

***A PRIORI* EMPIRICAL CALIBRATION FOR MULTIPLE-EVENT LOCATION METHODS**

Michael S. Antolik, Jason P. Erickson, Mark A. Tinker, Kevin D. Hutchenson,
Gordon D. Kraft, and Linda J. Owiesny

Quantum Technology Services, Inc.

Sponsored by Air Force Technical Applications Center

Contract No. F08650-01-C-0063

ABSTRACT

Multiple event location methods that explicitly solve for empirical path correction terms between the source region and receiver, such as Progressive Multiple Event Location, or PMEL (Pavlis and Booker, 1983), contain a singularity in the equations of condition that prevents the unique recovery of the path corrections and resolution of the absolute location of an event cluster. The standard practice for dealing with this singularity is to choose one event as a “calibration” event and to calculate the path corrections while holding its location fixed, or to use one or more master events. However, it is often the case that no independent information is available to constrain the locations of any of the cluster events, and the results obtained from the multiple event relocation may be highly dependent on the choice of calibration event. A significant improvement in these methods can be made, therefore, if an *a priori* test applied to data from an event cluster can aid in choosing a good calibration event.

In this paper we develop an *a priori* technique for selecting the best calibration events from a cluster and apply it to the PMEL method. We develop and test the technique using groundtruth data sets from nuclear test sites, mining blast data, and synthetic data. The procedure involves first the location of each member event independently using single event location (SEL) procedures, followed by the application of PMEL using each event as a calibration event. The results reveal a significant correlation between certain uncertainty parameters obtained from the SEL results and the mislocation (defined as the sum of the distances of all events in the cluster from groundtruth) obtained from PMEL. From the SEL parameters, we then define a calibration statistic (ϕ) that tends to have larger values for poor calibration events. ϕ is formed from a weighted sum of the deviations of the root-mean-square travel-time residual, the error ellipse strike, and the error ellipse eccentricity from the cluster mean of those quantities. The calibration statistic performs better than random chance in discriminating between acceptable and poor calibration events and performs about as well on the test data sets as with the development data set. It also works well using both teleseismic and regional phases.

The form of the dependence of PMEL mislocation on ϕ suggests that a critical value of ϕ can be used to identify poor calibration events. The best results are obtained with $\phi_{\text{crit}} = 0.25$. However, the best value of ϕ_{crit} can vary if the SEL uncertainty parameters (and, therefore, the ϕ -value) do not vary sufficiently between the member events of a cluster, and in some rare cases the correlation may break down. The level of scatter indicates that between 10–40% of the events in a cluster may be misclassified using this method, but in each case nearly all of the classification errors result from rejection of valid calibration events. The cost of making such an error is much less than using a calibration event that results in poor locations. The method also relies on having a sufficient number of member events within a cluster to calculate meaningful statistics, but the overall results suggest that this technique may aid the improvement of location estimates in regions that have not yet been calibrated.

OBJECTIVE

Multiple-event location algorithms are commonly used to jointly locate a set of clustered events while accounting for inaccuracies in the Earth model used to estimate the travel time along the source-receiver path. It is usually assumed that if the distance between the clustered events is small compared to the source-receiver distance, then the inaccuracies in the Earth model are the same for each event at all of the recording stations. Thus, by jointly locating the entire cluster, the necessary perturbations to the reference Earth model can be estimated and used to improve the accuracy of the resulting location estimates.

Two general types of multiple-event methods exist. The first is differencing methods such as Double Differences (DD) (Waldhauser and Ellsworth, 2000). These methods use residual differences between clustered events recorded at common stations to remove the effect of heterogeneity along the travel path. Although the DD method can in theory resolve the complete location of an event cluster (absolute location in space as well as correct relative locations between the events), the ability of the method to resolve absolute locations is subject to the same limitations as other multiple-event methods imposed by errors in the data and strong near-source velocity contrasts (Menke and Schaff, 2004; Michelini and Lomax, 2004). It also requires the presence of abundant nearby seismicity to achieve the most accurate relative locations.

The second general type of multiple-event method is the heterogeneity-based scheme where perturbations to the Earth model are explicitly solved for as part of the location process in the form of path correction terms. Examples of this type are the Joint Hypocenter Determination (JHD) method (Douglas, 1967) and the Progressive Multiple Event Location (PMEL) method (Pavlis and Booker, 1983). Both methods are effective in regions of sparse seismicity, and thus are quite useful in the nuclear explosion monitoring context. Unfortunately, they suffer from a singularity in the equations of condition that prevents the unique recovery of the path corrections and resolution of the absolute location of an event cluster. The standard method for dealing with this ambiguity is to choose one event whose location is very well known through independent means, and fix either the origin time or hypocenter location, or both. This allows the locations and origin times of all other events in the cluster to be tied to this calibration event, and the path corrections can then be uniquely determined.

