MINING EXPLOSION IDENTIFICATION AND DISCRIMINANT ASSESSMENT IN TWO UNIQUE REGIONS

Marie D. Arrowsmith^{1,2}, Brian W. Stump², Stephen J. Arrowsmith¹, and Michael A. H. Hedlin³

Los Alamos National Laboratory¹, Southern Methodist University², and University of California—San Diego³

Sponsored by the National Nuclear Security Administration

Award No. DE-FC52-03NA99510 Proposal No. BAA03-38

ABSTRACT

With the contribution of data from high-quality seismic stations throughout the globe, the event detection threshold is reduced to smaller magnitudes. At magnitude 4 and below, delay-fired explosions are often commonly observed along with natural seismicity. The addition of this class within an event catalog leads to several questions which must be addressed: (1) can delay-fired mining explosions be discriminated from earthquakes; (2) can the understanding of mining discriminants illuminate the explosion discrimination process in regions where there are no known nuclear explosions; and (3) are mining discriminants regionally dependent, and are there corrections to account for this dependence? A broader question is in understanding the ability to separate source and propagation path effects from regional data. Addressing this question utilizing regional datasets has direct application to the problems noted above but more generally to our ability to interpret regional seismic records from all types of events including earthquakes. To address these types of questions, we have assembled a database focused on two regions, the Western United States (WUS) and the Altai-Sayan (AS) region of Russia, which are both areas of prolific mining activity. As part of an extensive collaboration with the largest coal mine in the WUS, we have detailed shot information for ~1000 mining events, classified into six distinct blast types; we have limited information for events in the AS.

We have applied three discriminants to data from 11 stations and one array in the WUS. The first discriminant (amplitude ratios) yields station centric results, although the largest mining events separate from earthquakes that are <250 km from the mine; as the earthquake dataset expands spatially, discrimination performance degrades. 1D path corrections provide improvement, but additional calibrations are necessary to optimize this discriminant. The second discriminant (time-frequency) separates the larger types of blasts with the longest source duration at all stations. Smaller blasts do not discriminate because of the shorter shot durations. The third discriminant (time-of-day) may have a secondary role in the discrimination of an individual event but may be quite useful in assessing man-made seismic activity in a regional context. We have utilized waveform correlation techniques to better understand how factors such as mining blast type and location within the mine are manifested in the waveforms. Initial results show good correlation between blast types within two main pits; as the correlation threshold is increased, we are able to resolve spatial location within individual pits for the simplest types of mining blasts. In the AS, we calculated these same discriminants for ~ 260 earthquakes and ~ 850 mining events. The amplitude ratio discriminant shows significant overlap of the earthquake and mining populations. Certain events do separate, but the lack of ground-truth makes these events difficult to identify. Similar results are seen for the time-frequency discriminant. We do not know if the discriminant itself fails, or if the majority of our data points are from smaller shots that have shorter time durations. Time-of-day results are similar to the WUS in that presumed mining events fall within working hours and indicate the assessment utility of this tool. These unanswered questions illustrate the need for detailed ground-truth information. Future studies of mining discrimination, particularly where large datasets are to be acquired, should involve cooperation with mine operators in order to address ambiguities such as those identified in the AS study. Although we see mixed results with the amplitude ratio discriminant, there is more success with the time-frequency and time-of-day discriminants.

No discriminant individually is able to successfully act as a surrogate for a single-fired explosion. However, the three discriminants, when used in combination, can provide a means of defining a delay-fired population region that could be integrated into a model such as the Event Classification Matrix (ECM) to aid in identifying events that do not fall within traditional nuclear explosion or earthquake population bounds (Anderson et al., 2007). We have begun testing this methodology by using classification trees and Regularized Discrimination Analysis (RDA) to combine the three discriminants discussed above, which yield a statistical measure of the probability of correctly categorizing mining events.

OBJECTIVES

The two objectives for the final component of this study build on previous work (e.g., Arrowsmith et al., 2008) completed on developing both a comprehensive mining explosion dataset and a suite of mining explosion discriminants. These objectives are the following:

- Assessment of four mining explosion discriminants using ground-truth data in the WUS and AS. We assess four independent discriminants including amplitude ratios, time frequency, time-of-day, and waveform correlation.
- Assessment of techniques for discriminant combination. We examine two strategies for combining discriminants including classification trees and RDA.

