Bulletin of the Seismological Society of America, Vol. 96, No. 6, pp. 2368–2382, December 2006, doi: 10.1785/0120060039

Discrimination of Delay-Fired Mine Blasts in Wyoming Using an Automatic Time-Frequency Discriminant

by Stephen J. Arrowsmith,* Marie D. Arrowsmith, Michael A. H. Hedlin, and Brian Stump

Abstract Delay-fired mine blasts, which consist of a series of individual shots arranged in a grid pattern and detonated in sequence, can introduce spectral modulations into recorded seismograms. We can exploit spectral modulations to separate delay-fired mine blasts from the remaining event population, which includes single-fired mine blasts and earthquakes. Here, we enhance an existing algorithm (Hedlin, 1998) for the automatic discrimination of delay-fired mine blasts. A total of seven separate discriminants are computed, based on the spectrograms of recorded events. A feature-selection procedure is used to ensure that each discriminant is significant and contributes to the overall performance of the discrimination algorithm. The effect of input parameters on the methodology is explored. The choice of input parameters is made to maximize the mean Mahalanobis distance between the earthquake and delay-fired mine-blast populations. The technique is then applied to a dataset consisting of regional earthquakes and delay-fired mine blasts recorded at a station in Wyoming. The results show that the larger delay-fired mine blasts, the cast blasts, can be identified successfully by using this technique. The smaller mine blasts are not identified with this technique, although such events are of less interest in a nuclear-monitoring perspective. In a drop-one test, 89.5% of the events studied are successfully identified. Of the events that are misclassified, one is a cast blast and seven are earthquakes. The cast blast is misclassified because of noise on one component, which biased the value of a single discriminant. The earthquakes are misclassified because of a greater variance of the seven discriminants for the mine-blast population. The results suggest that this methodology is very successful at identifying cast blasts in Wyoming, and would be an extremely useful method to use as part of an integrated set of discriminants for the identification of small-magnitude regional events.

Introduction

A major challenge to effective monitoring of a Comprehensive Test Ban Treaty (CTBT) is the identification of events with body-wave magnitudes, m_b, less than 4.0. The advent of the International Monitoring System (IMS) provides an opportunity to detect and identify such events, which are lower in magnitude than they have been considered historically. Regional observations will comprise the bulk of detections for such lower-magnitude events and a new challenge will be the identification of mining explosions, which have magnitudes that fall into this category. The present challenge of seismic-event identification now includes the task of classifying not only earthquakes and single-fired explosions (nuclear or chemical), but also mining explosions, underground mining collapses, and rock bursts.

Many different types of explosions are used by the mining industry. A common technique employed in surface mining is delay firing, a technique in which individual boreholes in an explosive array are fired at different times in a regular sequence (Borg *et al.*, 1987). Typically, the explosive array consists of a number of rows, where the time differences between individual detonations on a single row (the intershot times) are much smaller than time differences between detonations of separate rows (the interrow times). The purpose of such a shot design is to maximize the fragmentation of the rock, while minimizing the ground motion (Dowding, 1985). In this study we analyze several different types of delay-fired mine explosion from a coal mine in Wyoming, including cast blasts, which are employed in surface mining to remove a layer of overburden material into an adjacent

*Present address: Los Alamos National Lab, EES-2, P.O. Box 1663, MS D401, Los Alamos, New Mexico, 87545; sarrowsmith@gmail.com

pit, blasts in the coal seams themselves (coal blasts), and parting shots, which are relatively small blasts of waste material.

Delay-fired mine explosions introduce time-independent spectral modulations into recorded seismograms. Several mechanisms have been proposed for explaining the origin of spectral modulations. These include the total duration of the entire set of explosives associated with delay-fired mining events, the dominant intershot time spacing, and the dominant interrow time spacing (Hedlin *et al.*, 1990). Intershot delays (which are 9 msec for the events discussed in this article) result in spectral modulations at very high frequencies, outside the bandwidth of regional observations. However, interrow delays are much longer (typically 200–1100 msec for the events discussed in this article) and lead to spectral modulations that can be recorded. Finally, the total durations of the shot sequences associated with some delay-fired mine blasts are conducive to causing spectral modulations (see discussion in Hedlin *et al.*, 1990). Spectral modulations have been exploited by Baumgardt and Ziegler (1988) and their time independence has been exploited by Hedlin *et al.* (1989, 1990) and Hedlin (1998) as a means of separating such delay-fired mine blasts from the earthquake population.

Hedlin (1998) introduced an algorithm based on the spectrograms of recorded data that could be used as an automated discriminant for delay-fired events. He applied the algorithm to data from several mining regions and showed that delay-fired mining events could be successfully separated from earthquakes by using this method. Although the algorithm in general, was successful in separating delay-fired mining events and earthquakes, it required some input parameters that were chosen in a somewhat ad hoc way. In this study we explore the effect of the choice of input parameters on the success of the discriminant. We enhance the methodology of Hedlin (1998) by using a more rigorous scheme for choosing values of the input parameters. We then show how such a scheme can improve the separation of earthquakes and cast blasts by applying the technique to data from a single station in Wyoming. The station selected for this study is located approximately 360 km from a coal mine in the Powder River basin, from which we have obtained ground-truth information for many explosions. In addition, the station is located in a region with a good distribution of natural seismicity, making this an excellent dataset for testing the methodology outlined in this article.

Methodology

Computation of Discriminants

The methodology developed in this study is designed to provide a set of discriminants for a single station that can be used to discriminate delay-fired mine blasts from other types of events. In this study we use a technique similar to the algorithm developed by Hedlin (1998) to compute a number of discriminants based on the spectrograms of recorded events. While the technique developed by Hedlin (1998) yielded nine discriminants for a single, three-component station, the methodology used in this study yields seven discriminants as described in the following text. Because there are some key differences between the two methods, we outline the full method used here for completeness. To aid in the explanation, Figures 1 and 2 illustrate the algorithm by using an example cast blast and an example earthquake.

The algorithm that we have developed is designed to exploit time-independent spectral modulations associated with delay-fired events. Therefore, the first step is to compute a spectrogram for each component of a three-component seismogram. The spectrogram durations begin at the first arrival (*Pn* or *Pg*) and extend through the *Lg*-wave coda (Fig. 1). The precise duration is set as a free parameter and is chosen automatically by the algorithm to optimize the separation of earthquakes and mine blasts. The durations of the spectograms for both the mine blast and earthquake in Figure 1 were 150 sec. The spectrograms are computed by using a short-time Fourier transform (STFT) with a 6.4-sec sliding Hanning window and 3.2-sec overlap. The choices of window length and overlap affect the trade-off between time and frequency resolution, and suitable values have been chosen for the time-frequency resolution of spectral modulations in this study. Although these parameters could be defined as free parameters, and optimized as outlined subsequently, we did not consider it necessary in this study. Each spectrogram is reduced to binary form by filtering each spectral estimate with two running-average filters that comprise different window lengths, and differencing the two filtered spectra (Fig. 1). The two separate window lengths (in hertz) of the running-average filters are set as a second set of free parameters, which are optimized automatically. The values of the two window lengths used for the running-average filters in Figure 1 were 0.39 Hz and 1.56 Hz. Locally high- and low-spectral values are then replaced with a 1 or 0, respectively, and the mean is removed. The purpose for converting each spectrogram to a binary spectrogram is to enhance the effect of time-independent spectral banding by whitening the spectra, and subsequently removing redundant signal. We have found that this step significantly improves the success of the algorithm in separating delay-fired mining explosions from earthquakes. Example binary spectrograms for a typical delay-fired cast blast and a typical earthquake are shown in Figure 1. Time-independent spectral modulations are clearly seen in the binary spectrogram for the cast blast, whereas the binary spectrum for the earthquake is essentially a random distribution of highs and lows. Clearly, it is straightforward to discriminate visually between the two types of events given these time-frequency displays. However, the algorithm outlined in this article can exploit the differences between the two types of time-frequency display automatically, so that the differences can be recognized as easily by numerical methods.

