VALIDATION AND GENERATION OF REFERENCE EVENTS BY CLUSTER ANALYSIS

E. R. Engdahl and Eric A. Bergman

University of Colorado, Boulder

Sponsored by Defense Threat Reduction Agency

Contract No. DTRA01-00-C-0032

ABSTRACT

High-resolution cluster analysis (multiple-event relocation) of earthquakes and other seismic sources is developed as a tool for assembling catalogs of reference events, especially those whose locations can be determined with an accuracy of 5 km or better [Ground Truth (GT) 5]. We use the Hypocentroidal Decomposition (HDC) method of Jordan and Sverdrup (1981), which is well suited to the rigorous statistical analysis required for this task. Candidate reference events typically arise from local seismic networks and from temporary deployments for aftershock studies that can yield very high-resolution hypocenters that, nevertheless, must be validated. We utilize arrival time data (as reported to the International Seismological Centre and to the U.S. Geological Survey's National Earthquake Information Center) at regional and teleseismic distances in the cluster analysis to validate candidate reference events, and in some cases, to generate new reference events.

HDC analyses have now been performed on a number of earthquake and explosion sequences in Eurasia and Africa, resulting in reference events with locations known to GT5 accuracy. In this paper we review and evaluate our analyses of these clusters to date, and address problem areas. In particular, we find that some candidate reference events cannot be validated because either the reported local network solutions are in error, or the coverage of reported arrival times used in the HDC analysis is not sufficient to constrain the locations. Some discrepancies may arise when local networks locate small precursors or low-energy early stages of rupture in larger earthquakes, while teleseismic stations record only the main pulse of energy release. We have found several cases in which there appear to be systematic biases in the time base used for local network solutions. In another case, we obtained "reference event" locations from two different sources for the same cluster. The two sets are similar enough that HDC cannot be used to discriminate between them, yet different enough to prevent either set from being accepted at GT5 accuracy. Our experiences highlight the importance of a thorough and many-faceted validation program for candidate reference events.

KEY WORDS: reference events, cluster analysis

OBJECTIVE

The primary objective of this research effort is to develop a comprehensive reference event database with validated travel-time information for regional seismic phases recorded by International Monitoring System (IMS) and surrogate stations in Asia and north Africa. This database can be used to support the calculation of regional travel-time curves and source-specific station corrections.

RESEARCH ACCOMPLISHED

Introduction

In this report, we discuss the results of cluster analyses on earthquake and explosion sequences in Eurasia and northern Africa using phase data reported to the ISC and NEIC. All of the events studied have magnitudes of 3.5 or greater. In most cases, reference event data are available from short-term portable seismograph deployments following the initiation of seismic activity. The HDC analyses produce new locations that are defined by "cluster vectors" in space and origin time, relative to the centroid, which is then located in the traditional manner to yield absolute locations and origin times. If one or more reference events are included in

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **OCT 2001**	2. REPORT TYPE	3. DATES COVERED **00-00-2001 to 00-00-2001**

4. TITLE AND SUBTITLE **Validation And Generation Of Reference Events By Cluster Analysis**	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) **University of Colorado, Boulder,Boulder,CO**	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 23rd Seismic Research Review: Worldwide Monitoring of Nuclear Explosions held in Jackson Hole, WY on 2-5 of October, 2001. U.S. Government or Federal Rights.

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)**	**12**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

the cluster, the centroid can be shifted to provide the optimal match to the reference locations, which brings all events in the cluster into close alignment with "ground truth". The expanded set of corrected hypocenters is then available to calculate source-station path corrections relative to the Earth model ak135 (Kennett *et al.*, 1995) for regional seismic phases recorded by IMS and surrogate stations in the region of interest. We follow with descriptions of individual sequences (see Figure 1) that highlight our experience in the analyses to date.

