

**IMPROVING LOCATION/DETECTION METHODS FOR MONITORING RESEARCH USING
LARGE-SCALE CORRECTION SURFACES, CROSS-CORRELATION TECHNIQUES AND GENETIC
ALGORITHMS**

Michael L. Begnaud¹, Charlotte A. Rowe¹, Lee K. Steck¹, and Woohan Kim²

Los Alamos National Laboratory¹ and Gyeongsang National University, Korea²

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

Contract No. DE-AC52-06NA25396

ABSTRACT

The availability of seismic bulletin sources presents a problem when preparing data sets for studies. With so many choices, which catalog should be used? We have developed a method of merging data from all available seismic bulletins into a single database of non-redundant phases for each event. With this new database, additional ground truth (GT) events are readily identified due to the merging of all possible arrivals for each event. The compilation of over 8500 GT25 or better events in Asia allows the generation of large-scale travel time correction surfaces. We have created Pg, Pn, P, Sg/Lg, Sn, and S surfaces for the 1382 current and historic stations that detected a GT event. The availability of correction surfaces for any and all stations in a large region permit relocations that result in greater accuracy and increased event clustering for entire seismic catalogs.

We have adapted a retroactive cross-correlation pick adjustment algorithm to function as a cross-correlation scanning detector. Application of this tool demonstrates the benefits of such procedures in areas having numerous, repeating events that give rise to highly similar waveforms. The adaptive, cross-coherency filtering method in our correlation detector enhances correlation success by comparing only the portions of waveform spectra exhibiting high cross-spectral coherency; this boosts the successful correlation of similar signals while eliminating the unintentional discarding of potentially useful parts of the spectrum which can occur when a priori filters are relied upon. In a recent application we applied the method to a swarm of earthquakes occurring in the middle Rio Grande Rift near Socorro, New Mexico, during October 2005. Initial analyst review of the data stream identified ~300 earthquakes in this seismic swarm, while application of our coherency-based cross-correlation scanner identified over 1600 repeating events.

Location of events under conditions of poorly understood seismic velocity models with non-optimal station distribution is an ongoing problem for monitoring efforts. To improve event locations, particularly for increased robustness in depth estimation, we are working with colleagues at Gyeongsang National University, Korea, on a genetic algorithm that simultaneously finds an optimal average one-dimensional (1D) velocity structure and then relocates hypocenters in this model. The genetic-algorithm method for hypocentral parameters (GA-MHYPO) searches for a best-fit model within prescribed limits to minimize theoretical and observed travel time differences, using a two-point ray-tracing method. Once the optimal model has been found for the input data, these and other nearby events are relocated using the model. We are testing the algorithm, using the best-quality events from the Himalayan Nepal Tibet Seismic Experiment (HIMNT) PASSCAL Instrument Center deployment in Nepal and southern Tibet and have obtained improved location statistics for the shallowest events, as well as minor improvements for deeper earthquakes.

OBJECTIVES

We have been investigating new and improved methods for seismic location and detection related to explosion monitoring. The objectives for these investigations are to improve methods of GT collection, data mining, and accuracy for event arrival picking, and to develop new genetic location and detection methods.

RESEARCH ACCOMPLISHED

Large-Scale Travel Time Correction Surfaces for Asia

We have developed a method of merging data from all available seismic bulletins into a single database of non-redundant phases for each event. With this new database, additional GT events are readily identified due to the merging of all possible arrivals for each event.

Many seismic location studies utilize catalog arrival information to produce travel time correction surfaces, but for a particular subset of stations and phases. In order to take full advantage of the improved locations that correction surfaces can produce for any area, it is necessary to create these surfaces for any and all current or historical stations within the region. We have been able to merge arrival information for over 150,000 events in the Asia region (Lon: 40°–125°, Lat: 10°–85°) from 1964 to present. Using this merged arrival set, we used the Bondár et al. (2004) ground truth criteria to find over 8500 existing and relocated events that pass GT25 or better (epicentral errors less than 25 km). We have created Pg, Pn, P, Sg/Lg, Sn, and S travel time correction surfaces in the Asia region for the 1382 stations that detected a GT event. Using the resulting correction surfaces, we have relocated all Asia events on a catalog scale. This included testing the effect of the corrected secondary phases on the relocations versus using first-P phases alone. The resulting locations are more consistent in their locations and errors, are more accurate, and display increased clustering around tectonic features in Asia. In addition, the resulting locations provide a catalog-scale base of improved locations for other types of regional studies.

Merging seismic catalogs

The abundance of local, regional, and global seismic catalogs presents a problem for location and GT event determination. Which catalog arrivals are best to use? The main global catalogs are: International Seismological Centre (ISC), Reviewed Event Bulletin (REB) from the International Data Center (IDC), and Earthquake Data Report (EDR) from the United States Geological Survey (USGS). Regional catalogs for Asia include the Kazakhstan National Data Center (KNDC) and the Annual Bulletin of Chinese Earthquakes (ABCE). In addition, there are many local bulletins and special studies available. Each catalog can have common stations/networks for the same event, but also with many stations/networks that are not common. By merging the catalog arrivals into a consistent, non-redundant set, we are able to test more events against the GT criteria and have all available arrivals for locating any given event (Figure 1).

