

LA-UR- 96-3533

Title:

AN EVALUATION OF GENERALIZED LIKELIHOOD RATIO OUTLIER DETECTION TO IDENTIFICATION OF SEISMIC EVENTS IN WESTERN CHINA

Author(s):

Steven R. Taylor, EES-3
Hans E. Hartse, EES-3

Submitted to:

DOE DELIVERABLE

RECEIVED

DEC 26 1996

OSTI

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER



Los Alamos
NATIONAL LABORATORY

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. The Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy.

Form No. 836 R5
ST 2629 10/91

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

**Portions of this document may be illegible
in electronic image products. Images are
produced from the best available original
document.**

An Evaluation of Generalized Likelihood Ratio Outlier Detection to Identification of Seismic Events in Western China

Steven R. Taylor and Hans E. Hartse

September 24, 1996

LAUR-96-XXX

CTBT R&D Program, ST482A

Geophysics Group, EES-3

Los Alamos National Laboratory

Los Alamos, NM 87545

Abstract

The Generalized Likelihood Ratio Outlier Detection Technique for seismic event identification is evaluated using synthetic test data and frequency-dependent P_g/L_g measurements from western China. For most seismic stations that are to be part of the proposed International Monitoring System for the Comprehensive Test Ban Treaty, there will be few or no nuclear explosions in the magnitude range of interest (e.g. $m_b < 4$) on which to base an event-identification system using traditional classification techniques. Outlier detection is a reasonable alternative approach to the seismic discrimination problem when no calibration explosions are available. Distance-corrected P_g/L_g data in seven different frequency bands ranging from 0.5 to 8 Hz from the Chinese Digital Seismic Station WMQ are used to evaluate the technique. The data are collected from 157 known earthquakes, 215 unknown events (presumed earthquakes and possibly some industrial explosions), and 18 known nuclear explosions (1 from the Chinese Lop Nor test site and 17 from the East Kazakh test site). A feature selection technique is used to find the best combination of discriminants to use for outlier detection. Good discrimination performance is found by combining a low-frequency (0.5 to 1 Hz) P_g/L_g ratio with high-frequency ratios (e.g. 2 to 4 and 4 to 8 Hz). Although the low-frequency ratio does not discriminate between earthquakes and nuclear explosions well by itself, it can be effectively combined with the high-frequency discriminants. Based on the tests with real and synthetic data, the outlier detection technique appears to be an effective approach to seismic monitoring in uncalibrated regions.

Introduction

Recent work has shown the effectiveness of various regional discriminants in western China such as P_g/L_g in different frequency bands (Hartse *et al.*, 1996). In the study of Hartse *et al.*, (1996), there was a limited dataset of known nuclear explosions, most of which were at relatively high magnitudes. Thus, it was possible to measure numerous seismic discriminants and qualitatively assess their efficacy by examining earthquake versus nuclear explosion separation on discrimination plots. For a majority of stations that are or will be part of the International Monitoring System (IMS) monitoring a possible future Comprehensive Test Ban Treaty (CTBT), there will be few or no

magnitudes (e.g. $m_b < 4$). This is in stark contrast to previous discrimination studies in the western United States which had numerous recordings of both earthquakes and NTS explosions (e.g. Taylor, 1996). Thus, the event identification task in new regions having little or no data from nuclear explosions becomes a problem in outlier detection rather than event classification. In new regions, the majority of well-recorded events may be drawn from a single population (e.g. earthquakes or industrial explosions). The problem will be to identify a clandestine nuclear explosion that hopefully will have different characteristics from the earthquakes and will therefore form a problem in outlier detection. The difference between event classification and outlier detection is illustrated schematically in Figure 1.

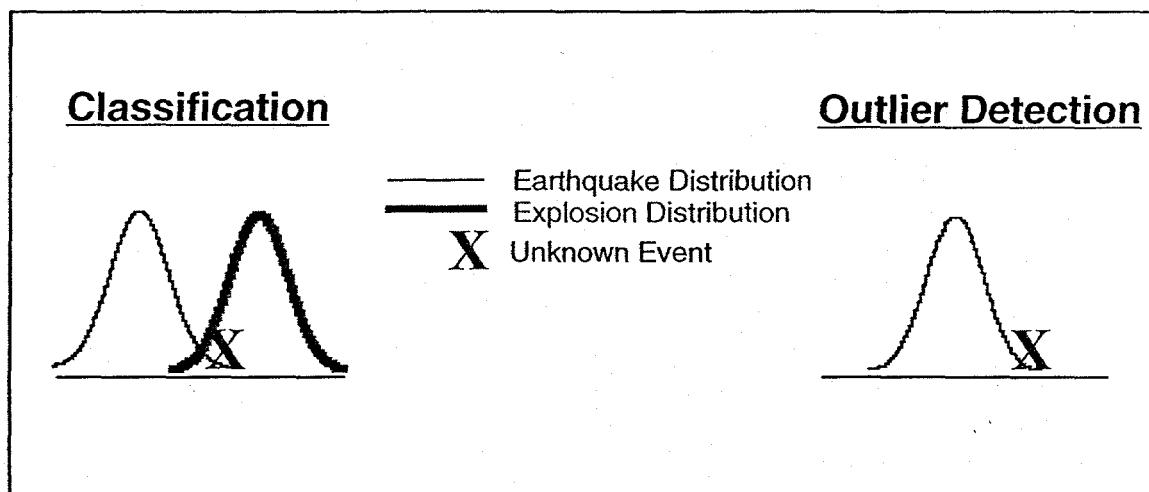


Figure 1. Schematic illustration showing difference between event classification and outlier detection. For event classification, the problem is to assign the unknown event "X" to either the earthquake or explosion population each having known characteristics. In a region having only data from known earthquakes, the problem of outlier detection is determining whether or not unknown event "X" belongs to the earthquake population or is from a different unknown population.

