IMPROVED GROUND TRUTH IN SOUTHERN ASIA USING IN-COUNTRY DATA, ANALYST WAVEFORM REVIEW, AND ADVANCED ALGORITHMS

Eric R. Engdahl[1], Eric A. Bergman[1], Stephen C. Myers[2], and Floriana Ryall[2]

University of Colorado[1] and Lawrence Livermore National Laboratory[2]

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

Contract No. DE-FC52-03NA99516[1] and W-7405-ENG-48[2]

ABSTRACT

We continue to improve the relative and absolute accuracy of locations for recent large earthquakes through analyzing clusters of earthquakes simultaneously with the Hypocentroidal Decomposition (HDC) method of multiple event relocation. Absolute locations of such clusters are calibrated with reference (or ground truth) event information from local, aftershock deployments or from nonseismic (e.g., InSAR) constraints, which provides independent evidence for the absolute location of one or more of the cluster events. We have used both local network data and InSAR analysis of co-seismic ground deformation for this purpose. There is also potential for using mapped fault zones to provide some constraint on absolute locations. When both location and origin time can be calibrated for a cluster through use of reference event information, we are able to estimate the true travel times to all reporting stations. These estimates are the basis for improved models of the crust and upper mantle, which in the future will permit far more accurate routine earthquake locations using regional seismic data.

Much development work has been done on the HDC code as it is applied to the ground truth problem, with emphasis on statistical rigor and robustness. One focus has been on an improved algorithm for shifting of the cluster hypocentroid to best fit the available ground truth data. We also now include the uncertainties of the ground truth data in this process, which yields more accurate estimates of the uncertainties of the final estimates of absolute (calibrated) locations. The other main focus has been on phase identification. A method based on probability density functions has been implemented, but much additional work is needed for full functionality. An improved (more robust) method for estimating empirical reading errors from cluster residuals has been implemented. Developmental aspects of this work are described.

We have established arrangements for gathering and assessing potential ground truth data and phase arrival-time data for events of interest from local and regional stations in Iran. Many new ground truth (GT5) events are now being obtained as an ongoing activity, and validation of these events is in progress. In addition, we have been able to obtain and apply waveform data and phase readings from digital stations. Example waveforms are shown. For events of interest these data are supplemented with regional and teleseismic phase arrival times carefully read by an expert analyst from seismograms in the waveform data base at Lawrence Livermore National Laboratory (LLNL).

Collaborative efforts continue with scientists at Cambridge and Oxford Universities on seismotectonic applications of cluster analysis and development of new ground truth constraints from geological and remote-sensing (InSAR) data. An example of one such an effort is the analysis of the 2002 Avaj (Changureh) earthquake sequence. A large main shock was followed by a number of smaller events, and a cluster of 17 events was formed. Analysis of the smaller events in the cluster was facilitated by analyst-reviewed phase readings and phase arrival times from stations in Iran. Results using the new HDC developmental code and these new data are shown along with some associated waveforms from digital stations.

OBJECTIVES

This research seeks to improve the database of ground truth information and velocity models useful for calibration in southern Asia with the following objectives: (1) Aggressive pursuit of in-country data acquisition, especially the collection of ground truth at GT5 level or better for events of magnitude 2.5 and larger recorded by dense local networks, including associated velocity models; (2) expanded analyst review of relevant regional waveforms for ground truth events by the comprehensive re-picking of phase arrival times from all available waveforms, with special attention to the regional phases Pg, P*, Pn, Sg, S* and Sn; and (3) application of advanced algorithms, such as those used for multiple event relocation, to refine and validate all available ground truth data, to achieve the optimal selection of data for analysis, to better understand the uncertainties of the results, and to handle the error budget as realistically as possible.

RESEARCH ACCOMPLISHED

Recent Developments of the Hypocentroidal Decomposition Method of Multiple Event Relocation for Application to Ground Truth Studies

Confidence Ellipses for Cluster Vectors

A review of the algorithm used in the hypocentroidal decomposition (HDC) code to derive confidence ellipses for cluster vectors (relative locations relative to the hypocentroid) led to modifications that make it more sensitive to the error budget of individual events. This algorithm was not fully developed in Jordan & Sverdrup (1981), and the HDC code originally used an approach that was based on the equivalent procedure for the hypocentroid. This approach had the effect of averaging the error budget over all events for scaling the ellipses. With more careful choices of parameters, confidence ellipses for "average" events change little, while confidence ellipses for "bad" events become larger (sometimes significantly), and those for very good events become slightly smaller. For ground truth work, in which the emphasis is on very well constrained events, there is little practical difference, but the new method is clearly more rigorous. There is no change to the estimates of cluster vectors themselves.

