x}}_i) \tag{2}$$

QDA probability formulas incorporate this distance measure and are used to classify observations. Figure 8 illustrates the regions with quadratic boundaries formed from the QDA discrimination rule.

4.2 Generalizations of LDA and QDA

Several generalizations of LDA and QDA have been created. Regularized discriminant analysis (RDA) provides a generalization of both LDA and QDA (Friedman, 1989; McLachlan, 1992). RDA uses the same general formulas as QDA, but replaces the S_i term with a covariance matrix that combines a weighted average of S_i and S along with an appropriately scaled multiple of the identity matrix. The best weighting and scaling (regularization) parameters are determined from the training data. Variably regularized discriminant analysis (VRDA) further generalizes RDA by allowing the weighting and regularization parameters to differ by group, based on the number of training samples for the group (Higbee, 1994). Once again, the best parameters are determined from the training data. Figure 9 illustrates the classification boundaries formed using VRDA.
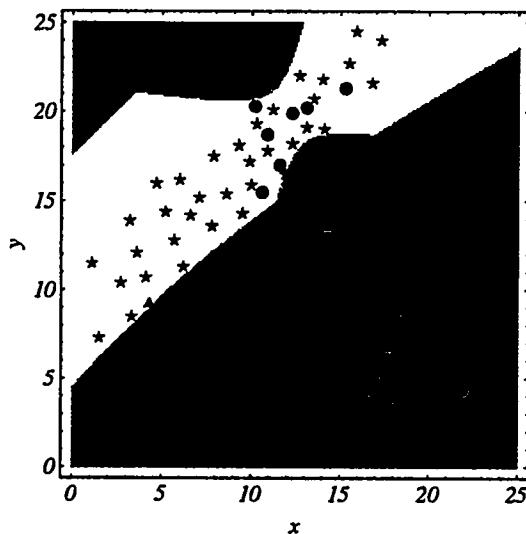
Figure 9: VRDA Regions   **Figure 10:** FDA Regions
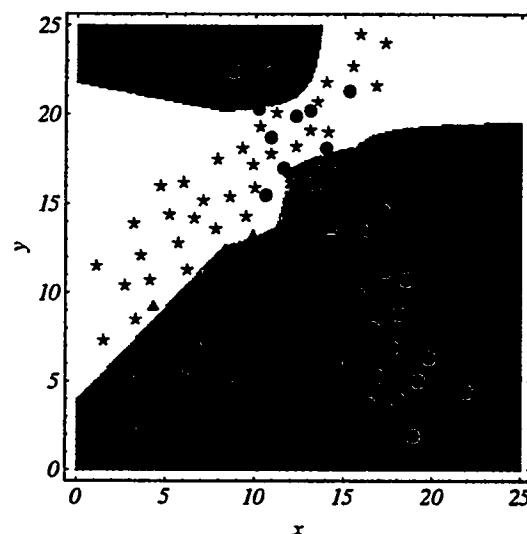
Flexible discriminant analysis (FDA) is a generalization of LDA (Hastie et al., 1994). FDA first reformulates LDA in a least squares linear regression framework and then substitutes non-parametric regression techniques in place of least squares regression to obtain new classification methods. A regression spline method such as multivariate adaptive regression splines (MARS) is one popular choice (Friedman, 1991) for use in the FDA algorithm. Figure 10 shows the classification regions with FDA using a MARS of degree two.

4.3 Logistic Discrimination

Logistic discrimination uses logistic regression to predict the probability that an event is an earthquake or explosion, conditional on a vector of seismic discriminants x. Specifically, the logistic regression model is

$$\Pr(\text{earthquake} \mid \mathbf{x}) \;=\; F(\beta'\mathbf{x}) \;=\; \frac{e^{\beta'\mathbf{x}}}{1 + e^{\beta'\mathbf{x}}} \tag{3}$$

