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

Contract No. DE-FC52-03NA99516[1] and DE-AC52-07NA27344[2]
Proposal No. BAA03-19

ABSTRACT

This research has the goal of developing in-country data sets that can be used to improve ground-based monitoring capabilities in southern Asia, in particular the region bounded by 20–44°N and 41–67°E, by providing information needed to develop and test more accurate travel-time models for seismic phases that propagate in the crust and upper mantle. We have also incorporated phase picks from an experienced analyst who reviewed waveforms of particular interest for specific events. These in-country arrival times and analyst-reviewed picks have been associated with known earthquakes reported by international agencies, combined with existing bulletin readings, and relocated using the Engdahl et al. (1998; EHB) methodology. Using in-country data, we have formed new events, mostly at lower magnitudes that were not previously included in standard global earthquake catalogs. This has resulted in a catalog of earthquakes in the region for the period 1923–2008 for events larger than about magnitude 2.5. Catalog events larger than about magnitude 4.0 have been highly reviewed. Events at lower magnitudes have been relocated with a standard procedure similar to the EHB procedure, but not all have been systematically reviewed.

The new catalog has been used to conduct a detailed analysis of historic and recently occurring event clusters (generally mainshock-aftershock sequences) using a multiple-event relocation technique and data sets of phase arrival times at distances from near-source to teleseismic. Absolute locations of such clusters are constrained using reference event information for one or more of the cluster events provided by local networks, aftershock deployments, or from non-seismic (e.g., InSAR or geological mapping) information. We have also developed a method for direct calibration of a cluster by using arrival time data only from local stations, with an appropriate crustal model, to locate the hypocentroid of the cluster. When both location and origin time can be calibrated for a cluster, we are able to estimate the unbiased 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 more accurate routine earthquake locations using regional seismic data.

We have performed hypocentroidal decomposition (HDC) calibration analyses on 27 earthquake clusters, containing 989 events, in the region. Of these, three clusters could not be calibrated at all. Twenty-two clusters contain at least one event with a calibrated location that meets GT5 criteria, a total of 549 GT5 events. We present a summary of the results of these calibration studies in the form of absolute travel-time information derived from the calibrated clusters, showing distance-dependence of travel times of different phases from different source regions. We also present summaries of empirical reading error determinations and of travel-time variability for different phases. Finally, we present a study of differences between arrival time picks made by an experienced analyst and those obtained from in-country and global bulletins.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **SEP 2008**	2. REPORT TYPE	3. DATES COVERED **00-00-2008 to 00-00-2008**

4. TITLE AND SUBTITLE **Improved Ground Truth in Southern Asia Using In-Country Data, Analyst Waveform Review, and Advanced Algorithms**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Lawrence Livermore National Laboratory,PO Box 808,Livermore,CA,94551-0808**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Proceedings of the 30th Monitoring Research Review: Ground-Based Nuclear Explosion Monitoring Technologies, 23-25 Sep 2008, Portsmouth, VA sponsored by the National Nuclear Security Administration (NNSA) and the Air Force Research Laboratory (AFRL)

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**8**	

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, Pb, Pn, Sg, Sb, and Sn; and (3) Application of advanced algorithms, specifically 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

Compilation of Arrival-Time Data and Relocation of Seismicity

A comprehensive catalog of all instrumentally recorded events, magnitude 2.5 and greater, during the period 1923-2008, for the region bounded by 20–44°N and 41–67°E, has been assembled. Available bulletin arrival-time data from in-country seismic networks in the region, as well as phase picks from an experienced analyst who reviewed waveforms of particular interest for specific events, have been compiled and, where possible, associated with arrival-time data from known earthquakes reported by international agencies. However, with the in-country data we have also formed many new events, mostly at lower magnitudes that were not previously included in standard global earthquake catalogs. This combined catalog of more than 25,000 events has been relocated using the Engdahl et al. (EHB; 1998) methodology. Epicenters from the resulting catalog are plotted in Figure 1A and the stations reporting arrival time data for the period 1923–2008 in Figure 1B. Significant scatter in the distribution of seismicity is evident since many events at lower magnitudes are poorly located and are not azimuthally well constrained, especially in parts of the region where there are few stations.