There are numerous multivariate event classification techniques ranging from traditional maximum-likelihood Gaussian classifiers (e.g. Hand, 1981) to neural networks (e.g. Haykin, 1993). The problem of outlier detection applied to seismic identification problems has recently been formulated by Fisk *et al.*, (1993) which is the technique that we follow in this report. We wish to test the outlier detection technique at a station having measurements from known nuclear explosions before applying it to stations having measurements from no known nuclear explosions. In this report, we first apply the outlier

detection technique to synthetic test data having known statistical characteristics. We then apply the technique to P_g/L_g measurements at WMQ. We also employ a feature selection technique based on a simple class separability metric in order to find an optimum subset of discriminants to use in the outlier detection.

Generalized Likelihood Outlier Detection

The details of the Generalized Likelihood Outlier Detection Technique are given in Fisk *et al.*, (1993; 1995) and Anderson (1984) and will be briefly reviewed in this section. In the discussion, we will assume that we have a large set of measurements from a population of Earthquakes (Q) and wish to use the outlier detection technique to identify suspect events that may be Nuclear Explosions (X). We assume for earthquake i , we have a set of discrimination measurements contained in vector \mathbf{v}_i , for N events from the earthquake population, π_Q , $[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N] \in \pi_Q$. We wish to test the hypothesis (H_0) that a new unknown event with measurement vector, \mathbf{v}_{N+1} belongs to π_Q versus the alternate hypothesis (H_1) that it does not (and hence belongs to some other unknown population). Conceptually, the hypothesis test is given by

versus

$$\begin{aligned} H_0: \quad & [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{N+1}] \in \pi_Q \\ H_1: \quad & [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N] \in \pi_Q, \mathbf{v}_{N+1} \notin \pi_Q \end{aligned} \quad (1)$$

and the Generalized Likelihood Ratio (GLR), λ , is given by

$$\lambda = \frac{L(H_0)}{L(H_1)}. \quad (2)$$

Where $L(H_0)$ is the likelihood of accepting the null hypothesis, H_0 . A large value for λ suggests that we accept the null hypothesis that the new event belongs to population Q . The main assumptions of the technique are that the data from each population are multivariate normal with the same covariance matrix.

The critical value for the GLR below which an event is declared an outlier is difficult to determine since its probability distribution is unknown (except for the case of normally and independently distributed measurements; Anderson, 1985). The critical value of the GLR, λ_α , below which an event is declared an outlier is therefore determined by the bootstrap technique at the α level of significance (e.g. Fisk *et al.*, 1993). Basically,

random samples are selected from the known population a number of times from which a histogram (probability distribution) of λ is constructed. From this, the cumulative probability distribution of λ can be used to find the critical value λ_α at the α significance level (false alarm rate).

Synthetic Test Case

In order to better understand the GLR Technique, we have conducted some simple test cases using synthetic data. We constructed a population of 500 "earthquakes" and 5 "explosions" from which 5 discriminants were generated and placed into vector \mathbf{v}_i for event i . The random data from each population are normally and independently distributed having equal variance for each discriminant. The only difference is that the mean value for each band is 0 for the "earthquakes" and progressively increases from 0 to 4 for the "explosions".

$$\begin{aligned}
 \mathbf{v}_i^{(j)} &= N\left(\mu_i^{(j)}, \sigma_i^{(j)}\right) \quad i = 1, 2, \dots, n^{(j)}; \quad j = X, Q \\
 \text{cov}(\mathbf{v}_i^{(j)}, \mathbf{v}_k^{(j)}) &= 0 \quad i \neq k \\
 \sigma_i^{(X)} &= \sigma_i^{(Q)} = 1 \\
 \mu^{(Q)} &= [0, 0, 0, 0, 0]; \quad n^{(Q)} = 500 \\
 \mu^{(X)} &= [0, 1, 2, 3, 4]; \quad n^{(X)} = 5
 \end{aligned} \tag{3}$$

The mean and 2-sigma values for each of the 5 discriminants are shown for the two populations in Figure 2. The separation between the "Explosions" and "Earthquakes" progressively increases from discriminant 1 to 5.