$$\Pr(\text{explosion} \mid \mathbf{x}) \;=\; 1 - \Pr(\text{earthquake} \mid \mathbf{x})$$
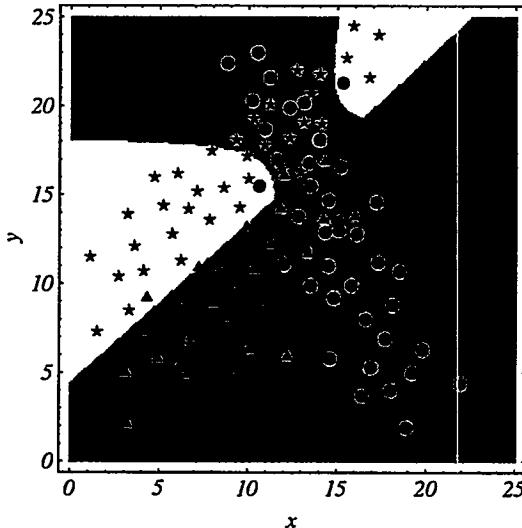
Figure 11: Logistic Discrimination Regions

where β is a vector of regression coefficients. Logistic regression can be viewed as tossing an earthquake/explosion coin where the probability of earthquake is $F(\beta'\mathbf{x})$. We classify an event with seismic measurements $\mathbf{x}$ as an earthquake if $\Pr(\text{earthquake} \mid \mathbf{x}) > \Pr(\text{explosion} \mid \mathbf{x})$, e.g., $\Pr(\text{earthquake} \mid \mathbf{x}) > 1/2$. A detailed discussion of the application of logistic discrimination can be found in McLachlan (1992). If we form a log–odds ratio, we have

$$log\frac{\Pr(\text{earthquake} \mid \mathbf{x})}{\Pr(\text{explosion} \mid \mathbf{x})} = \beta'\mathbf{x}, \tag{4}$$

which is the motivation for the term "logistic regression". The log–odds ratio between $\Pr(\text{earthquake} \mid \mathbf{x})$ and $\Pr(\text{explosion} \mid \mathbf{x})$ is a model reformulation known as a *logit* transformation. When generalized to more than two seismic sources, e.g., earthquake, single-point explosion, and ripple-fire explosion, these methods are known as logit analysis or logit discrimination (McLachlan, 1992). The parameters β are estimated using an iterative maximum likelihood procedure. Note that $\mathbf{x}$ can include multiple discriminants as well as interaction terms. Furthermore, categorical variables can be included since the model is formulated conditional on $\mathbf{x}$. An example is shown in Figure 11, again using simulated data.

4.4   Non-parametric Discrimination

Two well known non-parametric classification methods are kernel discrimination and K-th nearest neighbor (KNN) discrimination (Hand, 1981; Silverman, 1986; Scott, 1992; McLachlan, 1992). Kernel discrimination uses non-parametric estimates of the density (distribution) of each group (from the training data) to determine the likelihood or probability that an observation belongs to a particular group. Examples of non-parametric estimates of the density include the histogram and the kernel estimator. The kernel estimator can be viewed as an improved histogram. Instead of simple counts in non-overlapping bins (as in a histogram),

a kernel estimator uses a weighted count from a sliding window. The sliding window has two parameters: the kernel or shape (e.g., uniform, triangular, normal) that controls the weighting, and the kernel width which controls the length of the sliding window. Figure 12 shows decision boundaries based on a normal kernel estimator for the three groups of the example data. Note that the boundaries are highly nonlinear, offering greater flexibility. The kernel width is the major factor in determining the relative smoothness of the boundaries.
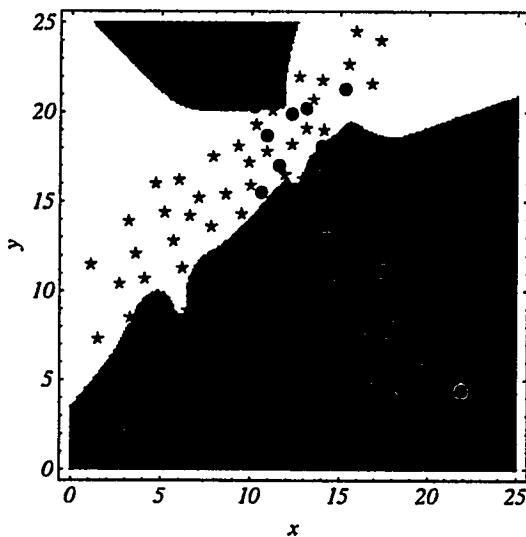

