A LOWER BOUND ON THE STANDARD ERROR OF AN AMPLITUDE-BASED REGIONAL DISCRIMINANT

D. N. Anderson¹, W. R. Walter², D. K. Carlson¹, T. M. Mercier¹, S. R. Taylor³

Pacific Northwest National Laboratory¹, Lawrence Livermore National Laboratory², and Rocky Mountain Geophysics, LLC³

Sponsored by National Nuclear Security Administration
Office of Nonproliferation Research and Development
Office of Defense Nuclear Nonproliferation

Contract Nos. DE-AC06-76RLO 1830¹³ and W-7405-ENG-48²

ABSTRACT

Wave path, magnitude, and signal processing corrections made to observed regional amplitudes fundamentally contain information regarding only the seismic source. These corrected amplitudes can then be used in ratios to discriminate between earthquakes and explosions. Source effects that are due to depth, focal mechanism, local material property and apparent stress variability that cannot easily be determined still remain in the signal. These effects establish a lower bound on the amplitude variability for new events, even after path and magnitude corrections are applied. We develop a general strategy to account for amplitude correction inadequacy by appropriately partitioning error. The proposed mathematics are built from random effects analysis of variance (ANOVA) and have application potential to a variety of amplitude correction theories, for example, see Taylor and Hartse (1998), Taylor et al. (2002), and Walter and Taylor (2002). The error components from random-effects ANOVA are the basis for a general station-averaged regional discriminant formulation. The standard error of the discriminant has a lower bound of amplitude correction error. The developed methods are demonstrated for a suite of Nevada Test Site (NTS) events observed at regional stations.
OBJECTIVES

In Taylor and Hartse (1998), Taylor et al. (2002), and Walter and Taylor (2002), the magnitude and distance amplitude correction (MDAC) technique corrects regional phase amplitudes independent of distance and magnitude. MDAC is a simple, physically based model that accounts for propagation effects, such as geometrical spreading and Q, and corrects amplitudes, assuming an earthquake model. MDAC P and S phase amplitude ratios can then be used to discriminate between earthquakes and explosions. Because of complex explosion source phenomenology, it is not obvious which discriminants or combination of discriminants will best separate earthquake and explosion populations. The MDAC technique provides the latitude to form a diversity of discriminants by mixing phases and signal-processing frequencies. Until now, no attempt has been made to obtain a realistic estimate of the error budget associated with MDAC amplitudes used to construct seismic discriminants.

The approach presented in this paper develops the mathematics to form a station-averaged discriminant, with the proper standard error, from multiple stations. Through MDAC we correct amplitudes and adopt the position that any remaining physical structure, not indicative of source, is a random model inadequacy effect. MDAC can be augmented with additional corrections and the basic multi-station model holds. This approach to discriminant formulation properly forms the variance of the discriminant with two components. Station noise is reduced through station averaging, and the model inadequacy variance component only decreases with improvements in physical path correction theory.

RESEARCH ACCOMPLISHED

Established signal-processing research treats amplitudes as lognormally distributed random variables; therefore, in log space, properly formed differences between amplitudes are Gaussian-based discriminants. The conceptual representation of the proposed model is

\[
\log(\text{Amplitude}) = \eta(\text{Source, Path}) + \text{Bias} + \text{EventEffect} + \text{Noise},
\]

where \( \eta(\text{Source, Path}) \) is MDAC; \( \text{Bias} \) is a constant effect that is due to source only; \( \text{EventEffect} \) is a random effect that varies from event to event and represents model inadequacy from effects such as depth, focal mechanism, local material properties, and apparent stress variability; and \( \text{Noise} \) represents measurement and ambient noise, also a random variable.

The mathematical statistics formulation of Equation (1) is

\[
Y_{ijk} = \eta_{ijk} + \mu_i + E_j + \varepsilon_{ijk},
\]

where \( Y_{ijk} \) is the log amplitude for source \( i = \{0, 1\} \), event \( j = \{1, 2, \ldots, m_i\} \), and station \( k = \{1, 2, \ldots, n_{ij}\} \). In other words, the source \( i \) event \( j \) amplitude observed by station \( k \) equals the sum of a known source/path effect (MDAC), a constant bias effect, a random event adjustment, and a noise effect particular to the source and event. Equation (2) is a mixed effect linear mode—see Searle (1971) for details.