The problem with this practice is that most often no independent information is available to constrain the locations of any of the events under consideration. Therefore, the choice of a calibration event becomes arbitrary. It also becomes problematic as to how to assess the uncertainty in the path corrections when estimating location error. Practical considerations usually mean that the *a priori* location estimates of the events in the cluster will differ in their quality, which may affect their utility as calibration events. The quality of these location estimates will depend on the station configuration that recorded each event, which may differ significantly, particularly if the time span of occurrence of the cluster is large. In addition, unless correlated relative arrival time picks have been made for the events, the random measurement error contained in the phase picks will most likely be different for each event. As a result, it has been found that the locations obtained for PMEL are often highly dependent on the particular choice of calibration event (Erickson et al., 2003). A significant step toward improvement of multiple-event location methods can be made, therefore, if an *a priori* test applied to data from an event cluster can aid in choosing the best calibration event. The main objective of this study is to investigate multiple-event location statistics and, if possible, to develop such a test.

Ideally, the “best” calibration event is that which places the locations of all events in a cluster closest to their actual, or “groundtruth,” locations. If the groundtruth location is known for one or more events of the cluster, then the obvious choice for calibration is already made, assuming that errors in the data are small. Therefore, any test designed to determine the best calibration event must rely on a statistic which does not depend on knowledge of groundtruth. This statistic should also be readily determinable from the available data before the relocation is attempted.

In the following section, we develop and test a technique for identifying poor calibration events from within an event cluster for multiple-event location methods. We use several groundtruth data sets from nuclear explosion test sites and mining sites as well as synthetic data. We apply the technique to the PMEL method, although it is applicable to all multiple-event location methods that require the use of a calibration or master event to resolve the absolute cluster location. The PMEL method, as outlined by Pavlis and Booker (1983), differs from other

27th Seismic Research Review: Ground-Based Nuclear Explosion Monitoring Technologies

heterogeneity-based methods in that the path corrections and hypocenter parameters are determined separately in a two-step, iterative inversion process rather than simultaneously.

RESEARCH ACCOMPLISHED

Test Procedure

The test procedure applied to each data cluster consisted of two parts. In step 1, each member event was relocated using standard single-event location (SEL) procedures. The groundtruth locations were used as the initial seed locations. We then calculated various statistics from the results for each member event. These included the root-mean-square (RMS) travel time residual, the area of the resulting confidence error ellipse, and the eccentricity and orientation of this ellipse. In addition, we also examined attributes such as the origin time error estimated for each event, the number of defining phases, the azimuthal gap in the station distribution, and the average *a priori* data error, or *deltim* parameter, assigned to the phases. This *a priori* error is used for data weighting in the inversions and also for calculation of the hypocenter error ellipses.

In step 2, we relocated the cluster multiple times using the PMEL method, each time selecting a different event for calibration. We then calculated a statistic which quantifies the overall mislocation for the events in the cluster. The mislocation is defined as the horizontal distance between each event and the corresponding groundtruth location. We assumed that depth had independently been determined from other means and fixed the depths of the events to the surface in both steps. Examination of each event was then carried out to check for possible correlations between the statistics calculated in step 1, and the mislocation statistic calculated after step 2.

Development of the Calibration Discriminant

A natural choice to apply these tests is the group of nuclear explosions from the Lop Nor test site. Highly accurate groundtruth locations have been published based on satellite imagery for most of the 21 known explosions conducted at this site (Fisk, 2002). The magnitudes for these events are all large and they were well recorded at teleseismic distances. However, little or no regional phase data are available. We used a subset of 18 events dating from 1978 until 1996 having between 6 and 14 first-arriving *P* phases. The distance range of the data was between 14° and 95°. The phase picks were defined using a waveform correlation (Erickson et al., 2003) basis, and a varying *deltim* was assigned to the phase picks to best represent the estimated measurement error. In order to obtain a more diverse set of error ellipses in the single-event locations, the analyst-defined *deltim* values were used even though no attempt was made to account for model error.