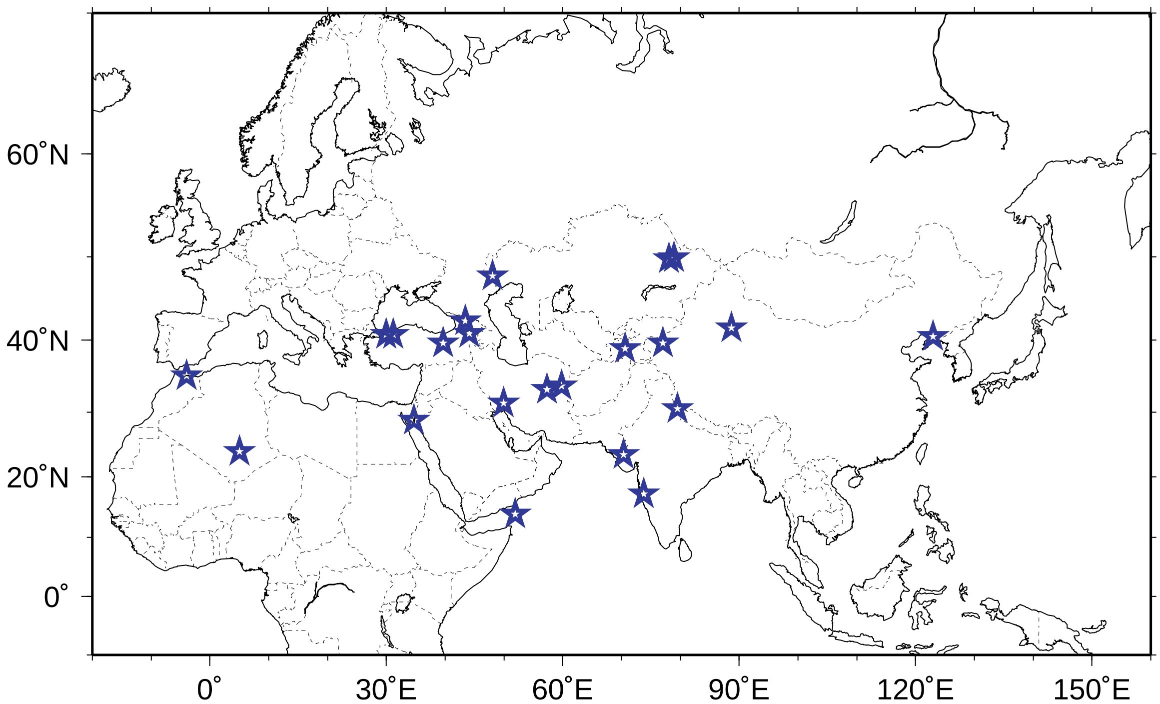

Figure 1. Locations of reference event clusters studied.

Validation Method

The approach being used is based on a high-resolution cluster analysis for relative event location, forming event clusters that contain one or more candidate reference events. The clusters are typically 50-100 km across and contain up to 100 or so events that have occurred since 1964 and that are well recorded at regional and teleseismic distances. The cluster is located in an absolute sense as if all the data were from a single event, using ak135. Obviously, this is subject to bias. To remove the bias, we shift the cluster in space and time to best match the reference events. The degree of consistency between the relative locations as determined by global arrival time data and the relative locations specified by the reference event data is one of the tests we use to validate candidate reference events.

Duezce, Turkey: An example of generation of ground truth events

This is a well-studied event on November 12, 1999, for which we were able to obtain 8 reference events, well recorded at regional and teleseismic distance, from studies made by Milkereit (2000) and Tibi (2001). A cluster of 32 events was formed from the mainshock and aftershocks through December 1999. We found very consistent shift vectors to the 8 reference events, resulting in 16 reference events of GT5 accuracy.

Figure 2 shows the progress in HDC analysis that leads to generation of new reference events.

Figure 2a shows an early relocation of 52 events, before poorly located events are removed and outlier residuals flagged. Green vectors show the change in relative location from the EHB starting locations. 90% confidence ellipses for relative location are shown in blue. Tic marks are at 10 km intervals. Figure 2b shows the final relocation of 32 well-constrained events. Figure 2c shows only the 8 original reference events (black number), and 8 "promoted" reference events (red number) that meet GT5 criteria.