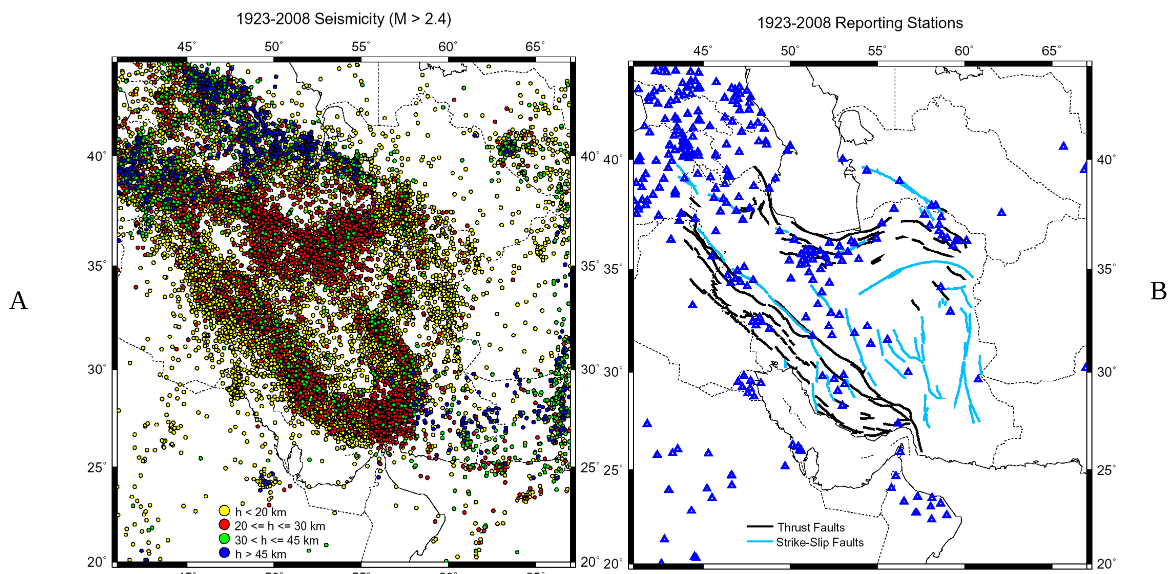


Figure 1. A) Seismicity map of the study region color coded by depth based on the new catalog assembled in this project. B) Stations reporting arrival time picks.