Figure 12: Kernel Discrimination Regions   **Figure 13:** KNN Discrimination Regions
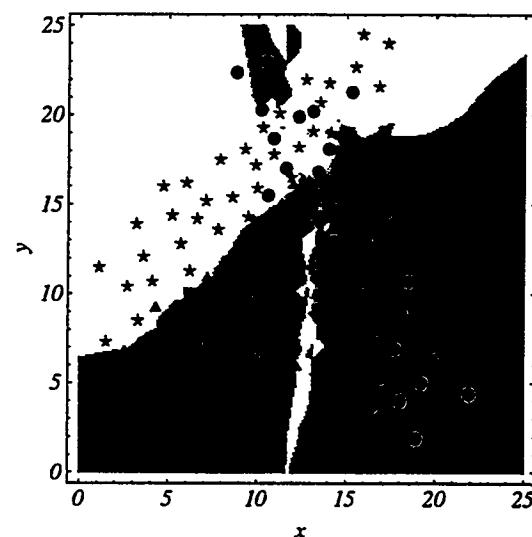
Figure 13 illustrates a discrimination method in which new events are classified according to a K-th nearest-neighbor (KNN) rule (McLachlan, 1992). With KNN, a candidate event is classified to the group which has the largest number of the nearest k points to the candidate event. The boundaries in this case are highly irregular, due to the lack of "smoothing". The decision boundary for the k-th nearest-neighbor rule will "smooth" as the value of k increases. The white region in Figure 13 is due to ties (equal number of circles and triangles in the k nearest neighbors).

4.5   Tree-Based Discrimination (CART)

The classification tree (Breiman et al., 1984) is a non-parametric method that seeks to partition a training sample of seismic discriminants into regions, each with a homogeneous event source. The end product of a classification tree is a collection of if–then questions (a decision tree) that can be applied to measured seismic discriminants. A classification tree can incorporate both continuous and categorical data. A classification tree is easily interpreted — an important feature when dealing with multi-dimensional discriminants with complex probability structures. Another important feature of a constructed classification tree is the ability to predict the source of a seismic event when some seismic measurements are not possible (surface-wave energy from a very deep earthquake) or simply missing (a broken seismometer).

