# EVENT IDENTIFICATION, ERROR PROPAGATION AND CALIBRATION ASSESSMENT

Dale N. Anderson[1], Stacey A. Hartley[1], Charles A. Lo Presti[1],
Alan C. Rohay[1], Steven R. Taylor[2], Aaron A. Velasco[2] and Chris J. Young[3]

Pacific Northwest National Laboratories[1]
Los Alamos National Laboratories[2]
Sandia National Laboratories[3]

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

Contract No. DE-AC06-76RLO-1830

## ABSTRACT

One component of nuclear explosion monitoring (NEM) research and engineering (R&E) is directed at the development of mathematical techniques that take full advantage of all information in a seismic signal. Regularized Discrimination Analysis (RDA) is a multivariate seismic event identification method that can be applied to a number of highly correlated regional discriminants. The parametric formulation of RDA includes Linear Discrimination (LDA), Quadratic discrimination (QDA) and Euclidean distance-based nearest-neighbor discrimination. We present methods to optimally select RDA parameters.

Error propagation is another focus area in the NNSA NEM R&E program. The detection and timing of seismic arrivals play a critical role in the ability to locate seismic events, especially at low magnitude. Errors can occur with the determination of the timing of the arrivals, whether these errors are made by automated processing or by an analyst. One of the major obstacles encountered in properly estimating travel-time picking error is the lack of a clear and comprehensive discussion of all of the factors that influence phase picks. We have developed a multivariate statistical model, experimental design, and analysis strategy that can be used in this study. We have embedded a general form of the International Data Centre (IDC)/U.S. National Data Center (USNDC) phase pick measurement error model into our statistical model. We can use this statistical model to optimally calibrate a picking error model to regional data.

We also present work on the development of statistical methodologies for comparing effects of station-specific correction surfaces on predicted seismic event locations and event location uncertainty from network model Monte Carlo simulation runs. Research and development work includes the investigation of Latin Hypercube Sampling (LHS) to design Monte Carlo simulation runs, the development of appropriate statistical models to describe travel-time correction surface errors, and the proper simulation of errors in phase identification and association processes. Also under investigation are statistics-based methods for visualizing and assessing differences between event locations and location uncertainty from different correction surfaces in network model simulations.

**KEY WORDS:** seismic identification, phase pick errors, error propagation, seismic network model simulation, surface correction assessment, event location.

## OBJECTIVE

The objectives of the PNNL statistics effort are to

- contribute to the development of optimal regional discrimination techniques that properly account for uncertainties in seismic signal processing, and
- develop efficient statistical metrics and visualization methods to measure or assess the improvement of location accuracy and precision as predicted using Monte Carlo network simulation that result from changes to velocity/travel-time calibration correction surfaces.

## RESEARCH ACCOMPLISHED

### Regional Discrimination

Ridge Discrimination techniques, first proposed by Smidt and McDonald (1976), were developed to address the problems associated with discrimination in high-dimension, co-linear settings. These methods are readily adaptable to linear, quadratic and outlier identification rules. Ridge Discrimination is a special case of Regularized Discrimination Analysis (RDA) developed by Friedman (1989). RDA includes LDA; QDA and Euclidean distance-based nearest-neighbor discrimination in its parameterization. These techniques can be used to transition from an outlier analysis approach for seismic identification to classical discrimination, as quality explosion calibration data are collected. Ridge Discrimination and RDA provide the statistical structure to model highly correlated seismic measurements. Omitting the correlation structure between seismic measurements in event identification can aggravate identification errors and give an erroneous impression of capability. With RDA, a large number of discriminants can be used and no *a priori* sub-selection of discriminants is necessary. We propose an information theory approach to the optimal selection of RDA parameters. These methods are based in the Kullback divergence index. Complete details of the proposed methods were submitted to the *Bulletin of the Seismological Society of America* for formal publication.