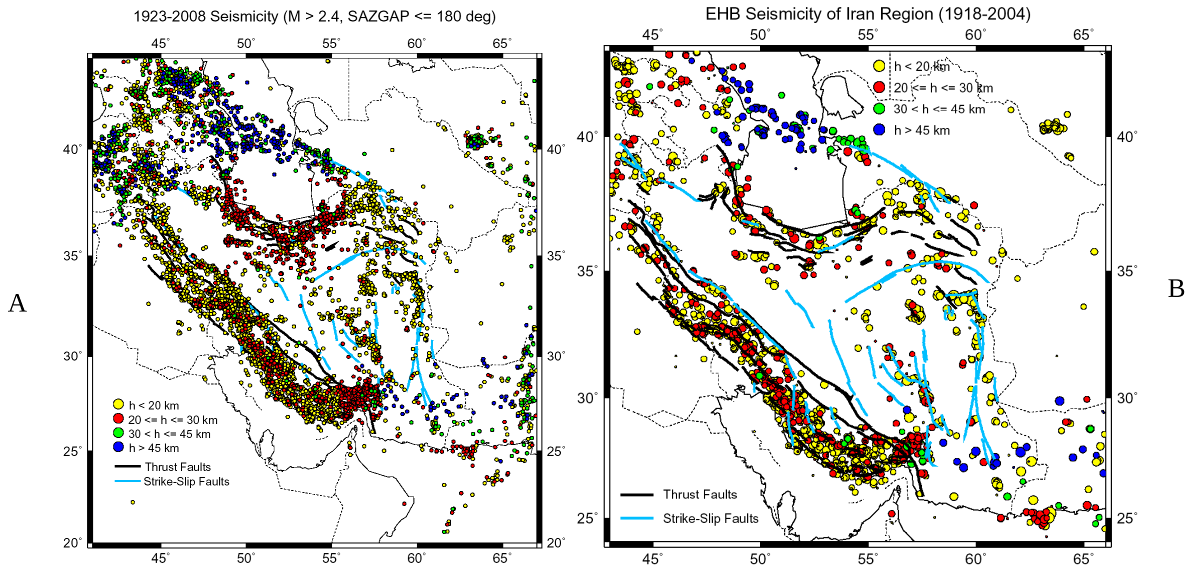


Figure 2. A) Events meeting a secondary azimuth gap criteria of less than 180 degrees when applied to station distributions at all distances. B) For comparison, a figure reproduced from Engdahl et al. (2006) for events where the same secondary azimuth gap criteria is applied only to stations at teleseismic distances (> 28 degrees).

In order to improve the location accuracy of the events shown in Figure 1A, we apply a modified EHB secondary azimuth gap criteria of less than 180 degrees when applied to event-station distributions at all distances. This reduces the database to over 7,000 events and significantly reduces the scatter. The seismicity distribution in the new map now shows a striking similarity to a map reproduced from Engdahl et al. (2006) showing events where the same secondary azimuth gap criteria are applied only to stations at teleseismic distances (> 28 degrees). However, the event depth distribution in Figure 1A requires further review (e.g., the depths in the Alborz region appear to be slightly overestimated).

Analyst Reviewed Events

Waveforms at regional distances in this region are often complex. For many events, this makes arrival time picks are difficult to make, especially for smaller magnitude events, resulting in reported times that often differ from picks made by an experienced analyst.

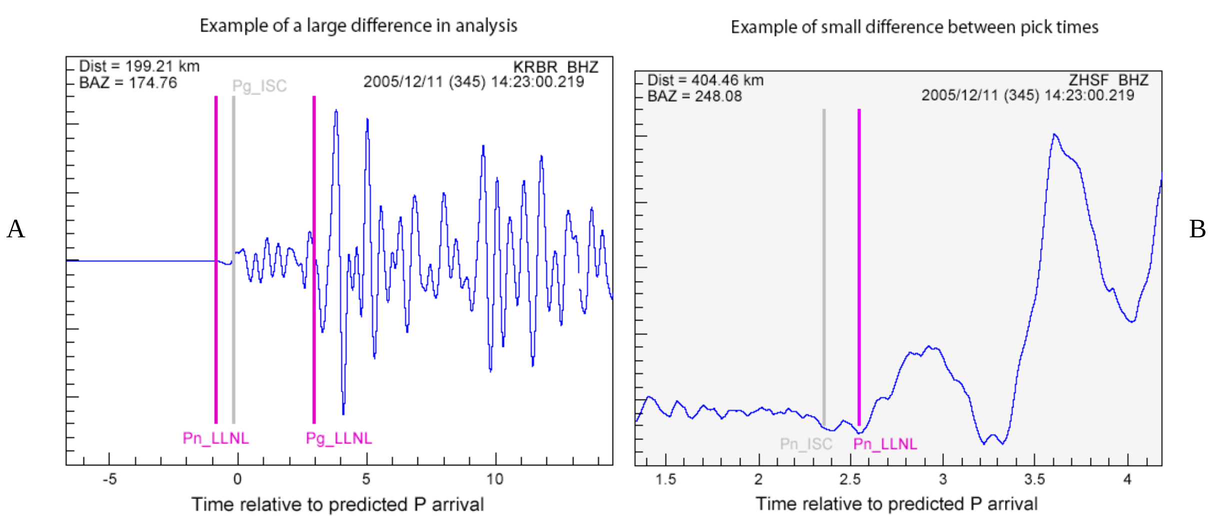


Figure 3. Regional waveforms read by an experienced LLNL analyst. Note the difference in time scales. Examples of A) Large difference in reported pick times, and B) Small difference in reported pick times.