No direct correlation could be found between the absolute size of the SEL error statistics and the PMEL mislocation statistic for the Lop Nor data. Figure 1 shows the size of the sum of the lengths of the mislocation vectors (normalized by the maximum for the cluster) plotted against the RMS travel time residual, the error ellipse size, the number of defining phases, and the average phase *deltim* from single-event location, when that event is used as a calibration event in PMEL. Although several choices for a mislocation parameter could be made, we found that there was virtually no difference between using the RMS length, the average length, and the sum of the mislocation vector lengths. We therefore discuss only the sum of the mislocation lengths in this study.

Several conclusions can be drawn from Figure 1. First, it can be seen that the majority of the events result in similar values of the mislocation statistic when used as calibration, but 6 of the 18 events result in a substantially larger mislocation. The median sum mislocation length in Figure 1 is 265 km, but the subset of 6 events with the largest mislocation values all have values exceeding 320 km. There is no relation between the number of defining phases used or the average *deltim* of an event and the mislocation statistic. This is somewhat surprising since these two parameters are often associated with location quality in standard single-event procedures. There is also no simple relation between the size of the travel time residual or the SEL error ellipse and the magnitude of mislocation. However, it is significant that 3 out of the subset of 6 events with the largest values of the mislocation statistic also have much larger SEL travel time residuals than the other events in the cluster. The other three events have RMS residual values lower than the average for the cluster. A similar pattern is seen in the area of the error ellipse, which is not unexpected since the SEL error ellipse size is partly determined by the size of the data residuals, in addition to the specific values of *deltim*.

Examining other variables, we find that all 3 of the events which have a large mislocation statistic but RMS residual values lower than the cluster average also have a SEL error ellipse whose strike differs from the cluster median by more than 20° (Figure 2). The median error ellipse strike is about 80° , and all but a handful of events cluster near this value. Of those events in this cluster, two belong to the above-mentioned subset of six events. These two events have among the largest values of RMS residual for the entire cluster. The other four events in this subset have SEL error ellipse strikes that differ from the median by a large value, only one of which also has a high RMS residual value (this event has both the highest normalized RMS residual and mislocation value for the entire data set). There is a clear correlation shown in Figure 2 between events having a large mislocation value and those having an SEL error ellipse whose orientation differs significantly from the cluster median. In addition, the correlation between the SEL RMS residual and the strike of the error ellipse is low, such that the combination of the two parameters enhance the capability of predicting whether a particular event will cause a large mislocation value when used as a calibration event for PMEL. The SEL error ellipse orientation is primarily a function of the station configuration, as well as the *a priori* data error (*deltim*). The data residuals, on the other hand, affect only the area of the ellipse. Thus it is not unusual for these two parameters to be uncorrelated.

Based on these results, we constructed a discriminant function from the SEL parameters designed to identify poor PMEL calibration events within a cluster. Figure 2 suggests that we need at least two parameters to do this. On further examination of the results, we chose three parameters; in addition to the two discussed above, we added a third (the eccentricity) which defines the shape or elongation of the SEL error ellipse. The Lop Nor data set also shows some correlation between the SEL ellipse eccentricity and the mislocation statistic, although it is not as strong as for the first two parameters. After examining the performance of discriminant functions involving various combinations of parameter weighting, we adopted the following as the calibration discriminant parameter ϕ :

$$\phi = \hat{r} + \frac{1}{2}\hat{s} + \frac{1}{2}\hat{e}, \quad (1)$$

where

$$\hat{r} = \frac{r - \bar{r}}{\max|r - \bar{r}|}, \quad \hat{s} = \frac{|s - \bar{s}|}{\max|s - \bar{s}|}, \quad \hat{e} = \frac{|e - \bar{e}|}{\max|e - \bar{e}|}, \quad (2)$$

and r , s , and e are the RMS residual in seconds, the strike of the SEL error ellipse, and the SEL error ellipse eccentricity, respectively. The overbars in equation (2) indicate the mean of that parameter over the cluster of events. Higher values of ϕ for an event tend to be associated with larger mislocation values when used as a calibration event. The variable \hat{r} is defined so that only positive deviations of the RMS residual from the cluster mean cause an increase in ϕ , whereas any deviation of s and e from the cluster mean will do so. This is because there is no reason to expect that deviations in the orientation or elongation of the SEL error ellipse of one sign or the other will correlate more strongly with larger mislocation values, as indicated by Figure 2. Note that defining ϕ as in equation (1) means that it does not depend on the choice of confidence level or the number of degrees of freedom assumed in calculating the error ellipse (Jordan and Sverdrup, 1981). This is because although the area of the error ellipse depends on \hat{r} , it does not explicitly enter into equation (1). The possible range of values of ϕ is -1.0 to $+2.0$.

