APPLICATION OF REGULARIZED DISCRIMINATION ANALYSIS TO REGIONAL SEISMIC EVENT IDENTIFICATION

Dale N. Anderson and Steven R. Taylor

Pacific Northwest National Laboratory
Los Alamos National Laboratory

Sponsored by U.S. Department of Energy
Office of Nonproliferation Research and Engineering
Office of Defense Nuclear Nonproliferation
National Nuclear Security Administration

Contract No. DE-AC06-76RLO-1830
Contract No. W-7405-ENG-36

ABSTRACT

We present a generalized multivariate seismic event identification method, Regularized Discrimination Analysis (RDA) [Friedman 1989], that can be applied to a large number of regional discriminants. RDA is readily adaptable to an outlier or classical identification approach to regional seismic identification. RDA is designed to address the problems associated with linear (LDA) and quadratic (QDA) discrimination in small-sample, high-dimensional settings. RDA includes LDA, QDA and Euclidean distance based nearest neighbor discrimination in its parameterization. RDA can be used to transition from an outlier analysis approach to seismic identification to classical discrimination as quality explosion calibration data are collected. Further, RDA provides the statistical structure to model highly correlated seismic measurements. We demonstrate the importance of including the correlation structure between seismic measurements in event identification. Not including this correlation structure in any identification framework can aggravate identification errors and give an erroneous impression of capability. With RDA, a large number of amplitudes from a Magnitude and Distance Amplitude Correction (MDAC) analysis [see Taylor et al. 1999] can be used and no a priori sub-selection of amplitudes (or discriminants) is necessary.

Key Words: discrimination, singular covariance matrix, outlier analysis.

OBJECTIVE

Not accounting for the dependence between individual seismic discriminants in any identification process can aggravate identification errors and give an erroneous impression of capability. For example, weight is positively correlated with height. An individual who is 6 feet tall and weighs 120 lb. might reasonably be viewed as unusual. However, taken alone, it is not unusual to find someone who is 6 feet tall, nor is it unusual for someone to weigh 120 lb. It is the inconsistency with the correlated behavior of height and weight that makes a 6 foot, 120-lb. person an *outlier*.

Two discriminants with a strong positive correlation provide redundant information about the source of a seismic event (strongly correlated discriminants vary, in a probabilistic sense, together). For example, if X and Y are discriminants and the correlation between X and Y is ρ, then the variance of $X+Y$ is

$$Var(X+Y) = {\sigma_X}^2 + {\sigma_Y}^2 + 2\rho\sigma_X\sigma_Y. \tag{1}$$

The $Var(X+Y)$ increases linearly in ρ. It is reasonable to conjecture that many regional discriminants will be positively correlated. Combining correlated discriminants with a sum, and computing the variance as if they are uncorrelated, may not be a good aggregation method. For a measured event discriminant $X = x$, a p-value can be computed as the conditional probability of observing a discriminant value equal to or more extreme than x. It is important to note that a p-value is a random variable because it is a function of a random variable. For observed discriminants X and Y, we can compute the p-values p_x and p_y and then aggregate this marginal information with the product $p_{Aggregate} = p_x\, p_y$. However, aggregating with a product can also be dangerous. For example, for bivariate normal random variables X, Y the variance of the product XY is

$$Var(XY) = \mu_Y{}^2\sigma_X{}^2 + \mu_X{}^2\sigma_Y{}^2 + 2\rho\mu_X\mu_Y\sigma_X\sigma_Y + (1+\rho^2)\sigma_X{}^2\sigma_Y{}^2. \qquad (2)$$

Here, the linear correlation between X and Y is ρ. $Var(XY)$ is a polynomial in ρ with no real-valued zeros ($Var(XY) > 0$) and a minimum at one of the values 1, -1 or $-\mu_X\mu_Y/\sigma_X\sigma_Y$. Values of ρ that increase $Var(XY)$ are governed by $-\mu_X\mu_Y/\sigma_X\sigma_Y$. This dependence will propagate into p-value calculations as well.

One of the main points of this paper is that seismic discriminants should be aggregated, to the best degree possible, with a statistical likelihood or probability model. A likelihood-based approach to combining discriminants provides a rigorous method to properly account for correlation between discriminants. The most desirable event identification framework would be composed of discriminants that are independent of each other, yet strongly indicative of the source of a seismic event. Independent discriminants contribute in a purely additive (orthogonal) way to the identification of an event, and never carry redundant information. For example, a principal components analysis (PCA) can be used to construct linear combinations of amplitudes that are orthogonal. In the traditional application of PCA, a subset (dimension reduction) of these linear combinations is used to construct a discrimination rule. The use of PCA in discrimination analysis has no limiting statistical deficiencies; however, we feel that the PCA approach presents some seismological concerns. Any feature selection analysis on amplitudes, including PCA, defacto constructs discriminants that may or may not have a known physical basis. Thus, the primary two or three PCA linear combinations may not be the best discriminants for a Mahalanobis distance based discrimination rule (see McLachlan, 1992, page 197 for a statistical basis for this observation.).