As shown in Figure 1, the effect of this change is seen primarily for the poorest-located events in a cluster, which tend to have larger confidence ellipses with the new algorithm. In the worst cases, an event located with a small number of readings that have large residuals, the difference can be a factor of 2 or 3 in area.

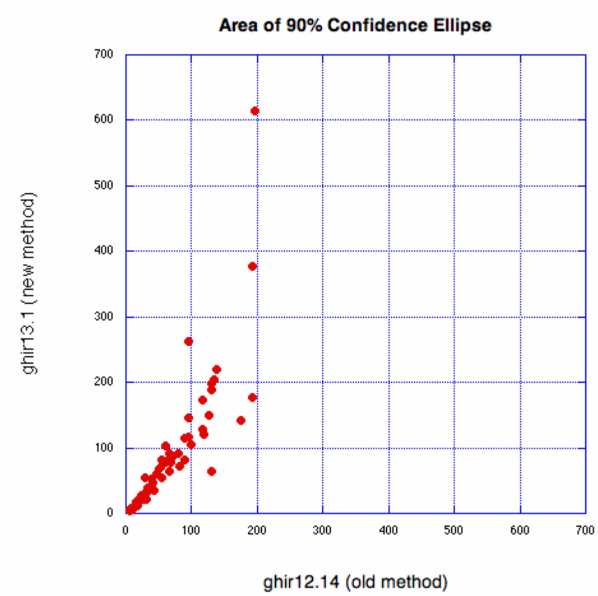

Figure 1. Comparison of the area of the 90% confidence ellipses with old and new algorithms (see text) for cluster vectors of the Ghir cluster.

Robust estimation of reading errors

One of the significant advances in location accuracy that HDC makes possible is the ability to make empirical estimates of reading error for individual station-phase combinations, based on the spread of path-corrected, normalized residuals. In the past, we made these estimates from the standard deviation (SD) of the residuals. Now we use a robust estimate of scale, "Sn," which does not need an estimate of central location and does not assume symmetric distributions (Croux and Rousseuw, 1992). The algorithm for Sn includes correction factors that make it nearly unbiased even at small sample sizes. A comparison of reading errors estimated from the same cluster data by SD and Sn in Figure 2 shows the expected effect of outliers, which cause excessively large estimates of SD. Sn is not so sensitive to outliers. Conversely, SD will yield unrealistically small estimates of spread when there are a small number of readings that happen to be close to one another. Sn is less sensitive to this situation.

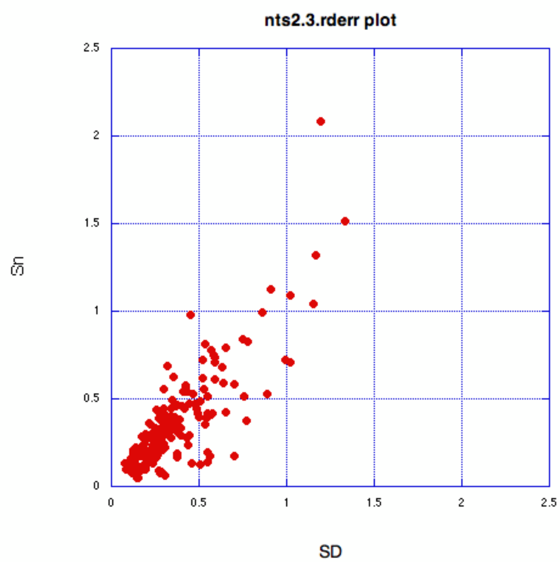

Figure 2. Comparison of empirical reading errors estimated by standard deviation and the robust Sn estimator.

GT Shifting Algorithm

In our ground truth work, we shift the cluster hypocentroid (which is biased by the use of an average global travel time model) to best fit a subset of cluster events with independently known locations. Formerly, we assumed that the ground truth data were all perfectly known. If there were more than one ground truth event, we averaged the shift vectors, ignoring the differing uncertainties of the associated cluster vectors. We now take into account the uncertainty of both the ground truth locations and the cluster vectors in this process.

Weighted Shift Vector

An estimate of the covariance matrix for the GT must be supplied. It can be derived from a single-event location procedure or derived from plausible uncertainties of a geographic estimate such as InSAR analysis of ground deformation. Because the cluster vector and GT are completely independent, their covariance matrices simply add. Each estimate of the shift vector (GT-HDC) is weighted inversely to the area of the 90% confidence ellipse of the combined (HDC+GT) estimation process. Origin time and depth calibration (if relevant) are handled in an equivalent univariate method.

