ANALYSIS OF HIGH-FREQUENCY Pg/Lg RATIOS FROM NTS EXPLOSIONS AND WESTERN U.S. EARTHQUAKES

Steven R. Taylor

Earth and Environmental Sciences Division
Los Alamos National Laboratory
University of California
Los Alamos, NM 87545

July 1, 1997

LAUR-95-3142
Published in
Bull. Seism. Soc. Am., 86, 1042-1053, 1996

ABSTRACT

The high-frequency P_g / L_g discriminant is studied between frequencies of 0.5 and 10 Hz using 294 NTS explosions and 114 western U.S. earthquakes recorded at four broad band seismic stations operated by Lawrence Livermore National Laboratory. The stations are located at distances of about 200 to 400 km from the Nevada Test Site (NTS). Event magnitudes ranged from about 2.5 to 6.5 and propagation paths for the earthquakes range from approximately 175 to 1300 km. The discriminant is shown to be very effective and displays improved separation between earthquakes and explosions as frequency is increased. Because of propagation effects, it is important to apply distance corrections directly to the amplitude ratios or to the magnitude-corrected amplitudes prior to computing the ratios. Multivariate discrimination analysis using both maximum-likelihood Gaussian classifiers and a backpropagation neural network show that approximately 95% of the events can be correctly identified. Both classification procedures were designed to handle missing data filled in using a nearest-neighbor algorithm. Except for a few notable exceptions, most of the earthquake misclassifications occur for $m_b < 4$, which is expected for events having reduced signal-to-noise ratios. All of the explosion misclassifications occur for $m_b > 4$ suggesting a source or near-source effect rather than an effect of poor signal-to-noise ratio. The explosions that were misclassified were typically of magnitude large enough to be classified correctly by m_b/M_S or Love wave energy. The main drawback of the P_g / L_g discriminant is that, because of signal-to-noise considerations and propagation effects, the number of measurements are reduced considerably at higher frequencies. It is expected that the problem will be amplified as magnitudes are reduced and event-receiver distances are increased.

INTRODUCTION

Recent renewed interest in monitoring a Comprehensive Test Ban Treaty (CTBT) has brought about a resurgence in the study of regional discriminants. At the time the classic discrimination paper of Pomeroy *et al.,* (1982) was written, broadband seismic systems were just coming on line in larger numbers. Thus, many of the discriminants tested used data from short-period (~1 Hz) and long-period bands. High-frequency (> 5 Hz) studies were only beginning in the mid 1980s. A number of high-frequency discriminants have recently been tested, some having mixed performance. For example, the high-frequency L_g spectral ratio discriminant taken in the 1-2 and 6-8 Hz bands showed good performance in the western U.S. down to relatively low magnitudes (m_b ~ 3; Taylor *et al.,* 1988; Walter *et al.,* 1995). In other geophysical regions, however, the performance of the L_g spectral ratio discriminant is disappointing (e.g. Baumgardt and Young, 1990; Chan *et al.,* 1990; Hartse *et al.,* 1995). In contrast, the high-frequency P/L_g or P_g/L_g discriminant appears to perform well in every region that it has been tested and shows a marked improvement over the P_g/L_g ratio taken at around 1 Hz (e.g. Walter *et al.,* 1995; Baumgardt and Young, 1990; Chan *et al.,* 1990; Dysart and Pulli, 1990; Chael, 1988).

In this paper, we extend the discrimination study of Taylor *et al.,* (1989) to include the high-frequency P_g/L_g discriminant. In that study, the P_g/L_g discriminant was tested at 1 Hz with disappointing results. Recent work of Walter *et al.,* (1995) using Nevada Test Site (NTS) explosions and earthquakes located on the NTS has shown the excellent performance of the high-frequency P_g/L_g discriminant. In many ways, this work parallels that of Walter *et al.,* (1995), except we extend the analysis to include earthquakes that are located in many areas of the western U.S. This provides a slightly better test of the discriminant in an actual monitoring situation and allows for the testing of propagation corrections.

We first describe the measurement of the amplitude ratios and the development of a magnitude scale based on the L_g amplitudes (used for plotting purposes). Distance corrections are then computed for each of the 6 different frequency bands used in the analysis. The discriminant is then qualitatively described and compared to other high-frequency discriminants. Finally, to examine the overall discrimination performance, different multivariate discriminants are computed. These include the traditional maximum likelihood Gaussian classifier and a backpropagation neural network.

DATA

The data used in this study were from 327 NTS nuclear explosions and 170 western U.S. earthquakes recorded by the Lawrence Livermore National Laboratory (LLNL) seismic observatory stations at Elko, NV (ELK), Kanab, UT (KNB), Landers, CA (LAC), and Mina, NV (MNV), located at distances of 200 to 400 km from the Nevada Test Site (Figure 1). For each event-station pair, the spectra were calculated from windowed P_n, P_g, and L_g phases. Group velocity windows were defined by t_1 and t_2, where $t_1 = \Delta / 6.0$ and $t_2 = \Delta / 5.0$ for the P_g phase, and $t_1 = \Delta / 3.6$ and $t_2 = \Delta / 3.0$ for the L_g phase (where Δ is the epicentral distance in km). The P_n window was selected manually and generally ranged in length from 4 to 5 seconds, starting from about 1 second prior to the P_n arrival time. To obtain a smoother spectrum, the P_n window was extended to 20 seconds by zero-padding of the data. Noise spectra were calculated in a 30 second window preceding the P_n arrival. To minimize numerical effects at high frequencies, the signals were differentiated to acceleration, windowed using a 10% cosine taper between the limits defined above and fast Fourier transformed. The resulting acceleration spectra were divided by f^2 to convert them to displacement spectra. If three-component data were available, the final spectra were an average of the vertical, radial, and transverse components for the P_g and L_g. In order to reduce the effect of noise on the signal spectra, only those frequencies for which the (S+N)/N level was greater than 2 were used (using the pre - P_n noise). The spectra for each phase were then sampled logarithmically at 41 points between 0.1 and 10 Hz. The sampled spectra were instrument corrected using the nominal instrument response (Jarpe, 1989). We then computed amplitudes in six different frequency bands (0.2-1; 1-2; 2-4; 4-6; 6-8; and 8-10 Hz) for each station by averaging over available measurements for a particular band. We also investigated the P_n/L_g discriminant, but because of the short time windows used for the P_n phase and poor signal-to-noise ratios, the performance was degraded relative to the P_g / L_g discriminant.