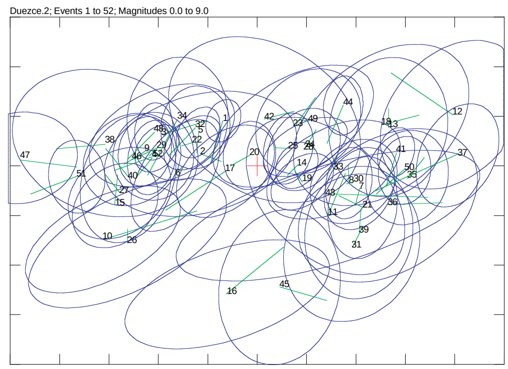

Figure 2a

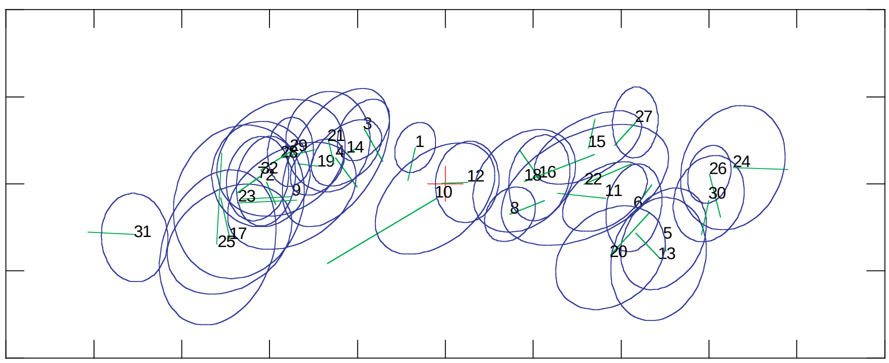

Figure 2b

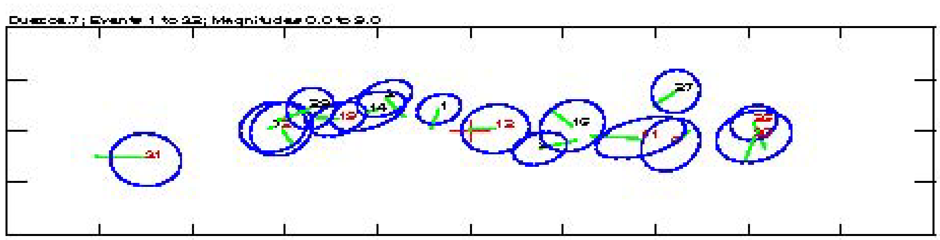

Figure 2c

Zirkuh & Dasht-e-Bayaz

This grouping arose from our study of the Zirkuh (Berberian *et al.*, 1999) earthquake and aftershocks of May-November 1997. The main shock is near the north end of a long fault, and most of the aftershocks, including the ones reported by an aftershock study as "reference events," are in the southern part of the fault zone, as much as 100 km from the main shock. However, several other large earthquakes have occurred near the northern end of the Zirkuh fault, where it intersects with the E-W trending Dasht-e-Bayaz fault (Berberian and Yeats, 1998). There were three major earthquake sequences in a three-week period from mid-November to early December

1979. We formed a cluster of larger events from these three sequences, the 1997 Zirkuh mainshock and a few of its aftershocks, and some individual events from other times, 36 events in total.

Although there are no reference locations available for any of these events, we have placed constraints on the absolute location of the cluster by utilizing the geometric constraints of the Zirkuh and Dasht-e-Bayaz faults, which are orthogonal to each other. The main events have well-constrained strike-slip mechanisms so the requirement that the main shock epicenters locate on the main faults (whose surface ruptures were well mapped right after the quakes) provides a strong constraint on the location of the cluster. Sixteen of the events in the cluster satisfy GT5 criteria on location, although origin times are still unconstrained.