A classification tree is constructed by recursively partitioning the discriminants in a training sample into regions that have progressively more homogeneous sources (classes). This tree-growing process tends to over fit to the training sample. To fit the structure of the training data, an overly complex tree is constructed first, and is then pruned using a measure that rewards predictive accuracy but penalizes large tree sizes (complexity). Pruning a tree essentially removes the least important splits or branches. From another viewpoint, the "accuracy versus complexity" measure used in pruning is designed to balance the homogeneity of the final regions and the complexity of the tree.
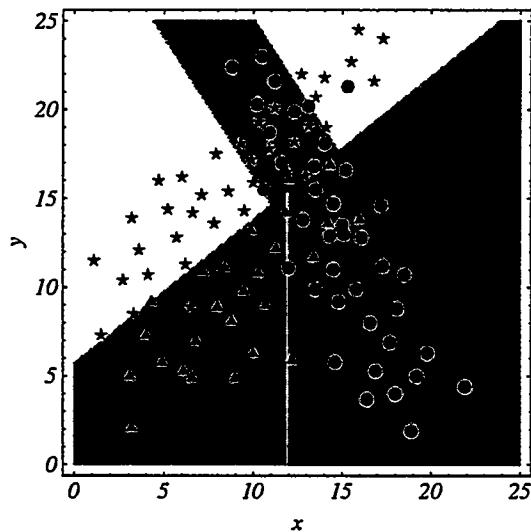
Figure 14: CART Regions   **Figure 15: CART Decision Tree**
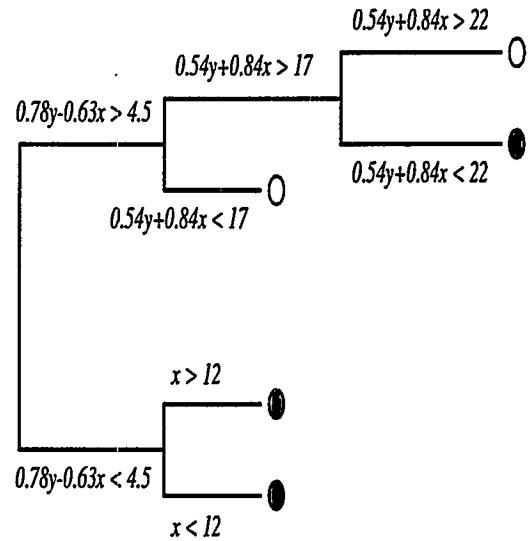
Figure 14 shows the classification regions resulting from the construction of a classification tree. Figure 15 is the associated decision tree. In order to predict whether an unknown observation belongs to the blue (circles), yellow (stars), or pink (triangle) group, one follows the path from the root of the decision tree to a leaf. For example, in Figure 15, if $.78y - .63x < 4.5$ and $x < 12$, then the event is classified as belonging to the pink (triangle) group.

5   Preliminary Evaluation of Classification Methods

The eight statistical classification methods that have been described in Section 4 (LDA, QDA, VRDA, FDA, logistic, KNN, kernel, and CART) can be compared and contrasted using the criteria outlined in Section 3. While all of the selection criteria should be considered, some criteria are of higher priority for the CTBT discrimination setting. Additionally, some important criteria such as "Amenable to Uncertainty Analysis" are reasonably satisfied by all eight statistical methods presented, and will not be considered further (though they should be considered carefully when looking at other possible methods such as neural net classifiers). Several major advantages and disadvantages (as viewed from a CTBT setting) for each of the eight methods are presented in Tables 3 through 10 on pages 21–23.

Table 3: LDA Advantages/Disadvantages

Advantages	Disadvantages
• The algorithm and the linear boundaries are simple to understand. • Small group sample sizes are handled (because of equal covariance assumption). • Robust to non-normality as long as distribution is symmetric and uni-modal. • Doesn't over fit to the training data. • All statistical parameters are automatically determined.	• Assumption of equal covariance and normality. • Missing values not handled within algorithm. • Sequential decision rules not provided.

Table 4: QDA Advantages/Disadvantages

Advantages	Disadvantages
• Different covariances allowed for each group. • Robust to non-normality as long as distribution is symmetric and uni-modal. • Doesn't over fit to the training data. • All statistical parameters are automatically determined.	• No small group sample sizes allowed. • Assumption of normality. • Missing values not handled within algorithm. • Sequential decision rules not provided.

Table 5: VRDA Advantages/Disadvantages

Advantages	Disadvantages
• Generalization of LDA and QDA that combines best properties of both. • Small group sample sizes are handled. • Robust to non-normality as long as distribution is symmetric and uni-modal. • Doesn't over fit to the training data. • All statistical parameters are automatically determined.	• Assumption of normality. • Missing values not handled within algorithm. • Sequential decision rules not provided.

Table 6: FDA Advantages/Disadvantages

Advantages	Disadvantages
• Generalization of LDA that relaxes the normality assumption. • Can handle multi-modal data. • Can be robust to training data outliers.	• Boundaries are not easily interpretable. • Statistical parameters not determined automatically. • Missing values not handled within algorithm. • Sequential decision rules not provided.