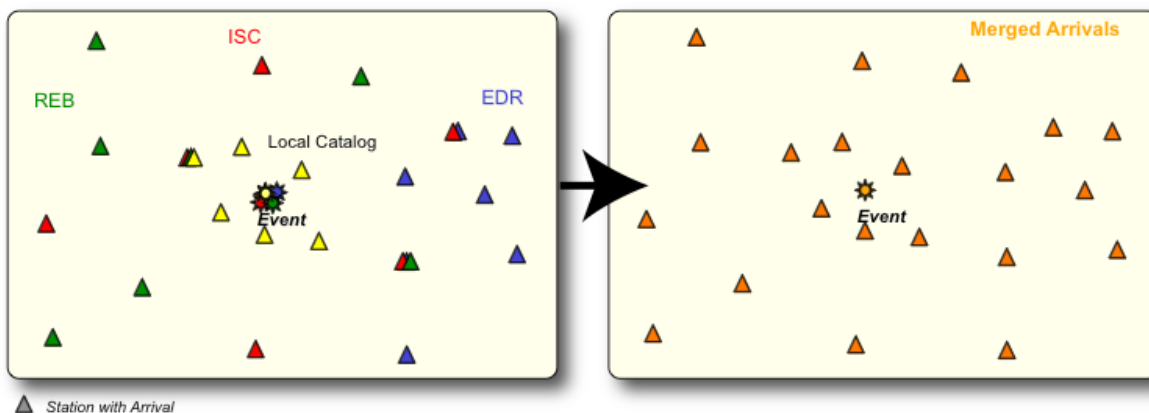


Figure 1. Cartoon example of catalog arrival merging for a single phase. Each local, regional, or global catalog is identified with a different color. Multiple triangles at a position indicate multiple catalogs listing an arrival for the same station. After merging based on an a priori ranking, each station has one arrival listed for a given phase.

Data mining for ground truth testing

After catalog arrivals have been merged and associated with an event, we process the events to mine for additional GT events, following the GT criteria of Bondár et al. (2004):

- Local networks (0–2.5°)—GT5, 95% confidence
 - At least 10 stations, all within 250 km, azimuthal gap <110°
 - At least 1 station within 30 km
 - Secondary azimuthal gap <160°
- Near-regional networks (2.5°–10°)—GT20, 90% confidence
 - Secondary azimuthal gap <120°
- Regional networks (2.5°–20°)—GT25, 90% confidence
 - Secondary azimuthal gap <120°
- Teleseismic networks (28°–91°)—GT25, 90% confidence
 - Secondary azimuthal gap <120°

A Perl script (*ARsorter*) searches through an event list and merges arrivals from ranked arrival sources together for the same event. Phases can be remapped to a consistent set. Option flags for only first-P or first-S phases can be specified. Arrivals from arrays are handled, keeping only one pick per phase per array as time defining (used for location). Travel-time tables are used to remap labeled first-P and first-S to regional name (i.e., Pg, Pn, etc.) and to look for major outliers.

We then process all events through the Bondár distance ranges, relocating with first-P arrivals only in those ranges. Events that then pass the criteria become GT events (Figure 2). Events with known/manually-set GT levels are not relocated; arrival associations are updated only.

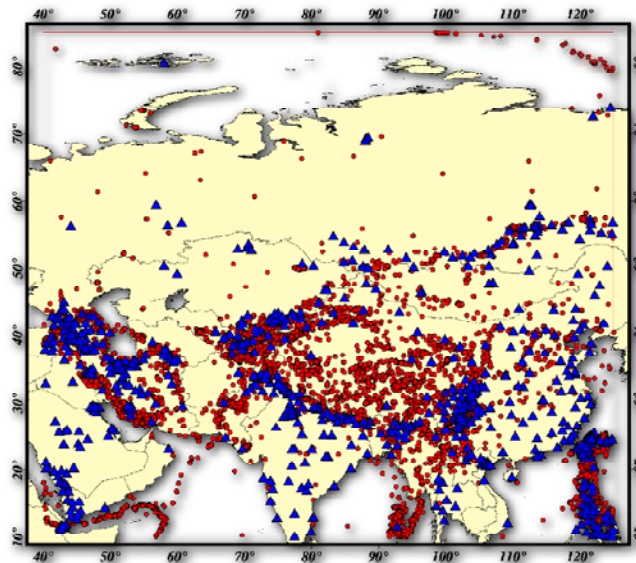


Figure 2. Stations (blue triangles) and ground truth events (red circles) for the Asia study region. A total of 1382 current and historical stations had GT event data from over 8500 possible GT25 and better events.

We select all current and historical single stations and arrays (1382 with GT data) from the Asia study region (Lon: 40°–125°, Lat: 10°–85°) (0). Array elements are not necessary as the reference stations provide the travel time correction. We have processed various global and regional catalogs for the entire globe in order to maximize the number of available GT25 and better events (over 8500) and their associated arrivals.