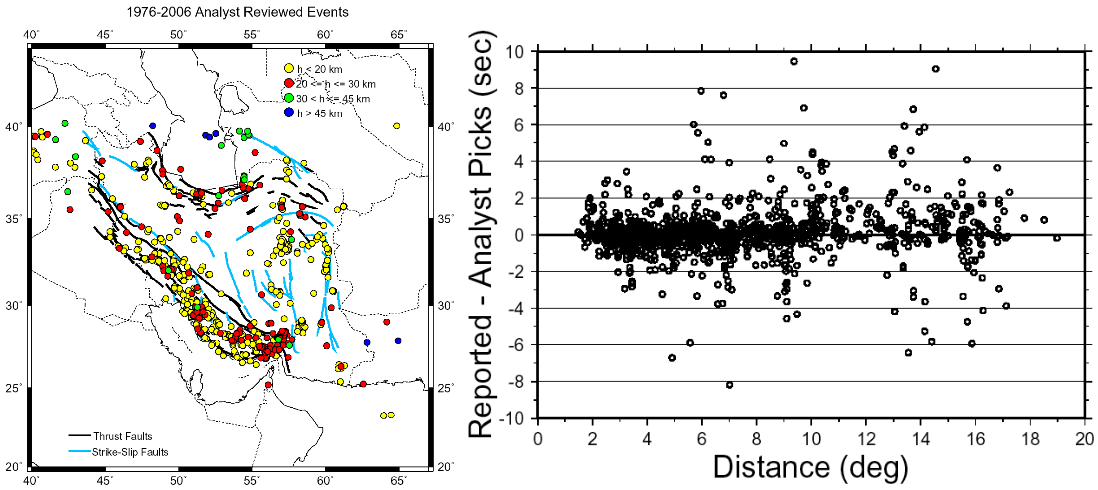


Figure 4. A) Analyst reviewed events, and B) Reported minus analyst picks as a function of distance.

Figure 3A is an example of a rather large late pick of the Pn phase by a regional network as it was reported to the International Seismological Centre (ISC) as compared to the obvious earlier pick made by an analyst. Figure 3B is an example of a Pn pick reported by a regional network that is slightly earlier than the analyst pick.

In Figure 4A are plotted all events for which an experienced analyst read arrival time picks from regional waveforms. The events sample a cross section of different tectonic regimes across the region. Figure 4B compares reported-versus-analyst pick times. In spite of the obvious outliers, the median (-0.06 s) and spread (0.51 s) are small, suggesting that more confidence can be placed in the picks reported by regional networks. However, analyst picks at distances of 12–17 degrees, at distances where there is increased complexity of waveforms due to multiple phase arrivals, appear to be slightly earlier than the reported picks.

Empirical Reading Errors

For each cluster analyzed in this study, we have conducted a careful "cleaning" exercise in which a robust estimate of the spread of residuals for each station-phase pair is made from the set of all observations of the given phase by the given station for the given cluster. We call this the "empirical reading error" for that station phase. The cleaning process consists of repeated estimates of empirical reading error, which is then used to identify outlier readings (typically starting with a 4.5σ cutoff), which are eliminated from the data set. This process is repeated with progressively smaller cutoff until the observed distribution of residuals satisfies a 3σ limit with the current estimate of empirical reading error. The number of samples available for this analysis varies for each station-phase pair, from 2 to the number of events in the cluster. Our estimate of empirical reading error includes traditional "picking error" but also includes all other effects which contribute to scatter of the residuals, including errors in relative location, unmodelled velocity variations in the source region, differences in practice among different analysts or the same analyst over time, changes or errors in timing equipment, other changes of instrumentation, and unreported changes in station location.

The number of station-phase pairs (in all clusters) for which we have determined empirical reading errors is summarized in Table 1.

Table 1. Number of empirical reading errors

	P	**S**
Crustal	195	106
Regional	1831	528
Teleseismic	8324	755

"Crustal" phases are Pg, Pb, Sg, and Sb. "Regional" phases are Pn and Sn. "Teleseismic" phases are first-arriving P and S. Histograms of the distributions of these phases are shown in Figure 5.