The \( E_j \) are modeled as independent Gaussian random variables with mean zero and variance \( \hat{\sigma}^2 \), whereas the \( \varepsilon_{ijk} \) are modeled as independent Gaussian random variables with mean zero and variance \( \sigma^2 \). Furthermore, \( E_j \) and \( \varepsilon_{ijk} \) are independent across all subscripts.
Statistical Properties of a Station-Averaged Amplitude

The statistical properties of the $Y_{ijk}$ in Equation (2) are written as $Y = \theta + \Omega$, where $Y \sim \text{MVN}(\theta, \Omega)$, where

\[
\theta = \begin{bmatrix}
\mu_i + \eta_{j1} \\
\mu_i + \eta_{j2} \\
\vdots \\
\mu_i + \eta_{jn_j}
\end{bmatrix}
\quad \text{and} \quad
\Omega = \begin{bmatrix}
\tau^2 + \sigma^2 & \tau^2 & \tau^2 & \cdots & \tau^2 \\
\tau^2 & \tau^2 + \sigma^2 & \tau^2 & \cdots & \tau^2 \\
\tau^2 & \tau^2 & \tau^2 + \sigma^2 & \cdots & \vdots \\
\vdots & \vdots & \vdots & \ddots & \tau^2 \\
\tau^2 & \cdots & \cdots & \tau^2 & \tau^2 + \sigma^2
\end{bmatrix}.
\]

MDAC corrections applied to $Y$ give

\[
X = \begin{bmatrix}
Y_{j1} - \eta_{j1} \\
Y_{j2} - \eta_{j2} \\
\vdots \\
Y_{jn_j} - \eta_{jn_j}
\end{bmatrix},
\]

which is multivariate Gaussian with mean vector $\mathbf{1}\mu$ and covariance matrix $\Omega$. The station-averaged amplitude, $\bar{X}_{ij} = \mathbf{1}^T X_{ij}/n_{ij}$, is Gaussian with mean $\mu_i$ and variance $\tau^2 + \sigma^2/n_{ij}$.

Excluding the $E_j$ term in the model in Equation (2) implies that $\eta$ is unbiased and that the variance of $\bar{X}_{ij}$ is $\sigma^2/n_{ij}$—the variance of $\bar{X}_{ij}$ becomes small as $n_{ij}$ increases. Omitting $E_j$ results in a model that is fundamentally inconsistent with physical basis because station averaging will not eliminate all noise corrupting a seismic source signal.

Model Properties

The expected value of the MDAC corrected amplitudes is $E[X_{0jk}] = \mu_0$ for earthquakes and $E[X_{1jk}] = \mu_1$ for explosions. The covariance matrix intuitively says that the variance of an amplitude is composed of model inadequacy and station noise. The correlation $\left(\tau^2 / (\tau^2 + \sigma^2)\right)$ between amplitudes implies that increased adjustment to $\eta$ to fit data increases dependency between stations observing an event. Minimal adjustment to $\eta$ is conceptually equivalent to the stations being incoherent, which in turn minimizes the variance of an amplitude through station averaging.

Discriminant Formulation

For this development, a discriminant is constructed from two different amplitudes (with different functions $\eta$ and parameter values $\mu$) that are statistically independent (uncorrelated). For specific regional phases, the discriminant equation can be represented with meaningful subscripts. For example, for a given event and source, the station average $\bar{P}_g$ is Gaussian with mean $\mu_{P_g}$ and variance $\tau_{P_g}^2 + \sigma_{P_g}^2/n_{P_g}$, and the station average $\bar{L}_g$ is Gaussian with mean $\mu_{L_g}$ and variance $\tau_{L_g}^2 + \sigma_{L_g}^2/n_{L_g}$. Note that if only stations observing a discriminant are used, then $n_{P_g} = n_{L_g}$; however, this constraint on discriminant construction is not necessary—using all available station amplitudes to construct a discriminant is theoretically sound if the MDAC correction is of good quality. The standard error of the $P_g$ versus $L_g$ discriminant is

\[
SE_{\bar{P}_g - \bar{L}_g} = \sqrt{\frac{\sigma_{P_g}^2}{n_{P_g}} + \frac{\sigma_{L_g}^2}{n_{L_g}}}. \tag{3}
\]