Correction surface generation

For the generation of correction surfaces, we use *Kbcit*, developed by Sandia National Laboratories (Ballard et al., 2004). We created two-dimensional travel time correction surfaces for P, S, Pn, Sn, Pg, and Sg/Lg phases for any available station in the region, producing 3666 individual surfaces. Phases are limited to beyond 2° because origin

depth will greatly affect points within 2° . For first-P and first-S surfaces, boundaries are placed at the change from Pn to P to limit “bleeding” of corrections across. Each phase uses the iasp91 model error surface (Kennett and Engdahl, 1991) for kriging the points, binned at 0.5° . Pg, Lg, S and Sn phases display far more scatter than Pn and P, most likely due to pick error.

Relocations and cluster analysis

To test the effects of the correction surfaces, we relocated any non-GT event in the study region having at least 4 Pn or P arrivals (first-P). No arrivals within 2° of a station were allowed because the correction surfaces were restricted to phases beyond 2° . Events were relocated using station elevation and ellipticity corrections. We performed three sets of relocations (first-P, first-P/S, all phases: first-P/S, Pg, Sg/Lg), limiting the defining phases for each set. The iasp91 model was used for relocations, with and without correction surfaces applied to determine how the surfaces affected origin clustering and phase residuals.

The resulting relocations (using *LocOO* from Sandia National Laboratories, [Ballard et al., 2004]) were binned into origin density grids using 0.05° cells having 25 km radii. The density grids, with and without correction surfaces, were differenced to help show areas of increased/decreased clustering, particularly around tectonic areas. Positive values (green) indicate shifts of iasp91 locations into those areas when using the correction surfaces. Note the areas on the northwest edge of Mongolia, west of the Tarim Basin, western India, and along the Himalayas where significant shifting of locations occurs when using the correction surfaces (Figure 3).

We plotted phase residual histograms for iasp91 locations and locations using correction surfaces for all P/S-type phases (Figure 4). The phase residual plots indicate a general reduction in residuals when using correction surfaces over iasp91 alone. Residual plots with distance (Figure 5) show dramatic bias removal for Sn and Lg phases.

Residuals for Pg and teleseismic S, however, do not show any real improvement using the correction surfaces.

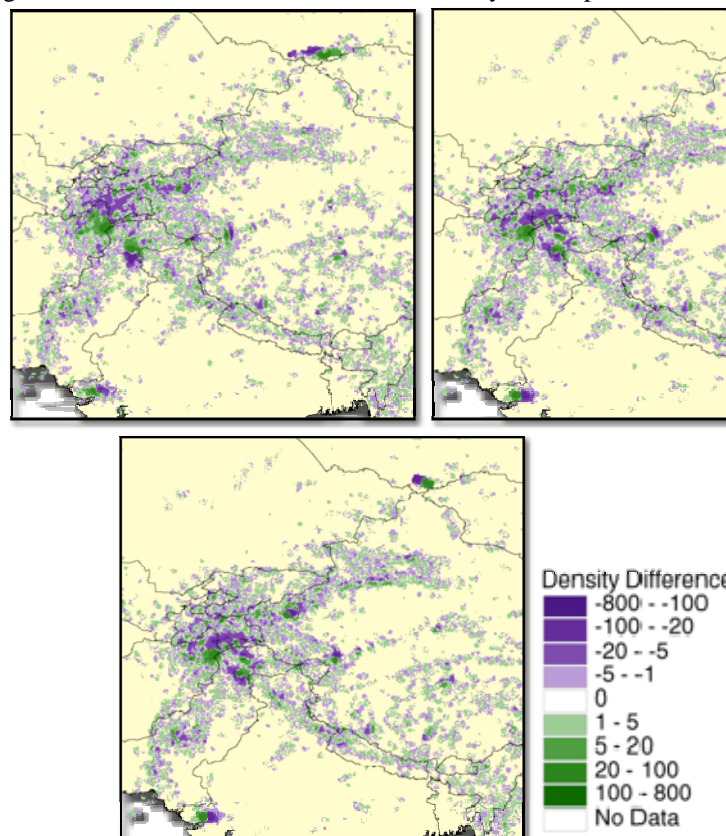


Figure 3. Origin density difference plots (density using corrections—iasp91 only) using different sets of defining phases for the western China region where the majority of events occurred: first-P (upper left), first-P/S (upper right), and all P/S (Pg, Pn, P, Sg/Lg, Sn, S). Purple indicates a reduction in origin density, green an increase in origin density. Note the shifting of linear patterns and large cluster shifts. The pattern of the shift for the prominent northeast cluster changes dramatically from first-P to using first-P/S or all P/S.