Uncertainty of Ground Truth Calibration

To estimate the uncertainty of the estimated "calibration shift" needed to bring the hypocentroid into optimal alignment with GT data, we add the combined HDC+GT covariance matrices, inversely weighted according to the area of the equivalent 90% confidence ellipses. This yields a covariance matrix that can be added to those of the cluster vectors and scaled to yield final estimates of the uncertainty of the calibrated absolute locations of all cluster events.

Phase Association

The clusters analyzed by HDC for our GT work have normally been preprocessed in a reviewed EHB (Engdahl et al., 1998) single-event algorithm which re-associates (or re-identifies) readings with phases in the ak135 model. Some additional editing of phase associations has been done by hand within the HDC processing. The EHB process utilizes a probabilistic method to associate readings with depth phases. As shown in Figure 3, the probability density functions (PDFs) for the phases of interest are determined from prior studies. For each reading, the relative probability of association with each depth phase is calculated, and a random number is generated to determine which phase is actually selected, with the appropriate probability. The HDC code has now adopted this approach to phase association for all readings. The algorithm operates on reading groups (all readings from a given station for an event), not individual readings. Phase association is done when data are first read and after the first iteration. If phase association is done after each iteration, even a few unpredictable changes in phase association can prevent convergence.

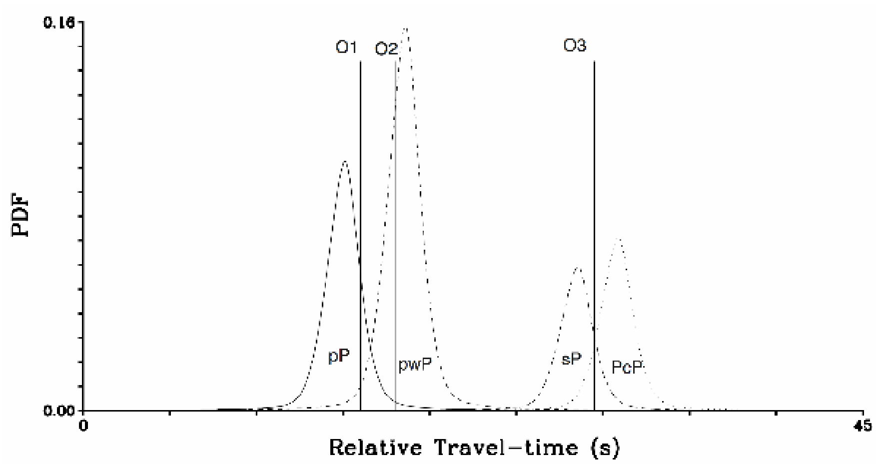

Figure 3. Example of EHB-style probabilistic phase association applied to depth phases.

Rule Set for Phase Association

A set of rules is being developed to incorporate seismological knowledge into this process. The basic idea is that the PDF approach is applied to all theoretical phases (ak135) within a large window (±10 s for P phases, ± 15 s for S phases and unidentified) around the observed arrival, but rules are used to eliminate some phases from consideration. The current rule set is as follows:

1) The initial reading of a station group is normally forced to associate with the first-arriving phase at that distance. This requirement is relaxed if the first reading is an S phase, if the theoretical first arrival would be Pdiff (these are often PKP phases), or if the first reading is very late with respect to the predicted arrival time. Such readings are not allowed to associate as depth phases.

2) A reading can only be associated with secondary arrivals of the same type (P or S, ignoring depth legs), if the phase type can be determined from the original dataset.

3) The first S reading is associated with the first-arriving S phase in ak135 at that distance. If the residual is large, however, an attempt is made to associate it with Lg (not in the ak135 model). If that fails, the reading is turned loose to associate with any eligible secondary phase using the PDF algorithm.

4) Association as a depth phase is only allowed if the corresponding parent phase has been associated.

5) An Lg phase identified by the operator is not automatically re-associated. Other secondary phases (S-type, unidentified, or P-type measured from horizontal components) may be associated with Lg. Lg is only a possible phase for association at distances less than 20° and for sources less than 35 km deep.