In the regional setting, a PCA will likely be based on earthquake data with no explosion data. This means that the PCA linear projections may do a poor job of combining explosion amplitude information, because $\Sigma_{EX} \neq \Sigma_{EQ}$ (in other words, important earthquake versus explosion discriminants may not be included in the final PCA linear amplitude combinations). If all the PCA linear combinations from Σ_{EQ} are used in an outlier analysis, then there will be no loss of information; however, this approach is conceptually equivalent to the regularized discrimination analysis (RDA) we present. RDA does not construct potentially controversial linear combinations as in a PCA. In a regional setting, the goal of independent discriminants, without a PCA type analysis is probably not possible because different seismic phases may share similar apparent source spectra and may overlap in time (e.g., Lg spectra may be contaminated by Sn coda). Figure 1 illustrates a fabricated model for two discriminants X and Y.

Figure 1a gives the bivariate ellipsoid that encloses 95% of the data from a particular source (the gray points). The Gaussian curves on the top and right sides of Figure 1a are the marginal densities for this model. The black point on the graph is clearly not a member of the population of gray points. However, neither of the marginal representations indicates that this point is unusual. Figures 1b and 1c show a transition from the region (black) that will include most all of the gray data to an outlier region. The circle, superimposed on Figure 1b, represents a source elimination rule that is constructed by assuming no correlation between X and Y. The ellipse, superimposed on Figure 1c, represents a source elimination rule that is constructed with the inclusion of correlation between X and Y. The most disturbing observation is the potential for identification errors when the no-correlation rule is used. In this case, the region outside of a decision rule defines false alarms and the darker region interior to a decision rule defines missed-explosions.

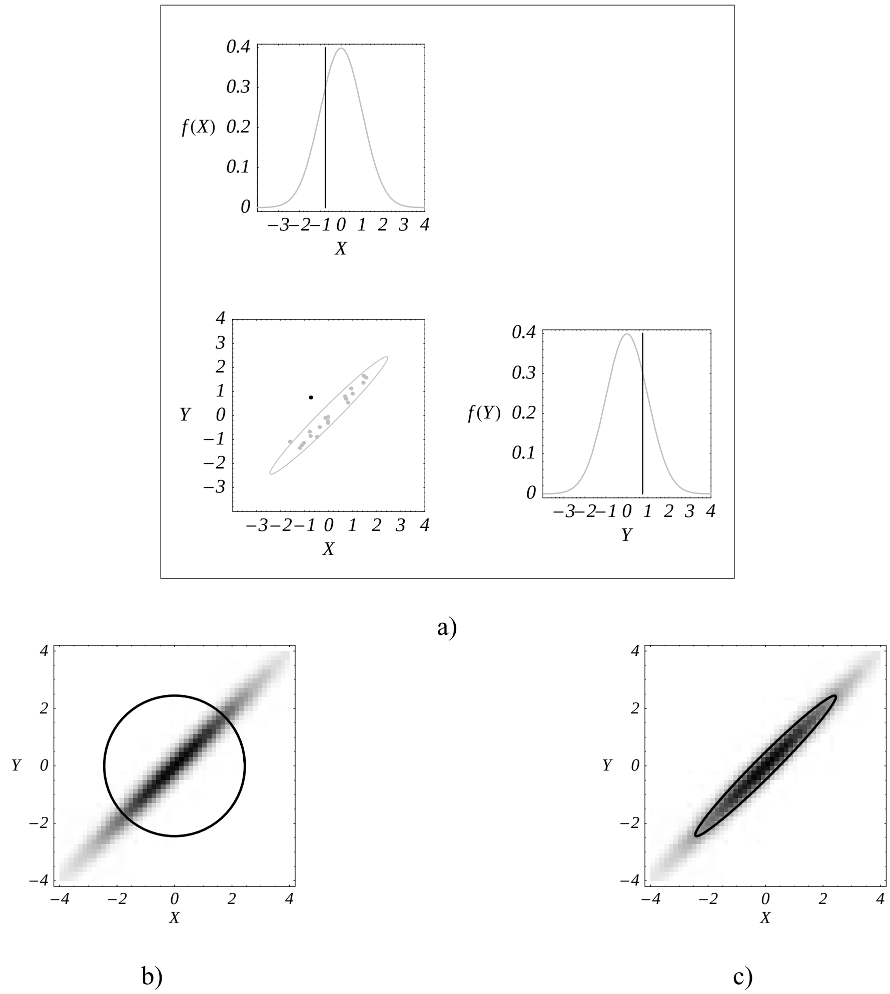


Figure 1. Fabricated model for discriminants X and Y and correlation and no-correlation based decision rules.