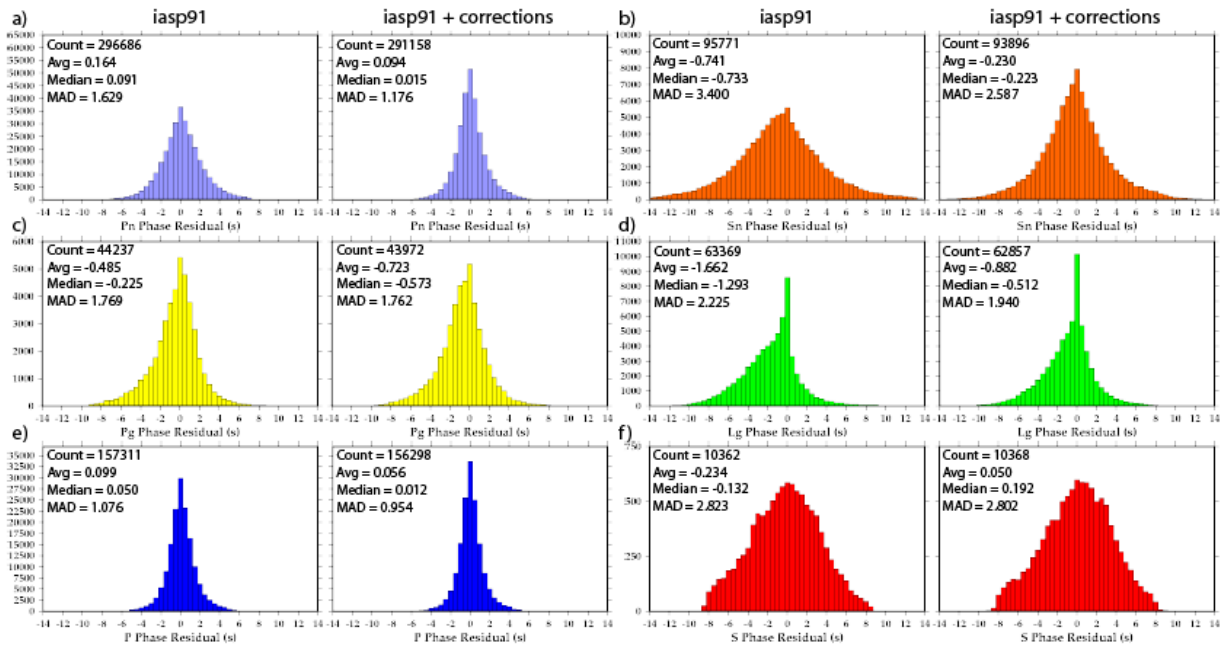


Figure 4. Phase residual histograms for relocations using iasp91 and iasp91 plus corrections for relocations using all P/S phases. a) Pn, b) Sn, c) Pg, d) Sg/Lg, e) teleseismic P, f) teleseismic S. All phases reduced average residual with decreased median and mean absolute deviation (MAD) values except for Pg and teleseismic S. These phases show virtually no change in the MAD values and slight degradation of the average/median values when using correction surfaces. Residuals for Sn show the most significant reduction in residuals when using corrections (69% for average, 24% MAD).

Cross-correlation scanning detector

Waveform cross-correlation (WCC) has been shown to significantly improve the relative hypocenter locations for events exhibiting similar waveforms (e.g., Aster and Rowe, 2000; Phillips, 2000; Rowe et al., 2002; Waldhauser and Ellsworth, 2000). Although relative event relocation has been shown to improve via WCC, it has also been shown that absolute event locations, which may be biased through overall pick bias at a given station, can be improved significantly through WCC-based phase repicking (e.g., Rowe et al., 2004).

We illustrate in Figure 6 the impact of pick bias. Shown are waveforms recorded at one station on the island of Montserrat, for a seismic swarm that occurred in 2005. The upper four panels illustrate initial phase pick scatter caused by emergent arrivals on a very noisy trace, and the improvement possible using WCC. Panels A) show waveform alignments (left) and resulting stack (right) aligned on preliminary picks. Panels B) show the realigned waveforms following WCC (left) and their associated stack (right). The lower panels show C) the Akaike Information Criterion (AIC) (Akaike, 1973) autopicker function (e.g., Leonard, 2000) applied to D) the realigned stack. Note on panel D the mean correlation pick (green line), which would be used for travel-time adjustments in all relative location schemes, is significantly later than the actual first arrival on the stack. The adjusted pick (red) based on the AIC autopicker, is closer to the correct P-wave arrival for this stack, and its contributing traces. Adjustment of all relative pick values by this offset will correct location bias in the event cluster, whose centroid may be significantly in error in terms of absolute location. This is particularly problematic for event families recorded by only a few stations; picking bias at one station can exert significant influence over the location.