Remaining Issues

Estimates of the probability density functions for some phases as a function of source depth and epicentral distance are available, but much work remains to be done. At present, the algorithm operates with identical PDFs for all phases. This cannot be improved incrementally—PDFs for all phases of interest must be updated at the same time to avoid introducing unwanted weighting of different phases. As our suite of GT-calibrated events grows, we will analyze the associated readings to derive more appropriate PDFs for phases of interest.

Ground Truth Data

Critical to our ground truth data discovery and acquisition process are collaborative arrangements that have been made with key organizations in southern Asia. These arrangements are built on exchanges that are mutually beneficial to the parties involved, usually based on our applying advanced techniques to refine locations of the host country's natural seismicity in return for access to in-country ground truth information. These arrangements provide a forum for gathering and assessing potential ground truth data, and collecting waveform and phase reading data for events of interest from local and regional stations. We are also in contact with several groups developing ground truth locations from InSAR-detected ground displacement and other satellite-based location methods that provide important constraints independent of seismic observations. Much new ground truth information is now being obtained from these sources as an ongoing activity. Validation of these data is in progress (Table 1 and Figure 4).

Validation through critical examination of the data and procedures that were used in the local network or InSAR location of a proposed ground truth event is an *internal* process. It is certainly of great value and, in some cases, adequate to guarantee ground-truth levels of accuracy. In many cases, however, an internal validation process of local network locations is highly susceptible to unavoidable uncertainties in the arrival time data and the local velocity structure. For example, there can be undocumented timing errors in the local network, incorrect station locations, incorrectly picked or mis-associated arrivals, and unrealistic estimates of reading error. A very difficult problem in many regions is the specification of a sufficiently accurate velocity structure for the local network location. Investigators rarely have enough information to control all these factors in a validation exercise, and a certain amount of faith is ultimately required in adding such events to a ground truth data set. Therefore, an *external* validation process, one that utilizes other information as a crosscheck on the reported or derived (using HYPOSAT) local network location, is highly desirable. We use HDC, a powerful algorithm for multiple event relocation, as a tool for discovery and validation of ground truth data. HDC is applicable in situations in which several candidate ground truth events and/or InSAR signals are located in a limited region, and in cases where other seismic activity in the area can be localized to known faults and other geologic features. The essence of the validation process is to compare the relative locations in space and time of events based on their ground truth locations, and the relative locations revealed by HDC. An added bonus of the validation process is the generation of additional ground truth events that are of GT5 quality. Although preliminary validation results are already available for nearly all these ground truth data, final results remain pending for most as we await analyst reviewed arrival-time picks and releases of International Seismological Centre (ISC) and United States Geological Survey (USGS) phase data for the more recent events.