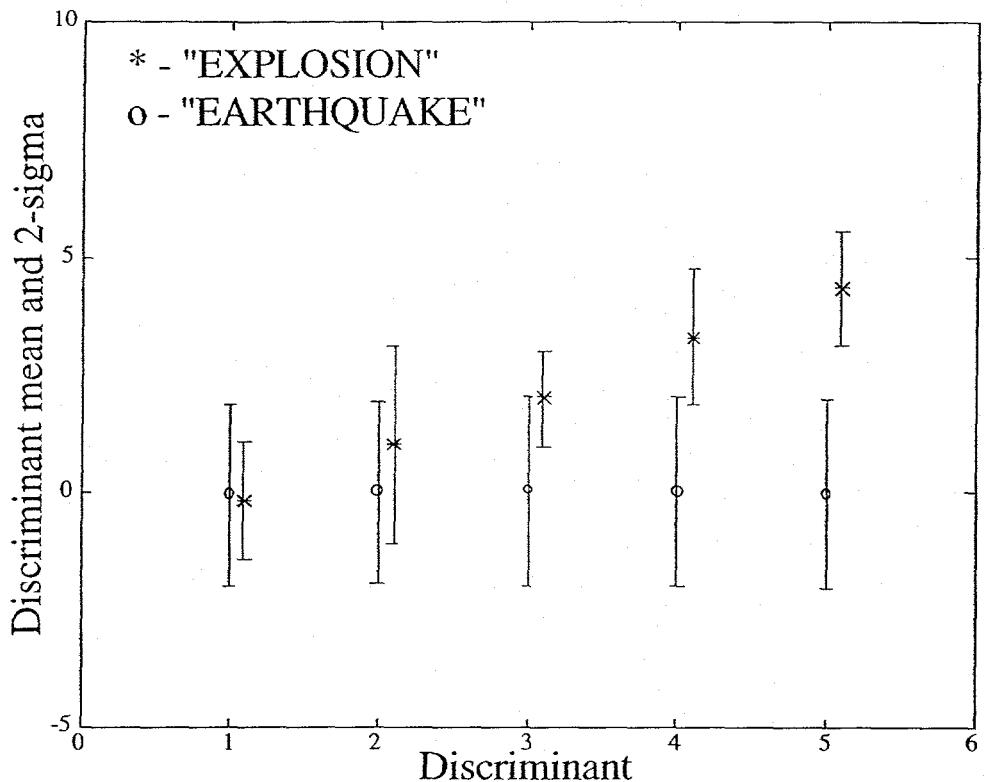


Figure 2. Illustration of mean and two sigma values for synthetic "Earthquake" and "Explosion" discriminant data. Note that error bars for the "Explosion" data deviate from the expected value of length 4 because of the small sample size (5).

As described above, the critical value for the GLR, λ_α , is computed using the bootstrap technique such that the false alarm rate is fixed. For the bootstrap example shown in Figure 3, we chose 100 random samples from the "earthquake" population and computed the GLR for each event using a leave-one-out procedure. This was performed 100 times resulting in 10000 GLR values whose distribution is shown in the histogram in the top portion of Figure 3. λ_α was then estimated from the cumulative frequency distribution shown in the lower portion of Figure 3 for $\alpha = 0.05$. Events with $\lambda < \lambda_\alpha$ are outliers and rejected as belonging to the "earthquake" population. α is the significance level of the test and $100\alpha\%$ of the "earthquakes" are falsely declared to be outliers. If a smaller significance level is selected, there will be fewer false alarms, but the likelihood of a missed violation is increased (i.e. falsely including an "explosion" in the "earthquake" population). In this case the critical value was estimated to be $\lambda_{0.05} = 2.7$.