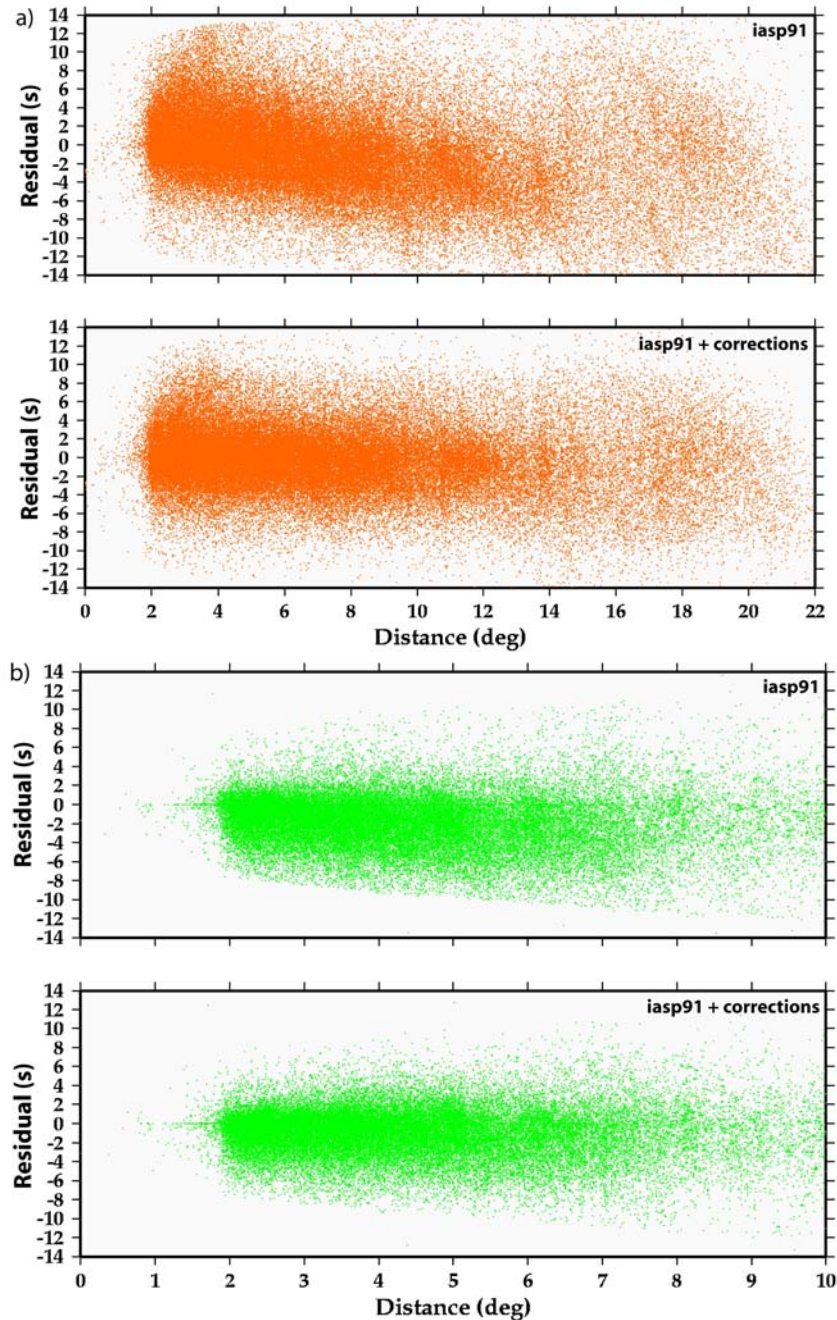


Figure 5. Examples of residuals with distance for (a) Sn and (b) Sg/Lg phases in relocations using all P/S phases, with and without corrections for iasp91. The residuals for these phases show a significant reduction in the trend with distance when using corrections.

The method of correlation detection (e.g., Harris, 2001; MacCarthy and Rowe, 2005; Rowe et al., 2005; Stankova et al., 2007) allows us to detect repeating seismic events within a waveform catalog by “scanning” the catalog with a master event or suite of basis events against which time-windowed segments of waveform data are correlated. In an application of this method to a seismic sequence that occurred in central New Mexico in 2005, the seismic cluster was identified by the data analyst as including ~300 earthquakes. Applying the WCC scanning detector (MacCarthy and Rowe, 2005) to the continuous digital waveform archive identified over 1600 events belonging to the cluster. At the same time, the WCC detector has applied a phase pick to the detected events, based on the pick for the master event. Waveforms (A) for the ~1600 detected events in the series, and the master waveform (B) used to scan the archive, are shown in Figure 7.