RESEARCH ACCOMPLISHED

Mining Explosion and Regional Earthquake Database

In order to assess discriminant performance, we assembled a database of waveforms for mining events in the US and Russia, discussed in detail in Arrowsmith et al. (2008). In-depth analysis has focused on ground-truth (GT) data in the WUS (Figure 1) and data collected as part of a joint collaboration with the AS Seismological Expedition (ASSE) (Figure 2).



Figure 1. WUS dataset used in the mining discrimination analysis. Left panel shows earthquakes (yellow circles) distributed across major geologic provinces (thick blue lines) and sub-provinces (thin blue lines). Stations are denoted by black triangles. The Black Thunder coal mine (red star) is in the Powder River Basin (red outline). The right panel illustrates GT data collected from Black Thunder; the top figure denotes the main types of shots performed at the mine, while the bottom figure shows specific shot information.



Figure 2. Data from the AS portion of the database (red outline). (Left) Seismic stations in the AS region. Green stations are IRIS (BRVK, KURK, MAKZ, TLY) or International Monitoring System (IMS) (ZAL). Blue stations are operated by the ASSE. We did not have access to ASSE seismic data. (Right) Events provided to the authors by the ASSE; we were given no information on how events were identified as either mining explosions (red circles) or earthquakes (yellow circles).

Magnitude and Distance Amplitude Corrected (MDAC) Spectral Ratios

Regional phase amplitudes are highly influenced by both propagation path and source size. In order to correct for these effects prior to forming ratios, we utilized the MDAC methodology described in Walter and Taylor (2002). The MDAC software will accept simple 1D attenuation (Q_0) models or 2D tomographic imagery; we used 1D averages based upon the geologic features and regions noted in Figure 1 to constrain reasonable attenuation boundaries. Full results obtained for Q_0 and the frequency scaling factor η can be found in Arrowsmith (2009). To correct for source size, we solved for stress drop, varying the values from 0.1 to 30 MPa for the earthquakes featured in Figure 1; we obtained an overall value of 0.1 MPa to best fit the data.

In order to test whether or not the MDAC methodology was adequately removing trends related to distance and source size, we used an F-test to compare two scenarios: the null hypothesis assumes a de-trended population and the alternative hypothesis assumes a trend to the data. The F-test considers the variance around each trend-line; if the tabulated F-statistic (F_{tab} ; from a standard table) is greater than the calculated F-statistic (F_{obs} ; from the data) at the 0.95 confidence level, we accept the null hypothesis that MDAC has successfully removed the trends. Figure 3 illustrates one result for station PD31 in the 6.0- to 8.0-Hz band. In this case, the trend related to the source has been removed, but the trend related to distance has not (associated Q_0 values are: Pn = 445, Pg = 345, Lg = 315), suggesting that the complicated paths seen in Figure 1 are not being well fit by a 1D attenuation model. Because the stress drop is very low, we forced the stress drop to be 3 MPa (Figure 4) and solved for attenuation (associated Q_0 values are: Pn = 240, Pg = 140, Lg = 140). While the attenuation model fits the data better, the values are below average for the region (see Arrowsmith [2009] for a comprehensive description of WUS attenuation, including references to other studies). The stress drop value of 3 MPa does not fit the data for Lg, but results in a de-trended population for both Pn and Pg. Both sets of results in Figure 3 and 4 illustrate the tradeoffs between source and path seen in the MDAC inversion, and the sensitivity of the population trend with the choice of parameters used in the correction process.



Figure 3. Corrected earthquake log(amplitudes) for Pn (denoted as P), Pg, and Lg at station PD31 in the 6.0–8.0 Hz frequency band for an MDAC stress drop of 0.1 MPa. The lefthand column is a function of m_b ; the righthand column is a function of distance (km). A regression line has been plotted for each population and the result of the F-statistic analyses are shown for each scenario, including F_{obs} , F_{tab} , and the associated p-value.



Figure 4. Corrected earthquake log(amplitudes) for Pn (denoted as P), Pg, and Lg at station PD31 in the 6.0–8.0 Hz frequency band for an MDAC stress drop of 3 MPa. The lefthand column is a function of m_b ; the righthand column is a function of distance (km). A regression line has been plotted for each population and the result of the F-statistic analyses are shown for each scenario, including F_{obs} , F_{tab} , and the associated *p*-value.