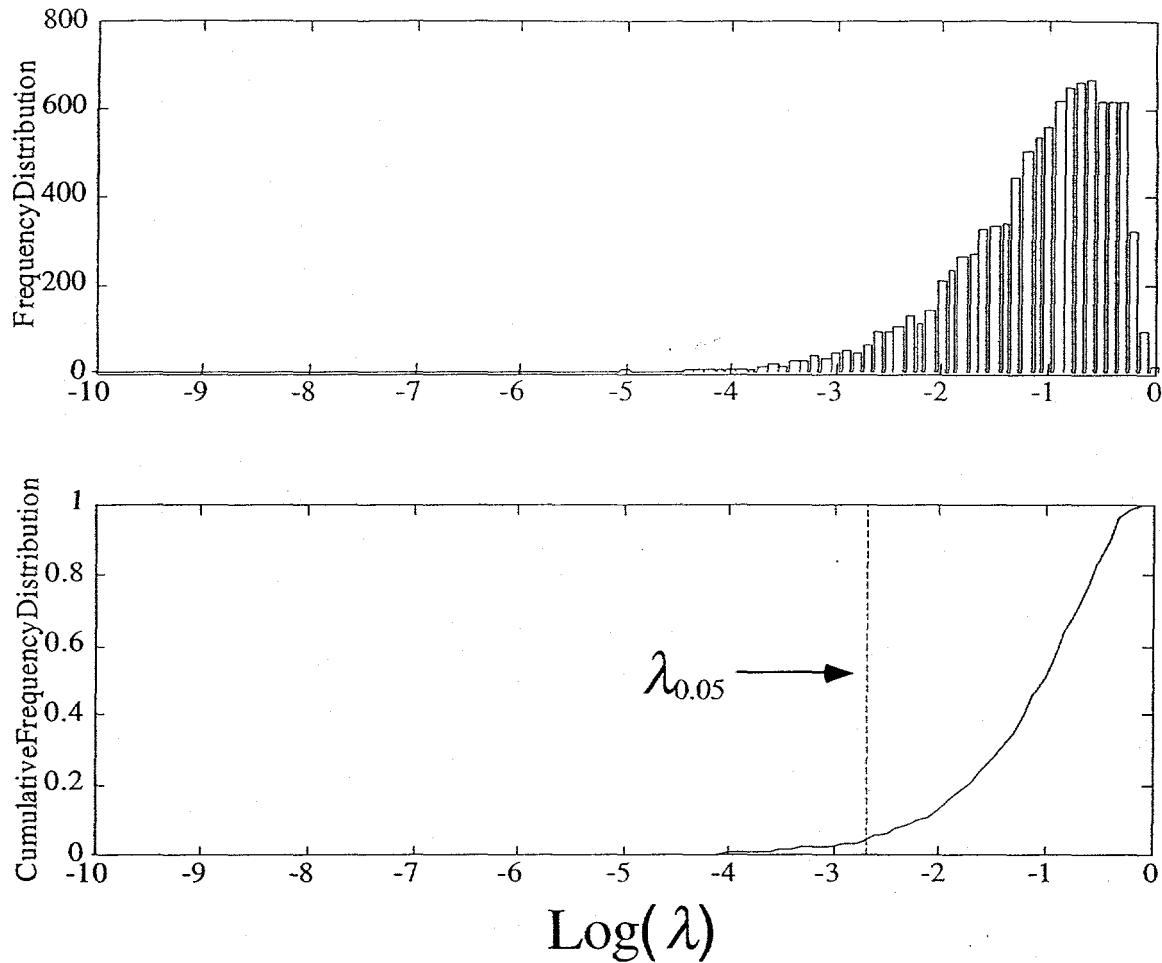


Figure 3. Histogram of the GLR, λ , for bootstrapped samples extracted from synthetic test data (top) and cumulative frequency distribution (bottom). Critical value for GLR is shown for $\alpha = 0.05$ significance level ($\lambda_{0.05} = -2.7$).

The GLR was then computed for the 500 “earthquakes” (using the leave-one-out technique) and the 5 “explosions”. The “explosions” were not included in the known set of “earthquakes”. The results are illustrated in the top portion of Figure 4, where the GLR is plotted versus the discriminant #5 (the best performing discriminant shown in Figure 2). The critical value for the GLR is also shown ($\lambda_{0.05}$). The “explosions” all have a low GLR well below $\lambda_{0.05}$. Approximately 5% of the “earthquakes” are also declared as outliers as expected from selection of the false alarm rate $\alpha = 0.05$. Note that the events with the largest GLR value have a discriminant value very near the mean value (0) for the “earthquake” population. As events deviate from 0 in both the positive and negative directions, the GLR typically decreases. This demonstrates that the GLR technique flags

both explosion-like earthquakes and "super" earthquakes. So in practice, a plot such as that shown in the top portion of Figure 4 could be used to eliminate the "super" earthquakes from further analysis, reducing the false alarm rate in half.

In the lower portion of Figure 4 we plot the GLR versus an estimate of the standard deviation for each event, $s = \sqrt{\sum_{i=1}^n (v_i - \bar{v})^2 / (n-1)}$, where $n = 5$. It can be seen that there is a progressive decrease in the GLR with increasing s . Events having a larger s , have one or more discriminant values that are far from zero and appear more like outliers. Interestingly, "earthquake" #81 (identified in Figure 4), is declared an outlier, but has a small s . Examination of v_{81} show that, by random chance, all of the discriminant values for this event are systematically high. Thus, it appears that the GLR outlier detection technique holds promise for flagging suspect events in a region that have characteristics different from a known population (e.g. earthquakes).