To reduce the binary spectrograms into discriminants

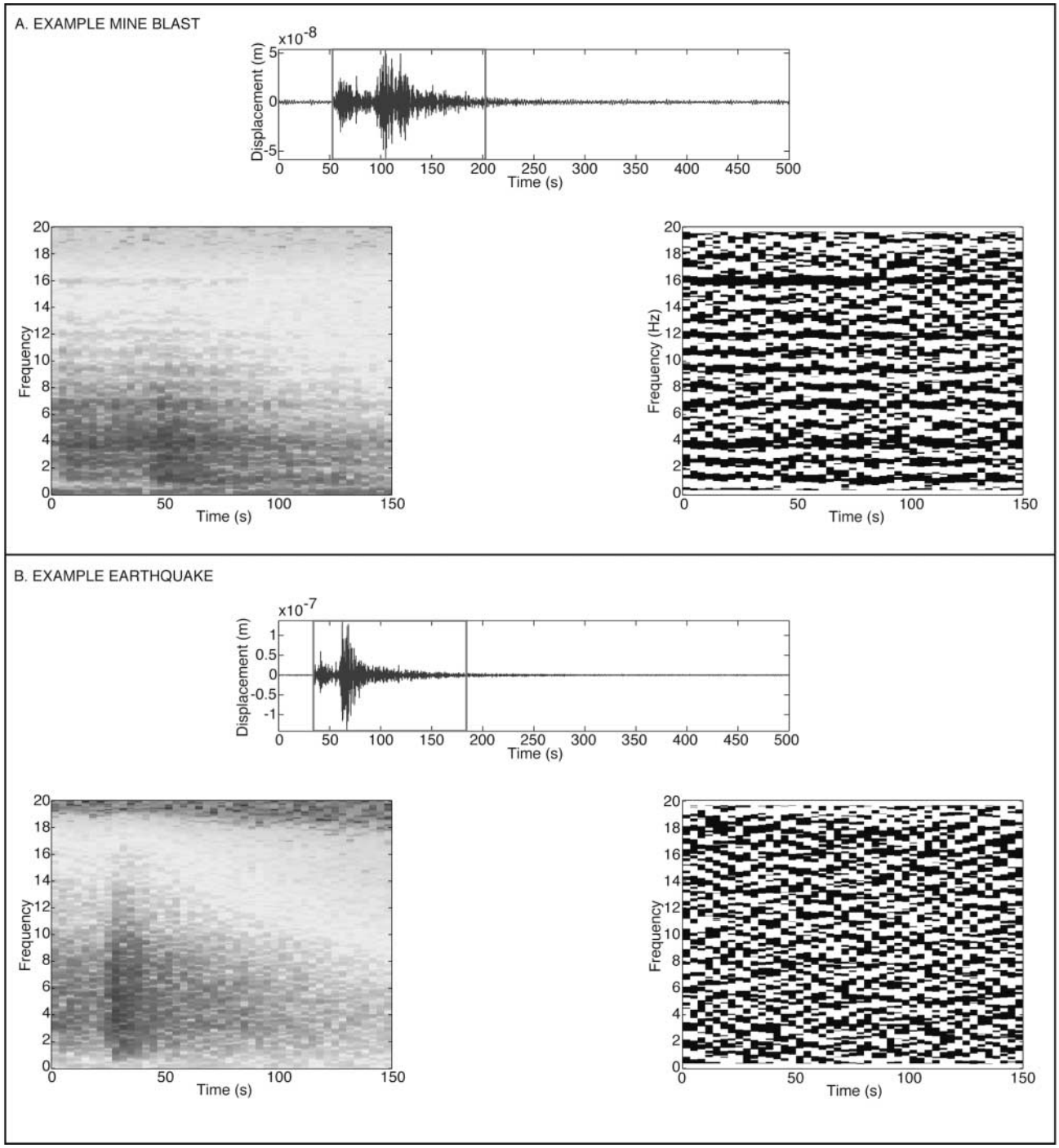

Figure 1. Input seismic waveforms with corresponding spectrograms and binary spectrograms for two example events. For each event, the portion of the waveform used in the computation of the spectrograms is shown. The first event, an example mine blast, exhibits clear spectral modulations that are time independent. The second event, which is a typical earthquake, shows no evidence of spectral modulations.

for automatic analysis, three types of discriminant are then computed: mean cepstral value, cross correlation, and autocorrelation. The estimates that go into the discriminants, which are derived from the binary spectrograms, are shown in Figure 2. The first type of discriminant is the mean cepstral value. To calculate the mean cepstral value we first compute the 2D Fourier transform of the binary spectrogram matrices, providing a 2D cepstral matrix. This can be considered to be an extension of cepstral analysis (Tribolet, 1979). In standard cepstral analysis a Fourier transform of

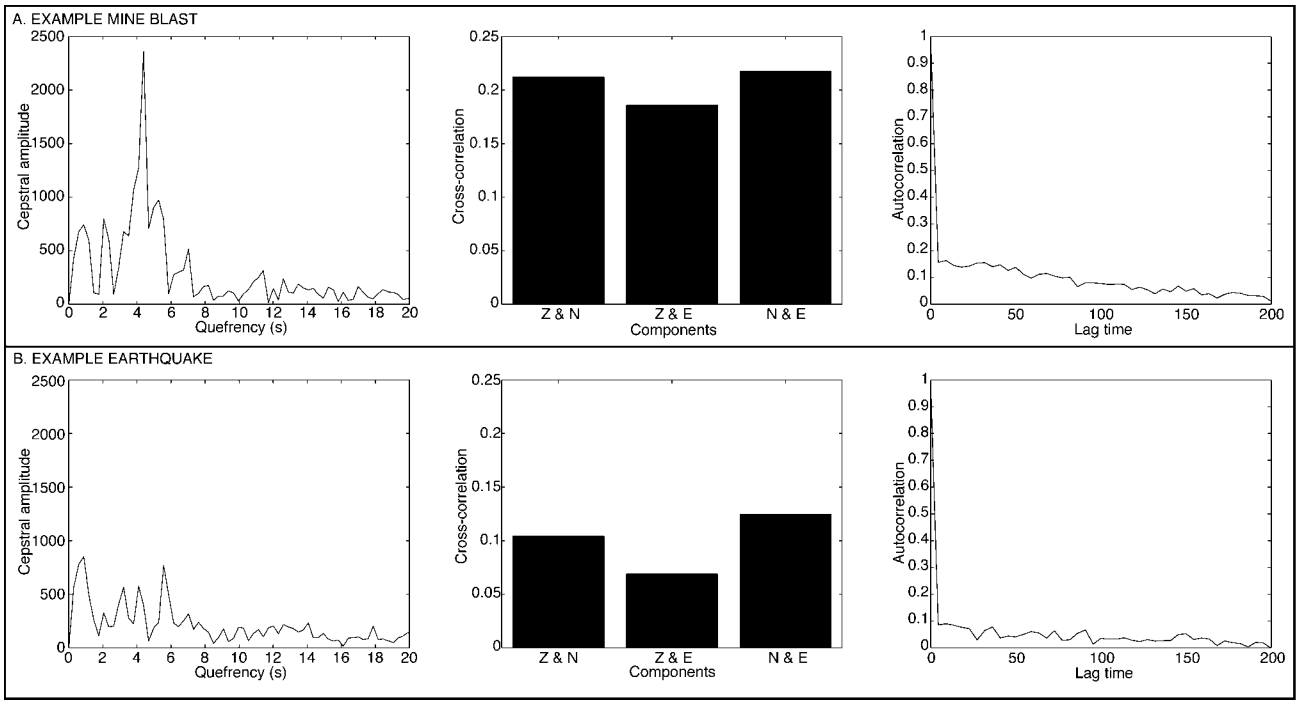

Figure 2. Estimates of cepstral amplitude, cross correlation, and autocorrelation for the two example events shown in Figure 1 (a, mine blast; b, earthquake). The left column shows the cepstral amplitude as a function of quefrency for the first 20 sec (north component only), the center column shows the estimates of cross correlation obtained for all three components, and the right column shows the estimates of autocorrelation as a function of lag time (north component only). Values of the discriminants, which are derived from these estimates, are given in Table 1.

the log of the amplitude spectrum is computed to highlight any regular spectral modulation regardless of its longevity. The independent variable is known as the "quefrency" and has units of time. In the form of cepstral analysis pursued in this article, a given point in the 2D cepstral matrix represents spectral modulation at a certain quefrency that is also periodic along the time axis at a given frequency. It is straightforward to isolate energy periodic in frequency and independent of time, yielding 1D cepstra (Fig. 2). The separate 1D cepstra from each of the individual components are then stacked. Because the high cepstral peaks observed in delay-fired events (e.g., Fig. 2) are associated with a source effect, they should correlate on all three components. However, cepstral peaks observed in earthquakes are typically uncorrelated on the three components, indicating time-independent spectral structure is not acquired during propagation. Therefore, by stacking the separate 1D cepstra, we improve the separation between earthquakes and delay-fired mine blasts because cepstral peaks observed in earthquakes will effectively be averaged out. Hedlin (1998) did not take this step, explaining why his technique yielded two additional discriminants to the technique described in this article. To calculate the mean cepstral value, we take the average value of the cepstrum over a window containing the low cepstral coefficients (e.g., over the window shown in Fig. 2). The window length used is set as a fourth free parameter in our algorithm, which is automatically set by the algorithm to optimize the separation of delay-fired mine blasts and earthquakes. The window length used to evaluate the cepstral means (which are given in Table 1) for the earthquake and mine blast shown in Figures 1 and 2 was 10 sec. We have found the mean value of the cepstrum to be a more successful discriminator than the maximum value (as used by Hedlin, 1998). This is because the filters used to convert spectrograms into binary form average cepstral peaks over a

Table 1
Values of Discriminants for the Example Mine Blast and Earthquake Shown in Figures 1 and 2

Discriminant	Mine Blast	Earthquake
Mean cepstrum	74.7127	64.1849
Cross correlation		
Z&N components	0.2121	0.1042
Z&E components	0.1857	0.0689
N&E components	0.2173	0.1248
Autocorrelation		
Z component	0.0977	0.0569
N component	0.1477	0.0550
E component	0.1098	0.0571

number of cepstral coefficients. The values of the mean cepstral discriminants computed for the example delay-fired mine blast and earthquake in Figures 1 and 2 are given in Table 1.