Figure 3 shows the mislocation statistic plotted against ϕ for the Lop Nor explosions. The horizontal red line represents an arbitrary threshold used to divide the data set into poor or valid calibration events for the purposes of evaluating performance of the discriminant. For this data set the division is easy to make because of the large separation between the poorest six events and the others in the cluster. However, for other data sets the application of this *a posteriori* threshold may not be as simple. The discriminant function does an excellent job of separating the subset of poor calibration events from the other events, although there is considerable scatter. The degree of success can be measured by the value of the squared point biserial correlation coefficient R^2 , which is 0.37. The value of R^2 indicates how much of the variation in mislocation is explained by the variation in ϕ . To demonstrate that consideration of the error ellipse shape aids in the discrimination process, we show in Figure 4 the plot of \hat{r} vs. the mislocation statistic. The correlation between \hat{r} and mislocation is obviously less in this case, and the value of R^2 drops to 0.28.

Other Test Data Sets

The reliability of the calibration discriminant was verified by testing the process on two independent groundtruth data clusters. The first cluster consists of mining blasts from the Powder River, Wyoming, region. The second consists of Nevada Test Site (NTS) explosions. The 13 blasts in the mining data set were associated with particular coal mines in the region based on data from a local station (Erickson et al., 2003) at a distance of 1° . The location of each event within the particular mine is not known, however. The area of each mining site varies from a few 10s to about 100 km^2 (Erickson et al., 2003). Thus, the precision of the groundtruth locations is probably 5 km or better. Each event was located using a sparse array of regional stations, with 5 phases per event.

The discriminant function for the Wyoming data is plotted against the sum mislocation in Figure 5. The events do not separate as readily into two populations as in the previous case. The choice of a cutoff value for the mislocation statistic is therefore not as clear as for the Lop Nor data set, but should be made so that a sufficient number of events is defined as falling into both categories. Choosing 90 km as the threshold value identifies 7 of the 13 events as poor calibration events, and the resulting value of R^2 is nearly the same as for the Lop Nor data, 0.38. We also note that the use of \hat{s} and \hat{e} does not aid in discrimination in this case; in fact, $R^2 = 0.50$ when \hat{r} is used as the only predictor variable. This is probably because the station configuration for all 13 events is identical, and the orientation and shape of the SEL error ellipses are determined only by the *deltim* parameter. Therefore, there is far greater similarity among the SEL error ellipses than for the Lop Nor data, which represent a much longer time span of occurrence and were recorded by a more diverse network of stations. In any case, it is evident that the size of the RMS residual is the most important indicator of the reliability of an event for calibration in both cases.

A diverse set of teleseismic, regional, and local phase data is available for NTS explosions. In order to better simulate a limited data scenario often encountered in monitoring situations, we restricted the analysis to first-arriving phases at far-regional distances ($4\text{--}15^\circ$). All of the explosions date from the early 1990s. For this data set, all phases have a default *deltim* of 0.5 s. The results are shown in Figure 6. Because of the constant default *deltim* parameter, the SEL error ellipses are quite similar and most of the ϕ values cluster between 0.5 and 1.0. However, a similar dependence of ϕ on mislocation is observed as for the other data sets. A separation threshold of 130 km identifies 5 of the 17 explosions as poor calibration events, and $R^2 = 0.30$. The fact that ϕ is just as successful in identifying poor calibration events from independent data sets is encouraging.

Synthetic Experiment

Because the amount of data with knowledge of groundtruth is limited, it is desirable to examine the performance of the calibration discriminant using synthetic data. We derived a synthetic cluster of 17 events having direct *P* phases recorded at 8 stations. The tests were carried out using a different version of the data in each case. First, synthetic arrival times were calculated at each station using a 1-D reference Earth model, and the locations were derived using this same reference model. A constant model error was added to the phases for each station in addition to a random measurement error. We also varied the weighting, or *deltim*, assigned to each phase. Three types of weighting were used. The first type simply involved an equal default value of *deltim* for each phase. The second type added a random perturbation to *deltim*, while in the third type, the *deltim* assigned to each phase was correlated to the data residual for each phase, as measured by the combination of model and measurement error. This latter type of situation may arise when the analyst is assigning *deltim* values to a data set of phase arrivals that have a low signal-to-noise ratio.