Zirkuh South

The bulk of the aftershocks for the 10 May 1997 Zirkuh earthquake in eastern Iran are in the southern end of the fault zone, and reference locations from an aftershock survey (M. Raeesi, MS Thesis, University of Tehran, 2000) are available for many of them. We analyzed a cluster of 42 events from the sequence and found many inconsistencies between the HDC results and the "reference" locations. Upon a closer reading of the thesis and discussions with the author and his advisor, we believe that further analysis of the aftershock survey data will be required in order to generate true reference event locations. There were numerous logistical problems with the survey (which was in a remote and dangerous part of the country) that may have compromised the results, and the fact that much of the aftershock activity occurred as much as 100 km away from the main shock epicenter led to unfortunate deployment decisions. The available reference material did not adequately document these problems. This is a good example of the need for a validation process for candidate ground truth events. We are pursuing additional data that may yet make it possible to use this sequence for ground truth analysis. There is a possibility to gain additional seismic readings through other agencies, and to use high-resolution maps of surface breakage to help constrain the location of the cluster.

Eilat, Gulf of Aqaba

A list of candidate reference events came from the JSOP project, combining data from several networks in the area. The main events of interest were a large earthquake on November 22, 1995, and its aftershocks. We formed a cluster of 37 events from this sequence and earlier events in the immediate area. As many as 13 events have been presented as reference events (Sweeney, 1998), but our validation exercise indicated numerous inconsistencies, probably because of the different organizations involved. Finally we kept only the main shock on November 22, 1995, as a reference event (using a location provided by A. Hofstetter, Geophysical Institute of Israel). It was the only event whose location from local and regional data met the GT5 criteria and for which HDC analysis provides sufficiently strong location accuracy. The cluster was shifted to best match this location and a total of 12 events in the cluster were promoted to reference event status at the GT5 level. Efforts are ongoing to obtain additional data (especially from Egyptian stations) that might bring some of the other events in the area up to reference event status. This is one of the larger clusters we have used, extending about 100 km in the north-south direction, so it would be especially desirable to obtain additional reference events to ensure that the "common path" assumption for cluster analysis is not being violated.

Garm, Tajikistan

Reference events for this region are obtained from a dense local seismic network (G. Pavlis, Indiana Univ.). We formed a cluster of 23 events between 1975 and 1984. They are all well recorded at regional and teleseismic ranges, and the HDC locations for most of them are close to or better than GT5. There are three good reference events, which give a consistent offset of the cluster. A fourth reference event was rejected because it was very inconsistent with the three others; either the HDC relocation or the reference location is bad. We will include additional events since 1984 to improve the station coverage and statistics.

Hoceima, Morocco

The reference events are part of a major swarm of earthquakes between 1994 and 1996. Most of the arrival time data for these events is at regional distances, however. Cluster analysis cannot be performed with such data, because it violates the assumption of common path errors. Our normal procedure is to use data only beyond 3°.

In this case, the cluster includes only 15 events (from May 1994 to June 1996) that can be located at GT10 or better by HDC, but none of them qualify for GT5. The cluster is very tight, about 20 km by 20 km; therefore, we could re-examine this cluster using data from shorter distances and include more events. Including more events would also improve the cluster vector statistics and reduce the size of the HDC confidence ellipses in some cases.

We used two reference events for this cluster, which were a revision of the locations reported by Sweeney (1998) and which gave very consistent epicenter shifts for the cluster. The difference in OT shifts between the two reference events was rather large, however, over a second. We learned subsequently that—because of timing problems in the network—the reported origin times for these reference events have been normalized (separately) by Department of Energy researchers to the ak135 reference model using reported regional and teleseismic residuals. Using this approach, the normalized origin time for each event becomes highly dependent on the stations reporting data for each event (576 & 200 stations). Therefore, although we can validate the two reference event locations, there is no independent information from this cluster on absolute travel times.

Jiashi, China

This is a massive swarm in western China, beginning in January 1997, that was studied by a temporary deployment of seismometers. We formed a cluster of 83 events that are well located with HDC. Although we were provided with a list of reference locations from the local network that included most of the cluster events, we discovered a serious problem. Only one event had been well located by the network, and the remainder had been located with a master event method that was apparently misapplied, as the pattern of epicenters was in substantial disagreement with the HDC results. Moreover we found a very large difference (nearly 3 seconds) between the reference solution of the master event and the HDC results. This is too large to be plausible as a manifestation of the Earth's heterogeneity. We suspect that there is a problem with the time base used for these Chinese temporary deployments, as we saw a similar phenomenon with the Xiyuan cluster (below). Time-keeping problems might also explain why the Chinese researchers used the master event method. We have recently learned (Wu Zhongliang, personal communication) that the China Seismological Bureau is installing a temporary BB network around Jiashi to study this swarm, and we hope to obtain additional reference event data here. If so, we will be able to promote a large number of events to GT5 status as a result of the HDC analysis.