The second type of discriminant that is derived from the binary spectrograms is based on the fact that time-independent spectral modulations observed in delay-fired mine blasts should be independent of component, because the modulations are a source effect. On the other hand, the essentially random binary spectrograms associated with earthquakes are uncorrelated on the three components. To exploit this fact, we compute the cross-correlations of the binary spectrograms on the Z&N, Z&E, and N&E components, where three-component data are available. The 2D zero-lag cross correlation between two binary spectrograms (denoted E and Z for the east–west and vertical recordings of the same event) is given by:

$$C_{EZ} = \frac{\sum_{i=1}^{nfreq}\sum_{j=1}^{ntime} E(i,j)Z(i,j)}{\sqrt{\left(\sum_{i=1}^{nfreq}\sum_{j=1}^{ntime} E(i,j)^2\right)\left(\sum_{i=1}^{nfreq}\sum_{j=1}^{ntime} Z(i,j)^2\right)}}, \quad (1)$$

where *nfreq* is the number of samples in frequency and *ntime* is the number of samples in time. Delay-fired mine blasts are associated with relatively high cross-correlation coefficients, compared with earthquakes (e.g., Fig. 2 and Table 1).

The third type of discriminant that is derived from the binary spectrograms is based on the fact that spectral modulations observed in delay-fired mine blasts are independent of time. In contrast, binary spectrograms computed for earthquakes are constantly changing with respect to time (Fig. 1). To exploit this difference we calculate the autocorrelation of each 2D binary spectrogram as a function of time lag. This calculation is performed on the Z, N, and E components respectively. The 2D autocorrelation for the recording of an event on the east–west component (denoted *E* as before) is given by:

$$A_E(k) = \frac{\sum_{i=1}^{nfreq}\sum_{j=1}^{ntime-k} E(i,j)E(i,j+k)}{\sqrt{\left(\sum_{i=1}^{nfreq}\sum_{j=1}^{ntime} E(i,j)^2\right)}}, \quad (2)$$

where k is the time lag. Delay-fired mine blasts are associated with high values of autocorrelation relative to earthquakes (at lag times greater than zero) (e.g., Fig. 2 and Table 1). The algorithm averages the autocorrelation at all lag times to yield three values of mean autocorrelation, one for each component. The values of the mean autocorrelation discriminant for the example mine blast and earthquake in Figures 1 and 2 are given in Table 1.

In total, this methodology yields seven discriminants (mean cepstral value, three values of cross correlation, and three values of autocorrelation), which can be used to identify delay-fired mine blasts. The values of the discriminants calculated for the example mine blast and earthquake shown in Figures 1 and 2 are given in Table 1.

Optimizing the Choice of Input Parameters Using Multivariate Statistics

The methodology described previously requires that four independent input parameters be defined: spectrogram duration in seconds (w), two separate averaging windows (in hertz) for computing the binary spectrograms (*sp1* and *sp2*), and the window length (in seconds) to use in evaluating the cepstral mean (*cep*). The choice of these input parameters affects the success of the method in separating delay-fired mine blasts from the remaining event population.

In this study the input parameters are chosen to optimize the separation of delay-fired mine blasts from the remaining event population. The objective function used is the mean Mahalanobis distance in seven dimensions, using all seven discriminants to maximize the separation between earthquakes and delay-fired mine blasts. The Mahalanobis distance (e.g., Manly, 2005) from an observation $x' = (x_1, x_2, \ldots, x_p)'$ to the center of a group i is defined as:

$$\begin{aligned} D_i^2 &= (\mathbf{x} - \bar{\mathbf{x}}_i)' \, \mathbf{C}^{-1} \, (\mathbf{x} - \bar{\mathbf{x}}_i) \\ &= \sum_{r=1}^{p}\sum_{s=1}^{p} (x_r - \bar{x}_{ri}) c^{rs} (x_s - \bar{x}_{si}), \end{aligned} \quad (3)$$

where c^{rs} is the element in the rth row and sth column of the inverse covariance matrix $\mathbf{C}^{-1}$. The Mahalanobis distance is a commonly used multivariate approach to discrimination as it weights the distance between two groups by the variability within the groups themselves. It is superior to the Euclidean distance because it takes into account the distribution of the points in each group and because the distances are scale invariant (i.e., not dependent on the scale of measurements). Standard metrics would assign equal weight to each discriminant and therefore we would be making an inherent assumption that all discriminants were of equal quality. Furthermore, using standard metrics, greater weight would be given to discriminants associated with larger scales. As with many multivariate quantitative methods, the Mahalanobis distance can solve for multiple dimensions simultaneously.

The seven discriminants defined earlier exploit similar properties of the binary spectrograms: the regular pattern of spectral scalloping, its time independence, and the correlation on all three components. Therefore, the discriminants are not completely independent of each other, providing some redundant information about the nature of the source. We perform a simple feature-selection procedure to ensure that the addition of each discriminant is significant, and therefore that each discriminant contributes to the overall separation between delay-fired mining explosions and the remaining event population. The feature-selection procedure

also identifies the best combinations of d discriminants to use (where $1 \leq d \leq 7$). We start by identifying the best single discriminant (using the Mahalanobis distance as a measure of discriminant quality). Next, we identify the best combination of d discriminants, where $d = 2,3, \ldots,7$. At each step, we calculate an F statistic as a guide to determine whether the addition of the new discriminant is significant. The F statistic (Hand, 1981; Taylor and Hartse, 1997) is given by:

$$F \sim \frac{(n - d - 1)n_1 n_2 (D_d^2 - D_{d'}^2)}{(d - d')\left[n(n - 2) + n_1 n_2 D_{d'}^2\right]}, \qquad (4)$$

where n is the total number of events, n_1 is the number of earthquakes, n_2 is the number of delay-fired mine blasts, d is the total number of discriminants, and d' is the number of discriminants for a particular iteration. $D_d^2 - D_{d'}^2$ is the difference in the Mahalanobis distance on the subset of d and d' discriminants. F is then compared with the tabulated F distribution for $(d - d')$ and $(n - d - 1)$ degrees of freedom to examine whether the addition of the new discriminant is significant. Another possible approach to dealing with potentially correlated discriminants would be regularized discrimination analysis (Friedman, 1989). However, this type of approach is beyond the scope of this study.

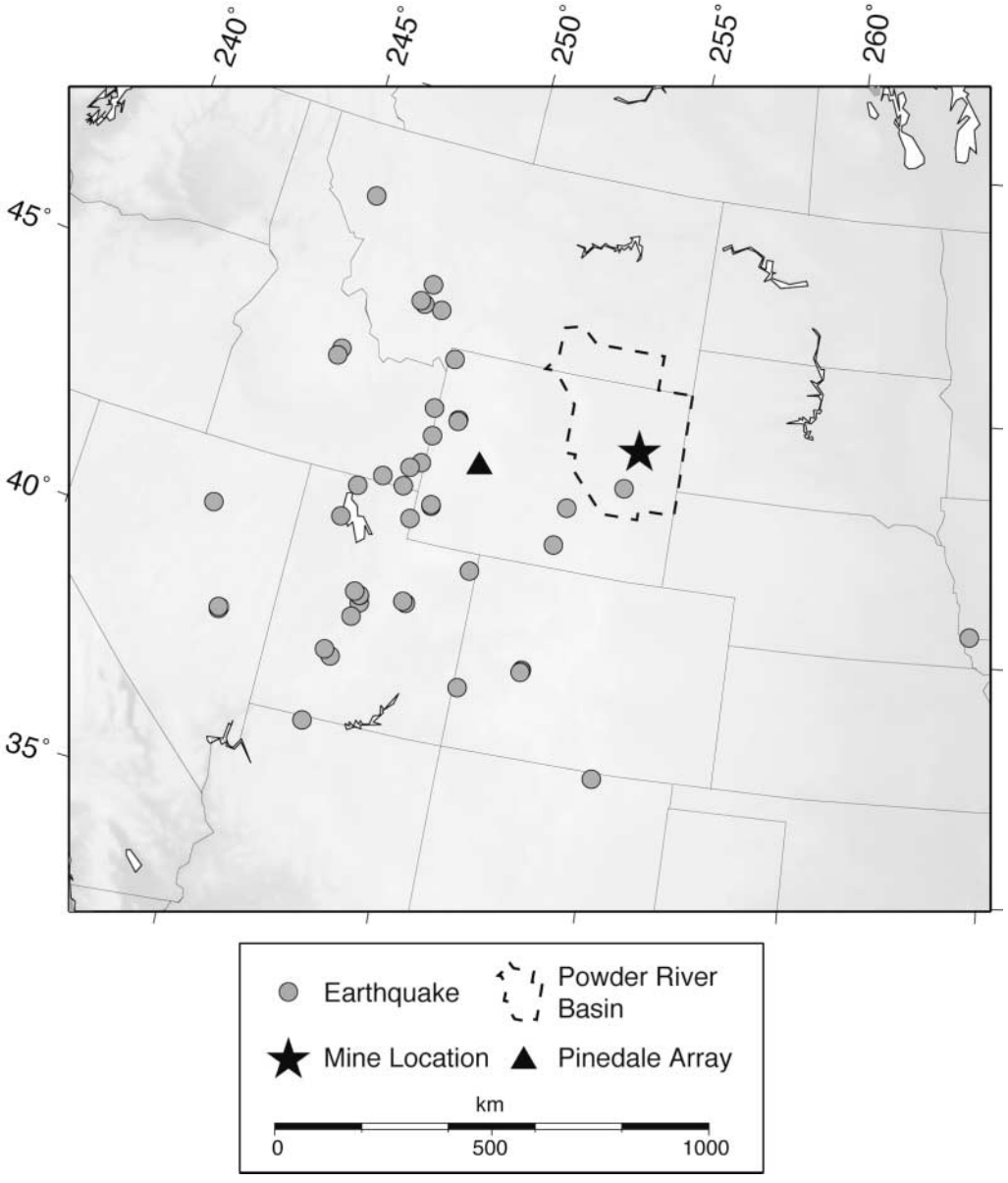

Figure 3. Map showing the locations of PD31 (black triangle), the mine location (black star), and locations of all 43 earthquakes (gray circles). The single mine location shown represents the locations of all 98 mine blasts. Background shading represents topography.