OUTLIER DETECTION ANALYSIS

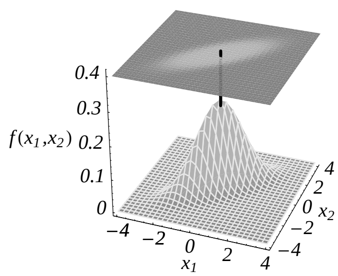
The methodology for outlier detection in the seismic context is well established (see Fisk et al. 1996; Taylor and Hartse 1997). Classical multipopulation discrimination methodologies may not be well suited to nuclear test monitoring for two main reasons. First, existing or planned seismic stations that will be used for monitoring have little or no nuclear explosion data on which to adequately characterize the statistical distribution of the nuclear explosion population. Note that industrial mining explosions are not necessarily a good surrogate upon which to base discriminants for nuclear explosions. Secondly, even if a set of nuclear explosion data exists, it is likely to be limited; that is, a small number of events, from a given test site, detonated under standard containment conditions (as opposed to potential evasive conditions). Such nuclear explosion data may not be suitable for deriving population statistics used in broad-area monitoring. Comparison of nuclear explosion data from different test sites or even from within a single test site (e.g. Nevada Test Site) illustrates the complexities of near-source nonlinear material properties and emplacement conditions on seismic discriminants (see Taylor 1991; Taylor and Denny 1991).

As noted in Fisk et al. (1996), for a large number of calibration earthquakes, the likelihood ratio outlier test statistic is essentially the multivariate normal density function (MVN(μ_{EQ}, Σ_{EQ})). Specifically, for a vector of discriminants $\mathbf{x}$, if

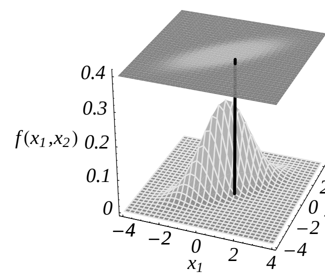
$$f_X(\mathbf{x}) = \left|2\pi\Sigma_{EQ}\right|^{-1/2} \exp\left\{-\frac{(\mathbf{x}-\mu_{EQ})'\,\Sigma_{EQ}^{-1}\,(\mathbf{x}-\mu_{EQ})}{2}\right\} \qquad (3)$$

is close to zero, then the data $\mathbf{x}$ indicate outlier, otherwise earthquake. Here, μ_{EQ} and Σ_{EQ} are estimated with the calibration data, thus establishing the outlier rule for future events. The term close to zero is defined by a critical value ξ that serves as a point of reference for evaluated values of $f_{\mathbf{X}}(\mathbf{x})$, and is determined by the tolerable false-outlier rate α. If μ_{EQ} and Σ_{EQ} are assumed known and $\mathbf{x}$ is a vector of p variables then

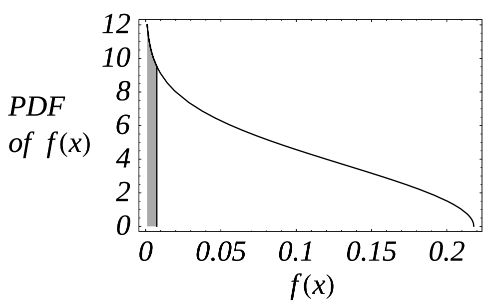
$$P\left(f_{\mathbf{X}}(\mathbf{x}) < \xi\right) = P\left((\mathbf{x}-\mu_{EQ})' \Sigma_{EQ}^{-1} (\mathbf{x}-\mu_{EQ}) > -2\ln(\xi\sqrt{\left|2\pi\Sigma_{EQ}\right|}\right). \tag{3.a}$$



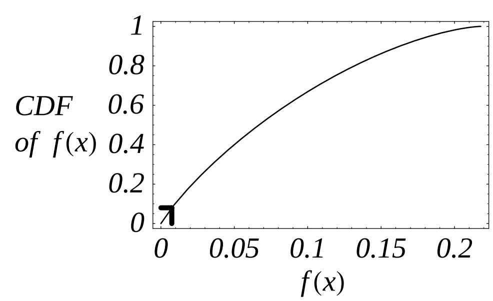
a) Discriminants x_1 and x_2 are interior to a $(1-\alpha)$% probability region which translates to a large likelihood value.



b) Discriminants x_1 and x_2 are outlier to a $(1-\alpha)$% probability region which translates to a likelihood value near zero.



c) Typical probability density function (PDF) of $f_{\mathbf{X}}(\mathbf{x})$ with a α% false-outlier region.



d) Typical cumulative distribution function (CDF) of $f_{\mathbf{X}}(\mathbf{x})$ with a α% false-outlier region defining the value of ξ.

Figure 2. Likelihood-based outlier rule. The rule is illustrated in two dimensions with discriminants x_1 and x_2. The rule is easily extended to p dimensions with a p dimensional likelihood function.

The random variable $(\mathbf{x}-\mu_{EQ})' \Sigma_{EQ}^{-1} (\mathbf{x}-\mu_{EQ})$ follows a chi-squared distribution with p degrees of freedom (Rencher 1998, Theorem 2.2F), thus

$$P\left(f_{\mathbf{X}}(\mathbf{x}) < \xi\right) = 1 - F_{\chi_p^2}\left(-2\ln(\xi\sqrt{\left|2\pi\Sigma_{EQ}\right|}\right);\ \xi \in \left(0, 1\ \sqrt{\left|2\pi\Sigma_{EQ}\right|}\right], \tag{3.b}$$