The occurrences of improbably small empirical reading errors (<0.15s) are dominated by cases in which there are only two or three samples of a station-phase pair that happen to be very close to each other. These distributions can be helpful in selecting a "default" reading error for cases in which it is not possible to make an estimate of empirical reading error from the distribution of residuals in a multiple-event analysis. However, the large range of empirical reading errors emphasizes the danger of assuming a single default picking error for all samples of the same phase in earthquake-location analysis, because data weighting normally depends on the assigned picking error. The problem is worse for smaller events with fewer observations. For a single-event location, there is no good solution other than using data for which a quantitative measure of picking error has been made at the same time as the pick itself.

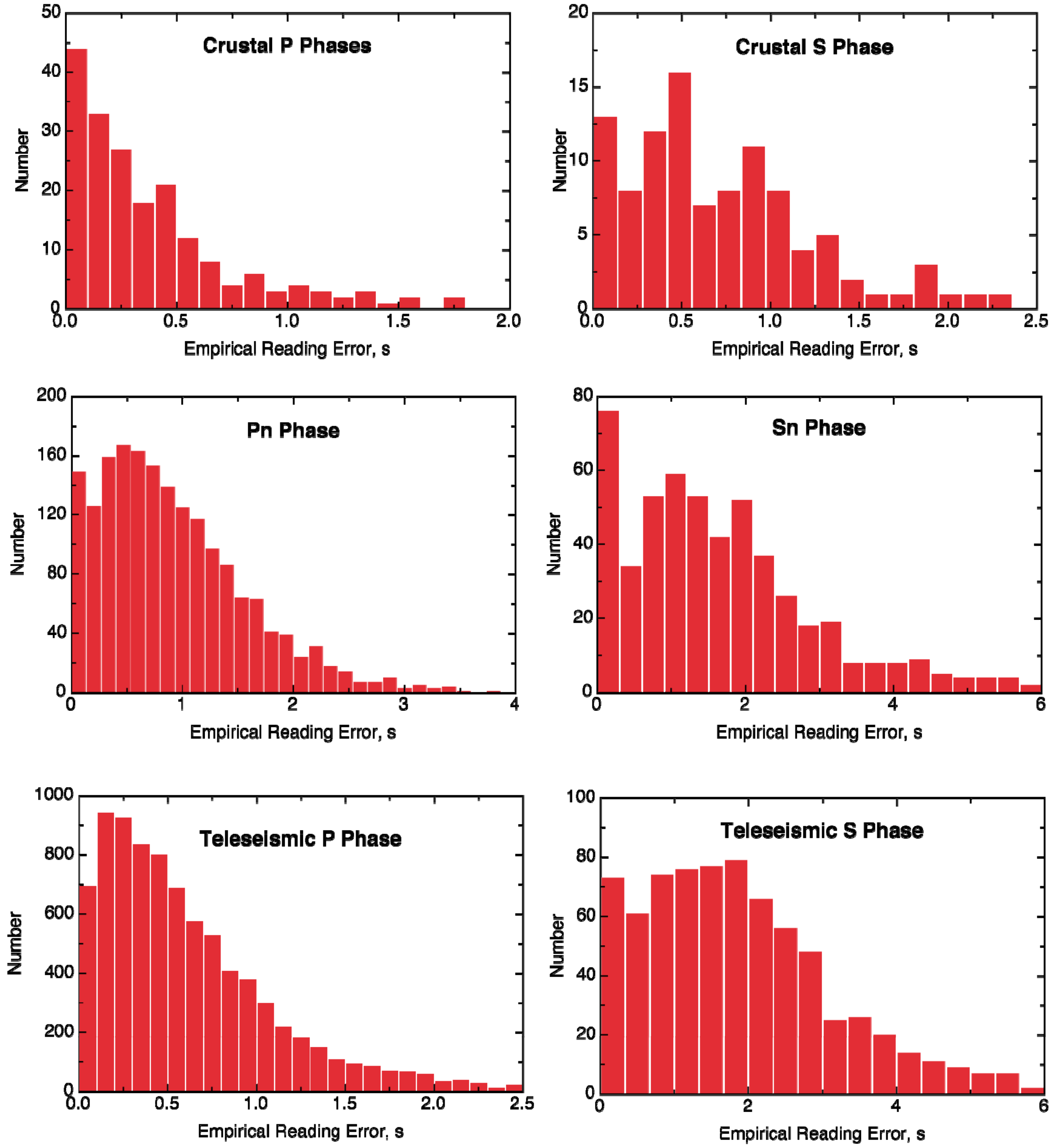


Figure 5. Histograms of empirical reading errors for crustal, regional, and teleseismic P and S phases, from 27 earthquake clusters in the study region. Note that scales are different for each histogram.