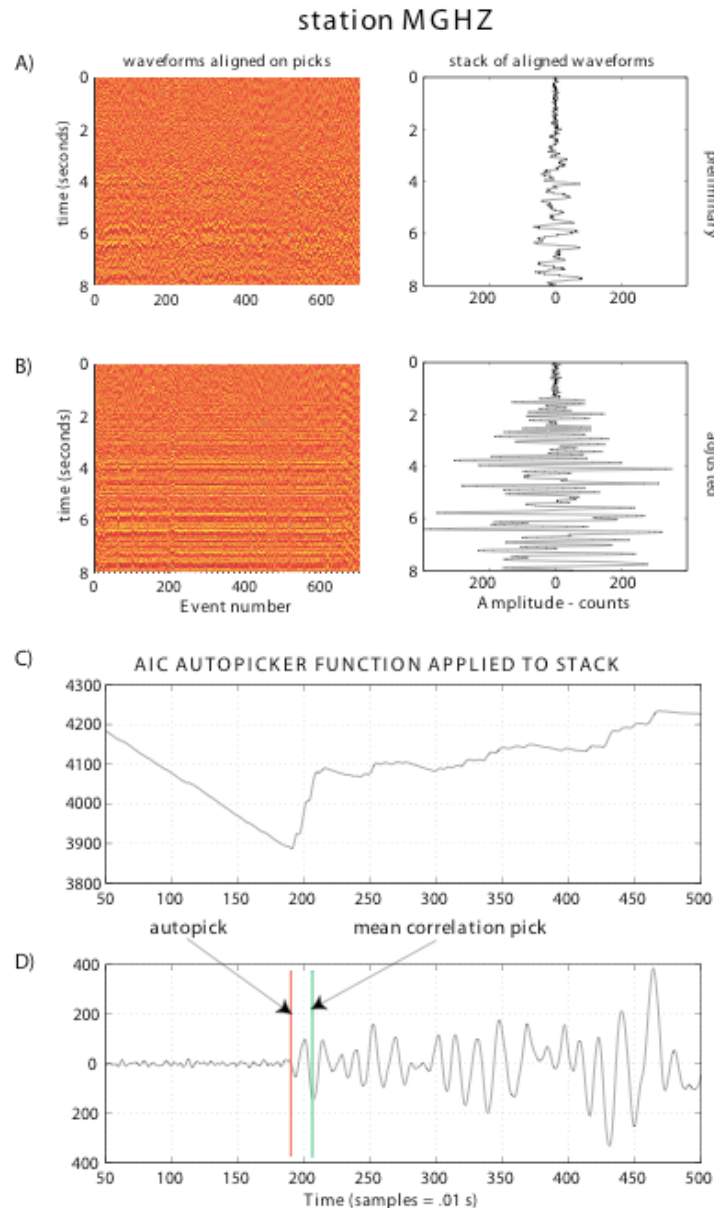


Figure 6. Illustration of correlation repicking and pick bias correction for absolute repacks. (A) 700 waveforms from a seismic event cluster aligned on preliminary picks (left) and stacked based on this alignment (right). (B) Traces re-aligned on WCC-adjusted picks (left) and stacked (right). (C) AIC autopicker function applied to the stack from B. (D) Stack from B, with mean correlation pick (vertical green line) and AIC pick (vertical red line) showing degree of pick bias at this station for these events.

Genetic Location Algorithm

The reliability of many seismic studies depends on the accuracy of hypocentral parameters. The hypocentral parameters determined by the conventional methods have errors although the P and S phase arrival time data may have no picking errors, if the true velocity structure is not used as the initial model. Unfortunately, the true velocity structure is not known for the most regions of Earth.

Using a genetic algorithm (GA) in earthquake location and velocity optimization problems, GA-MHYPO (Kim et al., 2006), searches for a global solution of the 1D velocity structure among the numerous models that are

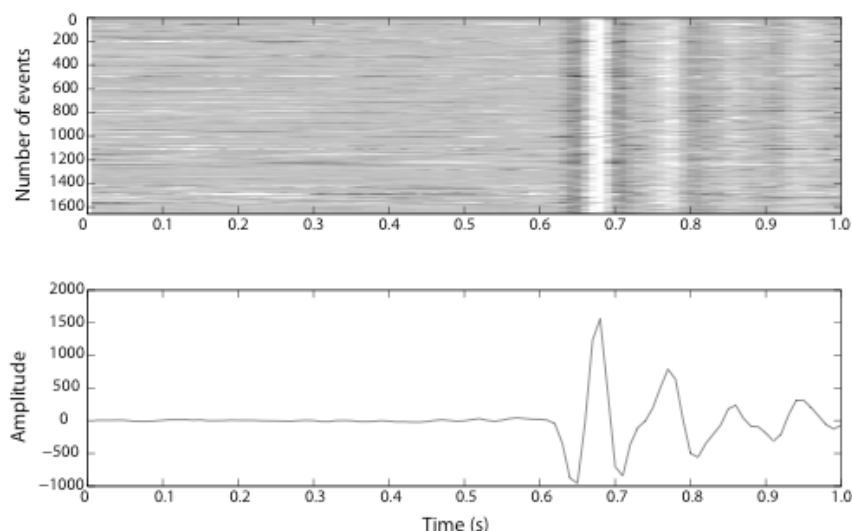


Figure 7. Earthquakes belonging to a seismic sequence occurring near Socorro, New Mexico, in October 2005. Top panel shows over 1600 waveforms recorded at station WTX, which were identified and picked using WCC detection scanning. Bottom panel shows master waveform used to scan the waveform archive. Only 300 of these earthquakes were identified during preliminary, manual analysis.

generated randomly within prescribed ranges. GA-MHYPO uses the velocity structure that provides the minimum travel-time difference between observed and calculated travel times for phases. The algorithm employs a two-point ray tracing method (Kim and Baag, 2002) to reduce error caused by ray paths between the source and receiver, and a weighting factor (Hahm et al., 2007) with respect to the take-off angle at a source to increase the accuracy of focal depth. Theoretically, GA-MHYPO can find the average true velocity structure between a source and receiver if the average true velocity structure was within the prescribed range and the phase data used has no picking error. Genetic algorithms in GA-MHYPO determine a global solution of 1D velocity structure to determine hypocentral parameters, while most existing methods use generally local solutions as the initial velocity model, which may differ from the true velocity. Therefore, the conventional methods have non-uniqueness problems depending on the basis of initial velocity models.