where $F_{\chi_p^2}(x)$ is the chi-squared cumulative distribution function with p degrees of freedom. To determine the value of ξ we solve for ξ in the equation $P\left(f_{\mathbf{X}}(\mathbf{x}) < \xi\right) = 1 - F_{\chi_p^2}\left(-2\ \ln(\xi\sqrt{\left|2\pi\Sigma_{\mathrm{EQ}}\right|}\right) = \alpha$. Thus an outlier rule based on $f_{\mathbf{X}}(\mathbf{x})$ is equivalent to an outlier rule based on $(\mathbf{x} - \mu_{\mathrm{EQ}})' \Sigma_{\mathrm{EQ}}^{-1} (\mathbf{x} - \mu_{\mathrm{EQ}})$. The outlier rule $f_{\mathbf{X}}(\mathbf{x})$, in two dimensions, is illustrated in Figure 2. Note that this rule is a one-sided test. In this paper, we use an outlier rule based on Equation 3 and illustrated in Figure 2.

RIDGE DISCRIMINATION AND REGULARIZED DISCRIMINANT ANALYSIS

The optimal regional discrimination method needs to be stable, robust and simple, and it should have a well-grounded physical basis. As regional seismic research continues, an optimal regional discrimination method will be developed. Ultimately, that method may use only two or three features from a seismic wave. *What is currently desirable is a technique that properly aggregates all available seismic information from a suite of phase measurements.* Classical discrimination does this because it is essentially based on the formation of a likelihood ratio, and statistical likelihood functions can properly combine phase amplitudes.

In the Gaussian case, the likelihood requires a covariance matrix of the phase amplitudes, and as discussed above these phase amplitudes may be strongly correlated. This will lead to a near singular covariance matrix that in turn will give a likelihood function with unstable statistical properties. There has been some very useful research on this problem within the statistics community (see Smidt and McDonald 1976; DiPillo 1976, 1977, and 1979; Randles et al. 1978; Loh 1995,1997; and Campbell 1980). In general terms, this research studied the utility of a ridge adjustment to the covariance matrix in discrimination analysis. Aki and Richards (1980) describe this type of adjustment as the stochastic inverse in seismological inverse problems. This adjustment is also similar to damped least squares (Aki and Richards 1980). In the statistical literature this approach is known as ridge discrimination. In a preliminary study, we have embedded these ideas into the event source elimination approach to regional discrimination. Our initial studies have produced some positive results.

Ridge Discrimination

Ridge discrimination was proposed as a method of addressing the problem of near-singular covariance matrices in Gaussian linear (LDA) and quadratic (QDA) discrimination. In ridge discrimination, the covariance matrix of the k^{th} group, used in a LDA or QDA application, is an additive combination of the sample covariance and the identity matrix. The weighting in this addition is governed by a smoothing parameter λ. A common λ is used across all groups. Formally, the ridge discrimination covariance matrix for the k^{th} group is

$$\Sigma_{\mathrm{k}}(\lambda) = (1-\lambda)\,\mathbf{S}_{\mathrm{k}} + \lambda \frac{tr(\mathbf{S}_{\mathrm{k}})}{p}\,\mathbf{I}\,\mathbf{;}\ \lambda \in [0,1]. \tag{4}$$

Here, $tr(\mathbf{S}_{\mathrm{k}})$ is the trace of the sample covariance matrix $\mathbf{S}_{\mathrm{k}}$ and p is the dimension of the amplitude vector. The covariance matrix $\Sigma_{\mathrm{k}}(\lambda)$ is essentially formed by adding a λ proportion of the average eigenvalue of $\mathbf{S}_{\mathrm{k}}$ to the diagonal elements of $\mathbf{S}_{\mathrm{k}}$. Equation 4 is equivalent to Equation 12.132 in Aki and Richards (1980) with ε (Aki and Richards) and λ playing analogous roles.

As a preliminary study of this approach, we have performed a leave-one-out Monte Carlo outlier analysis with highly correlated regional discriminant data. This type of Monte Carlo study can also be used to select optimal features for outlier detection. First, we fix a value of λ. With n earthquake events, a leave-one-out cross-validation involves n steps. For step i, the i^{th} event was removed from the earthquake data. This event was used as the test case and all other earthquake data were used to construct the covariance $\mathbf{S}_{\mathrm{EQ}}$ and the mean $\bar{\mathbf{x}}_{\mathrm{EQ}}$. We then construct the covariance $\Sigma_{\mathrm{EQ}}(\lambda)$. The $\mathbf{S}_{\mathrm{EQ}}$ and $\bar{\mathbf{x}}_{\mathrm{EQ}}$ are then used to generate a large number of simulated discriminants, and for each simulated data point we evaluate the multivariate normal (MVN) density using $\Sigma_{\mathrm{EQ}}(\lambda)$ and $\bar{\mathbf{x}}_{\mathrm{EQ}}$ (i.e., MVN($\bar{\mathbf{x}}_{\mathrm{EQ}}, \Sigma_{\mathrm{EQ}}(\lambda)$)). We need to do this because we use the MVN($\mu_{\mathrm{EQ}}, \Sigma_{\mathrm{EQ}}(\lambda)$) density as the outlier rule and these simulated data can be used to define a critical value ξ for the rule. The value ξ serves as the critical value to classify an event as earthquake or outlier. In our study, we use the 5^{th} percentile (α=0.05) of the MVN($\bar{\mathbf{x}}_{\mathrm{EQ}}, \Sigma_{\mathrm{EQ}}(\lambda)$) density values, gotten from the simulated discriminants, for the critical value ξ.