The results of these tests, not shown for issues of space, generally validate the results obtained using real data. The correlation between ϕ and the PMEL mislocation generally increases with the diversity of the error ellipses. Similar correlations to the real data examples are obtained where the *deltim* values are not constant. The PMEL mislocation obtained using a particular event for calibration has little dependence on ϕ if the station distribution for all events is similar and the *deltim* values are constant.

Method for Selection of Calibration Events

Having calculated ϕ for a particular data set, the question remains as to how to identify an appropriate calibration event or events, since in most applications groundtruth information will not be available. The level of scatter observed in these experiments suggests that a restrictive criterion should be used. However, it is probably not advisable to simply choose the event having the smallest value of ϕ , because it is not likely that this event will produce the best results. In Figure 5, for example, the event having the smallest ϕ -value has the 5th largest mislocation of the entire cluster. Application of the discriminant function only means that it is likely that the events resulting in the largest mislocation for a given cluster will not be found among those events having the smallest ϕ -values. The reason this is true is because the location error is stable for small values of ϕ , but its variance increases considerably at large values of ϕ (Figs. 3, 5, 6).

The results suggest two possible approaches for selecting calibration events. The first is simply to select a calibration event from among the smallest $k\%$ of the ϕ -values. An appropriate value of k might be 10–30%, which for the test cases described in this report would amount to selecting from the 3–6 events with the smallest ϕ -values. One problem with this approach is that it becomes quite restrictive if the number of events in the cluster becomes small. In that case, however, the validity of the statistical method is reduced, and there may be only a small range of mislocation and ϕ -values across all of the events of the cluster. If a cluster involves only a handful of events with an identical station distribution, then it will most likely not be possible to use this method to limit the choice of calibration event.

Another obvious selection criterion is simply to define a cutoff value of ϕ (ϕ_{crit}) and choose from among those events having a smaller ϕ than the cutoff. The drawback of this approach is that a cutoff value appropriate for one cluster might not be appropriate for another. However, the test cases described in this study suggest that the best value of ϕ_{crit} may be fairly stable. This value appears to be $\phi_{crit} \sim 0.25$. Applying this cutoff value to the Lop Nor data (Figure 3) would divide the cluster into 7 events appropriate for calibration and 11 that are not. All six of the events that produce relatively large mislocation values have $\phi > 0.25$. However, 5 events from among the group with small mislocation values also have $\phi > \phi_{crit}$. Thus, we have made 5 classification errors out of 18, for an accuracy rate of 72%. Note, however, the cost of the two types of error is not the same, i.e., it is more serious to choose an event for calibration that results in large mislocation than it is to reject a valid event. With the knowledge that one-third of the events are poor choices for calibration, we would expect an accuracy level of 56% on the basis of random chance.

For the Powder River data, a cutoff ϕ -value of 0.25 results in an accuracy rate of 85% (Figure 5). In this case one of the 7 events identified as poor calibration events is misclassified, while only one valid event is misclassified. Since a cutoff mislocation value of 90 km divides the data set nearly in half, only slightly better than 50% success could be expected by random chance. Although the correlation between the two statistics is better if we define $\phi = \hat{r}$ ($R^2 = 0.5$), the classification accuracy rate does not improve if we use an appropriately adjusted value for ϕ_{crit} . In the case of Figure 6, although the correlation is good the uniform *deltim* results in a small range in ϕ -value for the events in the cluster. Consequently, a larger ϕ_{crit} of 0.5 would improve the classification accuracy to 76%, although even if ϕ_{crit} were chosen as 0.25 none of the five very poor events would be classified as valid. For nearly all of the synthetic test cases that do not use a uniform *deltim*, $\phi_{crit} = 0.25$ appears to be a good choice of cutoff value. However, we caution that it may be advisable to test this value using synthetic data for the particular station distribution represented in the real data before using it to select calibration events.