MAGNITUDE AND DISTANCE CORRECTIONS

Although not directly used in the analysis, we derived a magnitude scale for the explosions and earthquakes based on the 1 Hz L_g spectral value. This magnitude scale is used for plotting purposes in this paper and is tied to the teleseismic m_b through the $m_b(P_n)$ of Denny *et al.*, (1987). For the earthquake dataset, we solved the equation

$$\log A_{ij} - m_b(P_n) = a \log \Delta_{ij} + b \qquad (1)$$

for a and b, where A_{ij} is the 1 Hz L_g spectral value (nm*s), and Δ_{ij} is the epicentral distance from for source i to receiver j. The values for a and b were calculated to be -2.8 $\pm$ 0.14 and -5.74 $\pm$ 0.39, respectively. These regression coefficients are then used to estimate a new magnitude by rearranging equation (1). Using this procedure the explosion magnitudes were biased low by 0.35 magnitude units because of the poor excitation of Lg relative to earthquakes. Thus, a source-type correction factor of 0.35 was added to the magnitudes of the explosions giving

$$m_b(Lg) = \log A_{ij} + 2.8 \log \Delta_{ij} + S_i + 5.74 \qquad (2)$$

where S_i is 0 for earthquakes and 0.35 for explosions.

The next step in the discrimination analysis is to derive distance corrections. To do this, we performed the same analysis as in equation (1) for each of the six frequency bands, only substituting the new magnitudes obtained from equation (2). For simplicity, we will refer to $m_b(Lg)$ as m_b throughout the rest of the paper. The residuals are then the distance corrected earthquake amplitudes used in the discrimination analysis discussed below. The regression coefficients derived from the analysis of the earthquake data were used to correct the explosion amplitudes. Figure 2 shows examples of the fits to the L_g amplitude (normalized for m_b) versus distance curves, linear fit, 95% confidence intervals for the regression line, and the distance-corrected amplitudes (residuals).

The distance correction could as easily be applied directly to the phase or spectral ratios. The advantage of correcting the measured amplitudes for a given frequency is that it enables one to quickly investigate different discriminants without having to continuously apply distance corrections (e.g. cross-spectral ratios). The approach of correcting measured phase amplitudes, however, would not be possible using data from poorly calibrated seismic networks. For amplitude ratios in a particular frequency band, the instrument effect cancels. However, for ratios involving measurements in different frequency bands, knowledge of the instrument response is critical.

To test the significance of the distance effect on the various ratios, we regressed the logarithm of the P_g /L_g ratios for the six different frequency bands and the logarithm of the L_g spectral ratio (taken in the 1 to 2 and 6 to 8 Hz frequency bands) against the logarithm of the distance. Examples are illustrated in Figure 3 along with the corresponding residuals and the regression coefficients are listed in Table 1. From Figure 3, it can be seen that the distance effects are significantly different for different discriminants. In general, the P_g /L_g ratio had very little distance dependence for this particular dataset and the distance correction has only a minor effect. In contrast, as discussed by Taylor *et al.,* (1988), the L_g spectral ratio has a strong distance dependence and the distance correction appears to help reduce scatter and separate the two populations. A test was made of the hypothesis that the amplitude ratio in a particular band is independent of the distance versus the alternative hypothesis that the amplitude ratio is dependent on distance. An F test was used to test whether the slope was zero. For each of the lowest four P_g /L_g bands (0.2 - 6 Hz), we rejected the null hypothesis in favor of the alternative hypothesis that the amplitude ratio is dependent on the distance at the 95% confidence level. For the two highest frequency bands (6 - 10 Hz), we could not reject the null hypothesis that the amplitude ratio is independent of distance. For the two highest frequency bands, the signal to noise levels are generally low at large distances. If we only include events having distances less than 600 km in the regression, the dependence on distance is greater and the regression coefficients are more nearly equal to those at 1 - 2 Hz (Table 2). However, the dependence of the P_g /L_g ratio is still small and the overall results of the discrimination study discussed below are not strongly dependent on the distance correction.

The reason for this frequency-dependent effect on the distance correction is unclear. One explanation is that the effect could be due to changes in attenuation mechanism with frequency. If at lower frequencies, intrinsic mechanisms were dominant, the L_g attenuation would be greater than that for the P_g and the amplitude ratio would increase with range. If at higher frequencies, scattering effects begin to dominate, the attenuation for each phase would become more similar and little distance effect would be observed for the amplitude ratio (cf. Taylor *et al.,* 1986).

P_g /L_g RATIO

Once the raw amplitudes are corrected for distance effects, we can compute cross-phase (e.g. P_g /L_g at 6-8 Hz), or cross-spectral ratios (e.g. P_g (1-2 Hz)/L_g (6-8 Hz) or L_g(1-2/6-8 Hz)). As discussed above, in this paper we focus mainly on the P_g /L_g phase ratio as a function of frequency. Figure 4 shows the P_g /L_g ratios for the six different frequency bands plotted as a function of magnitude. For each band, we chose to only include events having two or more readings for which we computed the log-average amplitude ratio. Figure 4 shows that the best separation is observed at higher frequencies.