Standardizing observed discriminants relative to the explosion population mean $(\mu_{P_g} - \mu_{L_g})$ and adjusting for uncertainty gives

\[
Z_{\bar{P}_g - \bar{L}_g} = \frac{(\bar{P}_g - \bar{L}_g) - (\mu_{P_g} - \mu_{L_g})}{SE_{\bar{P}_g - \bar{L}_g}}. \tag{4}
\]
The model in Equation (4) implies that $Z_{T_g - T_k}$ has different means for the explosion and earthquake populations and the same variance for both populations. Values of $Z_{T_g - T_k}$ below a decision threshold predict an earthquake as the source identification, otherwise an explosion is predicted.

**CONCLUSIONS AND RECOMMENDATIONS**

**Analysis Example**

The data used to illustrate the discriminant shown in Equation (4) are events at the NTS. Events were observed with combinations of four seismic stations: Kanab, Utah (KNB); Elko, Nevada (ELK); Landers, California (LAC); and Berkeley, California (CMB). MDAC amplitudes between 6 and 8 Hertz from these stations were averaged in the calculation of $Z_{T_g - T_k}$. The dataset consists of 56 earthquakes (EQ) and 159 explosions (EX), for a total of 215 events. Event magnitudes (Mw) ran between 2.6 and 7.1. The spatial distribution of events is given in Figures 1 and 2. Model performance is summarized in Table 1.

**Table 1. Performance of the MDAC ratios for NTS events. Columns represent predictions by the discrimination method; rows are the true classifications.**

<table>
<thead>
<tr>
<th></th>
<th>Predicted EX</th>
<th>Predicted EQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>EX</td>
<td>144</td>
<td>15</td>
</tr>
<tr>
<td>EQ</td>
<td>2</td>
<td>54</td>
</tr>
</tbody>
</table>

**CONCLUSIONS AND RECOMMENDATIONS**

The analysis demonstrates the performance of an MDAC regional discriminant derived from the model in Equation (2). The analysis also illustrates the sharp difference that is possible between model-based and empirical-based decision thresholds. An obvious but important observation is that the empirical-based threshold fits the tail of the observed data and is therefore strongly influenced by outlying calibration data. The model-based threshold uses the calibration data to fit model parameters that are more robust to calibration data. If solid theory is in place to support a model—e.g., Equation (2)—then a model-based decision threshold should prove to be more accurate in future event identification.

Regional amplitudes of varying or constant frequencies should be incorporated into statistically based discrimination frameworks if their underlying mathematical and probabilistic structure can be understood. The appropriate standard error to use for the regional MDAC amplitude ratios is one that incorporates both model inadequacy and ambient station noise. The random effects model presented above correctly leads to a standard error for station averaging with a lower bound that conforms to physical basis. The model also properly represents the correlation of station observed amplitudes when the theoretical amplitude correction is poor.

**ACKNOWLEDGEMENTS**

Rick Schult of the Air Force Research Laboratory has continually provided expertise, insight, and constructive comments regarding the development of discriminants. The authors also acknowledge Gordon Kraft of Quantum Technology Services, Inc., for his input toward discriminant formulation.
REFERENCES


FIGURES

Figure 1. NTS events and stations used in the analysis example. Explosions are red stars, earthquakes are yellow circles, and stations are pink triangles.
Figure 2. A closer view of NTS events used in the analysis example. Explosions are red stars and earthquakes are yellow circles.