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 research groups developing ground truth locations from InSAR-detected ground displacement and other satellite-based location methods, as well as geological field work, that provide important constraints on earthquake location that are independent of seismic observations. Much new ground truth information is now being obtained from these sources as an ongoing activity.

Location Calibration of Earthquake Clusters through Multiple Event Relocation

The HDC (Jordan and Sverdrup, 1981) method for location calibration 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. The HDC analysis includes further refinement of the data set by making empirical estimates of reading 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.

We have recently extended the calibration process to take into account the uncertainties in calibration data in estimating an optimal calibration shift for the cluster. We also estimate a term to account for the inconsistency between multiple calibration events. Our final estimate of local accuracy for events in calibrated clusters includes all these sources of uncertainty, as well as the uncertainty in relative locations derived from the HDC analysis.

We have also developed a method for the direct calibration of a cluster by using arrival time data only from local stations, with an appropriate crustal model, to locate the hypocentroid of the cluster. We have found a number of cases in which no individual earthquake in the cluster is well-recorded enough to be treated as a calibration event, but the cumulative local-distance readings of all events do provide sufficient location accuracy to calibrate the hypocentroid directly at GT5 or better. Many of the clusters presented below have been calibrated in this manner.

Summary of Calibrated Clusters

We have performed HDC calibration analyses on 27 earthquake clusters in the region that are summarized in Table 2. Not all clusters can be calibrated at GT5 or better levels of accuracy. Cluster boundaries and the locations of cluster events are plotted in Figure 6.

Cluster Name	Centroid Latitude	Centroid Longitude	Centroid Depth	No. of Events	GT5 Events	GT3 Events	Direct Calib.	Indirect Calib.	Calib. Level	Seism. Calib.	InSAR Calib.	Location Calib.	OT Calib.
Alborz	35.498	52.848	11.6	11	11	11	Y	N	1.6			Y	Y
Ardebil	38.093	47.891	9.9	16	0	0	Y	N	7.4			Y	Y
Avaj	35.552	49.035	7.9	17	16	1	N	Y	4.3	1		Y	Y
Baladeh	36.345	51.399	8	25	22	20	Y	N	1.9			Y	Y
Bam	28.942	58.346	9.7	22	12	0	N	Y	3.4		1	Y	N
Bastak	27.189	53.88	12.3	25	17	0	N	Y	3.5		1	Y	N
Darband	37.731	57.299	11.5	44	41	17	Y	N	2.9			Y	Y
Dasht-e-Bayaz	33.824	59.455	11.1	101	44	0	N	Y	4.7	3	3	Y	Y
Dinavar	34.571	47.303	20	17	0	0	N	N				N	N
Dorud	33.813	48.82	7	80	77	63	Y	N	1.8			Y	Y
Firuzabad	29.12	52.642	8.3	36	1	0	N	Y	13.3	1		Y	Y
Garmsar	35.293	52.209	10	16	16	15	Y	N	1.3			Y	Y
Ghir	27.982	53.638	15.1	67	51	23	N	Y	2.8	2	1	Y	Y
Igdir	39.162	44.157	14.7	71	47	1	Y	N	3.1			Y	Y
Jiroft	28.336	57.141	23.7	60	0	0	N	N				N	N
Kahak	34.47	50.89	15	15	15	6	Y	N	2.7			Y	Y
Kerman	30.065	57.611	12.8	55	0	0	N	N				N	N
Lagodekhi	41.847	46.294	14.9	44	36	19	Y	N	2.4			Y	Y
Nahavand	34.477	47.854	17.7	25	0	0	Y	N	8.4			Y	Y
Qeshm	26.875	55.925	14.3	43	38	18	N	Y	2.6	2	1	Y	Y
Rudbar	36.773	49.377	10.7	37	1	0	N	Y	8.0	1		Y	Y
Sefidabeh	30.681	60.5	8.4	7	7	6	N	Y	3.1	1	1	Y	Y
Shusf	32.10	60.006	6.7	7	4	1	N	Y	4.1		1	Y	N
Tabas	33.315	57.147	10.1	30	14	2	N	Y	4.4	2		Y	Y
Urmieh	38.445	45.23	13.3	26	25	0	Y	N	3.5			Y	Y
Zarand	30.691	56.828	11.5	49	44	7	N	Y	3.1	1	1	Y	Y
Zirkuh	32.991	60.123	11	43	10	0	N	Y	5.1	2		Y	Y
TOTALS				989	549	210	11Y	13Y		10	8	24Y	21Y