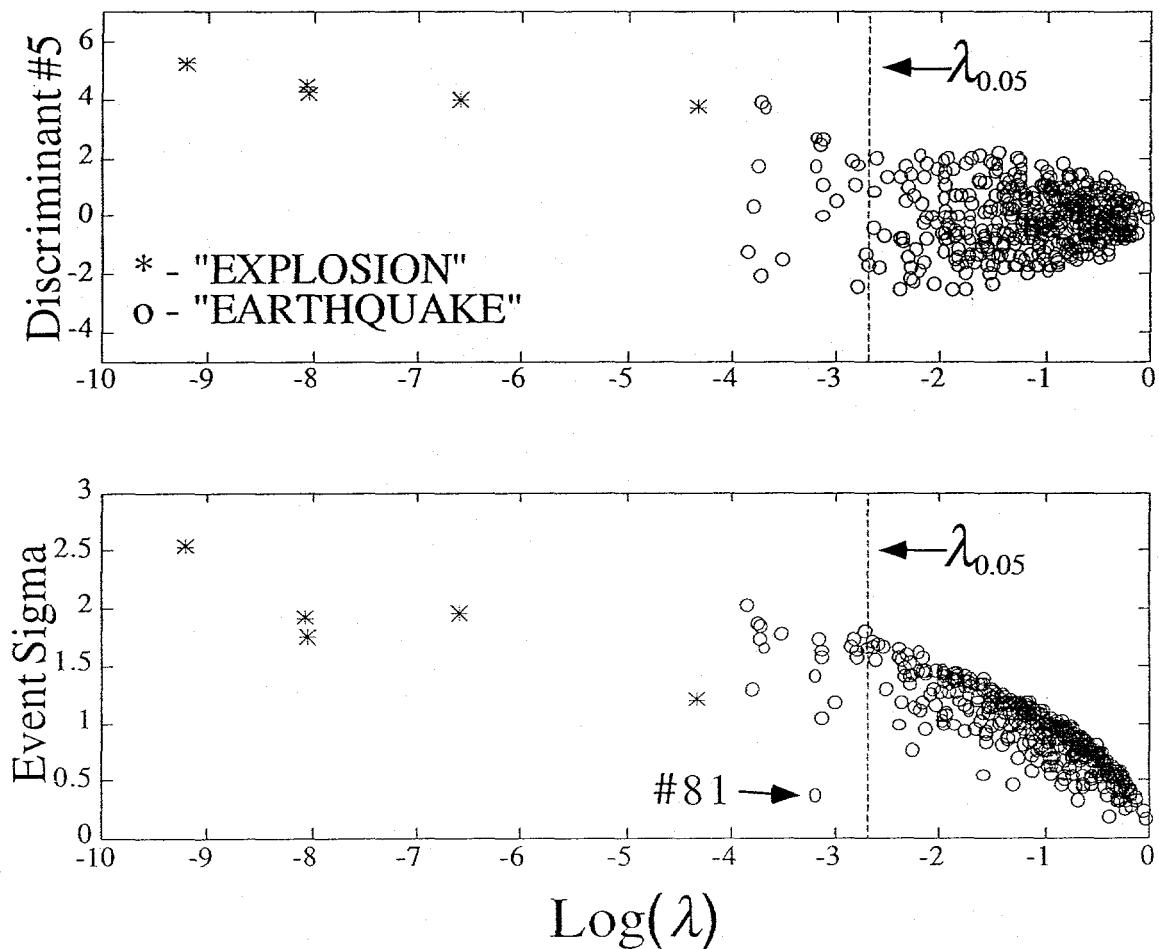


Figure 4. Generalized Likelihood Ratio, λ , versus discriminant #5 (see Figure 2) (top) and individual event sigma (bottom). Critical value for λ is shown for $\alpha = 0.05$ significance level. Event #81 is identified as an outlier not having a large sigma (see text for details).

Application of GLR Outlier Detection to P_g/L_g Data from WMQ

In this section, we apply the GLR outlier detection technique to P_g/L_g discrimination data from western China recorded at the Chinese Digital Seismic Network (CDSN) station WMQ in northwestern China. These tests are similar in many ways to the synthetic tests described above in that we have P_g/L_g in 7 different frequency bands for many earthquakes and just a few nuclear explosions. Details of the discrimination measurements and distance corrections are given in (Hartse *et al.*, 1996). For this dataset we had measurements from 156 known earthquakes (obtained from Preliminary Determination of Epicenter catalogs; PDE), 215 unknown events (presumed earthquakes and possibly some industrial explosions from Chinese State Seismological Bureau catalogs), and 18 known nuclear explosions (1 from the Chinese Lop Nor test site and 17 from the East Kazakh test site). In the analysis performed below, we use the PDE events as the “known” population.

We first perform a simple feature selection procedure in order to find the best combination of discriminants to use for the GLR outlier detection. Following Hand (1981), we compute the class separability measure, J

$$J = \text{tr} \mathbf{W}^{-1} \mathbf{B} \quad (4)$$

where \mathbf{W} is the within class scatter matrix and \mathbf{B} is the between class scatter matrix. Using the metric J (which is proportional to the Mahalanobis Distance), we perform an exhaustive search to find the best combination of discriminants. In other words, we first find the best single discriminant, followed by the best two, etc., by examining J for all possible combinations. A large value of J indicates large separation between the two populations being examined and small scatter within each population. A step-forward search was also performed that gave very similar results.

The WMQ dataset is incomplete in that there are missing measurements for many events due to poor signal to noise ratios or other factors. For the feature selection procedure, we included known earthquakes (92) and explosions (10) having complete measurements. The best performance in terms of misclassification rates and percentage of events having available measurements was the combination of a low-frequency P_g/L_g ratio

(0.5 to 1 Hz) with 2 or 3 ratios between 2 and 8 Hz. This was unexpected because the 0.5 to 1 Hz ratio by itself typically is characterized by very poor discrimination performance. However, because of the correlation structure, the low-frequency discriminant can be effectively combined with a high-frequency discriminant.

The results of the feature selection search based on examination of the J metric are shown in Figure 5. The filled circles indicate which band or bands were selected after each of the 7 steps. For example, after the first step, the highest frequency P_g/L_g band (4 to 8 Hz) was selected. This was expected, because it is well known that the P_g/L_g discriminant shows improved performance at higher frequencies (e.g. Hartse *et al.*, 1996). For step 2, it was found that the combination of a low-frequency band (0.5 to 1 Hz) with a high-frequency band (3 to 6 Hz) provided the largest J value (not significantly larger than the combination of 0.5 to 1 Hz and 4 to 8 Hz). For step 3, the largest value of J was found by combining the low frequency, with an intermediate- and high-frequency value. At this point it can be seen from Figure 5 that the value of J appears to level off indicating no real improvement in class separability. This is corroborated by calculation of an F statistic that provides the basis for determining whether the addition of an extra variable improves discrimination significantly (e.g. Hand, 1981; Figure 5). The rapid decrease in F and the leveling off of J between steps 3 and 4 suggest that it is only necessary to include 3 or 4 of the discriminants.