For both earthquakes and explosions, however, the most measurements occur in the 1-2 and 2-4 Hz frequency bands.

The improved separation between the earthquakes and explosions with increased frequency is further illustrated in Figure 5, showing the mean spectral ratio and one standard deviation for earthquakes and explosions. At lower frequencies (< 3 Hz), the mean spectral ratios are closer and the overlap is greater between the two populations. The spectral ratios show greater separation at higher frequencies and show no overlap at the one sigma level. However, because of propagation effects, most of the data points are recorded at the lower frequencies (Figure 6). Note that points are not included in Figure 6 either because of poor signal to noise, or because of data unavailability. This points out one of the limitations of the high-frequency discriminants: the performance is improved at high frequencies, but propagation effects can significantly reduce the number of high-frequency measurements.

The increase of the explosion P_g/L_g ratio with frequency is not totally understood. It is an effect that is also observed from the East Kazakh test site as well (Gupta *et al.*, 1992). One effect that could account for this is R_g to S (and L_g) scattering (Gupta *et al.*, 1992; Patton and Taylor, 1995) that would boost the L_g amplitudes at low frequencies (and lower the P_g/L_g ratio). It should be noted that simple 1-D elastic synthetic seismograms predict the same effect of the P_g/L_g ratio increasing with frequency for shallow explosions in velocity structures similar to NTS. In this case, the effect is due to either the frequency dependence of the non-geometric phase S^* or the near-surface low Q effect on pS that gets trapped in the crust as L_g (e.g. Lilwall, 1988). However, these simple, 1-D elastic calculations are unable to capture the complicated nonlinear effects associated with a nuclear explosion or the two- and three-dimensional near-source elastic propagation effects.

From Figure 4, it can also be seen that the P_g/L_g ratio increases with magnitude for all frequency bands and that better separation occurs for larger magnitudes. This was also noted by Taylor *et al.*, (1989) for the 1 to 2 Hz band and by Walter *et al.*, (1995) for the 6 to 8 Hz band. Because of the correlation of explosion yield (and hence magnitude) with depth, this at first glance appears to be a depth effect. However, Walter *et al.*, (1995) concluded that the increase of the P_g/L_g ratio with depth is actually a material effect related to the strength (measured by the product $\rho\alpha^2$ where ρ and α are the working point density and compressional velocity, respectively) and the gas-filled porosity (GFP). When they separated ratios for explosions detonated in high-strength, low GFP rocks or low-strength, high GFP rocks, the dependence on depth was negligible. Similarly, Gupta *et al.*, (1992) concluded that the higher-frequency L_g in the 3 to 7 Hz passband was dependent on source medium velocity. They hypothesized that for explosions detonated in low-velocity media, more energy from the free-surface *pS* conversion is trapped in the crust, enhancing L_g amplitudes. If this is the case, we would expect to see an apparent increase of the P_g/L_g ratio with depth, magnitude, and working-point velocity (and decrease with gas-filled porosity) for explosions at NTS (since these factors are all correlated).

Thus, for reasons that are still unclear, the explosions detonated in high-strength, low GFP rocks have a larger P_g/L_g ratio than those detonated in low-strength, high GFP rocks. This phenomenon cannot be due to a direct source effect on P_g and L_g because the two phases would be affected in a similar manner (assuming L_g is generated by near-source *P* to *S* conversions). A secondary source that depends on near-source material properties could explain the enhancement of L_g to P_g for the explosions detonated in high gas-filled porosity, low-velocity rocks. Possible candidates for secondary sources that would enhance high-frequency L_g and that could be related to near-source material properties are some type of shear source such as passive block motion (Patton, 1991) or cavity rebound (Jones *et al.*, 1993).

We show additional discriminants in Figure 7 to compare with the high-frequency P_g/L_g ratio. The high-frequency P_n/L_g ratio shows similar separation as the P_g/L_g ratio. However, because of poor signal to noise for the P_n phase, the ratio was measured for only a limited number of events. The scatter appears to be a little less for the P_n/L_g ratio which may be due to the lack of material dependence on the discriminant (as discussed by Walter *et al.*, 1995).

The L_g spectral ratio taken in the 1 to 2 and 6 to 8 Hz frequency bands shows good separation between the earthquakes and explosions (similar to that observed by Taylor *et al.*, 1988). However, there are a number of problems associated with the spectral ratio discriminant. As discussed by Taylor and Denny (1991), there is a strong material dependence on the spectral shape that causes complications with the explosion ratios. The rate of high-frequency decay is greater for explosions detonated in rocks having high gas-filled porosity, resulting in a higher spectral ratio (and better separation from the earthquakes). Although magnitude is not directly used in the multivariate discriminant analysis discussed below, there are a number of explosions in Figure 7 that overlap with the earthquakes at both low and higher magnitudes. This is probably due to the fact that these events are detonated in saturated media (the low magnitude explosions shown in Figure 7 are generally overburied and detonated below the water table) and the rate of high frequency decay is similar to that from earthquakes. Additionally, effects of the corner frequency with event size cause a systematic dependence of the spectral ratio discriminant with magnitude that is difficult to remove if it is desired to form a discriminant with a normal probability distribution for multivariate analysis. Lastly, the good separation between earthquakes and explosions for the spectral ratio discriminant is largely due to fairly unique geologic effects at NTS and evidence suggests that the discriminant may not be transportable to other regions (Baumgardt and Young, 1990; Chan *et al.,* 1990; Hartse *et al.,* 1995). Similar effects are observed for the P_g /L_g cross spectral ratio (P_g taken in the 1 to 2 Hz band and L_g taken in the 6 to 8 Hz band). Thus, for the multivariate discrimination analysis presented in the next section, we confine our analysis to the P_g /L_g ratio taken in the six different frequency bands.