Table 7: Logistic Discrimination Advantages/Disadvantages

Advantages	Disadvantages
• The method is reasonably simple to understand. • Statistical parameters are automatically determined (however, model form is not). • Handles both discrete and continuous variables.	• Becomes unstable when groups are widely separated. • Must specify number of terms and which terms (linear, quadratic, etc.) in model. • Missing values not handled within algorithm. • Sequential decision rules not provided.

Table 8: Kernel Discrimination Advantages/Disadvantages

Advantages	Disadvantages
• No distributional assumption. • Can handle multi-modal data. • Can be robust to training data outliers.	• Need to select kernel shape and width. • The boundaries are not easily interpretable. • Missing values not handled within algorithm. • Sequential decision rules not provided.

Table 9: KNN Advantages/Disadvantages

Advantages	Disadvantages
• The algorithm is simple to understand. • No distributional assumption. • Can handle multi-modal data. • Can be robust to training data outliers.	• Need to select k. • The boundaries are not easily interpretable. • Missing values not handled within algorithm. • Sequential decision rules not provided.

Table 10: CART Advantages/Disadvantages

Advantages	Disadvantages
• The algorithm and the linear boundaries are simple to understand. • Handles both discrete and continuous variables. • Supports a sequential decision rule. • Missing values are handled within the algorithm. • Can handle multi-modal data.	• Can over-fit to the training data. • Need to specify parameters for pruning the tree.

A more concise summary of the capabilities of the eight classification methods in comparison to several important selection criteria is presented in Table 11. In Table 11, each criterion is worded so that yes is good and no is bad. The table entries are Y for yes, N for no, or blank. A blank indicates that either a definitive answer is unavailable, or an answer depends on the application. As seen in Table 11, no single method is ideal for every criterion. There is often a trade-off between the criteria. One example of such a trade-off can be seen by comparing the entries (Y and N) in Table 11 for the criteria of "No/Few Assumptions" and "Works Well With Small Group Size".

6   Conclusions

To verify compliance with the CTBT, seismic measures must be incorporated in an appropriate multivariate statistical classification framework. Numerous issues to consider in selecting a classification method are outlined in this report. The issues or selection criteria are grouped into the categories of simplicity, robustness, applicability, and performance. Some of the most notable issues for the CTBT setting include:

• Physical interpretability of the method and the classification decision boundaries.

Table 11: Classification Method Comparison

Selected Issues/Criteria	LDA	QDA	VRDA	FDA	Logistic	Kernel	KNN	CART
Allows Both Discrete & Continuous Data	N	N	N		Y			Y
Missing Data Handled Directly	N	N	N	N	N	N	N	Y
Easily Understood Algorithm	Y			N	Y		Y	Y
Easily Interpretable (Linear) Boundaries	Y	N	N	N	N	N	N	Y
Works Well With Small Group Size	Y	N	Y		N	N	N	N
No/Few Assumptions	N	N	N	Y	Y	Y	Y	Y
Sequential Decision Rule Possible	N	N	N	N	N	N	N	Y
Allows Multi-modal Groups	N	N	N	Y		Y	Y	Y
Automatic Parameter Selection	Y	Y	Y	N		N	N	N

All issues/criterion stated so that Y is good and N is bad.

A blank entry indicates one of the following:

    1) A middle answer between Y and N is appropriate.

    2) Either Y or N is appropriate depending on setting or parameters.

    3) A definitive answer is unavailable.

- Proper handling of missing measurements.

- Incorporation of both discrete and continuous measurements.

- Classification accuracy.

The data from Taylor et al. (1989), augmented with 98 additional explosions, is used to illustrate the need for an appropriate multivariate approach that incorporates both seismic knowledge and statistical techniques in determining the appropriate seismic measures to use in classifying the source of seismic disturbances. The data is also used to show the importance of proper handling of missing observations. A classification method must either directly handle missing observations or appropriately compensate for their absence. The method of multiple imputation for incorporating the uncertainty due to missing observations was shown to help avoid an overly optimistic estimation of the classification error rate. Apparent error rates (those obtained by training and then testing on the same data) were shown to be poor estimators of the true classification error rate, as compared to cross-validated (hold one out) error rates.