Table 1. Ground Truth Data

Name	Date	Origin	Latitude	Long.	Depth	Mag.	Source	Status
Boroujen	78 07 03	01 35 25.9	31.702	51.193	25.1	4.3 (mb)	Local Net	Pending
Tabas	78 10 09 78 10 12	16 04 38.2 15 01 39.4	33.340 33.362	57.283 57.334	8.0 8.0	4.6 (mb) 4.9 (mb)	Local Net Local Net	Validated Validated
Korizon	79 12 02 79 12 11 79 12 16	21 09 34.1 02 16 58.1 22 35 35.1	34.047 34.023 34.094	59.669 59.684 59.491	4.0 4.9 5.1	4.1 (mb) 4.4 (mb) 5.0 (mb)	Local Net Local Net Local Net	Pending Pending Pending
Rudbar	90 07 24	07 27 16.6	36.651	49.769	12.0	4.3 (mb)	Local Net	Pending
Fork	90 11 21	03 42 34.6	28.315	55.498	12.0	4.6 (mb)	Local Net	Pending
Gorgan	91 10 17 91 10 17	15 41 48.1 19 16 25.6	36.156 35.948	53.367 53.358	22.0 22.0	4.3 (ML) 4.0 (ML)	Local Net Local Net	Pending Pending
Sefidabeh	94 02 23 94 02 23 94 02 23 94 02 24 94 02 26 94 02 28 94 03 02	08 02 02.5 11 54 30.6 22 45 15.5 00 11 10.1 02 31 08.8 11 13 51.5 14 57 16.9	30.870 30.837 30.902 30.881 30.817 30.884 30.882	60.520 60.510 60.470 60.473 60.530 60.527 60.470	7.0 9.0 9.0 9.0 9.0 7.0 9.0	6.1 (Mw) 5.5 (Mw) 5.3 (mb) 6.3 (Mw) 6.1 (Mw) 5.6 (Mw) 4.5 (mb)	Fit to InSAR Ground Displacement	Validated
Firuzabad	94 03 07	00 57 22.1	29.028	52.678	9.4	4.1 (mb)	Local Net	Pending
Bastak	97 05 05 97 09 18	15 11 52? 14 52 52?	27.130 27.083	53.881 53.492	5.2 3.5	5.4 (Mw) 5.0 (Mw)	InSAR InSAR	Pending Pending
Buyin-Zahradeg	97 06 07	20 29 47.3	36.170	50.480	8.0	3.9 (mb)	Local Net	Pending
Chakhu Sahlabad	97 06 20 98 04 10	12 57 32? 15 00 50?	32.28? 32.25?	60.00? 60.02?	2.0 6.7	5.6 (Mw) 5.8 (Mw)	InSAR InSAR	Pending Pending
Ghir	97 12 26 97 12 26 99 04 30	01 05 32.9 05 08 00.1 04 19 59.9	28.109 28.088 27.870	53.467 53.468 53.628	13.1 15.0 4.1	4.2 (mb) 4.1 (mb) 5.3 (Mw)	Local Net* Local Net* InSAR	Validated Validated Validated
Kerman	98 03 14 81 08 10 81 08 20	19 40 29? 21 29 19.1 19 02 08.7	30.14? 30.107 30.170	57.59? 57.688 57.473	9? 17.0 12.8	6.6 (Mw) 4.1 (mb) 4.6 (mb)	InSAR Local Net Local Net	Pending Pending Pending
Zagros	98 10 18	10 02 39?	28.6777	54.245	1.4	4.5 (Mw)	InSAR	Pending
Tabriz	99 08 19	04 33 11.3	38.513	46.559	10.1	4.5 (mb)	Local Net	Pending
Avaj	02 07 03	19 24 39.2	35.731	49.026	7.4	4.3 (mb)	Local Net*	Validated
Alborz	03 06 21 03 06 22	15 00 01.3 03 39 09.1	35.498 35.471	52.661 52.730	? ?	4.5 (mb) 4.3 (mb)	Local Net Local Net	Pending Pending
Bam	04 01 11	05 06 02.8	28.908	58.287	10.0	4.1 (mb)	Local Net	Pending
Baladeh	04 05 29 04 05 30 04 06 07 02 04 08	09 23 48.3 19 27 00.7 04 01 20.7 18 30 52.1	36.466 36.390 36.427 36.513	51.365 51.631 51.511 51.937	14.2 7.4 7.3 8.5	4.7 (mb) 4.4 (mb) 4.3 (mb) 4.1 (mb)	Local Net Local Net Local Net Local Net	Pending Pending Pending Pending
Zarand	05 02 22 05 05 14	02 25 22.1 18 04 54.8	30.774 30.806	56.736 56.991	14.1 14.1	5.7 (mb) 5.4 (mb)	Local Net Local Net	Pending Pending

* HYPOSAT Location

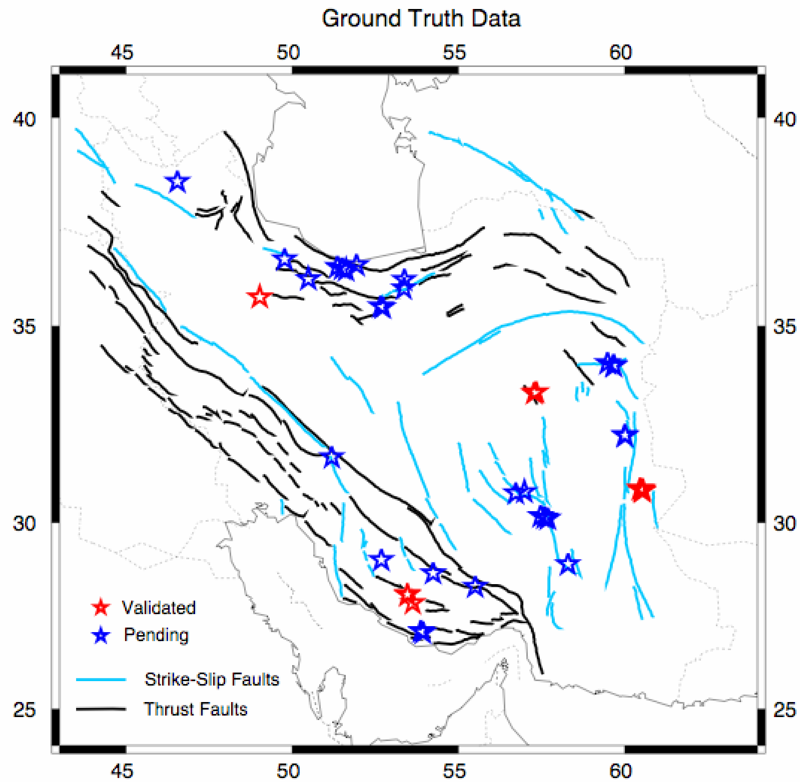

Figure 4. Ground truth data that have been validated or for which validation is pending.