Rather than selecting one single calibration event, it might be assumed that selecting a combination of more than one, or even all, of the events classified as acceptable calibration events would be preferred. However, our tests indicate that unless very accurate *a priori* knowledge of the event locations is available, this is not the case. The reason is that when an event is selected for calibration of PMEL, its location is held fixed until the final iteration, when it is allowed to move. Therefore the tendency is that the final location of the calibration event will be very close to its initial location. If the initial location is poor, then it will be restrained from moving toward its groundtruth location. Figure 7 shows the results when different initial locations are used as the seed in the PMEL process, using the Powder River cluster. If the initial locations are groundtruth, then in addition to achieving a much smaller total mislocation in every case, the total mislocation decreases as more events are added as calibration events. On the other hand, if the locations from SEL are used as the seed, the quality of the locations deteriorates as

27th Seismic Research Review: Ground-Based Nuclear Explosion Monitoring Technologies

the number of calibration events is increased. Therefore, in most cases it is probably not true that using more than one calibration event will gain an advantage over using a single event.

CONCLUSIONS AND RECOMMENDATIONS

We have shown that parameters describing the size, shape, and orientation of the error ellipse derived from standard single-event location techniques correlate with the degree of location error when a particular event is used for calibration of the PMEL method. We have used the 95% confidence error ellipse in our calculations, but the results are applicable for any confidence level and for any degree of freedom used (e.g., Jordan and Sverdrup, 1981) in the error calculation. By using a weighted combination of these parameters, it is possible to separate potentially poor calibration events from the remainder of the events in a cluster, but it is not possible to identify the absolute “best” event. The results depend on the diversity of the single-event error ellipses (which is reflected in the range of values of ϕ for a cluster), but in most of the test cases the method performs much better than random classification of the events as either poor or valid calibration events. The applicability of the method is also dependent upon the number of events in a cluster, since it relies on a sufficient number of data points from which to draw meaningful conclusions. The method will produce the best results if accurate estimates of the measurement and modeling uncertainty (*deltim*) for each phase are available.

The correlation between the PMEL location error and the single-event error ellipse statistic (ϕ) appears to be very nonlinear, which suggests that it is appropriate to separate the two event classes (poor or valid) using a cutoff value of ϕ . This selection criterion produces good results because the mean location error from PMEL appears to be quite stable at small values of ϕ but then increases dramatically once a threshold value is exceeded. The best threshold value found from most of the tests is $\phi_{crit} \sim 0.25$. The level of scatter, both for real and synthetic data, indicates between 10% and 40% of the events in a cluster may be misclassified using this criterion. However, it is significant that the majority of classification errors result from the rejection of valid calibration events. The cost of making such an error is much lower than using a calibration event that results in poor locations.

The best cutoff value for ϕ is stable across a wide range of test conditions, but is somewhat dependent on the particular station distribution and the phase data weights. It may be possible to calibrate the value of ϕ_{crit} for the seismicity in a particular region by performing synthetic experiments that replicate the typical distribution of stations prior to applying this method. The method outlined in this report should aid in producing the most accurate location estimates possible in a region that has not been calibrated.

REFERENCES

- Douglas, A. (1967), Joint epicenter determination, *Nature*, 215: 47-48.
- Erickson, J., A. Ortiz, M. Tinker, L. Owiesny, K. Hutchenson, and G. Kraft (2003), *Seismic location methods and capabilities*, AFTAC/TT Technical Report, Patrick Air Force Base, FL.
- Fisk, M.D. (2002), Accurate locations of nuclear explosions at the Lop Nor test site using alignment of seismograms and IKONOS satellite imagery, *Bull. Seism. Soc. Am.*, 92: 2911-2925.
- Jordan, T.H., and K.A. Sverdrup (1981), Teleseismic location techniques and their application to earthquake clusters in the south-central Pacific, *Bull. Seism. Soc. Am.*, 71: 1105-1130.
- Menke, W., and D. Schaff (2004), Absolute earthquake locations with differential data, *Bull. Seism. Soc. Am.*, 94: 2254-2264.
- Michelini, A., and A. Lomax (2004), The effect of velocity structure errors on double-difference earthquake location, *Geophys. Res. Lett.* 31: L09602, doi:10.1029/2004GL019682.
- Pavlis, G., and J. Booker (1983), Progressive multiple event location (PMEL), *Bull. Seism. Soc. Am.*, 71: 1753-1777.
- Waldhauser, F., and W.L. Ellsworth (2000), A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, *Bull. Seism. Soc. Am.*, 90: 1353-1368.