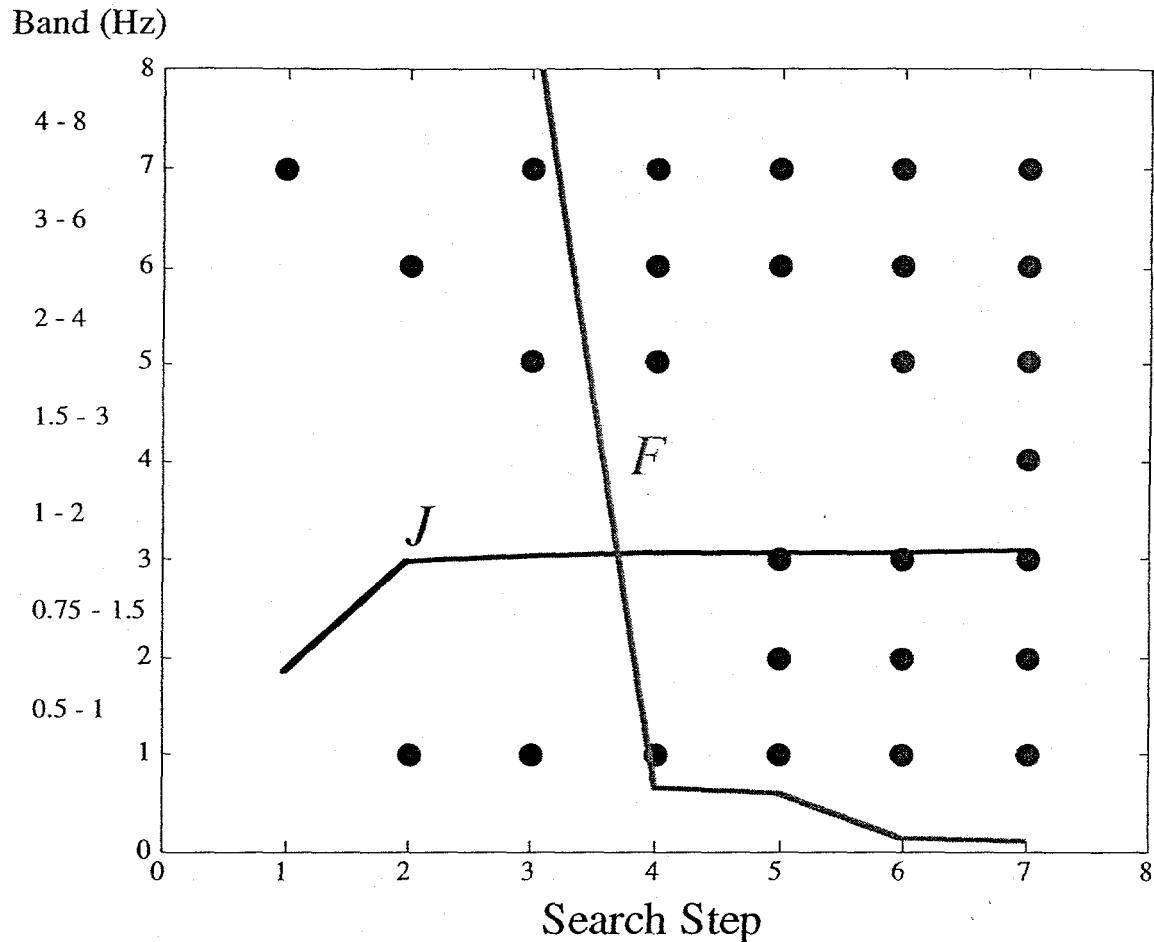


Figure 5. Results from feature selection search using the P_g/L_g ratio in 7 different frequency bands at WMQ. Filled circles indicate which discriminants were selected after each step. Also shown are the J and F values after each step (see text for details).

Using the 4 discriminants from step 4 (Figure 5), we computed the GLR for the earthquakes, explosions, and unknown events. The earthquakes from the PDE catalogs were assumed to be the known events for which the leave-one-out technique was used to compute the GLR. For events with missing data, we chose a subset of known earthquakes with measurements in at least the same bands and computed the GLR and recomputed the critical value for $\alpha = 0.05$. The results are illustrated in Figure 6, where the difference between the GLR and the critical value ($\lambda_\alpha - \lambda$) for each event is shown plotted versus event magnitude. We reject the hypothesis that an event belongs to the earthquake population if $\lambda_\alpha - \lambda > 0$ (i.e. an outlier is declared). The GLR technique appears to work very well, in that all of the explosions except one are declared as outliers. The explosions

with smaller values of $\lambda_\alpha - \lambda$ were all missing one or more measurements in the four frequency bands included. The explosion with the very large $\lambda_\alpha - \lambda$ is the Lop Nor explosion of September 29, 1988, which was characterized by a very large P_g/L_g ratio. Approximately 5% of the known earthquakes were declared to be outliers as expected by the significance level of the test ($\alpha = 0.05$). The one earthquake having a large value of $\lambda_\alpha - \lambda$ is actually the result of a processing error (due to a mislocation) and has a very small P_g/L_g ratio so is actually a "super earthquake".