Koyna Dam, India

Reference locations have been obtained from a local seismic network installed to monitor induced seismicity at the Koyna reservoir in 1967. Many events greater than magnitude 5 have been recorded over the past few decades. We initially worked with a list of 17 reference locations from the National Geophysical Research Institute (NGRI). The first one is the original 1967 event (and of questionable reliability), but the rest are all between 1993 and 2000. We formed a cluster of 31 well-recorded events, and found very consistent estimates of cluster offset from nine reference events.

This should have been one of our better ground truth examples, but soon after finishing the analysis, we learned that there is an alternative set of reference locations for these events, produced by a different group of seismologists in India. We performed the analysis again with the new reference locations (the two sets of reference events have some common events but are not identical) and found equally consistent estimates of cluster offset. The differences are sufficient to deny the possibility of declaring any of these events as GT5 events. We hope that it will be possible to reconcile the differences, but in any case this should be remembered as a cautionary tale about the reliability of so-called "reference events".

Racha, Georgia

The reference events come from a temporary seismic network (Fuenzalida et al, 1977b) that captured aftershocks of a large event in April 1991. The cluster events are from April-October 1991. The cluster included 47 events that could be located to GT10 or better. The cluster is elongated, about 70 km in an E-W direction, and composed of two sub-clusters. Some tests of relocating the sub-clusters separately were done, but we saw no evidence that smaller clusters improved the outcome. We used six reference events that gave a very consistent estimate of the cluster offset and 19 events of GT5 accuracy. We should be able to improve station coverage and statistics (and provide better coupling to IMS stations and surrogate stations) by adding more recent events to the cluster.

Sahara (Ahaggar Mountains), Algeria, nuclear explosions

This is a small cluster of 5 French nuclear tests in Algeria between 1962 and 1966 with reference event information published by Bolt (1976). As a validation experiment, we perform an HDC cluster analysis of the 5 events. The shifted HDC relocation is very good; the cluster is extremely tight, less than 10 km across; and we obtain a very consistent estimate of the shift vector, 5.6 km at 186°, -0.08 s origin time. The origin times are very close because these data were used with events elsewhere to set the baseline for the IASPI91 travel-time model. It is worth noting the arrival time data sources for these explosions: 620501 - ISS Bulletin: 631020 and 650227 - USC&GS Shot Report (most data read from original seismograms); and 651201 and 660216 - ISC Bulletin.

Spitak, Armenia

This is a small cluster of 11 earthquakes that are very well recorded at regional and teleseismic distances; the HDC locations are almost all better than GT5 quality in relative terms. The three reference locations are derived from a permanent regional network. The cluster consists mainly of an earthquake sequence in December 1988, with aftershocks continuing to May 1990, but there is also one event from January 1967 in the same area. This should have been an excellent ground truth resource, but the relative locations from HDC analysis are inconsistent with the reference locations. No two of the reference events are in agreement. Because of the very strong constraints on the HDC locations, we suspect a problem with the reference locations. As this cluster was formed from an active aftershock sequence, it may be that the local network and the global network (for HDC) are looking at different events which are nearly coincident in time, or the local network is picking up early low-energy precursors to larger events. Further investigation is required, possibly a review of waveform data.

Zagros, Iran

This is a cluster formed of seismicity over a fairly large area of the Zagros Mountains, motivated by a report of three reference events in the Sweeney Report, based on Asudeh (1983). The paper reported locations for three events in September 1976, based on a temporary deployment of seismic stations, but a careful reading indicates that the closest station to any of these events is at a distance greater than 150 km. Hence, the "reference" locations are probably no better than GT15. The HDC analysis for the cluster was very successful, yielding many events with relative locations at the GT5 level, so an effort to obtain true reference locations for a few events in the Zagros would be quite profitable. We are discussing this possibility with researchers of the IISEE in Iran, who operate seismic stations in the area.