### Explorations in Assessing Calibration Surfaces through Seismic Network Simulation

*Background.* For purposes of nuclear explosion monitoring, it is desirable to continuously improve calibration travel-time/velocity corrections in order to produce more accurate and precise predicted seismic event epicenters versus ground truth. Improvement may be expressed as reduction of the distance between the predicted epicenter and the true event epicenter, and also as reduction in the epicenter uncertainty taken as the area of the confidence ellipsoid area (CEA). Comparisons may be absolute, when ground truth is present and well understood, or comparisons may also be relative in the sense of comparing alternative correction surfaces.

Locating seismic events with depth depends on at least three receiving stations detecting at least four phases associated with the same event. Through iterative application of triangulation algorithms with an overall global velocity model and velocity or travel-time correction surfaces, the event can be positioned geographically and depth-wise within a small region (Sleep & Fujita, 1997). Seismic event locations are often expressed as a predicted epicenter surrounded by a 90% CEA. Methods for performing event location calculations are well documented (e.g., Bratt & Bache, 1988)

*Network Simulation Models.* One approach toward comparative assessment of correction surfaces is to use network simulation models such as NetSim (Sereno, Bratt, and Yee, 1990; Sereno, 1991) and its follow-on systems to generate simulated locations of synthetic events. Network simulation models are computer codes, which generate simulated phases at specific frequencies to express seismic events of specified magnitude and depth. They then model the propagation of these phases through the earth by applying specified velocity and travel-time correction surfaces, and model each receiving station's reception in the presence of local noise. This includes simulating the phase-picking process. Simulations may also account for station up-time reliability and local station signal-to-noise conditions. Once the simulated phases are detected and arrival times and perhaps bearings calculated, these systems solve for the predicted epicenter, depth, magnitude, and other parameters of interest. Since the input correction surfaces, noise models, and other factors are known only with uncertainty, it is current practice to generate random samples of inputs through Monte Carlo experiments to introduce uncertainty into various stages of the detection and location processes, perhaps through Latin HyperCube Sampling (McKay, Conover, and Beckman, 1979).

A very important principle in seismic network simulation is to consistently form realistic events in tune with sound geophysical science. From the viewpoint of event location, this implies that for any particular event, the simulation has the capability to select receiving stations in accordance with sound seismological practice. It needs to make reasonable decisions as to whether a particular simulated event is solvable in terms of parameters of interest. Consequently, it is useful to include criteria for deciding whether a particular data set represents events typically formed by standard practice. It is also critical that estimates of uncertainty for input variables be realistic.

*Azimuthal Gap Patterns to Assess Potential for Obtaining Stable Event Locations.* One way to determine whether an event is well located is to consider the distribution of azimuthal gaps around an event formed by the pattern of detecting stations. These are measures of the arcs between the bearings from the event-node to each station. A poorly located event, difficult to triangulate, may have a very large maximum azimuthal gap, greater than 250 or so degrees. A large maximum azimuthal gap indicates that all stations that detect the event have bearings in one quadrant around the event, thereby controlling location in only one direction. Another way to obtain a poorly located event is that all detecting stations have bearings in two opposite quadrants around the event such that the two largest azimuthal gaps be of the order of 150 degrees, so that location is still controlled in only one direction. For good location results, ideally the station bearings are distributed such that location is controlled from several directions, thereby allowing the most effective triangulation. Azimuthal gaps may be less useful when bearings as well as travel times are used to solve for the epicenter.

*Assessing Changes in Event Location and Uncertainty from Two Simulation Runs.* Two issues are *comparison metrics* and representative event *sampling* for performing assessments.

The goal for *comparison metrics* is to define measurements or counts which express change in detection accuracy and/or precision for events simulated at given grid nodes, individually and ensemble. Assessing accuracy pertains to measuring differences between estimated and input epicenters. Assessing precision pertains to measuring differences between confidence ellipsoid areas, and perhaps ellipsoidal shapes and orientations (strikes) as well.