MULTIVARIATE DISCRIMINATION ANALYSIS OF P_g /L_g RATIOS

In this section, we apply multivariate discrimination techniques in order to assess the discrimination capabilities of the combined P_g /L_g ratio in the six different frequency bands. Two different techniques are investigated, each giving similar results. The first is the traditional maximum likelihood Gaussian classification (e.g. Duda and Hart, 1973) and the second is a neural network approach (e.g. Haykin, 1994; Dowla *et al.,* 1990). Many of the statistical problems associated with earthquake/explosion discrimination are cogently described in Fisk *et al.,* (1993). First, we formulate the approach taken for the maximum likelihood classification.

The distance-corrected (averaged over 2 or more stations) logarithm of the P_g /L_g ratio taken in the six different frequency bands (Figure 4) are the parameters used in the multivariate analysis. For each event, we construct a feature vector, **v**, of length 6. If $p(\mathbf{v}|\theta_i)$ is the conditional probability of observing **v** from a population θ_i, then from Bayes Rule, the probability that an event is from population θ_i, given a measurement vector **v** is given by

$$P(\theta_i|\mathbf{v}) = \frac{p(\mathbf{v}|\theta_i)P(\theta_i)}{p(\mathbf{v})} \tag{3}$$

where for *n* classes

$$p(\mathbf{v}) = \sum_{j=1}^{n} p(\mathbf{v}|\theta_j)P(\theta_j) \tag{4}$$

Note that $p(\mathrm{x})$ is the probability density function, and $P(x)$ is the probability of an event x occurring

$$\int_{x_1}^{x_2} p(x)dx = P(x_1 \le x \le x_2) \tag{5}$$

Next, we form a discrimination function, $g_i(\mathbf{v}) = P(\theta_i|\mathbf{v})$, where $\mathbf{v}$ is assigned to class θ_i if $g_i(\mathbf{v}) > g_j(\mathbf{v})$ for all $i \ne j$. From (3) we define a new discrimination function

$$G_i(\mathbf{v}) = \ln g_i(\mathbf{v}) = \ln p(\mathbf{v}|\theta_i) + \ln P(\theta_i) - \ln p(\mathbf{v}) \tag{6}$$

and the equation for the decision boundary separating the populations is given by

$$G_i(\mathbf{v}) - G_j(\mathbf{v}) = 0 \tag{7}$$

For two populations, $\theta_1 = X$ and $\theta_2 = Q$ where X denotes nuclear explosions and Q earthquakes, we define the single discrimination function

$$G(\mathbf{v}) = G_X(\mathbf{v}) - G_Q(\mathbf{v}) \tag{8}$$

where, from (6) is

$$G(\mathbf{v}) = \ln \frac{p(\mathbf{v}|X)}{p(\mathbf{v}|Q)} + \ln \frac{P(X)}{P(Q)} \tag{9}$$

It should be noted that we can introduce misclassification costs into the problem where $C(\theta_j|\theta_i)$ is the cost associated with classifying an event in population θ_j when it is really from θ_i. Equation (9) becomes

$$G(\mathbf{v}) = \ln \frac{p(\mathbf{v}|X)}{p(\mathbf{v}|Q)} + \ln \frac{C(X|Q)P(X)}{C(Q|X)P(Q)} \tag{10}$$

This allows one to adjust the location of decision boundaries based on prior probabilities of occurrence and misclassification costs. If the second term in (10) is zero (e.g. equal misclassification costs and equal prior probabilities of occurrence) the total error rate is minimized. For example, one may assign a high misclassification cost for misidentifying an explosion, $C(Q|X)$, and the decision surface will move into the earthquake hyperspace and few explosions will be missed (assuming equal prior probabilities of occurrence for earthquakes and explosions). The deleterious effect of this, however, is that numerous false alarms will occur (earthquakes classified as explosions). In our analysis, we want to minimize the total error rate to examine discrimination capabilities, so we assume equal misclassification costs and prior probabilities of occurrence.

If we assume $p(\mathbf{v}|\theta_i)$ is multivariate normal

$$p(\mathbf{v}|\theta_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{v}-\mu_i)^T \Sigma_i^{-1}(\mathbf{v}-\mu_i)\right] \qquad (11)$$

where Σ_i is the d-by-d sample covariance matrix, and μ_i is the d-component mean vector for population i. For two-category classification, equation (10) becomes

$$G(\mathbf{v}) = \left[-\frac{1}{2}(\mathbf{v}-\mu_x)^T \Sigma_x^{-1}(\mathbf{v}-\mu_x)\right] + \left[-\frac{1}{2}(\mathbf{v}-\mu_Q)^T \Sigma_Q^{-1}(\mathbf{v}-\mu_Q)\right] + \frac{1}{2}\ln\frac{|\Sigma_Q|}{|\Sigma_X|} + \ln\frac{C(Q|X)P(X)}{C(X|Q)P(Q)}$$

(12)

resulting in a quadratic discrimination function in **v**.

If the simplifying assumption of equal covariances for the two populations, $\Sigma_X = \Sigma_Q = \Sigma$, is made, then equation (12) reduces to

$$G(\mathbf{v}) = \mathbf{v}^T\Sigma^{-1}(\mu_X - \mu_Q) - \frac{1}{2}(\mu_X + \mu_Q)^T \Sigma^{-1}(\mu_X - \mu_Q) + \ln\frac{C(Q|X)P(X)}{C(X|Q)P(Q)} \qquad (13)$$

resulting in a linear discrimination function.

Three considerations that must be addressed with the maximum likelihood (ML) discrimination approach are 1) whether the data are multivariate normal, 2) the covariance structure, and 3) how to deal with missing data. A chi-square test was made to test for normality of the log(P_g /L_g) ratios (e.g. Menke, 1984). In general, it was found that the 1-2, 2-4, and 4-6 Hz frequency bands were normally distributed at the 5% level of significance for both earthquakes and explosions. As discussed by Weisberg (1980), various transformations have been proposed to make data appear normally distributed (e.g. Box-Cox transformations). Application of those techniques, however, is beyond the scope of this paper.