__Validation of Ground Truth Data__

The HDC method for validation yields improved accuracy for both the relative and absolute locations of clustered earthquakes. The gist of the method is to use a multiple-event relocation method with regional and teleseismic phase arrival times to constrain relative locations of clustered earthquakes and then to calibrate the absolute location of the cluster by obtaining independent information on the absolute location of one or more members of the cluster. For each cluster, there is independent information on location that helps to calibrate the absolute locations. The Avaj cluster shown here, extracted from our new catalog of phase data for events in Iran, will illustrate our approach to this important problem. The HDC analysis includes further refinement of the data set by making empirical estimates of readings errors and using these estimates to help identify outliers. These steps yield significant improvements in accuracy and resolution for the relocations. Of course, the main benefit of HDC analysis is to largely remove the biasing effects (path anomalies) of lateral heterogeneity in the Earth, which permits much better resolution of the relative locations of cluster earthquakes.

In developing ground truth databases, we have previously depended almost entirely on arrival time picks reported in the catalogs of international agencies, such as the ISC and the USGS's National Earthquake Information Center (NEIC). While these picks have proven to be quite useful in the analysis of clustered ground truth events, where the statistical properties of source-station path anomalies can be determined, reading errors are often large and, of course, the picks cannot be confirmed. We address that deficiency by expanded analyst review of relevant waveforms that can be acquired for ground truth events in all countries of southern Asia. Our expanded analyst reviews will include the comprehensive repicking of phase arrival times from all available waveforms, with special attention to the regional phases Pg, P*, Pn, Sg, S*, and Sn. The product will be improved databases for all existing and newly discovered ground truth events of magnitude 2.5 and larger in southern Asia, including isolated single events in regions that are difficult or impossible to access. Through analyst review of waveform data, we will be able to update and replace catalog picks with analyst picks not only for new events but also for ground truth events in clusters that we already have studied. The goals will be to expand, further refine, and reduce the uncertainties (better statistics) in station path anomalies to those clusters, to eliminate or minimize reliance on catalog picks, and to restate the scale length of lateral heterogeneity in the region.

Avaj (Changureh): Calibration Using Aftershock Survey Data and Comparison with Geological Mapping and Remote Sensing Data

The cluster consists of the June 22, 2002, Mw 6.4 Changureh mainshock, aftershocks, and several earlier events, 17 events in total. The result of our HDC analysis of the Changureh earthquakes is shown in Figure 5. The relative locations of events are plotted with respect to the hypocentroid or geometrical center of the cluster vectors that describe the relative locations.

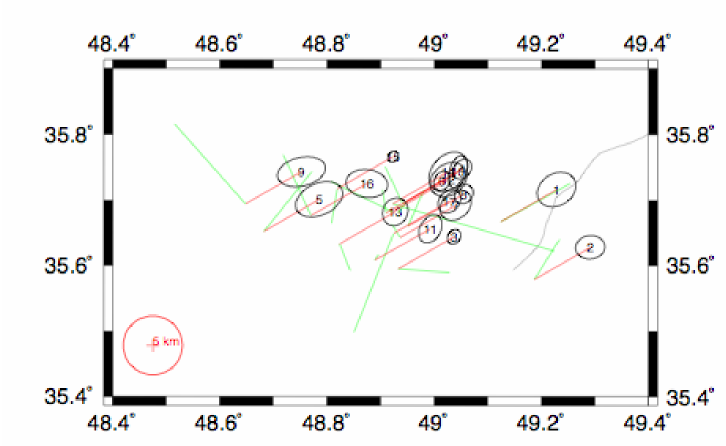

Figure 5. Relocated epicenters of the Avaj cluster. 90% confidence ellipses for relative locations (cluster vectors) are shown. Line to each event shows change in location from starting (EHB single event) location. Green segment is the change due to HDC analysis, red segment is the shift required to bring the cluster into agreement with the ground truth data for this cluster (event 12). Variable gray lines are rivers.