With regard to *sampling*, it is desirable to simulate events and detections in regions and with station networks associated with the calibrations under consideration. Moreover, since discrete regions of the globe may behave seismically in unique ways, tectonic provinces, surface roughness, and other geological characteristics also need to be considered (Ryaboy, 2000; Zhang, Lay, Schwartz, & Walter, 1996). This results in a need for sampling from the set of events, locations, and event-station pairs so that the results have minimal sampling biases, which could lead to false conclusions about the calibration surfaces.

To this end, the following grouping (or stratification) of events on the event-node grid is suggested. In some cases, the event grid-nodes only form study groups; in other cases the combinations of event grid-nodes and receiving stations taken together form study groups.

- Events chosen from inside or outside the convex hull of the station network
- Events chosen by surface topography or land areas versus oceanic
- Event-location/station subsets grouped into geophysically homogeneous regions such as tectonic provinces with regard to propagation path
- Event-location/station subsets grouped into regions where calibrations are changed vis a vis calibrations remain unchanged (control versus experimental regions)

Although the prime focus of a calibration improvement study would be to assess differences, the study might be more valuable if a control region is also available. A control region is a region with no calibration changes over propagation paths to a subset of stations, such that changes in predicted epicenters are not expected from the two runs. If changes are detected in a control region, contrary to expectations, this requires special evaluation, because it may mean the simulations were not working as expected.

*Numerical Stability in Computations to Solve for Location.* The event location process depends on solving the triangulation equations, which estimate the epicenter. These equations (Bratt & Bache, 1988) are often solved by maximum likelihood for the set of travel times, which jointly minimize residuals. In addition, some formulations also allow for inclusion of bearings from station to postulated event. This process of associating travel times (P-wave picks in time) to an event also has inherent uncertainty (and the possibility of misassociating picks from different events is not presently modeled). If the equations are well posed, the solution converges rapidly to a minimum-error solution, and the eigenvalues for the resulting covariance matrix (expressed as major and minor confidence ellipsoid axes) are such that the ellipse is not far from round. For the poorly posed cases, the equations are numerically unstable, the eigenvalues are grossly unequal, and the major and minor axes are also quite unequal. To obtain reliable results, it is critical to identify the cases that will not or did not converge to a reliable solution. This information can assist in understanding why the location process failed.

*Three Approaches toward Comparing Results from Network Simulation Runs: ensemble, paired event-node, and paired events.*

*Ensemble analysis:* The network simulation model is run N times (N large) and means, variances, and skewnesses are collected for key variables such as CEA and predicted epicenters at each event grid-node. Then, assessments of differences between two simulation runs are based on comparing distributions for the key variables expressed as means and variances at each event-node, without pairing.

*Analysis on individual event grid-node pairs*: The network simulation model is run N times at each event-node, same as above. The data at each event-node from the two calibrations are paired. The random number generator is not coordinated between the two runs so there is no control over calculation starting points. Assessment is performed based on the ensemble of mean differences between each event-node pair. The basic statistical test is the two-sample t-test at each event node to compare distributions of differences between pairs. Pairing event-nodes is a more powerful statistical test than comparing overall distributions in ensemble analysis.

*Individual events at event-node pairs*: The network simulation model is run N times, and data are collected for each individual event realization. Ideally, the random number generator is coordinated between the two runs so that results from each calibration for individual simulated events form pairs that can be directly compared, with all things equal except the experimental variables (calibrations). Assessment is performed based on the ensemble of individual event differences between the two calibration runs. The statistical test is the paired t-test at each event-node. Because the paired t-test is the most powerful statistical test in this situation, fewer simulation runs are needed to detect a given difference when individual simulated events are paired. One drawback with applying this approach is the nontrivial problem of coordinating the random number generator so that each simulated event has in effect the same starting point.

*A Collection of Analysis Tools for Assessing Difference Between Two Runs.* Two potential primary measures for assessing performance in locating simulated events are CEA and offset from CEA centroid to true location. The following lists of plots and tables is based on analysis and visualization of these two measures and related input parameters. It is not in any way complete.