To test for equal earthquake and explosion covariance matrices, we initially used the F-distribution to determine whether the individual variances were equal as suggested by Fisk *et al.,* (1993). This is not as rigorous as other statistical tests because equal variances are only an indication that covariance matrices are equal, and does not necessarily eliminate the possibility that they are unequal. For both filled (see discussion below) and unfilled data, we could not reject the hypothesis that the variances of the earthquake and explosion populations are equal for most frequency bands at the 95% level of confidence. A maximum likelihood ratio test (Appendix A; Anderson, 1984) was then applied to test the hypothesis that the earthquake and explosion covariance matrices are equal. A chi-square test was performed to test the value of the likelihood ratio and we rejected the hypothesis that that the earthquake and explosion covariance matrices are equal at the 5% level of significance. As will be further discussed below, however, no improvement was observed in discrimination performance under the assumption of unequal covariances.

One of the major problems faced in the multivariate discrimination analysis is how to deal with missing data values. Because of signal to noise problems or station unavailability, amplitude ratios are not available for all 6 bands for the majority of events. Out of the original 327 NTS explosions and 170 earthquakes processed, 294 explosions and 114 earthquakes have amplitude ratios in at least one band. Of these, 47% of the explosions and 49% of the earthquake ratios were missing, the majority of which were in the higher frequency bands (Figure 6). As discussed by Hand (1981), there are a number of approaches for handling missing data, none of which are without problems. Probably one of the most rigorous techniques is the Estimation-Maximization algorithm (Dempster *et al.,* 1977), which basically consists of iteratively solving for missing values by combining maximum likelihood estimates of data mean and covariance structure with regression analysis using available data. This is numerically quite intensive and we chose a simpler, ad hoc method using a nearest-neighbor technique (Duda and Hart, 1973). The nearest neighbor technique is a non-parametric method that basically consists of comparing an event having missing values with all other events to find those with the closest data structure and averaging to fill in the missing values. If $r_k(f_i)$ is the logarithm of the P_g/L_g amplitude ratio at frequency f_i for event k (having missing values to be filled in), and

$r_j(f_i)$ is the amplitude ratio for a comparison event *j* (having at least as many values as $r_k(f_i)$), then we calculate a measure of the distance between the measurements for the two events

$$\beta_{kj} = \frac{1}{m}\sum_{i=1}^{m}\left|r_k(f_i) - r_j(f_i)\right| \tag{14}$$

where *m* is the number of overlapping frequency bands for events k and j for which there are measurements. For each event k, a matching function, M_{kj}, is then computed relative to all other events *j* having amplitude measurements in overlapping frequency bands

$$M_{kj} = \frac{1}{1+\beta_{kj}} \tag{15}$$

The value of the matching score can vary from 0 to 1 and the values of the best matches are then averaged to fill in the missing values for event *k*. The main parameter to be adjusted is the number of events used to fill in the missing values. If too many events are used, the missing values are basically filled in with the mean of the event population which distorts the probability distribution of the filled data. Since filling in the missing data has the effect of reducing the variance, an F test was performed to determine if the variance of the filled and unfilled data were significantly different at the 95% level of confidence. Experiment showed that selecting the top 10% matches did not appreciably affect the distribution between the filled and unfilled data.

Figure 8 shows an example of the P_g /L_g ratio for the unfilled and filled data. For this frequency band, the number of points increased from 69 to 294 for the explosions and 54 to 114 for the earthquakes. Visual inspection of the two plots in Figure 8 suggests that the data structure is not significantly different between the filled and unfilled data. However, there is a slight indication of an increase in the number of points near the mean value for the explosions at a particular magnitude, and as discussed below, did not pass a chi-square test for normality at a high level of significance.

As mentioned above, a chi-square test was performed on both the filled and unfilled ratios for each frequency as a test for normality. The explosions generally passed the chi-square test for $\alpha = 0.05$ for all but the highest frequency band (8-10 Hz) and the earthquakes for the bands between 1 and 8 Hz. For reasons discussed above, the test for normality for the filled data was not as convincing. The filled explosion data passed the test for $\alpha = 0.01$ for frequencies between 1 and 6 Hz and the earthquakes between 1 and 4 Hz. Thus, in the multivariate discrimination analysis performed below, it must be kept in mind that the filled data are not normally distributed in the extreme frequency bands.

The multivariate discrimination was performed on the filled data by assuming both equal and unequal covariances for the earthquake and explosion data and the results were not significantly different for both cases. Thus, we assumed the simplest model of equal covariances (equation 13) with equal misclassification costs and equal prior probabilities of occurrence for the earthquakes and explosions. The leave-one-out method (Hand, 1986) was used to assess discrimination performance. Using leave-one-out, a discrimination function (e.g. using equation 13) is computed using all events except one. The one event is then classified using the discrimination function based on the remaining events and misclassification probabilities are tabulated.

The value of the discrimination function for each of the events having data assuming equal covariances using equation (13) is shown in Figure 9. The discrimination performance by assuming both equal and unequal covariances is listed in Table 2 where it can be seen that the difference between these two cases is negligible. In both cases, at least 95% of the events were correctly identified. This performance is similar to the multivariate study of Taylor *et al.,* (1989) (that did not include the high-frequency P_g /L_g ratio) and Dowla *et al.,* (1990) who used complete Q-corrected P_g and L_g spectra. However, in the present study, much fewer data were used than the two previous studies and the same discrimination performance was achieved. Interestingly, many of the misclassified explosions do not occur at small magnitudes (less than magnitude 4 as was observed in the study of Taylor *et al.,* 1989). Except for a few notable exceptions, most of the earthquake misclassifications occur for $m_b < 4$ which is expected for events having reduced signal-to-noise ratios. This is further illustrated in Figure 10 showing the cumulative probability of misclassification as a function of magnitude for earthquakes and explosions. All of the explosion misclassifications occur for $m_b > 4$ suggesting a source or near-source effect rather than an effect of poor signal to noise. The explosions that were misclassified were typically of magnitude large enough to be classified correctly by m_b/M_S or Love wave energy. Detailed analysis of the misclassified events will be the subject of future work.