Eight statistical classification methods (LDA, QDA, VRDA, FDA, Logistic, KNN, Kernel, and CART), representing a wide range of possible approaches, are explained and discussed in relation to the selection criteria outlined in this report. A summary of several important selection criteria for each of the eight methods reveals that no one method is best on all issues. However, the CART classification approach appears to satisfy many important issues with regard to the CTBT setting. Further analysis, based on more realistic operational data, is currently underway to better assess the important issue of classification accuracy for

the various methods. After this analysis is completed, a recommendation for an appropriate CTBT classification framework will be made, based on the analysis and the conclusions from this report.

The evaluation of statistical classification methods in this report not only helps in determining an appropriate CTBT operational classification algorithm, but is also helpful in determining which classification methods to include in an "analyst tool box" to be used to understand new regions and to examine seismic events of interest. Examining the results from several different classification algorithms and understanding the strengths and weaknesses of each can help in better understanding and characterizing events of interest.

References

Anderson, D. N., Hagedorn, D. N., Higbee, K. T., Miller, N. E., Redgate, T., and Rohay, A. C. (1995). Statistical classification methods for seismic discrimination. Poster presentation at the Fall American Geophysical Meeting, San Francisco, CA.

Anderson, K. K. (1995). Assessing regional discriminants with incomplete data. Technical Report PNL-10804, Pacific Northwest National Laboratory, Richland, WA.

Blandford, R. R. (1995). Regional seismic event discrimination, monitoring a comprehensive test ban treaty. NATO Advanced Studies Institute Series E, NATO.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and Regression Trees*. Chapman & Hall, New York, NY.

DOE (1994). Comprehensive test ban treaty research and development FY95-96 program plan. Technical Report DOE/NN-0003, United States Department of Energy.

Friedman, J. H. (1989). Regularized discriminant analysis. *Journal of the American Statistical Association*, 84(405):165–175.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *Annals of Statistics*, 19:1–66.

Hand, D. J. (1981). *Discrimination and Classification*. John Wiley & Sons, New York, NY.

Hastie, T., Tibshirani, R., and Buja, A. (1994). Flexible discriminant analysis by optimal scoring. *Journal of the American Statistical Association*, 89(428):1255–1270.

Higbee, K. T. (1994). Variably regularized discriminant analysis. In *American Statistical Association 1994 Proceedings of the Statistical Computing Section*, pages 263–268, Alexandria, Virgina. American Statistical Association.

McLachlan, G. J. (1992). *Discriminant Analysis and Statiscial Pattern Recognition*. John Wiley & Sons, Inc., New York, NY.

Michie, D., Spiegelhalter, D. J., and Taylor, C. C., editors (1994). *Machine Learning, Neural and Statistical Classification*. Ellis Horwood, New York, NY.

Rubin, D. B. and Schenker, N. (1986). Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statistical Association*, 81(394):366–374.

Scott, D. W. (1992). *Multivariate Density Estimation*. John Wiley & Sons, New York, NY.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, New York, NY.

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.