**Table 1**. Displays showing differences or "movement" in primary measures for events. These may be expressed in terms of change in measured values or as percent change in measured values.

| Item to Plot/Tabulate | What It Shows/Interpretation |
|---|---|
| Change in mean CEA | Improvement in location precision |
| Change in Relative Standard Deviation | Display variability in CEAs, a measure of stability |
| t-test statistics or alpha probabilities | Displays statistically significant changes |
| Percent of events below 1000 km2 | Associated with external criterion for a "good" event |
| Major/Minor axis ratio | Expresses stability in detecting station configuration |
| Change in Offset | Shows whether predicted events from new calibrations are closer to true location |
| Percent of CEAs actually covering true event locations | Gross measure of accuracy |

**Table 2.** Secondary Measures involving counts of events satisfying various criteria

| Item to Plot | What it Shows/Interpretation |
|---|---|
| Proportion of well-located events to total events | Provides an idea of quality of station coverage |
| Proportion of events which have low (good) covariance matrix condition numbers | Provides an idea of quality of numerical processing, and ultimately station coverage. (This is expected to be high) |
| Coverage: proportion of events inside convex hull of stations | Provides an idea of quality of station coverage |

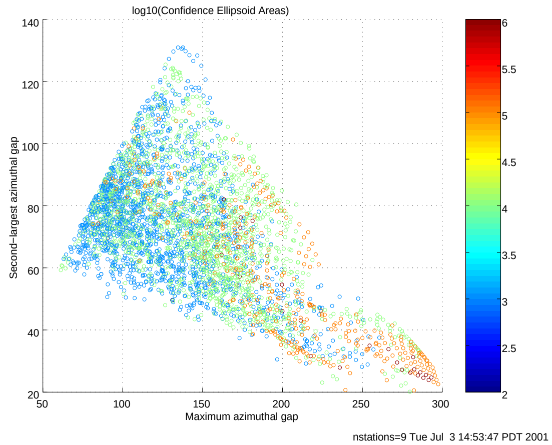


**Figure 1:** Plotting mean 90% confidence ellipsoid areas at each event-node against the largest and second-largest azimuthal gaps based on the nine closest stations for calibrations of CALIB1 and CALIB2. These are theoretical, not based on detections. Although large CEAs are found throughout, the largest CEAs are most likely found where the maximal azimuthal gap > 250°.

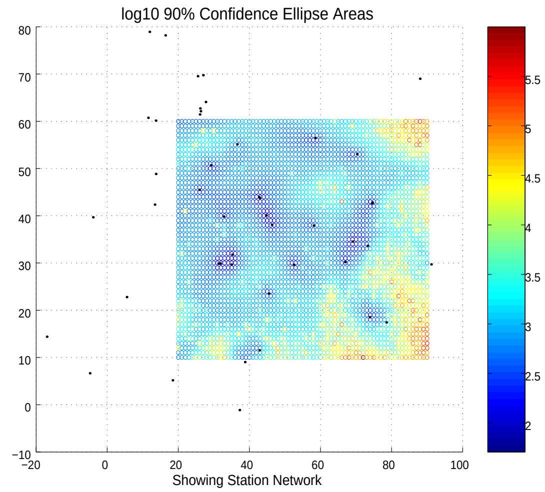


**Figure 2**: Plotting mean 90% confidence ellipsoid areas geographically at with the station network shown based on CALIB1 and CALIB2. The largest CEAs lie outside or on the edge of the station convex hull, where maximum azimuthal gap is expected to be largest. CEAs (shown in $\log_{10}$ scale) range from ~50 $km^2$ to over 250,000 $km^2$.