GA-MHYPO determines the hypocentral parameters accurately under ill conditions, because this method searches and uses the true or closer to average true velocity, and reduces computational error caused by ray path and focal depth. The ill conditions include limited number of phase data, large azimuthal gap, and noisy data with large picking error due to poor signal-to-noise ratios. We tested GA-MHYPO in a comparison with HYPOINVERSE (an example of conventional location methods) to investigate the relative location improvement for synthetic data both without, and with, picking error under a variety of situations, varying number of arrivals, azimuthal gap, and focal depth. The computational results show that the focal depth and horizontal errors determined by GA-MHYPO are less than 0.1 km for the error-free picks, and 1 km for the data with moderate picking error. Under ill conditions, the accuracy of hypocentral parameters obtained from GA-MHYPO in our test is improved by more than a factor of twenty for error-free data or a factor of five for the data with pick error, compared with those obtained by HYPOINVERSE.

The GA-MHYPO algorithm was applied to the Himalayan Nepal Tibet Seismic Experiment (HIMNT) data where 1D models have been previously calculated (Monsalve et al., 2006). Events were grouped by sites in order to process co-located events together (Figure 8). The genetic algorithm in GA-MHYPO searches for a best-fit 1D velocity profile within a prescribed range for the number of layers assigned. Dividing the depth profile into relatively small layers (5 km each for HIMNT data) permits the algorithm to search for an optimal velocity profile by perturbing each of these layers. We can see that the resulting velocity model for Group A and B (Figure 9) displays significantly more detail than the model arrived at by Monsalve et al. (2006). In the absence of additional information, the simpler model is generally best, but because of the concentration of events below 45 km, as well as from 0 km to 20 km, both the shallowest and deepest part of the velocity structure can be better characterized through use of GA-MHYPO. Between 25 and 45 km depth there are few hypocenters available to constrain the layers; hence, the values at these depths are less robust. Note in particular the significant velocity inversions found at

about 55 and about 75 km by GA-MHYPO. The large number of hypocenters contributing to these layers provides some confidence that the inversions are real, rather than an artifact arising from high-pick uncertainty for a few events. These velocity lows suggest lower crust and/or upper mantle complexity in this region of significant tectonic convergence, and may indicate the superposition of crust/mantle transition zones for the Indian and Eurasian plates.

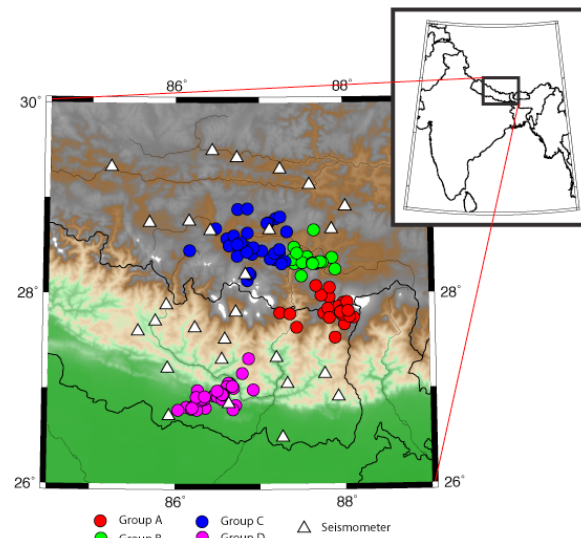


Figure 8. Events and stations from the HIMNT, with events subdivided into groups for testing with the GA-MHYPO algorithm.

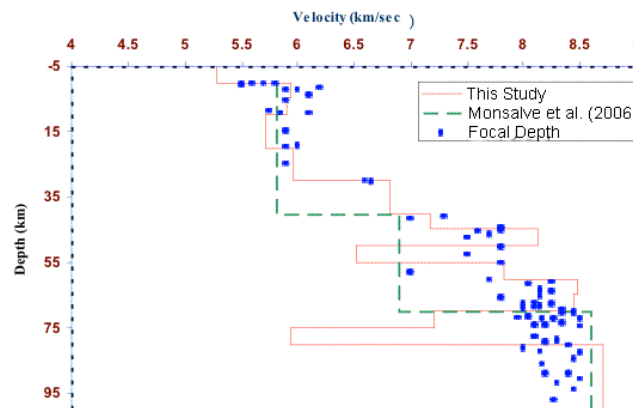


Figure 9. Velocity/depth profile using results from GA-MHYPO for event Groups A and B together. Events are plotted to show concentrations with depth. The velocity profile from GA-MHYPO has similarities to the previous Tibet profile (Monsalve et al., 2006) but is able to resolve low-velocity zones at depth.

CONCLUSIONS AND RECOMMENDATIONS

Merging seismic catalogs allows for more arrivals for relocation procedures and increases the number of GT events that can be mined from the database. The use of empirical travel time correction surfaces can be applied to ALL stations in an area, whether currently operational or historical, and permits the relocation of events across any time boundary. Relocating events in Asia using the catalog-scale correction surfaces results in significant translation and clustering of events, especially in the western China/Mongolia region. Analysis of phase residuals suggests the correction surfaces affect P, Pn, Sn, and Lg the most noticeably and that secondary phases “need” corrections if used for location.