Table 2. Summary of ground-truth cluster data. The location of the hypocentroid in geographic coordinates and depth fixes the absolute hypocenters of cluster events. "Calibration level" is an estimate of the uncertainty of the calibration of the cluster hypocentroid, to which the uncertainty of the relative location (cluster vector) of individual events must be added in order to obtain the absolute location uncertainty for each event.

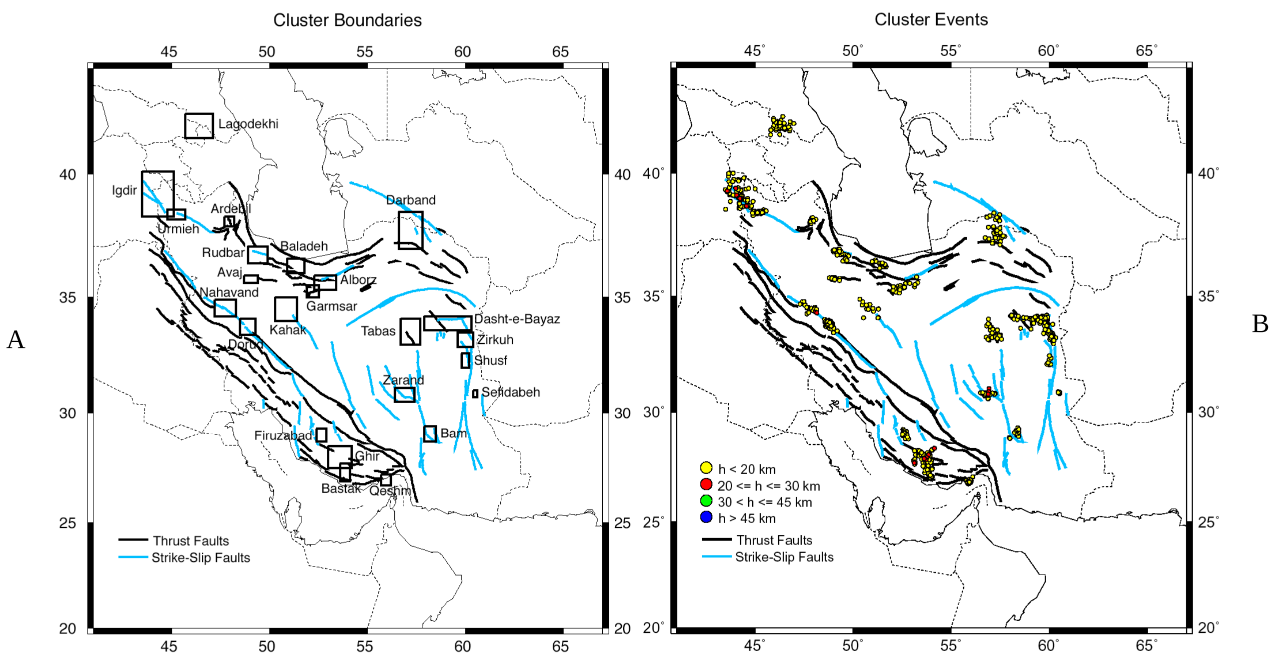


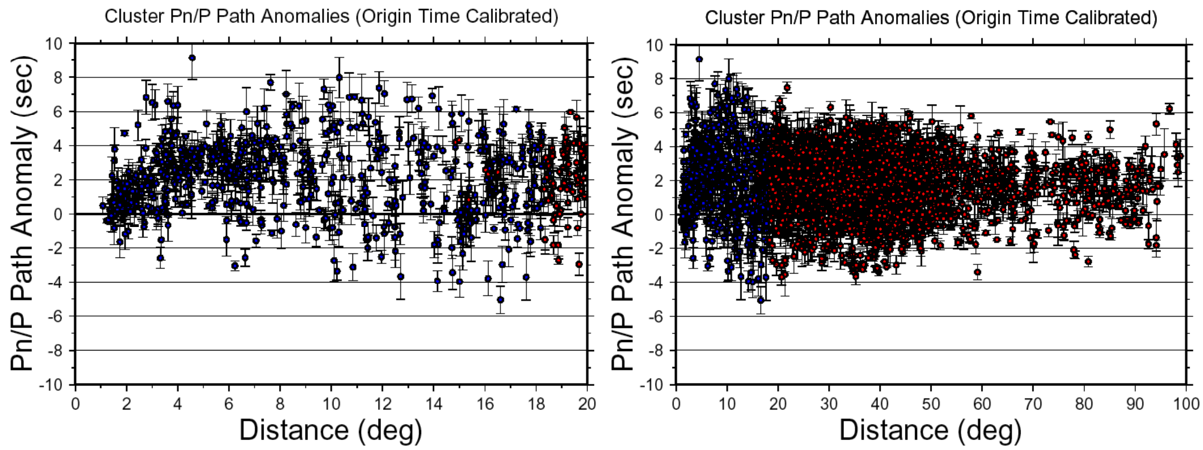
Figure 6. Earthquake clusters studied for ground truth calibration. A) Boxes show the region covered by each cluster, with the name we have assigned. B) Distribution of cluster events color-coded by depth.

Summary of Travel-Time Anomalies

When both location and origin time can be calibrated for a cluster, we are able to estimate the unbiased 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 more accurate routine earthquake locations using regional seismic data. We use the calibrated cluster arrival time data to infer empirical path anomalies (relative to the global model ak135) from each cluster source region to surrounding seismic stations. The path anomalies can be the result both of variations in bulk velocity and differences in ray-path geometry caused by lateral heterogeneity.

We have combined the empirical path anomalies of all the clusters that are calibrated in both location and origin time to produce a summary plot of Pn/P arrivals as a function of epicentral distance (Figure 7).

A



B

Figure 7. Empirical path anomalies, relative to ak135, and spread for Pn (blue) and P (red) phases from 21 earthquake clusters that are calibrated both in location and origin time. A) Regional distances. B) Regional and teleseismic distances.