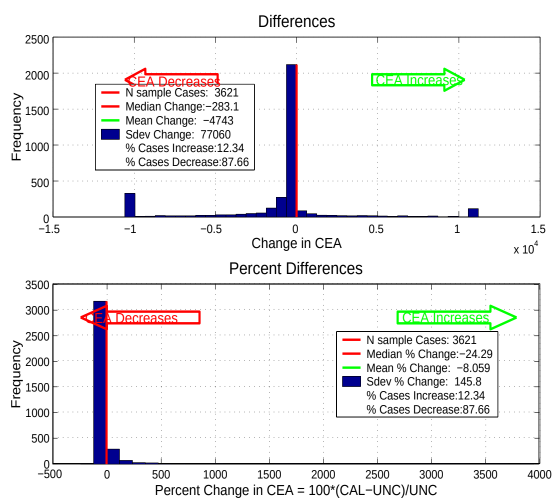


**Figure 3**: Summary histogram plots of differences between mean 90% CEAs from CALIB1, a baseline calibration and CALIB2, a new calibration. These plots allow assessment of "movement" of sets of CEAs. In this case, fully half the data set of mean CEA's decreased in size by at least 24.3% as a result of new calibrations.

| | | *CALIB2* | |
|---|---|---|---|
| | *Threshold* | *<1000* $km^2$ | *>1000* $km^2$ |
| *CALIB1* | *<1000* $km^2$ | 1019 *(28.14%)* | 15 *(0.41%)* |
| | *>1000* $km^2$ | 592 *(16.34%)* | 1995 *(55.09%)* |

**Figure 4**: Tabulating the "movement" of mean 90% CEAs across the 1000 $km^2$ threshold from estimates from two calibration surfaces. In this case, there were 592 instances in which the CEA for CALIB1 was initially > 1000 $km^2$ and with CALIB2 became small enough to cross the threshold. There were 15 instances in which CEAs initially smaller than the threshold were made larger as a result of CALIB2.

*Materials and Methods for Generating Data.* The CAT network simulation model was used to generate case study data for the plots shown in this report applying two calibrations here termed CALIB1 and CALIB2, with the goal of determining whether CALIB2 gives more precise results than CALIB1. The data shown are based on 4.25 mb event magnitude calibration runs. The confidence ellipsoid areas are means of N=1000 realizations per event-node. Also available were corresponding standard deviations of CEAs.

*Discussion.* Figure 1 illistrates how the two largest theoretical azimuthal gaps seem to be associated with larger mean 90% CEAs, at least for this case. Figure 2 plots the same set of CEAs geospatially; and illustrates how the larger CEAs lie outside the convex hull of stations, where station coverage is less optimal. These two plots together illustrate a way to demonstrate relative station coverage over a region of interest. Figure 3 presents an ensemble of event-pair differences between mean CEA estimates at each event-node from two calibrations applied in pre-test/post-test order. Both absolute differences and percent difference relative to CEAs from the pre-test calibrations are shown with summary statistics. In this case, almost 88% of the event-nodes show decreases in mean CEA. Regions with increases in CEA as a result of applying calibrations can be pinpointed by appropriate geospatial plots. Figure 4 analyses changes in mean 90% CEAs as a result of applying the two calibrations in pre-test/post-test order. Here the statistic of interest is the proportion of mean CEAs from the ensemble of event-nodes crossing (becoming larger or smaller than) 1,000-km$^2$ 90% CEA threshold.

Any of these displays can be partitioned into subsets, either based on a priori stratification, or a posteriori observation, in order to explore fine-structure of response to calibrations. Examples are events in different geographic regions, or events detected by certain stations, or events-station pairs for which seismic travel paths cover certain regions of interest in order to explore fine-structure of response to calibrations.