A backpropagation neural network was also used in the multivariate discrimination analysis. Details of the backpropagation algorithm are given in Dowla *et al.,* (1990). The main modification to the study of Dowla *et al.,* (1990) is utility of a Levenberg-Marquardt learning rule that shifts from a gradient descent method to a Gauss-Newton method for adjusting the weights as an error minimum is approached (Haykin, 1994). Experiments were undertaken with different network architectures involving hidden layers, and, as with the study of Dowla *et al.*, (1990), a simple network appeared to perform very well. The network consisted of one input layer with six inputs feeding into a single neuron having an output through log sigmoid nonlinear transfer function giving an output between 0 and 1 (corresponding to earthquake and nuclear explosion, respectively; Figure 11).

The leave-one-out approach was used to estimate discrimination performance on the same filled matrix used for the maximum-likelihood classifier (ML) discussed above. Random initial weights were used and weights were adjusted until a predetermined error goal was achieved. The final results are very similar to those using the ML classifier and are listed in Table 2. There was a strong correlation between the output of the neural network and the discrimination function. For this case, the ML classifier was almost a factor of 10 faster than the neural network (although no attempt was made to optimize the training of the neural network). This is in contrast to the study of Dowla *et al.*, (1990) which used complete spectra sampled at 41 points. In that case, the covariance matrix was very large and significant computer time was consumed in computing its inverse so the neural network was much faster than the ML classifier.

It should be noted, that in the simplest sense, the fundamental structure of the backpropagation neural network shown in Figure 11 is very similar to that of the maximum-likelihood classifier. To see this, we note that the maximum-likelihood classifier assuming equal covariances (equation 13) is of the form

$$\begin{aligned} G(\mathbf{v}) &= \mathbf{w}^T\mathbf{v} + \mathbf{w_o} \\ &= \sum_{i=1}^{d} w_i v_i + w_0 \end{aligned} \tag{16}$$

Equation (16) is the form of the simplest neural network called the single-layer perceptron used for two-category classification (e.g. Duda and Hart, 1973; Haykin, 1994; Figure 11). The structure of the perceptron is very similar to that of the backpropagation network used in this study. The principal differences are that the backpropagation learning algorithm uses a nonlinear log sigmoid transfer function (as opposed to a hard limit transfer function that outputs either a 1 or 0, depending on the output from the neuron of the perceptron) resulting in a different learning rule. For the ML classifier, the w_i are the coefficients of the discriminant function to be solved for. For the perceptron, the w_i are the synaptic weights (solved for by a given learning rule) that weight each input prior to summation (Figure 11). In both the ML classifier and the perceptron, the bias term, w_0, controls the location of the hyperplane separating the two populations.

Despite the similarities, there are a number of fundamental differences between the perceptron and the ML classifier that can give each an advantage or disadvantage depending on the problem to be solved. The way in which each algorithm solves for the weights is different. For the ML classifier, the weights are computed from the covariance structure of the data assuming a normal distribution. This involves computation and inversion of the data covariance matrix and can be a burdensome problem for large datasets (e.g. for the case of complete seismic spectra in Dowla *et al.,* 1990). The perceptron is nonparametric in that it makes no assumptions regarding the underlying probability distribution of the data. As discussed above, some of the seismic discriminants are not convincingly Gaussian. However, the overall performance of the ML and neural network classifiers was not significantly different (Table 2). A recursive learning rule is used for the perceptron where the weights are iteratively updated according to the rule (in its simplest form)

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \rho_n \sum_{\mathbf{v} \in \mathrm{E}_n} \mathbf{v} \tag{17}$$

where E_n is the set of events misclassified by $\mathbf{w}_n$ and ρ_n is the learning rate. Because the perceptron uses a hard limit transfer function, it only works well when the two populations are linearly separable and have essentially no overlap. This is not a problem, however, for nonlinear neural networks like that used in this study. Because of the recursive learning rule, the main advantage of the perceptron is that the storage requirements are greatly reduced over the ML algorithm.

CONCLUSIONS

The high-frequency P_g /L_g discriminant has been studied using NTS explosions and western U.S. earthquakes. The events used in the analysis are essentially the same as those used in previous studies (e.g. Taylor *et al.,* 1989; Dowla *et al.,* 1990). The study of Taylor *et al.,* (1989), however, did not incorporate the high-frequency P_g /L_g discriminant. The study of Dowla *et al.,* (1990) implicitly had the high frequency ratio in the sampled phase spectra and examining the weights of the neural network demonstrated that the network used this information (in addition to spectral shape) as part of its classification.

However, the data set was very large and would be inefficient to use for regional monitoring on a world-wide scale. In contrast, the high-frequency $P_g\ /L_g$ discriminant is comprised of relatively few measurements and provides discrimination performance that is as good as the previous studies (generally 95% correct classification). It is expected that if additional discriminants were added (e.g. m_b/M_s), performance would be further improved. As discussed above, the explosions that were misclassified using the high-frequency $P_g\ /L_g$ discriminant were of magnitude large enough to be classified by m_b/M_s or Love wave energy. The main drawback of the discriminant is that, because of signal-to-noise considerations and propagation effects, the number of measurements are reduced considerably at higher frequencies. It is expected that the problem will be amplified as magnitudes are reduced and event-receiver distances are increased. Future studies will involve analysis of the explosions having anomalous high-frequency $P_g\ /L_g$ ratios in order to obtain an improved understanding of the physical basis of this important discriminant.