The empirical path anomalies for individual clusters show evidence both for departures from the average earth model used for reference (ak135), and for lateral heterogeneity. When the path anomalies are combined however, most structure is lost in a cloud of impressive width, a range of 10–12 seconds for P phases over the regional and teleseismic distance range. Even here there is a suggestion of about 2-3 seconds baseline offset from ak135 in the study region. This can be accounted for with a crustal structure that is both thicker (40–45 Km Moho depth, vs. 35 km for ak135) and slower in bulk velocities. It is clear, however, that accurate earthquake location in this region will require the use of crustal models that are more specific to the source regions.

CONCLUSIONS AND RECOMMENDATIONS

We have compiled a comprehensive catalog of all instrumentally recorded events, magnitude 2.5 and greater, that have occurred during the period 1923-2008 for the region bounded by 20–44°N and 41–67°E. The catalog includes substantial numbers of phase readings at in-country seismograph stations, as well as phase picks made by an experienced analyst who reviewed waveforms of particular interest for specific events. From this catalog we have acquired new or improved ground truth events and calibrated earthquake locations in the region, based on detailed multiple event relocation and the use of calibration data, both from local seismic network data and from InSAR and geological data. As new data have been acquired the quality and quantity of calibrated earthquake locations in this region has improved significantly.

REFERENCES

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. Am*. 88: 722–743.

Engdahl, E.R., J. A. Jackson, S. C. Myers, E. A. Bergman and K. Priestley (2006). Relocation and assessment of seismicity in the Iran region, *Geophys. J. Int*. 167: 761–778.

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. Am*. 71: 1105–1130.