Table 2
Blast Types for the Large Coal Mine in the Powder River Basin, Wyoming

Blast Type	Description	Minimum Yield (lbs)	Maximum Yield (lbs)
Cast overburden (cast blasts)	Overburden is cast and removed into adjacent empty pit.	300,000	2,500,000
TS overburden	Blasts in overburden material to be excavated by shovels loading into trucks	100,000	600,000
Coal–main	Blasts in the main coal seam, which is 60–70 ft in thickness	20,000	200,000
Coal–upper	Blasts in the upper coal seam, which is 10 ft in thickness.	2000	10,000
Parting	Blasts of waste material layer between the upper and lower coal seam, ranging from 0 to 40 ft in thickness	200	500

Dataset

We have chosen to test the methodology on data from a single three-component station (PD31) that constitutes part of the Pinedale seismic array (PDAR) in Wyoming. PD31 is the only broadband station in the PDAR array, with a sampling rate of 40 Hz. In a future article we will discuss the development of the technique for arrays, and its subsequent application to the full Pinedale array. PD31 is located 367 km from one of the largest coal mines in the Powder River Basin, a region containing ~1 trillion tons of coal (Fig. 3). In addition, natural regional seismicity from Wyoming, Utah, Montana, and Idaho is recorded at PD31. Through cooperation with the mine, we have obtained ground-truth information comprising origin time and blast-type information for several events. There are five categories of event, which range in yield from 2.5 million to 200 Ibs (Table 2). The events typically have intershot times of 9 msec and interrow times that range from 200 to 1100 msec. The dataset comprises 98 regional recordings of mine blasts that could confidently be associated with the origin and location information from the mine (and generated sufficient signal to be seen at PD31). We have obtained origin times and locations for regional earthquakes from the U.S. Geological Survey (USGS) catalog and from earthquake catalogs compiled by the University of Utah and the Montana Bureau of Mines and Geology. There is a risk that some of these events may actually be mining events from other mines. The different catalogs use different criteria for categorizing events as mining events or earthquake. The USGS classifies events as mine blasts based on five criteria: location (whether an event occurs near or at a known mine location), time of day (mine explosions typically occur during local daylight hours), seismic waveforms (various criteria are used that include the relative amplitudes of seismic phases, the presence or absence of *S* or *Rg* and whether there is an emergent beginning of phases due to ripple firing), if events are not reported as felt (the calculated magnitudes of seismic events in some mining districts are large enough that, if the events were earthquakes, they would probably have been felt at nearby towns), and independent knowledge of operators of regional seismic networks. The other catalogs rely to a greater extent on information from mine operators and on location. Although these tests are comprehensive, they are not conclusive and there is always the possibility of misidentifying an event as an earthquake. We have obtained waveform data for 43 regional earthquakes that generated sufficient signal to be recorded at PD31.

We have picked regional *Pn, Pg*, and *Lg* arrivals for each event in the dataset at PD31. For the purpose of the algorithm described in this article, only the first onset time (*Pn* or *Pg*) is required. As discussed previously, the spectrogram duration is set as an input parameter and is chosen automatically to optimize the separation of delay-fired mine blasts from the remaining event population.

Results and Discussion

Tuning and Application of the Methodology to the Wyoming Dataset

For each class of mine blast described in Table 2, we have computed spectra using the Multitaper method (Thompson, 1982). We observe clear spectral scalloping for the largest blasts ("cast overburden" or "cast blasts") but little or no scalloping for the smaller blast types (Fig. 4). By applying the methodology discussed earlier to the dataset of 98 delay-fired mine blasts and 43 earthquakes recorded at PD31 we observe similar results. To maximize the separation of each blast type from the earthquake population, the values of the input parameters have been optimized separately for each type of blast. We find that the cast blasts separate well from the earthquakes, and that the truck-shovel (TS) overburden blasts show some separation, but the smaller blast types do not separate from earthquake population. Figure 5 summarizes this finding for a single discrim-

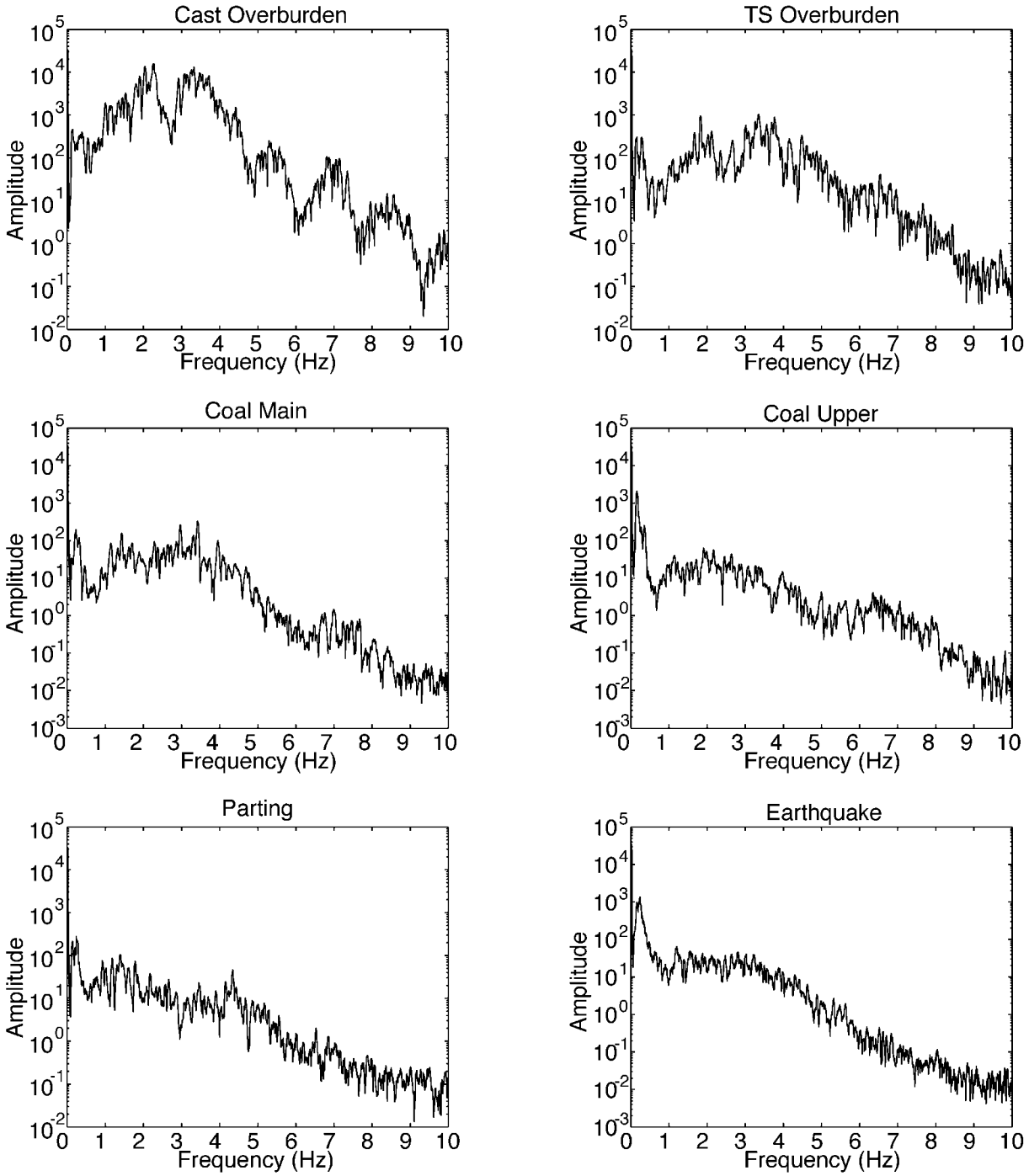

Figure 4. Typical spectra for each of the six types of event studied (Table 2). Clear spectral scalloping is observed for the cast-overburden blasts and there is evidence of scalloping in the TS-overburden blasts. The spectra for the other types of mine blast look more similar to the earthquake spectra.