If the MVN($\bar{\mathbf{x}}_{\mathrm{EQ}}, \Sigma_{\mathrm{EQ}}(\lambda)$) density is greater than ξ, when evaluated with discriminants from an unknown event, then we would conclude the event is an earthquake. If the MVN($\bar{\mathbf{x}}_{\mathrm{EQ}}, \Sigma_{\mathrm{EQ}}(\lambda)$) density is smaller then ξ, then the data are in the extreme regions of the density support and we would call the event an outlier. We evaluate the test case with this rule. Repeating this process for all n of the earthquake data gives a leave-one-out cross-validated error rate for the fixed λ value. We then fix another value of λ and repeat the cross-validation analysis.

The data used in this preliminary study consist of amplitudes of Pn, Pg, Sn and Lg taken in seven different frequency bands and corrected for source and propagation effects (see Taylor et al. 1999). The data consist of 412 earthquakes and 4 explosions. For each source, the data are normalized to the low frequency Lg.

Four sets of seven amplitudes are analyzed in this paper. The seven amplitudes were selected from the twenty-seven amplitudes (28-1=27, low frequency Lg was used to normalize the amplitudes). Additionally this data set has four explosions that were used as test cases in all of the steps of the cross-validation study. In our study, we compute a cross-validated false-outlier rate; however, we cannot reasonably estimate a missed-explosion rate because we have only the four explosions.

We have noted that a very small value of λ can be used to get a covariance $\Sigma_{\mathrm{EQ}}(\lambda)$ with an acceptable condition number (ratio of max to min eigenvalues). In Figure 3, we summarize the distribution of the test values $f_{\mathbf{X}}(\mathbf{x})$ with trade-off plots. The ordinate is the 5th percentile of $f_{\mathbf{X}}(\mathbf{x})$ evaluated at each test amplitude data point and the abscissa is the interquartile range (IQR) of $f_{\mathbf{X}}(\mathbf{x})$. Each point is labeled with its corresponding λ value. Because all of the test values are computed from earthquake amplitudes, we want a distribution of test values that will optimally indicate earthquake. In terms of the trade-off plots in Figure 3, we want the 5th percentile of the outlier test statistic, $f_{\mathbf{X}}(\mathbf{x})$, to be as large as possible. Note that a small increase in the value of λ from zero generally decreases the IQR and increases the 5th percentile up to an optimum. These are desirable distributional properties because these two features indicate that the distribution is shifting away from zero and the variability of the distribution is decreasing. This property is further illustrated with fabricated boxplots in Figure 4, and a summary plot with regional data in Figure 5. Eventually the λ values cause the 5th percentile to move toward zero with a continued mild decrease in the IQR. Again, this is not desirable because small test values indicate outlier. The main points in this discussion are that a mild increase in λ away from zero will give a covariance $\Sigma_{\mathrm{EQ}}(\lambda)$ with an acceptable condition number, and that there is an optimal value of λ that gives test value ($f_{\mathbf{X}}(\mathbf{x})$) distribution properties that minimize false-outlier rates. As shown in Figure 5, for an optimal λ, the ability to detect the explosions as outliers to the earthquake population is excellent. Thus in this preliminary study, we have observed that λ can have an optimal value that achieves, in the mean square error sense, a minimal false-outlier rate. Other research in support of this observation can be found in Peck et al. (1988). These methods show promise as a way to address the problem of high correlation among seismic discrimination measurements. Ridge discrimination is especially appealing when it is generalized to regularized discrimination analysis.