We have used the corrected data (using 0.1 MPa stress drop and associated attenuation) to form phase, spectral, and cross-spectral ratios for the eight stations featured in Figure 1. In order to find the optimal ratio at each station, we have used the Mahalanobis distance, a multivariate statistical approach. The Mahalanobis distance weights the distance between two populations based upon the population variability. We pre-selected 12 discriminants, and determined the best amplitude ratio by maximizing the Mahalanobis distance between the earthquake and mining explosion populations. This analysis was performed for four cases: all earthquakes and all mining explosions, all earthquakes and only cast blasts, earthquakes within 250 km of Black Thunder and all mining explosions, and earthquakes within 250 km from Black Thunder and only cast blasts. Results for all stations are summarized in Table 1. The results are very station-centric, and there is high dependence on path when considering earthquakes in the entire region. However, when earthquakes within 250 km of the mine are considered, discrimination performance is greatly enhanced at most stations, making the amplitude ratio discriminant a promising one for mining explosions.

 Table 1. The best-performing discriminants for WUS stations for the four scenarios described in the text. The top line in each cell indicates the discriminant used and the bottom line indicates the corresponding Mahalanobis distance value. Results are color-coded by general geologic regime, as defined by the authors based upon geologic provinces featured in Figure 1.

Station	Quakes	Quakes	Quakes (250 km)	Quakes (250 km)
	All Mining	Cast Blasts	All Mining	Cast Blasts
AHID	Pn(6-8)/Lg(6-8)	Pn(6-8)/Lg(6-8)	Pg(6-8)/Lg(6-8)	Pg(6-8)/Lg(6-8)
	0.4572	0.0338	18.2054	18.2054
BOZ	Pn(4-6)/Pn(1-2)	Pn(4-6)/Pn(1-2)	Pn(6-8)/Lg(6-8)	Pn(6-8)/Lg(6-8)
	1.5995	1.7336	67.0649	63.4165
BW06	Pg(6-8)/Lg(1-2)	Pg(6-8)/Lg(1-2)	Pg(6-8)/Pg(1-2)	Pg(6-8)/Lg(6-8)
	0.6104	0.9644	53.2512	30.3084
CBKS	Pn(4-6)/Lg(4-6)	Pn(4-6)/Lg(4-6)	Pn(4-6)/Lg(4-6)	Pn(4-6)/Lg(4-6)
	1.1324	1.0315	11.0146	9.3419
ISCO	Pg(4-6)/Pg(1-2)	Pg(4-6)/Pg(1-2)	Lg(6-8)/Lg(1-2)	Lg(6-8)/Lg(1-2)
	1.2961	1.3241	88.7212	90.1181
MSO	Pg(6-8)/Pg(1-2)	Pg(6-8)/Pg(1-2)	Pg(4-6)/Pg(1-2)	Pg(4-6)/Pg(1-2)
	1.6058	1.6058	6.7304	6.7304
PD31	Pn(4-6)/Pn(1-2)	Pn(4-6)/Lg(4-6)	Pn(1-2)/Lg(1-2)	Pn(1-2)/Lg(1-2)
	0.3745	0.1433	9.1353	4.6621
RSSD	Pg(6-8)/Lg(6-8)	Pg(6-8)/Lg(6-8)	Pg(6-8)/Lg(6-8)	Pg(6-8)/Lg(6-8)
	7.9723	15.5696	99.0786	90.261

Time Frequency

The time frequency discriminant utilizes the unique spectral signature of ripple-fired mining events as a function of time and has been described in detail in previous publications (e.g., Arrowsmith et al., 2006; 2007; 2008). Larger blasts (cast blasts) discriminate very well from the earthquake population, and station location seems to play no role in the success of the discriminant, consistent with the fact that this discriminant is strongly related to the distinctive spectral shape resulting from the long source duration. Smaller blasts merge with the earthquake population, and as there are many small blasts in the dataset, it would be easy to assume the discriminant was not working if we did not have the additional explosion features summarized in Figure 1 (right panel). Being able to constrain the type of shot, as well as other types of information on the blasting configuration is invaluable in assessing discriminant success (Figure 5).



Figure 5. Comparison between AS time frequency results (left) and WUS time frequency results for five stations (right). In the AS, we have no ground-truth information to constrain event type, while in the WUS, we have information such as that shown in the right panel of Figure 1 (events are color-coded as in that panel). In the WUS, we have determined that the larger blast types, such as cast blasts, are essentially the only types of mining explosions that can effectively be discriminated using this technique. Because we do not have detailed GT information in the AS region, we do not know if the discriminant itself fails, or if the shots simply are too small in size (and have a short time duration) to effectively discriminate from the earthquake population.