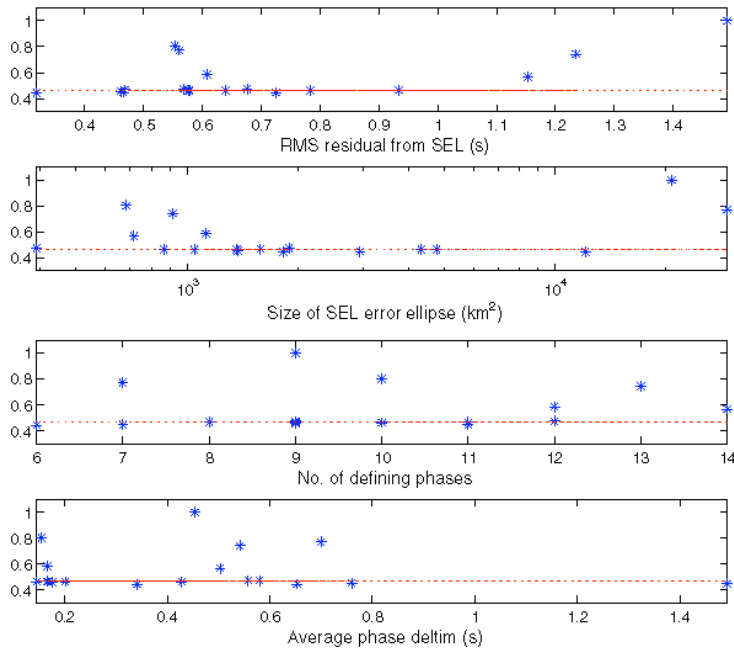


Figure 1. Mislocation normalized to cluster maximum plotted vs. SEL statistics for Lop Nor explosions. Each blue symbol represents one event, plotted at the mislocation value obtained when using that event as calibration for PMEL. The red line is the median mislocation for the cluster.

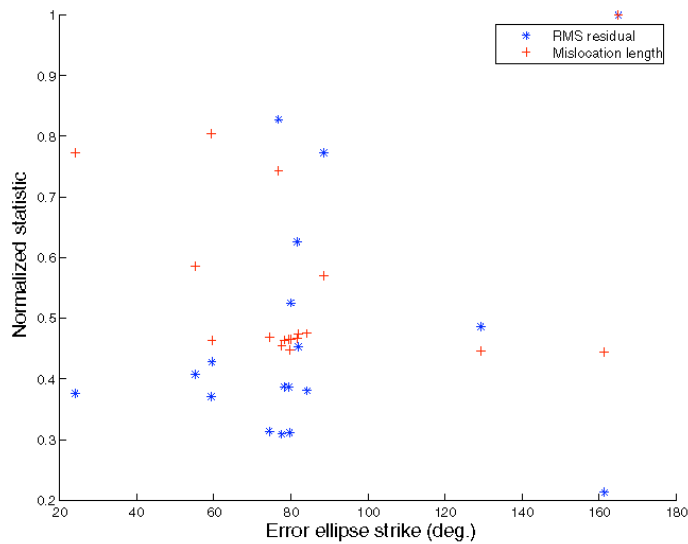


Figure 2. Normalized PMEL mislocation (red) and SEL RMS time residual (blue) for each event of the Lop Nor cluster plotted against the strike of the SEL error ellipse. The majority of events have an error ellipse strike near 80°.

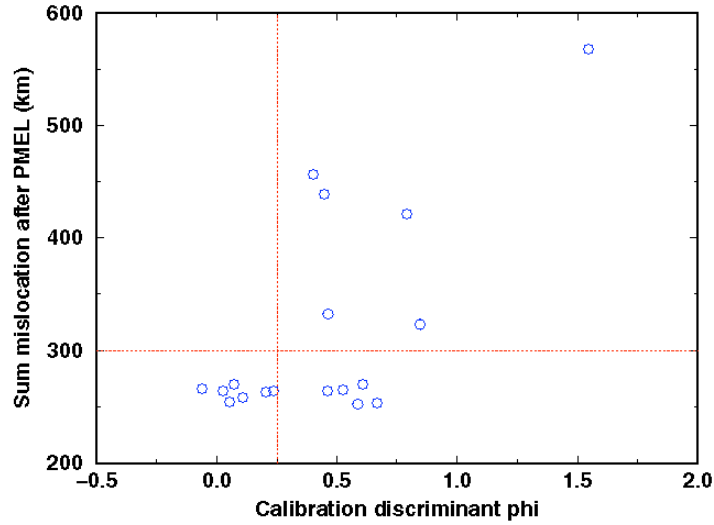


Figure 3. Relationship between the PMEL mislocation and the calibration discriminant function ϕ for the Lop Nor data. The horizontal line represents the *a posteriori* threshold applied to distinguish between good and poor calibration events, while the vertical line denotes $\phi_{crit} = 0.25$ (see text). Events that are correctly classified using these parameters lie in the lower-left and upper-right quadrants of the plot.