The classification techniques used in this study will not be directly applicable to many regions of the world for CTBT monitoring because of the deficiency of nuclear explosions in the training set. Thus, the problem will be one of outlier detection (rather than classification) in which the training data consist of a single class (e.g. earthquakes). In this case, a hypothesis test is undertaken to determine whether a given event belongs to the same population as the training set (e.g. Fisk *et al.*, 1993).

APPENDIX A: Testing for Equality of Covariance Matrices

We briefly review the likelihood ratio test done to test whether the earthquake and covariance matrices are equal, details can be found in Anderson (1984). Using a maximum likelihood ratio criterion, we wish to test the null hypothesis

$$H_0 : \Sigma_X = \Sigma_Q = \Sigma \tag{A1}$$

versus the alternate hypothesis

$$H_1 : \Sigma_X \neq \Sigma_Q \tag{A2}$$

The likelihood function for the null hypothesis is

$$L_0 = \frac{1}{(2\pi)^{dN/2} |\Sigma|^{N/2}} \exp\left[-\frac{1}{2} \sum_{i=1}^{2} \sum_{k=1}^{n_i} (\mathbf{v}_{ik} - \mu_i)^T \Sigma^{-1} (\mathbf{v}_{ik} - \mu_i) \right] \tag{A3}$$

where $N = n_x + n_Q$ and d is the number of discriminants (dimension of Σ). The likelihood function for the alternate hypothesis is

$$L_1 = \prod_{i=1}^{2} \frac{1}{(2\pi)^{dn_i/2} |\Sigma|^{n_i/2}} \exp\left[-\frac{1}{2} \sum_{k=1}^{n_i} (\mathbf{v}_{ik} - \mu_i)^T \Sigma_i^{-1} (\mathbf{v}_{ik} - \mu_i) \right] \tag{A4}$$

The likelihood ratio for testing the null hypothesis is given by

$$\lambda_0 = \frac{\max L_0}{\max L_1} \tag{A5}$$

As shown by Anderson (1985), a new test statistic can be formed, $\lambda_c = -2\ln\lambda_0$, that has a chi-square distribution with $d(d+1)/2$ degrees of freedom. Using the chi-square test, we can reject H_0 if $\lambda_c \geq \lambda_\alpha$, where λ_α is the critical value such that the chi-square test has a significance level of α (i.e. α is the probablility that H_0 is rejected when it should be accepted).

For the filled P_g /L_g amplitude ratio data, λ_c was calculated to be 206.5 which is greater than $\lambda_\alpha = \chi^2(0.95, 21) = 32.7$ and we rejected the hypothesis of equal covariances at the 5% level of significance.

ACKNOWLEDGMENTS

Discussions with Bill Walter, Howard Patton, and Tom Weaver are greatly appreciated. Special thanks to Kevin Anderson for statistical advice. The paper was reviewed by Hans Hartse and Kevin Anderson. 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.

Baumgardt, D.R., and G.B. Young, Regional seismic waveform discriminants and case-based event identification using regional arrays, *Bull. Seism. Soc. Am.*, 80, 1874-1892, 1990.

Chan, W.W., R. Baumstark, and R.K. Cessaro, Spectral discrimination between explosions and earthquakes in central Eurasia, Air Force Geophysics Laboratory, GL-TR-90-0217, 36pp, 1990.

Chael, E.P., Spectral discrimination of NTS explosions and earthquakes in the southwestern United States using high-frequency regional data, *Geophys. Res. Lett.*, 15, 625-628, 1988.

Dempster, A.P., N.M. Laird, and D.B. Rubin, Maximum likelihood estimation from incomplete data via the EM algorithm, *J. Royal Stat. Soc.*, 39, 1-38, 1977.

Dowla, F.U., S. R. Taylor, and R.W. Anderson, Seismic discrimination with Artificial Neural Networks: Preliminary results with regional spectral data, *Bull. Seism. Soc. Am.*, 80, 1346-1373 1990.

Duda, R.O., and P.E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, 482pp, 1973.

Dysart, P.S. and J.J. Pulli, Regional seismic event classification at the NORESS Array: Seismological measurements and the use of trained neural networks, *Bull. Seism. Soc. Am.*, 80, 1910-1933, 1990.

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, 69pp, 1993.

Gupta, I.N., W.W. Chan, and R.A. Wagner, A comparison of regional phases from underground nuclear explosions at East Kazakh and Nevada Test Sites, *Bull. Seism. Soc. Am.*, 82, 352-382, 1992.

Hand, D.J., *Discrimination and Classification*, Wiley, New York, 218pp, 1981.

Hartse, H.E., S.R. Taylor, and G.E. Randall, Regional Event Discrimination in Western China, Trans. Am. Geophys. Un., 76, S204, 1995.

Haykin, S., *Neural Networks*, Macmillan College Publishing Co., Inc., Englewood Cliffs, NJ, 695pp., 1993.

Jarpe, S.P., Guide for using data from the LNN Seismic Monitoring Network, *Lawrence Livermore National Laboratory, Livermore, CA*, UCID-21872, 5pp, 1989.

Jones, E.M., F.N. App, and R.W. Whitaker, Ground motions and the infrasound signal: A new model and the discovery of a significant cavity rebound signal, *Los Alamos National Laboratory, Los Alamos, NM*, LAUR-93-861, 1993.

Lilwall, R.C., Regional m_b/M_S, L_g/P_g amplitude ratios and *Lg* spectral ratios as criteria for distinguishing between earthquakes and explosions: a theoretical study, *Geophys. J.*, 93,137-147, 1988.

Menke, W., *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press, Orlando, Florida, 259pp, 1984.