inant (the cepstral mean). For the purposes of illustration, the values of the input parameters used in Figure 5 have been optimized for the separation of the cast blasts from the earthquakes. The objective function used was the Mahalanobis distance calculated with all seven discriminants. Different values of input parameters do not improve the separation of the smaller blast types from the earthquakes, but worsen the separation of the cast blasts from the earthquakes. It is clear that this technique is quite successful for identifying the larger cast blasts and TS overburden blasts, but that it cannot be used as a discriminant for the smaller blast types. However, from a nuclear-monitoring perspective, the large cast

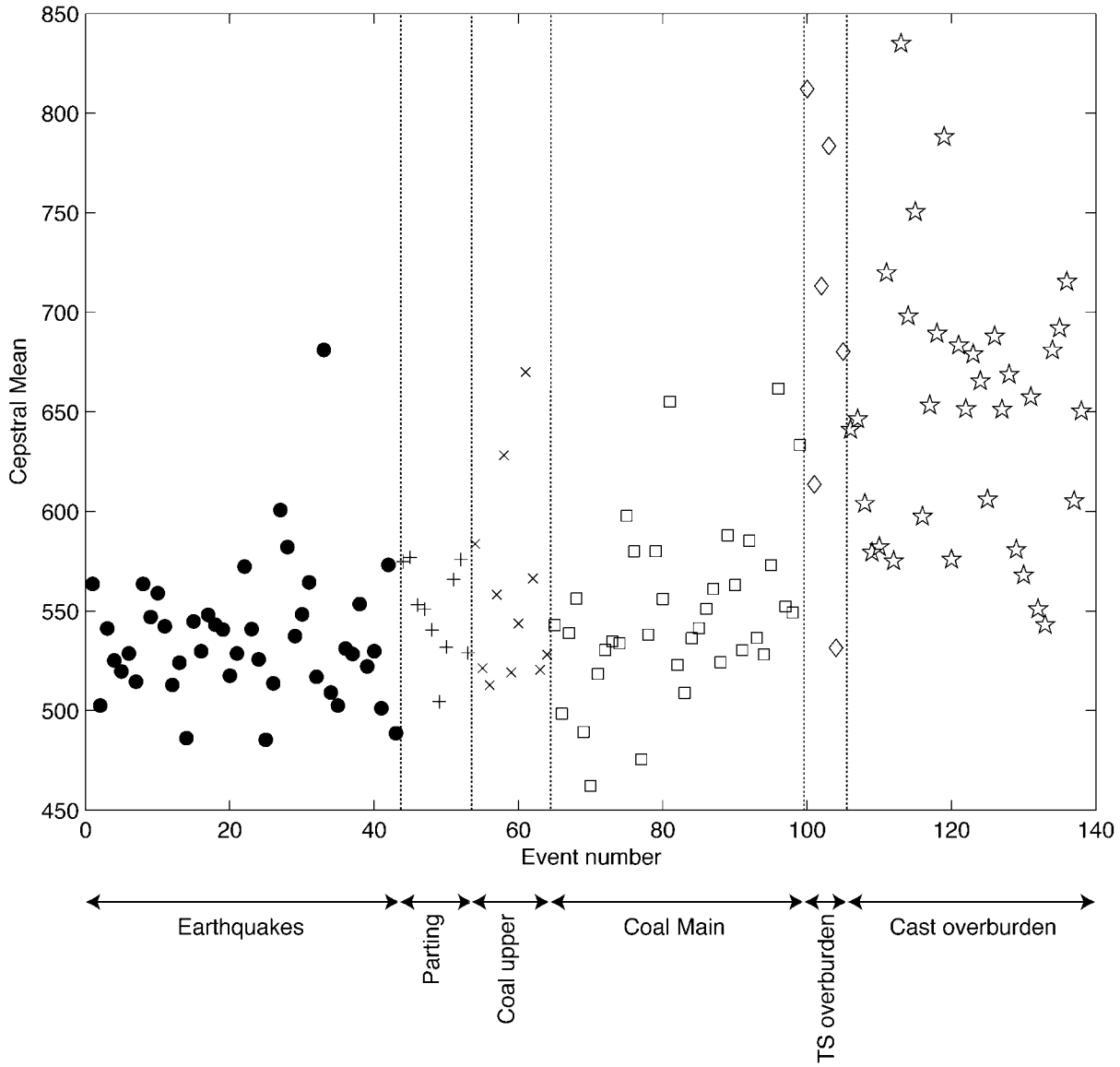

Figure 5. Values of the cepstral mean discriminant computed for each type of event. Earthquakes (black circles) typically have values between 500 and 575. Parting shots (plus signs), blasts in the upper coal seam (crosses), and blasts in the main coal seam (open squares) do not separate from the earthquake population. TS-overburden shots (open diamonds) show some evidence of separation. Cast-overburden shots (open stars) are clearly associated with larger values of cepstral mean than the earthquake population.

blasts are the sources of primary interest because they are most likely to be events that could cause a false alarm of the monitoring system. The smaller-magnitude events that do not separate from the earthquake population by using this technique would be too small to be identified as potential nuclear tests.

To illustrate how the values of the input parameters are optimized to separate the mine blasts from the earthquakes, we summarize the approach for the cast blasts next. For the cast blasts, the dataset consists of a total of 76 events (43 earthquakes, 33 cast blasts). We have tested a total of 1350 different combinations of input parameters. For each combination of input parameters, our objective function is the mean Mahalanobis distance between the earthquake and cast-blast groups, using all seven discriminants in the distance calculation (Fig. 6). Reasonable bounds for the four different input parameters (*w, sp1, sp2*, and *cep*) were defined and all possible combinations of parameters were tested (using a specified discretization). Although it may be possible to further improve the result by testing further combinations of input parameters (and refining the discretization), the computation quickly becomes very expensive. For

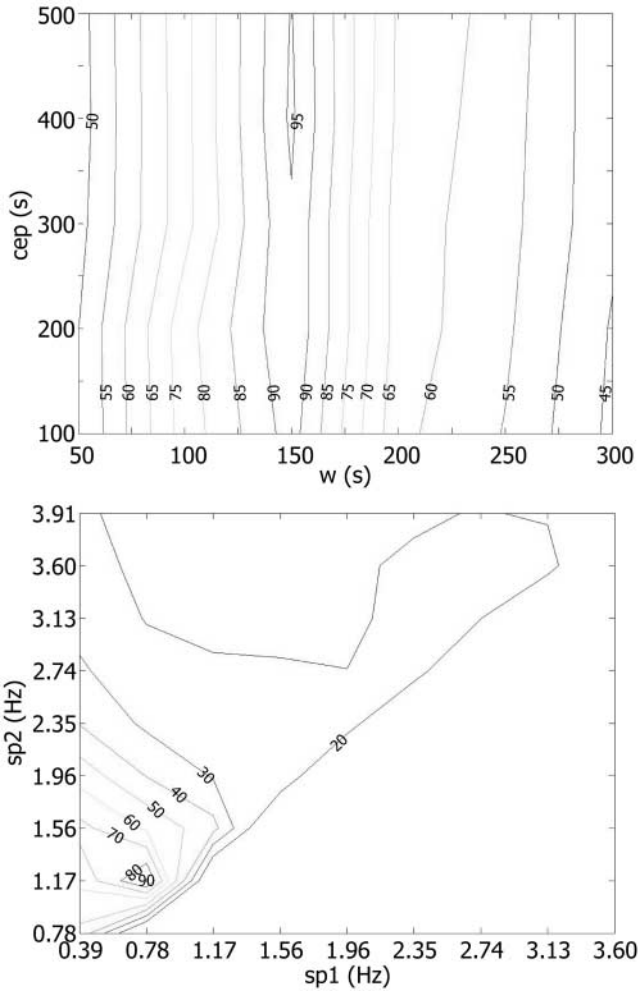

Figure 6. Cross sections of the objective function (Mahalanobis distance) that show the optimal input parameters. (Top) Mahalanobis distance as a function of *cep* and *w* with *sp1* and *sp2* held at their optimal values. (Bottom) Mahalanobis distance as a function of *sp1* and *sp2* with *cep* and *w* held at their optimal values. Note the requirement that $sp2 > sp1$.

1350 combinations of input parameters, the computation took ~120 hours on a Mac G5 (Dual 2 GHz processors). For the purpose of this article we aim to show simply the effect of the choice of input parameters, and how this choice may be optimized. Figure 6 shows cross sections of the objective function (i.e., the mean Mahalanobis distance) that illustrate how the procedure converges on the optimal input parameters. The calculated optimal input parameters are: $w = 150$ sec, $cep = 400$ sec, $sp1 = 0.78$ Hz, and $sp2 = 1.17$ Hz.