The Time-of-Day Discriminant

The time-of-day discriminant is extremely useful as a means of verifying catalog consistency (Figure 6) and evaluating the general trends of an active seismic region (Figure 7). While this discriminant alone cannot be used to identify a specific event, it has the potential to help characterize new regions where man-made events are prevalent. When combined with a geographic context, such as in our study, and that of MacCarthy et al. (2008), we gain the ability to focus on areas of specific interest that have an abundance of daytime activity.



Figure 6. Time-of-day distributions with respect to month and hour of day for the AS dataset for both probable mining explosions (left) and earthquakes (right). Surface reflections of the histograms show trends, where lighter colors indicate larger numbers of events. The concentration of mining events between 10:00 and 17:00 local time as opposed to the random distribution of earthquakes give us confidence in the catalog data provided to us by the ASSE.



Figure 7. Time-of-day map for events between 8 am and 6 pm local time in the WUS. Event locations are mapped into 1° bins. The daytime count is listed above the nighttime count, and bins are color coded by the percentage of daytime to nighttime events. Areas of high daytime activity, which do not correspond to known mining regions, are discussed in Arrowsmith (2009).

Waveform Correlation for the Black Thunder Mine

A clustering method based on regional waveforms has been developed and applied to a set of ground-truth data from Black Thunder; this mine has spatial dimensions exceeding 5 km (Figure 7) and a variety of source blasting practices provide a basis for investigating the strength of these tools in separating source timing and location effects. Preliminary results at a number of stations suggest that a hybrid technique utilizing multiple correlation cutoffs and a clustering algorithm might be useful in separately identifying location and source timing effects. At modest levels of correlation, source locations are empirically shown to dominate. Using these results, increasingly larger correlation cutoffs identify nearby events with similar source timing effects. Figures 8 and 9 illustrate the result of changing the correlation coefficient threshold from 0.7 to 0.6.



Figure 7. Google Earth satellite image of the mining region, showing close-up of North pits and distances between mining clusters.



Figure 7. Clustering results for 60 seconds of waveform at RSSD, filtered at 2–4 Hz, with a correlation coefficient of 0.7. Events are colored according to the scheme in Figure 1. The various pit identifiers are listed in the left column.



Figure 8. Clustering results for 60 seconds of waveform at RSSD, filtered at 2–4 Hz, with a correlation coefficient of 0.6. Events are colored according to the scheme in Figure 1. The various pit identifiers are listed in the left column.

Discriminant Combination

Combining discriminant measures has been shown to increase overall classification power; however, difficulty lies in combining multiple measures in a statistically sound way while accounting for problems such as missing data. In a comprehensive report, Anderson et al. (1996) investigated eight classification methods for combining discriminants. Current discussions with Dr. Dale Anderson about our dataset have affirmed two methods, classification trees and RDA (Anderson and Taylor, 2002) to be the most suitable for merging discriminant measures. The classification tree considers each discriminant value, partitioning event types into homogenous regions. The process is iterative, and can be branched or pruned based upon the number of discriminants, events, and probability levels achieved at each step in the iteration.

We follow the procedure outlined in Anderson and Taylor (2002) to establish RDA decision rules for identifying events. RDA is a method developed by Friedman (1989) to deal with highly correlated discriminants, and potentially small training samples. RDA encompasses two standard discrimination techniques, linear and quadratic discrimination, and builds a weighted-average covariance matrix. This covariance matrix is dependent on two smoothing parameters, γ and λ . Figure 3 of Anderson and Taylor (2002) illustrates the effect of adjusting γ and λ in order to best define decision regions that fit the true probability of the event classes. An unknown event can then be identified using these decision regions in combination with the Mahalanobis distance metric.

To begin defining decision regions for mining explosions, we have integrated data at all the WUS stations into one single database of measurements consisting of event ID, time-frequency discriminant value, time-of-day, Pg/Lg (6–8 Hz) measurement, geologic region, source type (SEQ, OEX, MEX), where SEQ is a shallow earthquake, OEX is a mining explosion other than cast blast, and MEX is a cast blast. We use a classification tree to characterize probable earthquakes based on time-of-day. Events that fall within the range of hour GMT 18 through 24 (the time when most mining events occur at Black Thunder) are considered unknown events and are set aside for further analysis using RDA.