U.S. Congress (1988). Seismic verification of nuclear testing treaties. Office of Technology Assessment OTA-ISC-361, U.S. Congress.
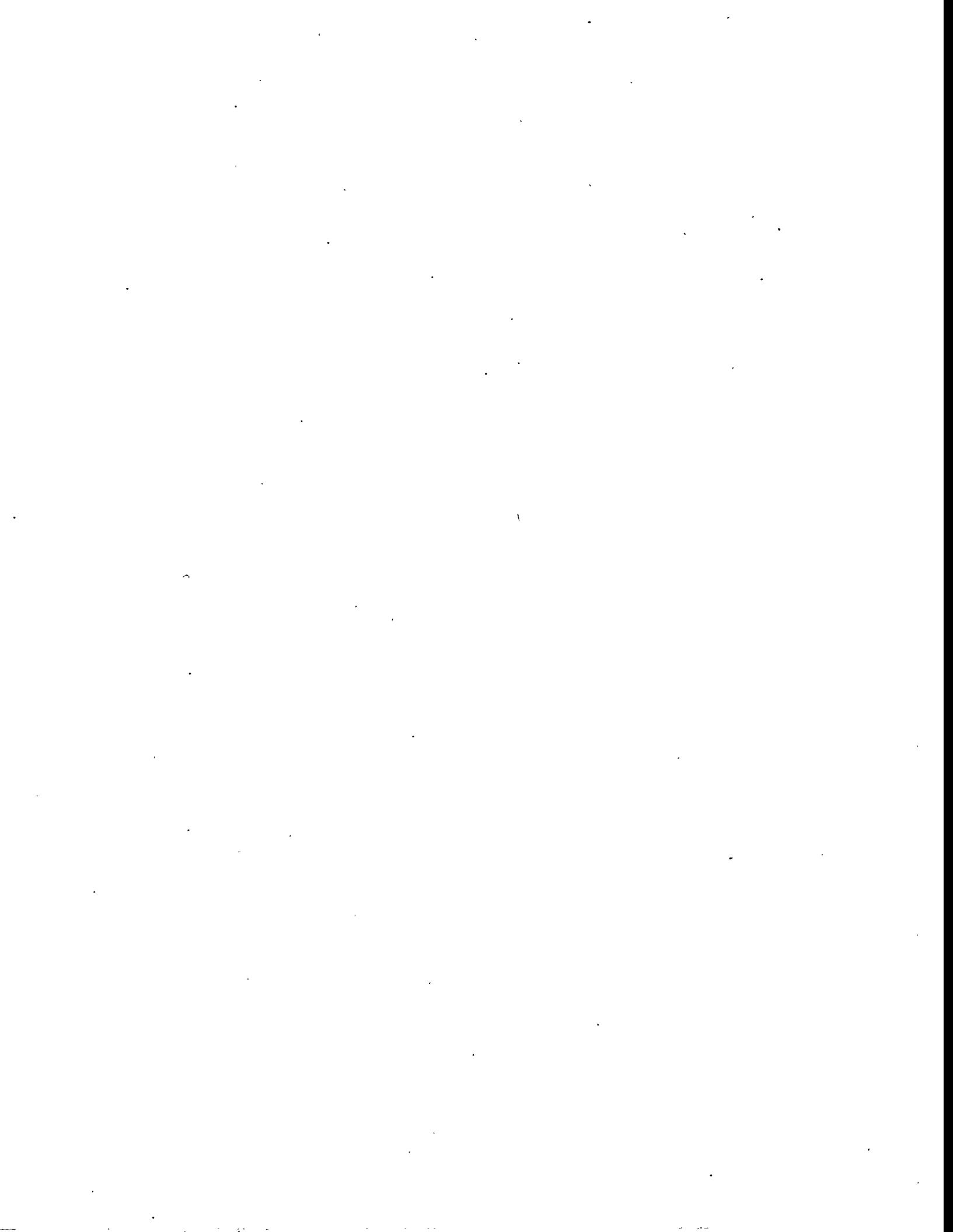
Distribution

**No. of
Copies**

**No. of
Copies**

Offsite

7 Air Force Technical Application Center
   1030 South Highway A1A
   Patrick AFB, FL 32925-3002

     N.C. Anderholm, DOB
     E. Herron, TTD
     V. Hsu, TTR
     C. McBrearty, TN
     F.F. Pilotte, TT
     D.R. Russell, TTR
     R. Schult, TTR