The results of the feature-selection procedure, which was described previously, are summarized in Figure 7 for the discrimination between the earthquake and cast-blast groups. The feature-selection procedure is used to determine the optimal combination of discriminants, and to ensure that all seven discriminants provide complimentary information. As discussed earlier, the procedure first finds the best single discriminant, followed by the best combination of two discriminants, the best combination of three discriminants, and so on. The top panel in Figure 7 indicates which discriminant or discriminants are selected after each step in which a new discriminant is added. The corresponding Mahalanobis distance and *F* statistic are also shown for each step to illustrate whether the addition of a new discriminant is significant. Clearly, the best discriminant is the autocorrelation on the east component (auto(E)), followed by the autocorrelation on the north component (auto(N)) and mean Cepstrum (mean Cep) in descending quality. Of the remaining discriminants, cross correlation between the north and east components (cc(N&E)) is the best followed by autocorrelation of the vertical component (auto(Z)), cross-correlation between the vertical and north components (cc(Z&N)), and finally cross correlation between the vertical and east components (cc(Z&E)). The overall Mahalanobis distance between the earthquakes and cast blasts increases as each new discriminant is added, but the final two discriminants (cc(Z&N) and cc(Z&E)) do not result in a significant increase in the Mahalanobis distance. This observation fits with the *F* statistics that show that five of the seven discriminants improve the discrimination significantly, but that the poorest two discriminants do not. Therefore, cc(Z&N) and cc(Z&E) could have been removed from the discrimination procedure without degrading the overall performance. However, the inclusion of the two discriminants does not degrade the Mahalanobis distance between the two groups and we have included them in our results.

The values of the discriminants that have been calculated using the optimum choice of input parameters are shown in Figure 8. Values of the cross-correlation coefficient are typically 0.06–0.16 for earthquakes and 0.14–0.24 for cast blasts. Values of the autocorrelation are typically 0.06–0.1 for earthquakes and 0.1–0.2 for the cast blasts. A couple of earthquakes, for example event 33, are associated with high values for a number of discriminants. Such values are more consistent with cast blasts, suggesting they may be misidentified in the earthquake catalogs.

Drop-One Event Identification Test

To test the ability of this methodology to successfully discriminate earthquakes and cast blasts in real time, we have performed a drop-one event identification test. Using the optimum choice of input parameters determined previously (Fig. 6), we have calculated the Mahalanobis distances (using all seven discriminants) from a single unknown event to both the earthquake and cast-blast groups, which comprise a total of $N - 1$ events (where $N = 76$ is the total number of events). The unknown event is then classified as being a member of whichever group it is closest to (using the Mahalanobis distance from the event to both groups). We have performed this computation *N* times, changing the unknown event each time. Using this approach, 89.5% of the events

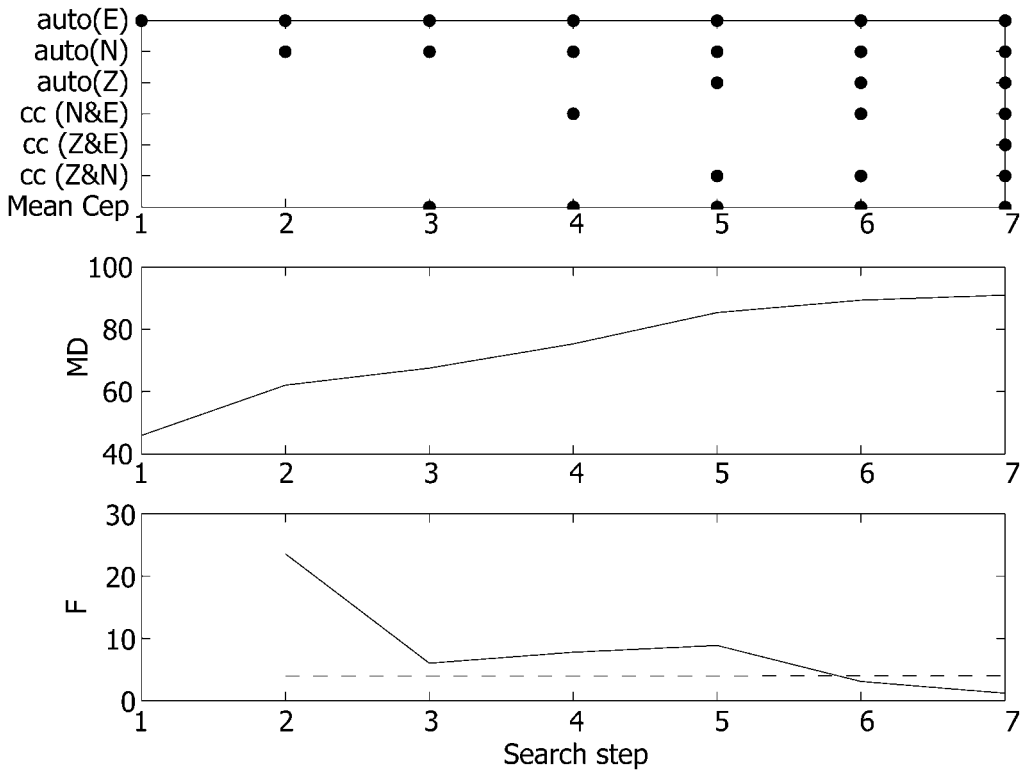

Figure 7. Results of the seven-step feature-selection procedure. Each step represents the optimum combination of d discriminants ($1 \leq d \leq 7$). (Top) Dots indicate the optimum combination of discriminants for each search step. (Middle) Mahalanobis distance between the earthquake and cast-blast groups calculated for each search step (using the appropriate number of discriminants). (Bottom) Calculated F statistic from equation (4) (solid line) and corresponding tabulated value of the F distribution for $(d - d')$ and $(n - d - 1)$ degrees of freedom (dashed line).

were correctly identified as either mine blasts or earthquakes and 10.5% of the events were incorrectly identified (Fig. 9).

The results shown in Figure 9 demonstrate that for cast blasts the Mahalanobis distances to the earthquake population become much larger than Mahalanobis distances to the cast-blast population. For earthquakes, Mahalanobis distances to the cast-blast population are generally greater, but not by such a large amount. Therefore there are more earthquakes that are misclassified as cast blasts (7) than cast blasts misclassified as earthquakes (1). This is because the amount of variation in the cast-blast population (reflected in the covariance matrices in equation 3) is greater than the amount of variation in the earthquake population. Therefore, whereas this method is very successful at correctly identifying cast blasts, it may result in the incorrect classification of earthquakes as cast blasts.

We have analyzed in detail the eight events that are misclassified in the drop-one event identification test. The single cast blast that was misclassified did exhibit spectral scalloping (Fig. 10), suggesting that the methodology should have successfully identified it as a cast blast. The reason that the event was misclassified was because of noise on the vertical-component recording (which contained 37% more noise than the north and east components). In fact, for a total of six of the seven discriminants, the event was closer to the cast-blast population than the earthquake population (Fig. 10). The only discriminant for which the earthquake population was closer was the autocorrelation on the vertical component. This illustrates the importance of the covariance matrices in weighting each discriminant by its associated variances and covariances when computing the Mahalanobis distance (equation 3). Discriminants with low variances are given a greater weight in the Mahalanobis distance calculation, while the correlations with other discriminants (measured by the covariances) also play an important role. Despite the fact only one discriminant (autocorrelation on the vertical component) places the event in the earthquake population, this discriminant is associated with very low variances of both the earthquake and cast-blast populations (e.g., Fig. 8). This effect is illustrated by the inverse covariance matrices of both the earthquake and cast-blast populations (Fig. 11), which effectively weight the relative importance of each discriminant (equation 3). From a statistical perspective it is logical to assign a greater weight to the autocorrelation on the vertical component as there is very little variation among all sources but a good separation between