In regions without good station coverage or well-known velocity structure, it is more difficult to obtain accurate regional event locations. Moreover, the standard CEA may lead to an impression of more accurate event location than actually is the case (Kadinsky-Cade et al., 1995). Much depends on the choice of appropriate phases for given regions, as well as reliability in detecting and picking phases. Seismic event location is complicated by the presence of uncertainty. Many of the inputs are simply not well known, and are subject to quasi-random processes, which means that sometimes events will be detected, sometimes not. Sometimes enough stations detect the event to allow triangulating an event location, sometimes not. The degree to which the signal is visible above noise may also be random, and the weaker the signal, the more uncertain will be the predicted location. To consider the impact of these uncertainties is one reason to perform network simulations with Monte Carlo techniques.

**Uncertainty in Regional Phase Picks**

The detection and timing of seismic arrivals play a critical role in the ability to locate seismic events, especially at low magnitude. Errors can occur with the determination of the timing of the arrivals, whether these errors are made by automated processing or by an analyst. One of the major obstacles encountered in properly estimating travel-time picking error is the lack of a clear and comprehensive discussion of all of the factors that influence phase picks. We have developed a multivariate statistical model, experimental design, and analysis strategy that can be used to study possible factors that need to be modeled to properly study phase arrival time picking errors. We have embedded a general form of the International Data Centre (IDC)/U.S. National Data Center (USNDC) phase-pick measurement error model into our statistical model. We can use this statistical model to optimally calibrate a picking error model to regional data.

## CONCLUSIONS AND RECOMMENDATIONS

A sampling of statistical and visualization tools for comparing simulation results from two calibrations is briefly described. These tools can be based both on primary measures directly output by the simulation runs, or based on counts of categorized input or variables. To obtain reliable results, it is critical that computed results are obtained through numerically stable computations, that appropriate stratification has been applied, either regionally or by other variables of interest, to avoid sampling biases, and that realistic uncertainties are applied to inputs to Monte Carlo runs. The most powerful statistical tests are tests based on pairs of event grid-nodes and on pairs of individual event realization from the two calibrations, if all other factors are equal.

## REFERENCES

Bratt, S. R., and T. C. Bache (1988). Locating Events with a Sparse Network of Regional Arrays. Bull. Seism. Soc. of Am., Vol. 78, No. 2, pp 780-798, April.

Claassen, J. P. (1994). The Monitoring Potential of the Open Worldwide Networks. SAND93-1931, Sandía National Laboratories, Albuquerque NM.

Kadinsky-Cade, K., R. O. Joh, A. Dainty, and J. Cipar (1995). Practical Issues in Location of Small Events: Poor Station Coverage and Poorly Known Velocity Structure. Proceedings of the 17th Annual Seismic Research Symposium on Monitoring a Comprehensive Test Ban Treaty, 12-15 September 1995, PL-TR-95-2108, Phillips Laboratory, Hanscom AFB, MA.

McKay, M. D., W. J. Conover, and R. J. Beckman (1979). A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics, Vol 221, pp. 239-245.

Ryaboy, V. (2000), IMS Calibration in North America and Northwestern Eurasia Utilizing 3-D Crustal and Upper Mantle Velocity Models to Refine Location of Regional Seismic Events. 22nd Annual DoD/DOE Seismic Research Symposium, Volume II, pp 349-358.

Sereno, T. J. (1991). Simulation of the Detection and Location Capability of Regional Seismic Networks in the Soviet Union. Annual Technical Report SAIC-91/1061, Sci. Appl. Intl. Corp., San Diego, CA.

Sereno, T. J., Bratt, S. R., Yee, G. (1990). NETSIM: A Computer Program for Simulating Detection and Location Capability of Regional Seismic Networks. Annual Technical Report SAIC-90/1163, Sci. Appl. Intl. Corp., San Diego, CA.

Sleep, N. H. and K. Fujita (1997). Principles of Geophysics. Blackwell Science Publishing, Malden, Massachusetts.

Zhang, T-R., T. Lay, S. Schwartz, and W. R. Walter (1996), Variation of Regional Seismic Discriminants with Surface Topographic Roughness in the Western United States. Bull. Seism. Soc. of Am., Vol. 86, No. 3, pp 714-725, June.