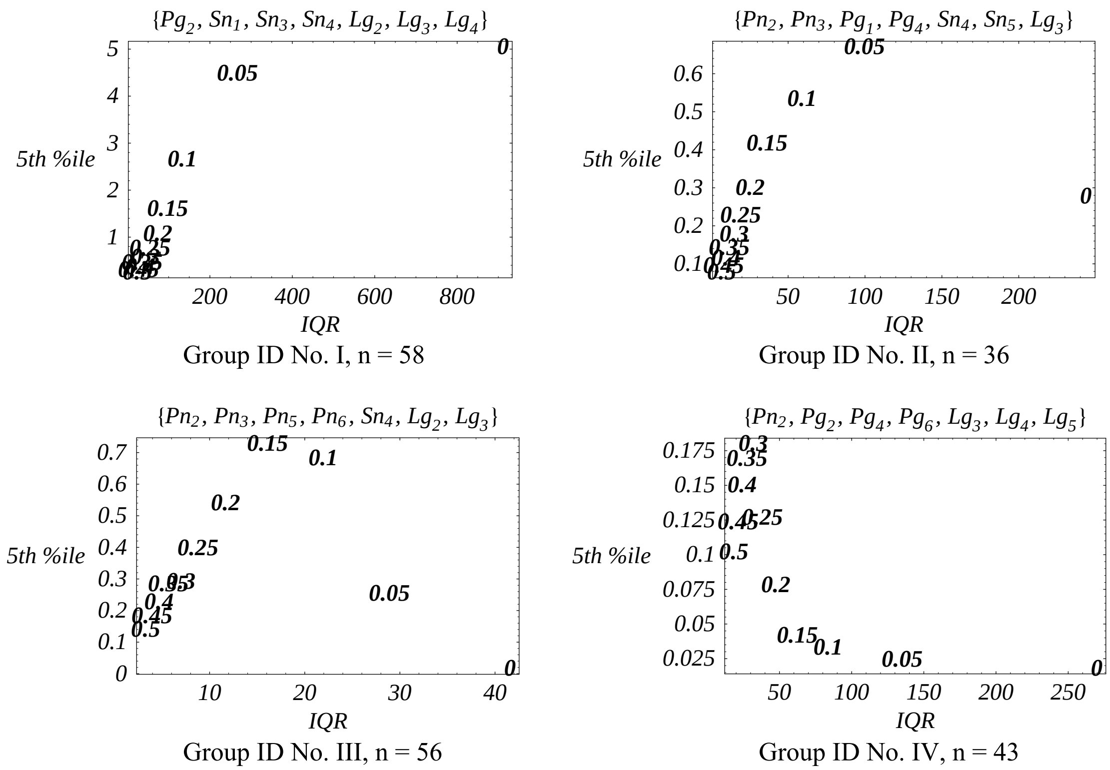


Figure 3. Trade-off plots of the test value data (MVN($\mu_{EQ}, \Sigma_{EQ}(\lambda)$)) evaluated with the cross-validation earthquake amplitudes. The value of λ appears in red, and the number (n) of cross-validation data used to construct the plot appears at the bottom of pair of plots.

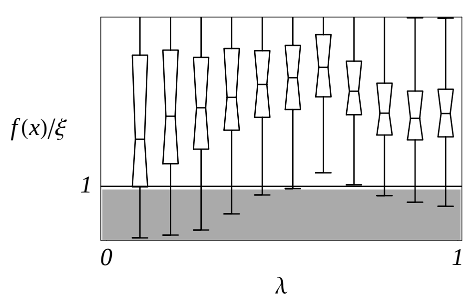


Figure 4. Illustration of the distributional properties of the test statistic $f_{\mathbf{X}}(\mathbf{x})$ as a function of λ. An increase in the value of λ away from zero generally decreases the IQR and increases the 5th percentile up to an optimum. These are desirable distributional properties because these two features indicate that the distribution is shifting away from zero and the variability of the distribution is decreasing. Eventually the λ values cause the 5th percentile to move toward zero with a continued mild decrease in the IQR. The shaded region represents the critical region that defines outlier.

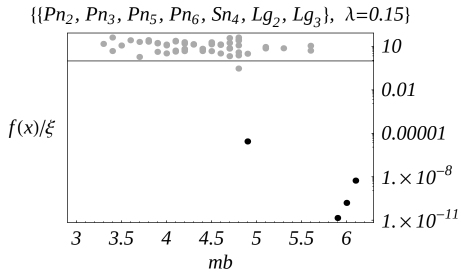


Figure 5. Test statistic $f_{\mathbf{X}}(\mathbf{x})$ plotted against mb for Group No. III in Figure 3. In this summary plot, all earthquake events were used to estimate μ_{EQ} and $\Sigma_{EQ}(\lambda)$. The optimal λ value is 0.15. Based on these data, this value will give a minimum IQR and a 5th percentile that is optimally above zero.

Regularized Discrimination Analysis

Regularized discrimination analysis (RDA) was proposed by Friedman (1989) as a method of discrimination to address applications with highly correlated discriminants and small training samples for some classification groups. Friedman s generalization of ridge discrimination involves the construction of a weighted-average covariance matrix

$$\mathbf{S}_{\mathrm{k}}(\gamma) = (1-\gamma)\,\mathbf{S}_{\mathrm{k}} + \gamma\,\mathbf{S}\,;\gamma \in [\mathbf{0},\mathbf{1}]. \tag{5}$$

Here, $\mathbf{S}_{\mathrm{k}}$ is the computed covariance matrix for kth group, and $\mathbf{S}$ is the pooled covariance matrix. Note that $\mathbf{S}_{\mathrm{k}}$ may be singular due to a small number of training data or strongly correlated variables for the kth group. $\mathbf{S}_{\mathrm{k}}(\gamma = 0)$ is computed from the kth group data alone (QDA) and $\mathbf{S}_{\mathrm{k}}(\gamma = 1)$ is a pooled covariance (LDA). RDA uses a two-parameter formulation of a covariance matrix in forming discrimination rules. With $\mathbf{S}_{\mathrm{k}}(\gamma)$ defined above, the RDA covariance matrix is

$$\Sigma_{\mathrm{k}}(\lambda,\gamma) = (1-\lambda)\,\mathbf{S}_{\mathrm{k}}(\gamma) + \lambda\,\frac{tr(\mathbf{S}_{\mathrm{k}}(\gamma))}{p}\,\mathbf{I}\,;\lambda \in [\mathbf{0},\mathbf{1}], \gamma \in [\mathbf{0},\mathbf{1}]. \tag{6}$$

Note that this is simply the ridge discrimination formulation with $\mathbf{S}_{\mathrm{k}}$ replaced by $\mathbf{S}_{\mathrm{k}}(\gamma)$. Here, λ and γ are the same values across all groups. Higbee s (1994) generalization allows λ and γ to change from group to group.