WCC has been shown to significantly improve the relative hypocenter locations for events exhibiting similar waveforms. Although relative event relocation has been shown to improve via WCC, it has also been shown that

absolute event locations, which may be biased through overall pick bias at a given station, can be improved significantly through WCC-based phase repicking.

GA-MHYPO, a genetic earthquake location algorithm, searches for a global solution of the 1D velocity structure among the numerous models that are generated randomly within prescribed ranges. The algorithm uses the velocity structure that provides the minimum travel-time difference between observed and calculated travel times for phases and determines the hypocentral parameters accurately under ill conditions. Our models for the HIMNT region suggest the existence of two high velocity anomalies that may represent two Mohorovicic discontinuities in the India/Asia continental collision zone, each deriving from one of the two converging plates.

REFERENCES

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, in *2nd International Symposium on Information Theory*, B. Petrov and F. Csaki (Eds.), pp. 267–281.
- Aster, R. C. and C. A. Rowe (2000). Automatic phase pick refinement and similar event association in large seismic datasets, in *Advances in Seismic Event Location*, C. H. Thurber and N. Rabinowitz (Eds.), Amsterdam: Kluwer, pp. 231–263.
- Ballard, S., M. Chang, J. Hipp, L. Jensen, R. Simons, and L. Wilkening (2004). The 2004 Knowledge Base parametric grid software suite, in *Proceedings of the 26th Seismic Research Review: Trends in Nuclear Explosion Monitoring*, LA-UR-04-5801, Vol. 1, pp. 211–220.
- Bondár, I., S. C. Myers, E. R. Engdahl, and E. A. Bergman (2004). Epicentre accuracy based on seismic network criteria, *Geophys. J. Int.* 156: 483–496.
- Hahm, I. K., W. Kim, J. M. Lee, and J. S. Jeon (2007). Determining the hypocentral parameters of local earthquakes using take-off angle as weighting factor, *Geosciences Journal* 11: 39–49.
- Harris, D. B. (2001). Subspace techniques for detecting repeating events (abstract), *Seis. Res. Lett.* 72: 245.
- Kennett, B. L. N. and E. R. Engdahl (1991). Traveltimes for global earthquake location and phase identification, *Geophys. J. Int.* 105: 429–465.
- Kim, W. and C.-E. Baag (2002). Rapid and accurate two-point ray tracing based on a quadratic equation of takeoff angle in layer media with constant or linearly varying velocity functions, *Bull. Seism. Soc. Am.* 92: 2251–2263.
- Kim, W., I. K. Hahm, S. J. Ahn, and D. H. Lim (2006). Determining the hypocentral parameters for local earthquakes in 1-D using genetic algorithms, *J. Geophys. Res.* 166: 590–600.
- Leonard, M. (2000). Comparison of manual and automatic onset time picking, *Bull. Seism. Soc. Am.* 90: 1384–1390.
- MacCarthy, J. K. and C. A. Rowe (2005). Automatic scanning detection for characterization of dome-related seismic swarms at Mount St. Helens, and their evolution through time (abstract), *Eos Trans. AGU*, AGU Fall Meeting.
- Monsalve, G., A. Sheehan, V. Schulte-Pelkum, S. Rajaure, M. R. Pandey, and F. Wu (2006). Seismicity and one-dimensional velocity structure of the Himalayan collision zone: Earthquakes in the crust and upper mantle, *J. Geophys. Res.* 111: B10301, doi:10.1029/2005JB004062.
- Phillips, W. S. (2000). Precise microearthquake locations and fluid flow in the geothermal reservoir at Soultz-sous-Forets, France, *Bull. Seism. Soc. Am.* 90: 212–228.
- Rowe, C. A., R. C. Aster, B. Borchers, and C. J. Young (2002). An automatic, adaptive algorithm for refining phase picks in large seismic data sets, *Bull. Seism. Soc. Am.* 92: 1660–1674.
- Rowe, C. A., J. K. MacCarthy, F. Giudicepietro, S. Scarpetta, and S. R. Taylor (2005). Seismic event identification using scanning detection: a comparison of denoising and classification methods (abstract), *Eos Trans. AGU*, AGU Fall Meeting.
- Rowe, C. A., C. H. Thurber, and R. A. White (2004). Dome growth behavior at Soufriere Hills Volcano, Montserrat, revealed by relocation of volcanic event swarms, *Jour. Volc. Geotherm. Res.* **134**, 199–221.
- Stankova, J., S. Bilek, C. A. Rowe, and R. C. Aster (2007). Characteristics of the October 2005 microearthquake swarm and reactivation of seismic swarms over decade-scale time periods near Socorro, New Mexico, *Bull. Seism. Soc. Am.*, in review.
- Waldhauser, F., and W. L. Ellsworth (2000). A double-difference earthquake location algorithm: Method and application to the Northern Hayward Fault, California, *Bull. Seism. Soc. Am.* **90**, 6, 1353–1368.