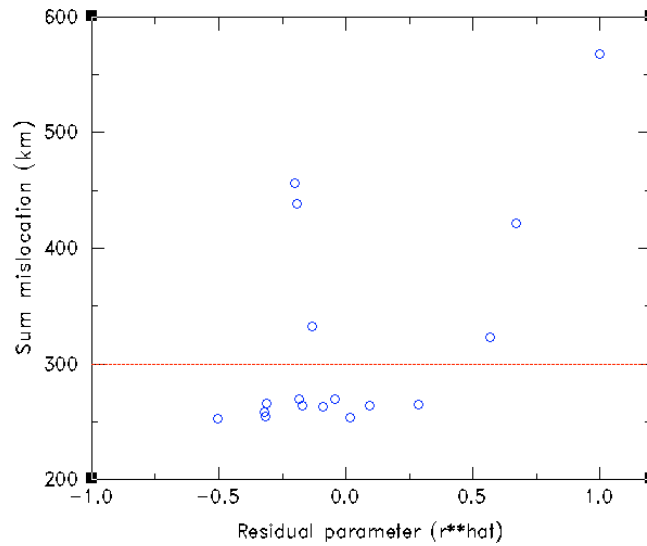


Figure 4. Relationship between the PMEL mislocation and the RMS residual parameter from equation (2) for the Lop Nor data. The correlation is lower than in Figure 3.

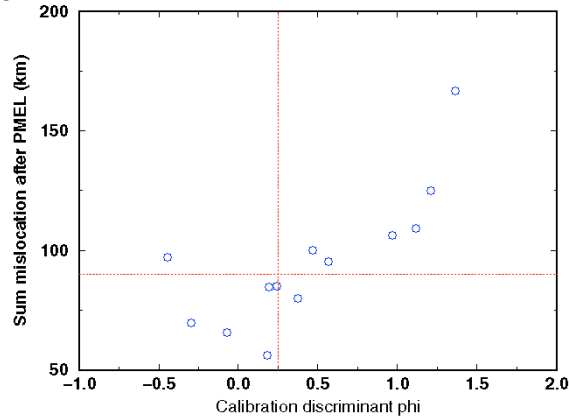


Figure 5. Same format as Figure 3 but showing results for the Powder River, WY, mining data.

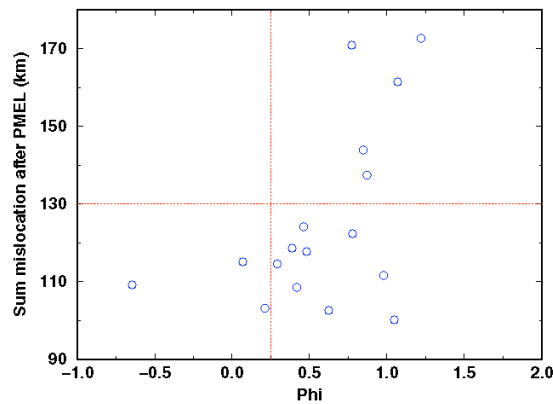


Figure 6. Same format as Figure 3 with results from the cluster at NTS. The discriminant values cluster in a narrow range of ϕ . The optimum ϕ_{crit} for this data is larger than 0.25.

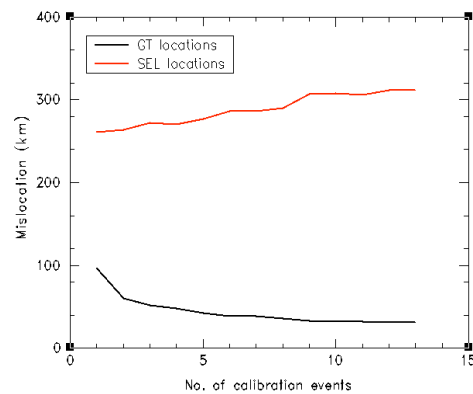


Figure 7. Results of applying different numbers of PMEL calibration events to the Wyoming mining blast data. The calibration events are applied in the order of increasing ϕ -value. The black curve was obtained using groundtruth as the initial seed locations and the red curve obtained using the individual SEL locations as seed. The total mislocation decreases as more calibrations are added if groundtruth is used but increases for the latter case.