We searched over γ and λ space using 80% of the dataset to select optimal values, where optimization is achieved by minimizing the false discovery rate (i.e., the false positive rate, given H₀ = explosion). For each matrix of values, we select the minimum γ and λ pair. If there is more than pair that fits the condition, we choose the value that is closest to the origin. We then used 20% of the dataset to test values and make predictions on our ability to categorize events, repeating 100 times for the best γ and λ . The initial results indicate the mean probability of correctly categorizing a mining explosion given a mining explosion is approximately 0.75.

Although further work is needed, the three discriminants, when used in combination, show the potential for defining a delay-fired population region that could be integrated into a model such as the ECM to aid in identifying events that do not fall within traditional nuclear explosion or earthquake population bounds (Anderson et al., 2007).

CONCLUSIONS AND RECOMMENDATIONS

The optimum discriminant for mining explosions would uniquely categorize all mining events, and clearly separate them from earthquakes and, more importantly, underground nuclear explosions. Such a discriminant does not appear to exist; the complex and variable nature of mining explosions, and complexity of path effects at regional distances ensure that this is a difficult problem. Despite this, we have outlined a suite of independent discriminants—each of which provides a different constraint on source physics—that, in combination, can correctly categorize events as mining explosions with a probability of approximately 0.75. Although further work is required to improve upon this result, this study shows that there is considerable promise in developing, if not the optimum discriminant, at least a practical approach that can bring strategic value to operational monitoring needs.

ACKNOWLEDGEMENTS

Data was provided by the IRIS Data Management Center and supplemented by Los Alamos National Laboratory (LANL) and Lawrence Livermore National Laboratory. Additionally, LANL provided computing and advising support throughout the project. Arch Coal gave us access to invaluable data from Black Thunder and Dale Anderson assisted with discriminant combination efforts. The National Science Foundation's Graduate Research Fellowship program provided supplemental assistance for the completion of the companion PhD dissertation (Arrowsmith, 2009) that resulted from this work.

REFERENCES

- Anderson, D. N., D. Fagan, M. Tinker, G. Kraft, and K. Hutchenson (2007). A mathematical statistics formulation of the teleseismic explosion identification problem with multiple discriminants, in *Proceedings of the 29th Monitoring Research Review: Ground-Based Nuclear Explosion Monitoring Technologies*, LA-UR-07-5613, Vol. 1, pp. 526–530.
- Anderson, D. N. and Taylor, S. R. (2002). Application of regularized discrimination analysis to regional seismic event identification, *Bull. Seismol. Soc. Am.* 92: 2391–2399.
- Anderson, D. N, K. K. Anderson, D. N. Hagedorn, K. T. Higbee, N. E. Miller, T. Redgate, and A. C. Rohay (1996). Statistical Classification Methods Applied to Seismic Discrimination, *Pacific Northwest National Laboratory Report*, PNNL-11192US, i–27.
- Arrowsmith, M. D. (2009). Mining Explosion Identification as an Application to Treaty Verification, PhD dissertation, Southern Methodist University, 519 p.
- Arrowsmith, M. D., S. J. Arrowsmith, B. W. Stump, and M. Hedlin (2008). A suite of discriminants for ground-truth mining events in the Western U.S. and its implications for discrimination capability in Russia, in *Proceedings of the 2008 Monitoring Research Review: Ground-Based Nuclear Explosion Monitoring Technologies*, LA-UR-08-05261, Vol. 1, pp. 544–553.
- Arrowsmith, S. J., M. A. H. Hedlin, M. D. Arrowsmith, and B. W. Stump (2007). Identification of delay-fired mining events using seismic arrays: Application to the PDAR array in Wyoming, USA, *Bull. Seismol. Soc. Am.* 97: 989–1001.
- Arrowsmith, S. J., M. D. Arrowsmith, M. A. H. Hedlin, and B. Stump (2006). Discrimination of delay-fired mine blasts in Wyoming using an automatic time-frequency discriminant, *Bull. Seismol. Soc. Am.* 96: 2368–2382.
- Friedman, J. (1989). Regularized discrimination analysis, J. Am. Stat. Assoc. 84: 165-175.
- MacCarthy, J. K., H. E. Hartse, M. Greene, and C. Rowe (2008). Using waveform cross-correlation and satellite to identify repeating mine blasts in Eastern Kazakhstan, *Seismol. Res. Lett.* 79: 393–399.
- Walter, W. R. and S. R. Taylor (2002). A Revised Magnitude and Distance Amplitude Correction (MDAC2) Procedure for Regional Seismic Discriminants: Theory and Testing at NTS, *Lawrence Livermore National Laboratory*, UCRL-ID-146882, 1–13.