Tabas, Iran

The cluster of 35 events is based on the aftershock sequence of the Tabas-e-Golshan earthquake of September 16, 1978, in eastern Iran, plus additional events in the area through 1992. Berberian (1982) discussed the tectonic significance of the locations for hundreds of aftershocks, but the actual hypocentral data are contained only in his thesis. Referring to the thesis, only two events are part of the cluster of larger earthquakes that can be well located with the HDC method. Unfortunately, we discovered inconsistencies in these reference data. The thesis contains a master table of all events and a table of the "best-located" events, but the hypocenters of the two "reference" events are different in the two tables. Moreover, the locations from the master table are in significantly better agreement with the HDC results than the locations from the "best-located" table. Hence, we have used the latter events to shift the cluster, resulting in 12 reference events of (tentatively) GT5 quality. But the origin times also seem to be biased relative to the origin times resulting from the HDC analysis. We are investigating these issues with the help of Berberian.

Erzincan, Turkey

Three reference events were identified in the Sweeney report from a study by Fuenzalida *et al.* (1977a). The cluster events are from October 1976-April 1992 and included only 9 events. However, the reference events gave a consistent estimate of the shift vector as 9.0 km at 240°, -2.05 s in origin time which yielded 6 events with an accuracy of GT5 or better.

Lop Nor, China, nuclear explosions

Source parameters for Lop Nor explosions that are available from various sources appear to be conflicting. Hence, we decided to perform a careful cluster analysis of 20 events in three closely spaced source regions at the Lop Nor test site. All could be located to GT5 or better and all were tied to the origin times and locations (based on satellite imagery) of four events provided to us by Terry Wallace of the University of Arizona. These events gave us a very consistent estimate of the shift vector as 2.0 km at 289°, +0.82 s in origin time. Of the 11 events in common with Wallace's original JED analysis only 2 events differed by more than 2 km from his locations (94/10/07 at 2.5 km and 96/06/08 at 3.8 km). Of course, all of the origin times agreed to within 0.2 sec. We also compared our results to the JED locations published by Gupta (1995). Of the 13 events in common with his analysis 7 events differed by more than 2 km from our locations, with one differing by 7.7 km. Also, since he used the JB model for his analysis, his origin times were all about 2.5 sec earlier than ours.

Gulf of Aden

These are large events with CMT solutions, whose locations were obtained by a normal non-linear waveform inversion, but with the location constrained to a chosen bathymetric feature (Pan *et al.*, 2000). Cluster analysis provides a means of testing the accuracy of the constrained locations that have the potential for improving our understanding of regional variations in travel times. Probably the largest source of uncertainty comes from the choice of which feature to constrain the location to. This is not easy when there are many like features (ridges and transforms) close together. In addition, there is the problem of intra-plate events that occur close to the plate boundary.

Cluster analysis of 55 events in this region that included six reference events from the Harvard compilation resulted in shifted HDC locations for 18 events with an accuracy of GT5 based on the size of the HDC confidence ellipse. However, since the location of the Harvard reference events is based on sea floor topography with resolution not better than 5 km, the GT5 shifted HDC locations can't really be much better than GT10 in the absolute sense. Most of these 18 events are on the transform segment of the cluster. However, one of the Harvard reference events moves away from the pattern of other nearby ridge-axis events and we suspect that this is actually an intra-plate event close to the ridge.

Azgir, Kazakhstan, nuclear explosions

The HDC procedure is compared to the procedure of Joint Hypocentral Determination (JHD; Dewey, 1972, 1989) using seven closely spaced underground nuclear explosions near Azgir with ground truth information reported by Sultanov *et al.* (1999). In this comparison, as reported by Israelsson *et al.* (2001), HDC and JHD are applied to the same arrival time data to validate computational consistency of event locations and source-station travel time corrections, and to compare the scaling of error ellipses by the different data weighting of the two methods. The resulting HDC and JHD epicenters are generally consistent with a maximum separation of 1.7 km and with error ellipses that overlap ground truth within about 0.5 km. HDC and JHD error ellipses have slightly different orientations and ellipticity, and the JHD ellipses are on average slightly larger than HDC ellipses with the semi axes being about 10% longer. The median of the origin time differences is about 0.1 sec with a maximum difference of 0.2 sec. These analyses confirm that Sultanov's ground truth information for the Azgir explosions is self consistent. The arrival time data, which was provided from sources other than ISC/NEIC, provide especially useful regional travel times.