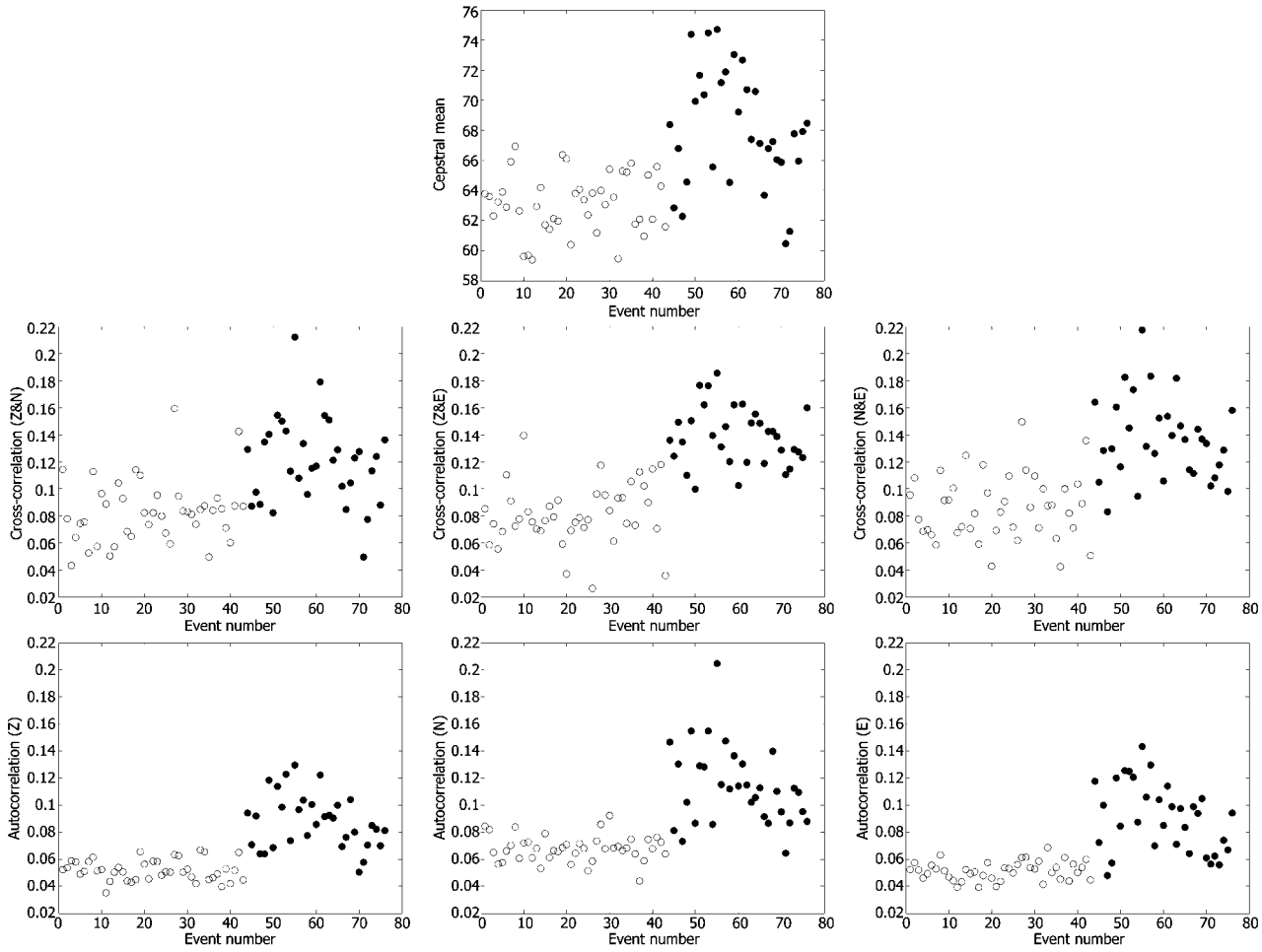

Figure 8. Values of each discriminant for all the earthquakes (white circles) and cast-overburden blasts (black circles).

the two groups. Note that the use of a poor discriminant (i.e., poor separation between the groups) with low variances of both groups could potentially swamp the distance calculation with noise and should be avoided. In this study, the misclassified cast blast is a consequence of unusually high noise on the vertical component, which biases the autocorrelation on the vertical component in this case. This problem could be overcome by correcting for noise on each component, which we will discuss in more detail in a forthcoming article.

The seven earthquakes that were misclassified occurred in different source regions. Therefore they are not associated with propagation-induced spectral scalloping, as observed by Sereno *et al.* (1985a,b). Furthermore, their spectra do not exhibit spectral scalloping as was observed for the misclassified mine blast. The reason for the earthquakes being misclassified is related to the larger variation in the cast-blast population (as discussed earlier). By combining this technique with additional discriminants (e.g., amplitude ratios, correlation, m_b/M_s), this problem should be overcome.

Portability to Different Regions

The methodology developed in this study is designed to identify delay-fired mining explosions in regional seismic datasets. As outlined earlier, the application of the method to a dataset of regional recordings of cast blasts and earthquakes in Wyoming has proven very successful. Note that this method would be portable to any region where delay-fired mining explosions are practiced. In support of this, Hedlin (1998) studied regional waveforms in six different mining regions and found evidence for time-independent spectral banding in each region. However, the optimal input parameters determined in this study would not necessarily be portable to different regions. The optimal choices of input parameters *sp1*, *sp2*, and *cep* are related to the typical blasting practices in a particular region; the spectrogram duration (w) is related to the distance of the recording station from a mining region. To apply this methodology in a new region, a set of known delay-fired mining explosions and earthquakes would be required with which to train the input parameters.

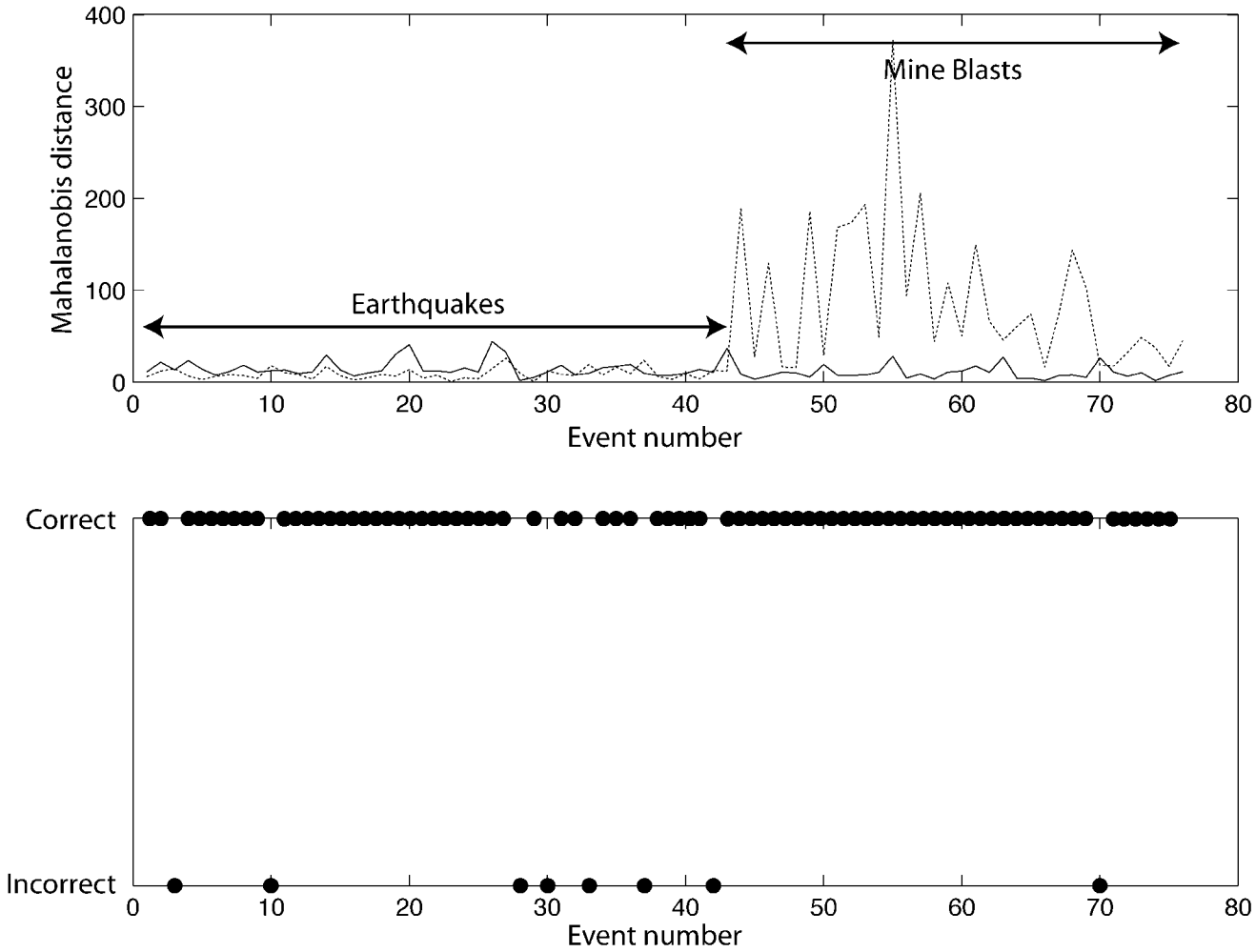

Figure 9. Results of the drop-one event identification test. (Top) Mahalanobis distances from earthquake (dashed line) and cast-blast (solid line) populations for each of the 76 events tested. Events 1–43 are earthquakes, and events 44–76 are cast blasts. (Bottom) Indicates for each event if it was correctly identified or not.