RDA theory can potentially be integrated into an outlier analysis approach to event identification or used as a classical seismic discrimination method. For outlier analysis, pooling ($\mathbf{S}$) could potentially occur across seismic stations within a geophysically homogeneous region. Here, each station may have observed a small number of seismic events. An RDA type covariance would be constructed for each station (Equation 6). We are researching techniques of optimally choosing values for λ and γ for $\Sigma_{\mathrm{k}}(\lambda,\gamma)$ when only earthquake data are available.

There are some very appealing features of RDA. As noted in Friedman (1989), RDA reduces to quadratic discrimination for values of $\lambda = 0$, and $\gamma = 0$. For $\lambda = 0$, and $\gamma = 1$ RDA reduces to linear discrimination. Other extremes in the RDA parameters give nearest neighbor and weighted nearest neighbor type discrimination methods. In the nearest neighbor case, $\lambda = 1$, and $\gamma = 1$ which gives $\Sigma_{\mathrm{k}}(1,1) = tr(\mathbf{S})/p\,\mathbf{I} = \omega\,\mathbf{I}$, ω a constant. For the weighted nearest neighbor case, $\lambda = 1$, and $\gamma = 0$ which gives $\Sigma_{\mathrm{k}}(1,0) = tr(\mathbf{S}_{\mathrm{k}})/p\,\mathbf{I} = \omega_k\,\mathbf{I}$, ω_k a constant. In both of these cases, the discrimination function is based on Euclidean distance rather than Mahalanobis distance. However in the weighted nearest neighbor case the kth term in the discrimination function, corresponding to the kth group, is weighted with ω_k. These observations are summarized in Figure 6. RDA also addresses some of the inadequacies with QDA. In particular, QDA usually requires larger sample sizes than LDA and is quite sensitive to model violations (Friedman 1989).

RDA provides a rich and adaptable family of discrimination methods that appear to be very applicable to the regional seismic problem. RDA readily provides a statistical framework to use MDAC amplitudes (see Taylor et al.

1999) in regional seismic discrimination. We also note that the $\Sigma_k(\lambda,\gamma)$ can be used as the covariance matrix in negative evidence methods (Anderson et al. 1999). In fact, RDA can be the foundation for any enhanced discrimination framework that is based on the use of statistical likelihood functions. For the classical discrimination problem (calibration data for all seismic sources), optimal values of λ, and γ are identified with cross-validated error rates. The details of this procedure can be found in Friedman (1989).

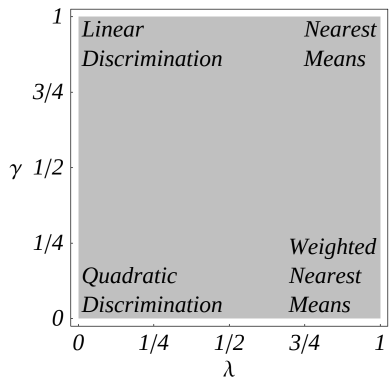


Figure 6. The relationship between regularized discrimination analysis (RDA) and some other classical discrimination methods.

CONCLUSIONS AND FURTHER DEVELOPMENTS

We have illustrated the importance of properly modeling the correlation structure of seismic measurements in the seismic event identification problem. If this correlation is not captured in the mathematics of a discrimination method, then both false-alarm and missed-explosion rates can be aggravated. Empirically we have noted that the false-alarm errors can be seriously increased when discriminant correlations are not properly modeled. In the regional discrimination problem, seismic measurements will be strongly correlated (e.g., Lg spectra may be contaminated by Sn coda and amplitudes may be constructed with frequency bands that overlap). This poses another problem in that an estimate of a covariance matrix for these measurements will be near singular. The true, unknown covariance matrix is itself near singular. Ridge discrimination and its generalization, RDA, provide a statistical method that can perform well in the presence of highly correlated seismic measurements. These methods properly combine measurements through a covariance matrix and are mathematically adaptable to a variety of regional seismic identification settings. We believe that regularized discrimination provides a reasonable solution to the regional discrimination problem and provides the flexibility to adapt to a future maturation of seismic event identification.

If an outlier detection algorithm is based on the use of statistical likelihood functions, then in an operational setting, missing data may be accounted for using detection thresholds with negative evidence methods (Anderson et al. 1999). However, outlier detectors are constructed with ground truth earthquake data, which can have left-censored data due to poor signal-to-noise, particularly at high frequencies. Woodward et.al (1999) show the utility of filling missing ground truth earthquake data with the EM algorithm with a marked decrease in error rates. Anderson and Phillips (1999) have developed an approach to incorporate censoring thresholds associated with missing spatial data into Kriging parameter estimates. We are adapting the work of Woodward et.al (1999) and Anderson and Phillips (1999) to the outlier detector approach presented in this paper. In particular, future developments include;

- the development of methods to incorporate amplitude censoring thresholds into the EM algorithm to fill missing earthquake training data,
- the development of a method of optimally choosing the RDA parameters λ and γ using only earthquake training data (RDA concepts integrated into outlier analysis),
- the comparison of optimal λ and γ determined with explosion and earthquake training data with optimal λ and γ determined only with earthquake training data.