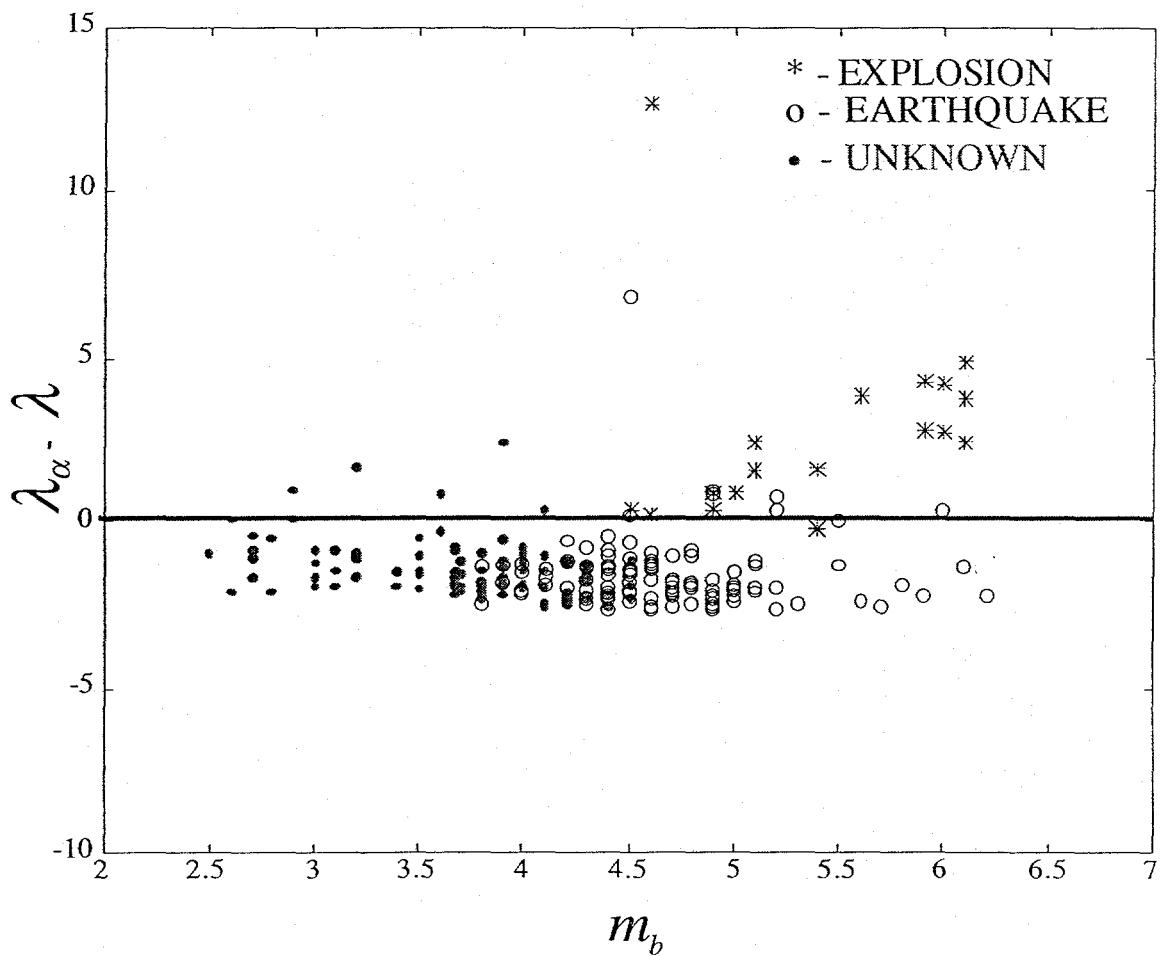


Figure 6. Results from application of GLR outlier detection to P_g/L_g ratio in 4 different frequency bands at WMQ for nuclear explosions, known earthquakes, and suspected earthquakes (unknown). The difference between the critical value for $\alpha = 0.05$ and the GLR, $(\lambda_\alpha - \lambda)$, is plotted versus event magnitude. If $\lambda_\alpha - \lambda > 0$ and event is declared an outlier from the known earthquake population (see text for details).

Finally, we performed the GLR outlier test on the discriminants for each of the search steps from the feature selection procedure described above (Figure 5). We computed the probability of including a nuclear explosion in the earthquake population, $P(Q|X)$, for each step and the percentage of missing data for each of the three populations. From this figure, it can be seen that there is a tradeoff between the number of missed explosions and the percentage of events that could be processed. The best discriminant is the P_g/L_g ratio in the highest frequency band (Step 1) showing no missed explosions. However, because of propagation effects at the higher frequencies, fewer measurements were available (33% of the explosions had no measurement in this band due to poor signal to noise). When three or four discriminants are included, a larger number of events can be examined. However, as discussed above, some of these events were still incomplete and were not correctly identified. $P(Q|X)$ appears to progressively increase as more discriminants are added and is highest when all 7 bands are included, showing that it is not useful to include many poor discriminants into the procedure. The progressive increase in $P(Q|X)$ as more discriminants are added is partially due to the fact that some events at each step have some missing values.