1 Air Force Office of Scientific Research
   110 Duncan Avenue
   Suite B115
   Bolling AFB, DC 20332-0001

     S.K. Dickinson, NL

5 Arms Control & Disarmament Agency
   3020 21st St., NW
   Room 5499
   Washington, DC 20451

     R. Cockerham
     P. Cordon
     S. Day
     O.J. Sheaks
     L. Tanzini

1 Arms Control Intelligence Staff
   NPIC
   P.O. Box 70967, Southwest Station
   Washington, DC 20024-0967

     M. Begley

1 Arms Control Intelligence Staff
   ACIS/TMG/NTT, Rm 6T11 NHB
   Washington, DC 20505

     J. Filson

1 Arms Control Intelligence Staff
   DCI/ACIS
   Washington, DC 20505

     L. Turnbull

1 Center for Monitoring Research
   AFTAC/TT/CMR
   1300 N. 17th St., Suite 1450
   Arlington, VA 22209-2308

     R.R. Blanford

1 Defense Nuclear Agency
   6801 Telegraph Road
   Alexandria, VA 22310

     D.A. Linger

1 FCDNA/FCTTS
   1680 Texas Street SE
   Kirtland AFB, NM 87117-5669

     R.E. Reinke

2 Maxwell Laboratories
   S Cubed Division
   P.O. Box 1620
   La Jolla, CA 92038-1620

     T.G. Barker
     K.L. McLaughlin

No. of Copies

2 Nuclear Treaty Programs Office
1901 N. Moore St., Suite 609
Arlington, VA 22209

    R. Alewine
    S. Bratt

6 Phillips Laboratory
Geophysics Branch
Earth Sciences Div., PL/GPE
29 Randolph Road
Hanscom AFB, MA 01731-3010

    A. Dainty
    D.G. Harkrider
    R.S. Jih
    K. Kadinsky-Cade
    J.F. Lewkowicz
    D.T. Reiter

2 U.S. Department of Energy
Forrestal Building
1000 Independence Avenue, SW
Washington, DC 20585

    L.A. Casey, NN-20
    L. Evanson, NN-40.2A

1 U.S. Department of the Interior
U.S. Geological Survey
Reston, VA 22092

    W.S. Leith, MS 928

14 Lawrence Livermore National Laboratory
P.O. Box 808
Livermore, CA 94551

    M.D. Denny, MS L-205
    W. Dunlop, MS L-200
    L.A. Glenn, MS L-200
    P. Goldstein, MS L-205
    W.J. Hannon Jr, MS L-205
    D.B. Harris, MS L-205
    K.M. Mayeda, MS L-205

No. of Copies

    K.K. Nakanishi, MS L-205
    H.J. Patton, MS L-205
    A.S. Ryall, MS L-205
    A.T. Smith, MS L-205
    P. Sokkappa, MS L-195
    W.R. Walter, MS L-205
    J.J. Zucca, MS L-205

9 Los Alamos National Laboratory
P.O. Box 1663
Los Alamos, NM 87545

    F.N. App, MS F659
    W.M. Brunish, MS F659
    A.H. Cogbill, MS F659
    M. Hodgson, MS D460
    D.C. Pearson, MS C335
    G.E. Randall, MS C335
    B.W. Stump, MS C335
    S.R. Taylor, MS C335
    T.A. Weaver, MS C335

8 Sandia National Laboratory
P.O. Box 5800
Albuquerque, NM 87185

    D.R. Breding, MS 0655
    E.P. Chael, MS 0655
    J.P. Claassen, MS 0655
    P.B. Herrington, MS 0655
    R.G. Keyser, MS 0979
    J.R. Trujillo, MS 1138
    L.S. Walker, MS 0979
    C.J. Young, MS 0750

No. of
Copies

Onsite

38 Pacific Northwest National Laboratory
P.O. Box 999
Richland, WA 99352

D.N. Anderson, K5-12
K.K. Anderson, K5-12
D.M. Boyd, K7-02
D.L. Eggett, K5-12
T.R. Fox, K6-48
D.N. Hagedorn, K5-12 (3)
K.T. Higbee, K5-12 (5)
E.F. Mercer, K6-40 (18)
N.E. Miller, K5-12
T. Redgate, K5-12
B.A. Pulsipher, K5-12
A.C. Rohay, K9-48
F.M. Ryan, K7-70
R.A. Warner, K6-48
Information Release, K1-06