REFERENCES

Aki, K. and P.G. Richards (1980). *Quantitative Seismology: Theory and Methods*. W.H. Freeman and Company, San Francisco.

Anderson, D.N., S.R. Taylor and K.K. Anderson (1999). Discrimination Information in Phase Amplitude Thresholds with Application to Western China Regional Data , *Bulletin of the Seismological Society of America*, 89, pp. 648-656.

Anderson, K.K. and W.S. Phillips. (1999). A Simple Procedure for Kriging Left-Censored Data. , PNNL-SA-31023, Pacific Northwest National Laboratory, Richland, Washington.

Campbell, N.A. (1980). Shrunken Estimators in Discriminant and Canonical Variate Analysis , *Applied Statistics* , 29, pp. 5-14.

Clemen, R.T. and R.L. Winkler (1985). Limits for the precision and value of information from dependent sources , *Operations Research*, 33, pp. 427-442.

DiPillo, P.J. (1976). The Application of Bias to Discriminant Analysis , *Communications in Statistics — Theory and Methods*, A5(9), pp. 843-854.

DiPillo, P.J. (1977). Further Application of Bias to Discriminant Analysis , *Communications in Statistics — Theory and Methods*, A6(10), pp. 933-943.

DiPillo, P.J. (1979). Biased Discriminant Analysis: Evaluation of the Optimum Probability of Misclassification , *Communications in Statistics — Theory and Methods*, A8(14), pp. 1447-1457.

Fisk, M.D., H.L. Gray, and G.D. McCartor (1996). Regional Event Discrimination without Transporting Thresholds , *Bulletin of the Seismological Society of America*, 86, pp. 1545-1558.

Friedman, J.H. (1989). Regularized Discriminant Analysis , *Journal of the American Statistical Association*, 84, pp. 165-175.

Higbee, K.T. (1994). Variably Regularized Discriminant Analysis , *Proceedings of the Statistical Computing Section*., American Statistical Association, Alexandria, Virginia. pp. 263-268.

Lehmann, E.L. (1983). *Theory of Point Estimation*. John Wiley & Sons, New York.

Loh, Wei-Liem. (1995). On Linear Discriminant Analysis with Adaptive Ridge Classification Rules , *Journal of Multivariate Analysis*, 53, pp. 264-278.

Loh, Wei-Liem. (1997). Linear Discrimination with Adaptive Ridge Classification Rules , *Journal of Multivariate Analysis*, 62, pp. 169-180.

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

Morrison, D.G. and D.C. Schmittlein (1991). How many forecasters do you really have? Mahalanobis provides the intuition for the surprising Clemen and Winkler result , *Operations Research*, 39, pp. 519-523.

Peck, R., L.W. Jennings and D.M. Young (1988). A Comparison of Several Biased Estimators for Improving the Expected Error Rate of the Sample Quadratic Discriminant Function , *J. Statist. Comput. Simul.*, 29, pp. 143-156.

Randles, R.H., J.D. Broffitt, J.S. Ramberg and R.V. Hogg (1978). Generalized Linear and Quadratic Discriminant Functions Using Robust Estimates , *Journal of the American Statistical Association*, 73, pp. 564-568.

Rencher, A.C. (1998). *Multivariate Statistical Inference and Applications*. John Wiley & Sons, New York.

Smidt, R.K. and L.L. McDonald (1976). Ridge Discriminant Analysis , University of Wyoming Department of Statistics, Technical Report No. 108, Laramie, Wyoming.

Taylor, S.R. (1991). Regional seismic observations from NTS explosions, in Explosion Source Phenomenology , *AGU Monograph 65*, Taylor, S.R., H.J. Patton , and P.G. Richards, eds., pp. 185-196.

Taylor, S.R., and M.D. Denny (1991), An Analysis of Spectral Differences Between NTS and Shagan River Nuclear Explosions , *J. Geophys. Res.*, 96, pp. 6237-6245.

Taylor, S.R., and H.E. Hartse (1997). An Evaluation of Generalized Likelihood Ratio Outlier Detection to Identification of Seismic Events in Western China , *Bulletin of the Seismological Society of America*, 87, pp. 824-831.

Taylor, S.R., A.A. Velasco, H.E. Hartse, W.S. Phillips, W.R. Walter, and A.J. Rodgers (1999), Amplitude Corrections for Regional Seismic Discriminants. , LAUR-99-3040, Los Alamos National Laboratory, Los Alamos, New Mexico.

Woodward, W.A., S.R. Sain, H.L. Gray, and M.D. Fisk (1999). Outlier Testing When Some Data are Missing , *21st Seismic Research Symposium*., 21-24 September, Las Vegas, Nevada., U.S. Departments of Energy and Defense, pp. 674-682.