Pg/Lg GLR Outlier Test

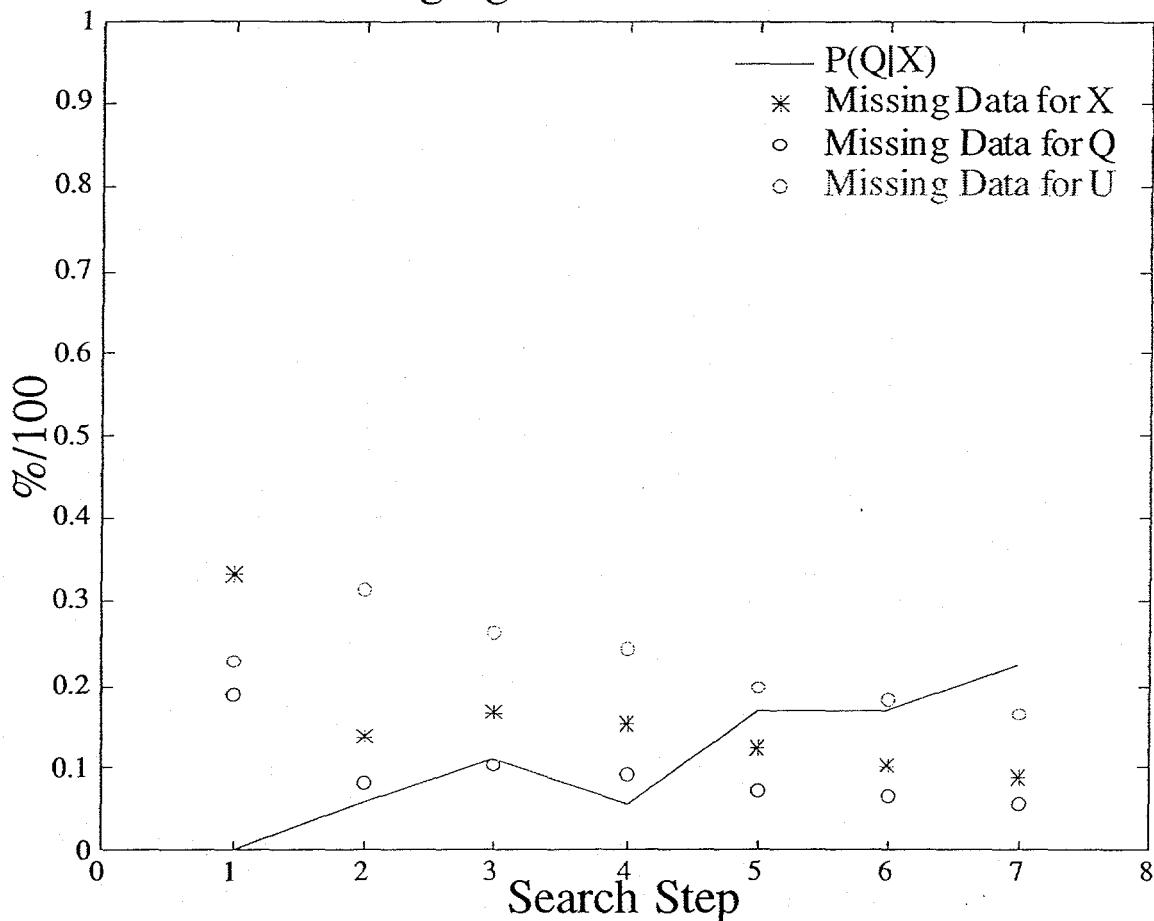


Figure 7. Results from application of GLR outlier detection to P_g/L_g ratios at WMQ for nuclear explosions, known earthquakes, and suspected earthquakes. Search step indicates which combination of amplitude ratios were used (from feature selection procedure; Figure 5). $P(Q|X)$ are missed violations (i.e. nuclear explosions, X, included in the earthquake, Q, population). The percentage of missing measurements is indicated for each of the three populations (U - presumed earthquakes).

Conclusions

The Generalized Likelihood Ratio Outlier detection technique is shown to be a reasonable alternative to traditional seismic event classification methods when no calibration information from nuclear explosions is available. A simple feature selection procedure suggests that it is useful to combine a low-frequency (0.5 to 1 Hz) P_g/L_g ratio with higher-frequency ratios in order to improve outlier detection (or discrimination) capabilities. It is expected that discrimination performance will improve as additional discriminants are added (e.g. spectral ratios, $m_b - M_s$, P_g/L_g in higher frequency bands) which will be the subject of future work.

ACKNOWLEDGMENTS

Helpful discussions with Mark Fisk are appreciated. We also thank W. Scott Phillips for obtaining WMQ data for events in the SSB catalog and for reviewing the manuscript. This work is performed under the auspices of the U.S. Department of Energy by Los Alamos National Laboratory under contract W-7405-ENG-36.

REFERENCES

- Anderson, T.W., *Introduction to Multivariate Statistical Analysis*, Wiley, New York, 675pp, 1984.
- Fisk, M.D., H.L. Gray, and G.D. McCartor, Applications of generalized likelihood ratio tests to seismic event identification, Phillips Laboratory, Hanscom Air Force Base, MA, PL-TR-93-2221, 69 pp, 1993.
- Fisk, M.D., H.L. Gray, and G.D. McCartor, Regional event discrimination without transporting thresholds, submitted to *Bull. Seis. Soc. Am.*, 25pp, 1996.
- Hand, D.J., *Discrimination and Classification*, Wiley, New York, 218pp, 1981.
- Hartse, H.E., S.R. Taylor, W.S. Phillips, and G.E. Randall, Regional event discrimination in central Asia with emphasis on western China, *Los Alamos National Laboratory, Los Alamos, NM*, LAUR-96-2002, 45 pp, submitted to *Bull. Seism. Soc. Am.*, 1996.
- Haykin, S., *Neural Networks*, Macmillan College Publishing Co., Inc., Englewood Cliffs, NJ, 695pp., 1993.
- Taylor, S.R., Analysis of high-frequency Pg/Lg ratios from NTS explosions and western U.S. earthquakes, *Bull. Seism. Soc. Am.*, 86, 1042-1053, 1996.