The Changureh mainshock (#3) is at the southeastern edge of the 1992 seismic activity. Events 1 and 2 are from 1967 and 1984. These results support a fault model in which rupture initiated in the southeast end of the rupture zone and propagated unilaterally to the northwest for 20-25 km (Walker et al., 2005). This finding would be consistent with the source time function duration of about 5 seconds, derived from body wave modeling by the Cambridge group.

We obtained arrival time data for one of the cluster events (#12) from an aftershock study by a group of IIEES scientists and also a crustal model derived from the aftershock study of Changureh (Farahbod, personal comm.). We relocated the event using the HYPOSAT code of Schweitzer (2001) and used this location to calibrate the absolute location of the HDC cluster. The mislocation vector is rather typical of those we have found in southern Asia, 10.8 km at an azimuth of 60°, with an origin time shift of -1.87 seconds.

<u>Regional Path Anomalies and In-Country Seismograms</u>

We use the calibrated cluster arrival time data to infer empirical path anomalies (relative to the global model ak135) from the Avaj source region to surrounding seismic stations. Figure 6 shows the results for Pn and P phases at regional distances. There is broad consistency of path anomalies at most azimuths. The early arrivals at stations in Saudi Arabia reflect propagation across the Arabian shield. The path anomalies can be the result both of variations in bulk velocity and differences in ray paths caused by lateral heterogeneity.

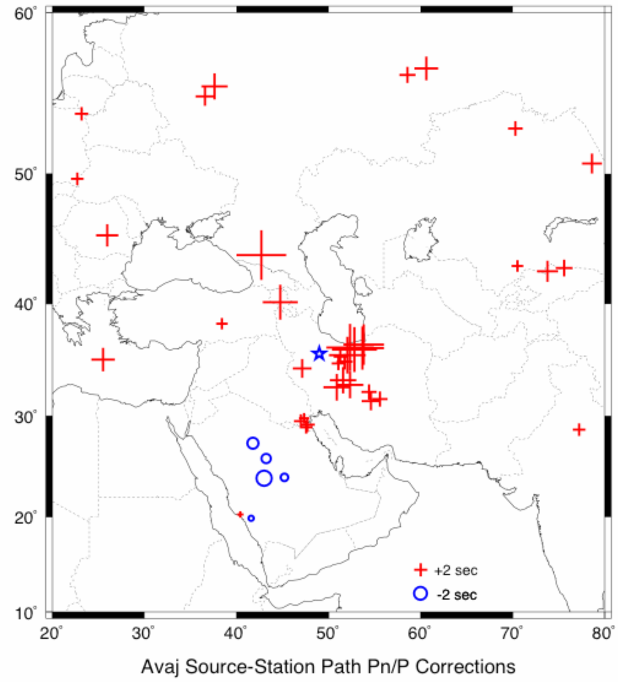

Figure 6. Empirical path anomalies (relative to ak135) for Pn and P phases from Avaj (star).

Figure 7 is a location map of digital stations that have operated or are currently operating in Iran. Stations of the Iranian Seismological Telemtry Network (ISTN) are separated into sub-networks whose data are compiled into a single database of phase data and waveforms. Example waveforms extracted from this database for the Avaj cluster are displayed in Figure 8.

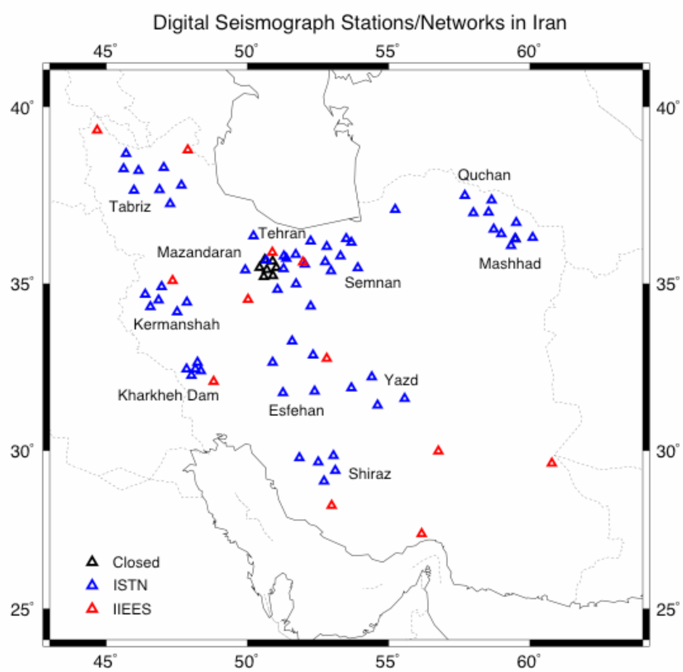

Figure 7. Digital seismograph stations that have operated or are presently operating in Iran.