Conclusions

The identification of small-magnitude regional events as part of a CTBT is a current challenge that requires a fully automatic technique for the identification of delay-fired mine blasts. Such a technique could be used as part of an integrated suite of discriminants that would include additional algorithms for the identification of single-fired mine blasts, explosions, and earthquakes. In this article we have enhanced the fully automatic algorithm proposed by Hedlin (1998). Our new algorithm provides seven separate discriminants for separating delay-fired mining explosions from the remaining event population. We have applied the new algorithm to a dataset of regional delay-fired mine blasts and earthquakes from Wyoming. Our results show that the larger delay-fired mine blasts, which include both the cast blasts and TS overburden blasts, can be identified successfully by using this technique. The smaller mine blasts are not identified by using this technique. However, for the purpose of nuclear explosion monitoring we are mainly interested in the larger cast blasts, which have a greater potential to be misidentified as nuclear tests. We have shown how the choice of input parameters can have a crucial effect on the success of the discriminant for identifying cast blasts. Of the seven discriminants, the autocorrelation on the east component provides the best separation between earthquakes and cast blasts. The remaining discriminants, in descending order of quality, are autocorrelation on the north component, Cepstral mean, cross correlation between the north and east components, autocorrelation on the vertical component, cross correlation between the vertical and north components, and finally cross correlation between the vertical and east components. Five of the seven discriminants are significant in terms of improving the separation between the two event classes, but the inclusion of the remaining two discriminants does not degrade the overall separation. In a drop-one test,

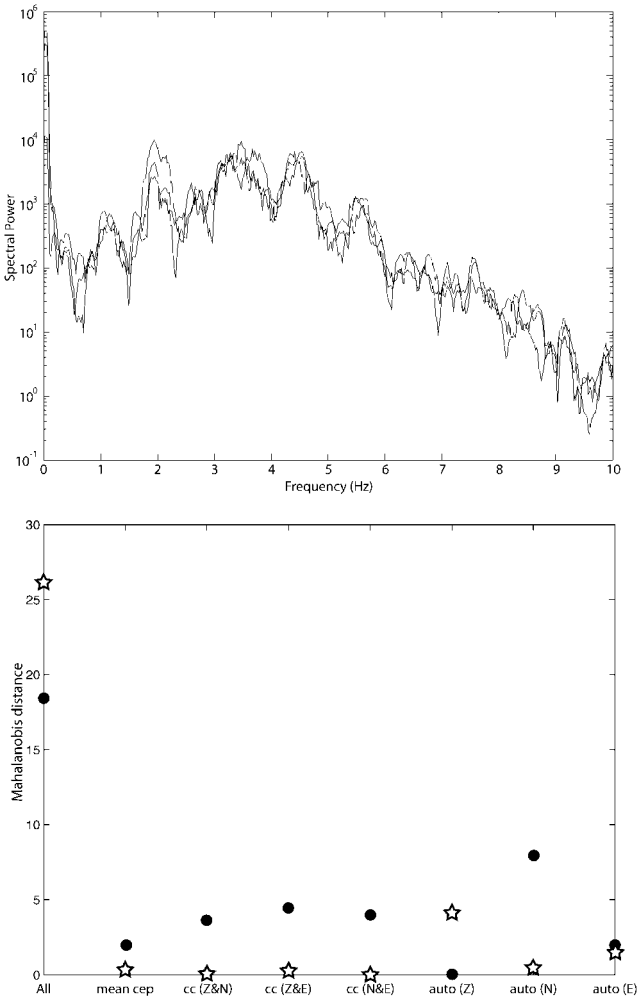

Figure 10. (Top) Vertical, north–south, and east–west components showing spectra of the one cast blast that was misclassified in the drop-one event identification test. The spectra show clear evidence of spectral scalloping (similar to the cast-overburden blast in Fig. 4). (Bottom) Mahalanobis distances to the earthquake (solid circles) and cast-blast (open stars) populations from the misclassified mine blast for all the discriminants together and for each discriminant separately.

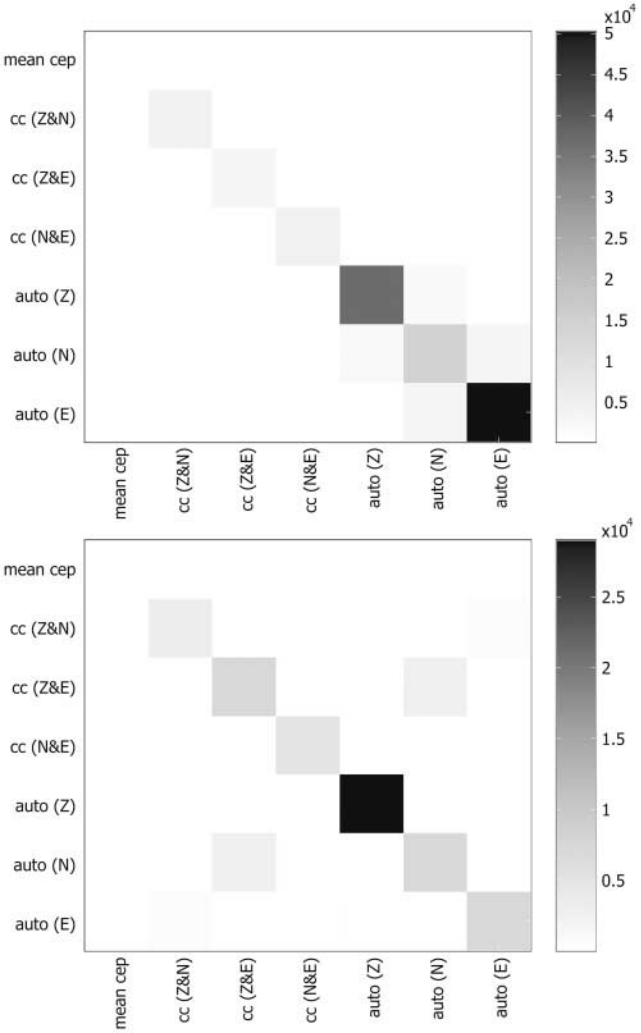

Figure 11. Inverse covariance matrices for the earthquake population (top) and cast-blast population (bottom), not including the one misclassified cast blast. Dark colors indicate low variances and covariances, which result in a greater weight in the Mahalanobis distance calculation. The distance from the earthquake population is dominated by autocorrelations on the east and vertical components, whereas the distance from the cast-blast population is dominated exclusively by the autocorrelation on the vertical component.

the method presented in this article successfully identifies 97% of cast blasts and 84% of earthquakes. The misclassified earthquakes are a consequence of the greater variances associated with the cast-blast population, which effectively decreases the Mahalanobis distances of earthquakes from the cast-blast population. It is also possible that some earthquakes may actually be delay-fired mine blasts, which were misidentified in the earthquake catalogs. The one misclassified cast blast is caused by noise on a single component, and would not have been misclassified if we were to weight each component by signal-to-noise ratio. Overall, these results suggest that the method outlined in this article would be a very useful part of a CTBT monitoring technology.

Acknowledgments

We are extremely grateful to Steve Beil at Arch Coal for providing mine shot information. We are also very grateful to Steve Taylor of Rocky Mountain Geophysics for his help with the feature-selection algorithm. We would also like to thank Dale Anderson, Keith Koper, and the two anonymous reviewers for their constructive criticisms, which helped to greatly improve the manuscript. The research was funded by National Nuclear Security Administration (NNSA) contract numbers DE-FC52-03NA99510 and DE-FC52-03NA99511. Data were obtained from the Incorporated Research Institutions for Seismology Data Management Center (IRIS DMC).

References

Baumgardt, D. R., and K. A. Ziegler (1988). Spectral evidence for source multiplicity in explosions: application to regional discrimination of earthquakes and explosions, *Bull. Seism. Soc. Am.* **78,** 1773–1795.

Borg, D. G., R. F. Chiappetta, R. C. Morhard, and V. A. Sterner (1987). *Explosives and Rock Blasting*, Atlas Powder Co., Dallas, Texas.

Dowding, C. H. (1985). *Blast Vibration Monitoring and Control*, Prentice-Hall, Englewood Cliffs, New Jersey.

Friedman, J. H. (1989). Regularized discriminant analysis, *J. Am. Stat. Assoc.* **84,** 165–175.

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

Hedlin, M. A. H. (1998). A global test of a time-frequency small-event discriminant, *Bull. Seism. Soc. Am.* **88,** 973–988.

Hedlin, M. A. H., J. B. Minster, and J. A. O. Orcutt (1989). The time-frequency characteristics of quarry blasts and calibration explosions recorded in Kazakstan, USSR, *Geophys. J. Int.* **99,** 109–121.

Hedlin, M. A. H., J. B. Minster, and J. A. O. Orcutt (1990). An automatic means to discriminate between earthquakes and quarry blasts, *Bull. Seism. Soc. Am.* **80,** 2143–2160.

Manly, B. F. J. (2005). *Multivariate Statistical Methods: A Primer*, CRC Press LLC, Boca Raton, Florida.

Sereno, T. J., and J. A. Orcutt (1985a). Synthetic seismogram modeling of the oceanic Pn phase, *Nature* **316,** 246–248.

Sereno, T. J., and J. A. Orcutt (1985b). Synthesis of realistic oceanic Pn wave trains, *J. Geophys. Res.* **90,** 12,755–12,776.

Taylor, S. R., and H. H. Hartse (1997). An evaluation of generalized likelihood ratio outlier detection to identification of seismic events in western China, *Bull. Seism. Soc. Am.* **87,** 824–831.

Thompson, D. J. (1982). Spectrum estimation and harmonic analysis, *IEEE Proc.* **70,** 1055–1096.

Tribolet, J. M. (1979). *Seismic Applications of Homomorphic Signal Processing*, Prentice-Hall Signal Processing Series, Prentice-Hall, New York.

Institute of Geophysics and Planetary Physics
Scripps Institution of Oceanography
University of California
San Diego, California 92093-0225
hedlin@epicenter.ucsd.edu
(S.J.A., M.A.H.H.)

Department of Geological Sciences
Dedman College
Southern Methodist University
P.O. Box 750395
Dallas, Texas 75275-0395
plutonium@gmail.com
bstump@smu.edu
(M.D.A., B.S.)

Manuscript received 22 February 2006.