Balapan/Degelen, Kazakhstan, nuclear explosions

Coordinates for many of the nuclear explosions at the Balapan and Degelen test sites have been determined using a combination of LANDSAT and SPOT images (Thurber *et al.*, 1993; Thurber *et al.*, 1994; P. Richards, Lamont Doherty Geological Observatory). Origin times are also known very well on some of these events. An integration of these sources results in 100 explosions at the Balapan test site and 152 explosions at the Degelen test site that are of accuracy GT1 or better for which we have assembled the associated reported arrival-time data.

Chamoli, India

Proprietary locations from a temporary network provided six reference events that could be used in cluster analysis. The station azimuth gap for these solutions is less than 100 degrees and independent relocations using the same data are consistent. However, local data from other sources have yet to be integrated, so that our source conservatively estimates the accuracy of these events as GT5-GT10.

A provisional cluster of 48 events yielded 20 events with an accuracy of GT5 based on the size of the HDC confidence ellipse. However, taking into account the uncertainty in reference event locations, we can only provisionally estimate the accuracy of these locations asGT5. Interestingly enough, this recent Chamoli sequence seems to have filled a gap where no previous earthquakes during the 1964-1999 period seem to have occurred.

Bhuj (Republic Day) earthquake; Gujarat, India

Proprietary locations from an 8-station aftershock deployment that are considered accurate to +/- 1-2 km, as reported on at the 2001 Seismological Society of America meeting, have been made available to the project. However, integration of arrival time data from other deployments and permanent network stations operated by Indian institutions, as well as improvements to the local velocity model, need to be made before final reference event locations can be determined. Nevertheless, 6 of the 8 events from the proprietary data set were recorded well enough at regional and teleseismic distances to be included in a provisional Bhuj cluster.

Results of a provisional cluster analysis of 51 events that included the Bhuj main shock and all of the larger aftershocks were quite consistent. HDC locations with respect to the 6 reference events were very systematic. Hence, the shifted HDC locations probably provide the only reliable source of information on the location of events in this sequence that occurred before any temporary networks were installed. In all, 16 of these locations were of GT5 quality, but further improvement can be expected with additional data. The earthquake sequence is unusual in that there is a wide range of depths (8 - 30 km), confirmed not only by local network hypocenters but by depth phases as well. Hence, a final piece of processing could be to assign (for events which were set at an optimal depth of 18 km) more appropriate depths as indicated by the distribution in space of the local network hypocenters.

Izmit, Turkey

There were dense deployments of temporary networks for both the Izmit and Duezce events that occurred along an extended fault system. From locations provided by A. Hofstetter (Geophysical Institute of Israel), we were able to identify three reference events in the Izmit segment of the fault system that included the main shock and two large aftershocks. A cluster of 20 events resulted in 8 events of GT5 quality that are tied to only 2 of the reference events. The inconsistent shift vector and the early offset of the main shock origin time to the cluster result lead us to believe that the reported reference event was a small sub-event of the complex main shock which was not seen at regional distances. Accurate fault maps for the region may help to improve the resolution of these results.

Xiuyan, China

The Xiuyan earthquake occurred in a region where the maximum intensity contours are so tightly constrained that the location accuracy of the main shock could be estimated by the Chinese as GT2. In addition, a local seismic network of analog and digital was being operated during a period that overlapped with the earthquake sequence. Four events reported by Xu (2001) were recorded well enough to be used as reference events, and arrival time information to stations in China were provided to the project by Xu.

Cluster analysis of 30 events across the aftershock region led to the following observations: (1) We cannot validate the reference event locations to any better than GT10 because either there are errors in these locations or the station coverage used in the HDC analysis is not sufficient to constrain our locations. In most cases, it is a combination of both of these problems; (2) Origin times indicated by our cluster analysis using regional and teleseismic data reported by ISC/NEIC are systematically about 4 sec later than the origin times of the reference events. Hence, the median path anomalies to all stations used in the cluster analysis are positive. Chinese seismological stations have timing synchronous with UNT with an accuracy better than tens of msec. Reconciliation of these origin time differences may be related to inconsistent time standards globally and needs to be investigated.