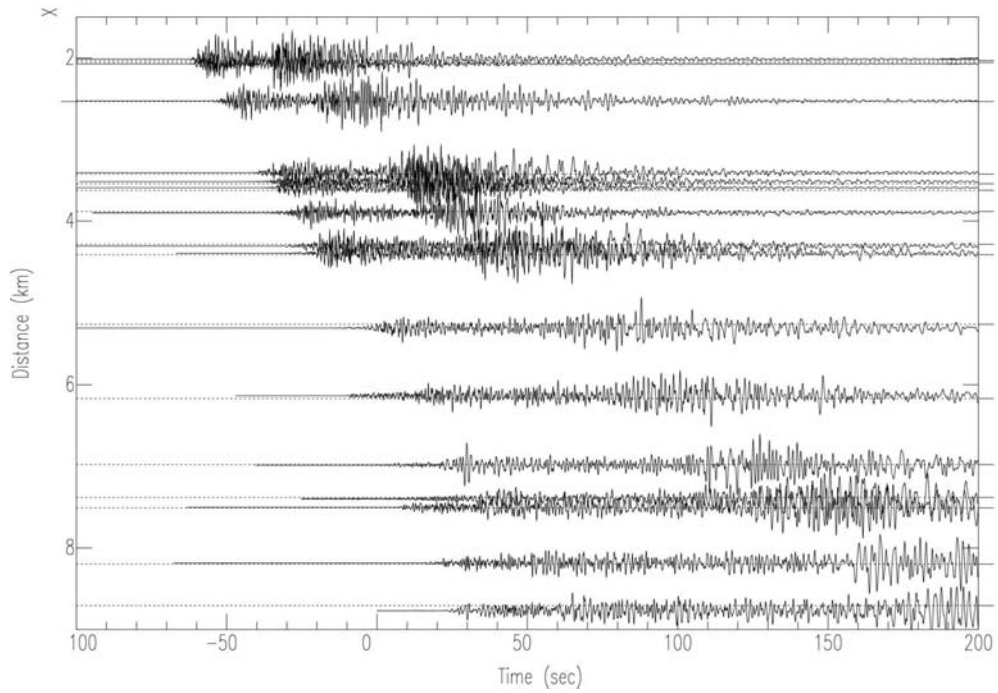

Figure 8. Sample record section of ITSN regional seismic stations, for the Avaj mainshock.

CONCLUSION(S) AND RECOMMENDATION(S)

We have developed several new ground truth events in southern Asia, based on detailed multiple event relocation and use of reference events, both from local seismic network data and from InSAR data. The use of analyst-reviewed picks is extremely helpful in some circumstances, and the practice should be expanded. The greatest value comes from having the analyst read seismograms from stations that were not reported in the standard global catalogs. We are continuing to develop resources for local network data inside Iran and expect these efforts to lead to new ground truth events and resulting data on empirical path anomalies that will substantially improve location capabilities in this region.

ACKNOWLEDGEMENTS(S)

We are grateful to Eric Fielding, Richard Walker and James Jackson at Bullard Laboratories, Cambridge University, for providing data, figures, and advice used in this study.

REFERENCE(S)

Croux C., and P. J. Rousseuw (1992),Time efficient algorithms for two highly robust estimators of scale, *Computational Statistics, V1*, Dodge and Whittaker, ed., *Physica-Verlag, Heidleberg*: 411–428.

Engdahl, E. R., R.D. Van der Hilst and R.P. Buland (1998), Global teleseismic earthquake relocation with improved travel times and procedures for depth determination, *Bull. Seism. Soc. Amer.*, v. 88: 722–743.

Jordan, T.H. and K.A. Sverdrup (1981), Teleseismic location techniques and their application to earthquake clusters in the south-central Pacific, *Bull. Seis. Soc. Amer.*, 71, 1105-1130.

Schweitzer, J. (2001), HYPOSAT - an enhanced routine to locate seismic events, *Pure appl. geophys.* 158: 277–289.

Walker, R. T., E. Bergman, J. Jackson, M. Ghorashi, and M. Talebian (2005), The 2002 June 22 Changureh (Avaj) earthquake in Qazvin province, northwest Iran: epicentral relocation, source parameters, surface deformation and geomorphology, *Geophys. J. Int.*, 160: 707–720.