Patton, H.J., Seismic moment estimation and the scaling of the long-period explosion source spectrum, *Explosion Source Phenomenology, AGU Monograph* , 65, 171-183, 1991.

Patton, H.J. and S.R. Taylor, Analysis of spectral ratios from NTS explosions: Implications for the source mechanisms of spall and the generation of Lg waves, *Bull. Seism. Soc. Am.*, 85, 220-236, 1995.

Pomeroy, P.W., W.J. Best, and T.V. McEvilly, Test ban treaty verification with regional data-a review, *Bull. Seism. Soc. Am.*, 72, S89-S129, 1982.

Taylor, S.R., B.P. Bonner, and G. Zandt, Attenuation and scattering of broadband P and S waves across North America, *J. Geophys. Res.*, **91**, 7309-7325, 1986.

Taylor, S.R., N.W. Sherman, and M.D. Denny, Spectral discrimination between NTS explosions and western U. S. earthquakes at regional distances, *Bull. Seism. Soc. Am.*, **78,** 1563-1579, 1988.

Taylor S.R., M.D. Denny, E.S. Vergino, and R.E. Glaser, Regional discrimination between NTS explosions and western U.S. earthquakes, *Bull. Seism. Soc. Am.*,79, 1142-1176, 1989.

Taylor, S.R. and M.D. Denny, An analysis of spectral differences between NTS and Shagan River nuclear explosions, *J. Geophys. Res.*, 96, 6237-6245, 1991.

Taylor, S.R. and F.U. Dowla, Spectral yield estimation of NTS explosions, *Bull. Seism. Soc. Am.*, 81, 1292-1308, 1991.

Walter, W.R., K.M. Myeda, and H.J. Patton, Phase and spectral ratio discrimination between NTS earthquakes and explosions Part 1: Empirical observations, *Bull. Seism. Soc. Am.*, 85, 1050-1067, 1995.

Weisberg, S., *Applied Linear Regression*, Wiley, New York, 283 pp, 1980.

Vergino, E.S. and R.W. Mensing, Yield estimation using regional $m_b(P_n)$, *Bull. Seism. Soc. Am.*, 80, 656-674, 1990.

TABLE 1

REGRESSION COEFFICIENTS FOR Pg/Lg VERSUS $\log_{10}$(distance - km)

Band (Hz)	n	slope	interc ept	corr. coeff.	F

0.2 - 1	42	-0.73	1.25	0.21	-0.50	13.20
1 - 2	338	0.42	-1.55	0.24	0.29	31.50
2 - 4	285	0.36	-1.26	0.22	0.25	18.78
4 - 6	235	0.23	-0.84	0.23	0.15	5.56
6 - 8	185	0.10	-0.47	0.27	0.05	0.54
6 - 8*	170	0.45	-1.36	0.24	0.23	9.27
8 - 10	112	-0.02	-0.18	0.26	-0.01	0.02

* - regression performed only using events with distance less than 600 km

TABLE 2
PERFORMANCE OF DIFFERENT DISCRIMINATION ALGORITHMS

Algorithm	$P(Q\|X)$	$P(X\|Q)$	$P(X\|X)$	$P(Q\|Q)$	n_x	n_Q
$\Sigma_X = \Sigma_Q$	0.017	0.053	0.983	0.947	294	114
$\Sigma_X \neq \Sigma_Q$	0.041	0.044	0.959	0.956	294	114
ANN	0.022	0.053	0.978	0.947	294	114

FIGURE CAPTIONS

Figure 1. Map of the western U.S. showing western U.S. earthquakes used in this study, the Nevada Test Site (NTS) and the four broadband seismic stations operated by Lawrence Livermore National Laboratory.

Figure 2. $\log(L_g) - m_b$ for earthquakes plotted as a function of the logarithm of the distance, Δ, for 1-2 Hz (upper left) and 6-8 Hz (lower left) and fit from linear regression. Residuals from each of the regressions shown in right-hand column of the figure. The residuals are used as the distance-corrected amplitudes in subsequent analysis.

Figure 3. Logarithm of the P_g/L_g ratio and L_g spectral ratio for earthquakes plotted versus the logarithm of the distance, Δ, and linear fit with 95% confidence interval (left). Residuals representing distance-corrected ratios shown in the right column.

Figure 4. Logarithm of P_g/L_g ratios for six different frequency bands plotted as a function of magnitude.

Figure 5. Mean of the logarithm of P_g/L_g ratios and 1 standard deviation for six different frequency bands for earthquakes and explosions.

Figure 6. Histogram showing number of events used for P_g/L_g amplitude ratios as a function of frequency for explosions (top) and earthquakes (bottom). Note that a total of 294 explosions and 114 earthquakes were available for analysis.

Figure 7. Examples of different distance-corrected discriminants; 6 - 8 Hz P_g/L_g ratio (upper left), 6 - 8 Hz P_n/L_g ratio (upper right), L_g spectral ratio 1 - 2 over 6 - 8 Hz, P_g (1-2 Hz)/L_g (6-8 Hz) cross spectral ratio.

Figure 8. 6 to 8 Hz P_g/L_g discriminant for raw unfilled data (top) and data filled using nearest neighbor rule (bottom). See text for details.

Figure 9. Discrimination function plotted versus magnitude using equation (13). Zero line indicates division between explosions, $(G(\mathbf{v}) > 0)$, and earthquakes, $(G(\mathbf{v}) < 0)$, by assuming equal covariance matrix misclassification costs and prior probabilities of occurrence. Explosions below the line are classified as earthquakes (missed violations) and earthquakes above the line are classified as.explosions (false alarms).

Figure 10. Cumulative probability of misclassifying an explosion or earthquake as a function of magnitude for the case assuming equal covariances.

Figure 11. Schematic diagram showing network architecture of backpropagation neural network used to discriminate earthquakes and explosions on the basis of P_g/L_g amplitude ratios in six different frequency bands.