CONCLUSIONS AND RECOMMENDATIONS

High-resolution cluster analysis is being applied to earthquake sequences and to some nuclear explosion sites in Asia and north Africa for which one or more of the associated events is known to an accuracy of 5 km or better (GT5). These analyses produce new locations (relative to the centroid of the cluster) and 90% confidence ellipses that in some cases, after shifting the centroid to best match reference event locations, sharpen the spatial relationship of earthquake sequences to known faults.

Our initial work with candidate "reference events" suggests that considerable care must be taken to ensure reliable results. Many aftershock studies and temporary seismograph deployments in remote areas suffer from logistical, operational, and analytical difficulties that may compromise the quality of the computed locations. Such problems are seldom apparent in published papers and abstracts. In many cases it will be necessary to gain access to raw data and analysis records—and most importantly, to gain the cooperation of the original researchers—to confirm the reliability of "reference events" offered by the seismological community in these regions.

The Hypocentroidal Decomposition method of cluster analysis has proven to be very well suited to the requirements of ground truth validation exercise, but we have also found a number of areas in which additional development of the method is needed. Enhancements to the algorithm are needed to implement a more appropriate statistical model for reading errors, and for dealing with outliers. A statistically rigorous procedure for optimally matching the HDC cluster with reference event locations is needed. Further research is also needed on some aspects of the application of the HDC method. For example, we need a better understanding of how to choose a minimum epicentral distance for data to be used in HDC analysis, and a protocol for deciding how large a cluster may be without compromising the analysis.

REFERENCES

Asudeh, I. (1983), ISC mislocation of earthquakes in Iran and geometrical residuals, Tectonophysics, 95, 61-74.

Berberian, M. (1982), Aftershock tectonics of the 1978 Tabas-e-Golshan (Iran) earthquake sequence: a documented active 'thin- and thick-skinned tectonic' case, Geophys. J. R. astr. Soc., 68, 499-530.

Berberian, M. and R. S. Yeats (1998), Patterns of historical earthquake ruptures in the Iranian Plateau, Bull. Seismol. Soc. Am.

Berberian, M., J.A. Jackson, M. Qorashi, M.M. Khatib, K. Priestley, M. Talebian, and M. Ghafuri-Ashtiani (1999), The 1997 May 10 Zirkuh (Qa'enat) earthquake (Mw 7.2): faulting along the Sistan suture zone of eastern Iran, Geophys. J. Int., 136, 671-694.

Bolt, B.A. (1976), Nuclear explosions and earthquakes: The parted veil, W.H. Freeman.

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., 88, 722-743.

Fuenzalida, H., L. Dorbath, A. Cisternas, H. Eyidogan, A. Barka, L. Rivera, H. Haessler, H. Philip, and N. Lyberis (1977a), Seismic source study of the Racha-Dzhava (Georgia) earthquake from aftershocks and broadband teleseismic body-wave records: an example of active nappe tectonics, Geophys. J. Int.,130, 29-46.

Fuenzalida, H., L. Rivera, H. Haessler, D. Legrand, H. Philip, L. Dorbath, D. McCormack, S. Arefiev, C. Langer, and A. Cisternas (1977b), Seismic source study of the Racha-Dzhava (Georgia) earthquake from aftershocks and broadband teleseismic body-wave records: an example of active nappe tectonics, Geophys. J. Int.,130, 29-46.

Gupta, V (1995), Locating nuclear explosions at the Chinese test site near Lop Nor, Science and Global security, 5, 205-244.

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.

Kennett, B. L. N., E.R. Engdahl and R.P. Buland (1995), Constraints on seismic velocities in the Earth from traveltimes, Geophys. J. Int., 122, 108-124.

Pan, J., M. Antolik and A. Dziewonski (2000), Locations of mid-oceanic earthquakes constrained by sea-floor bathymetry, Eos Trans. Amer. Geophys. Union, 81, F868.

Milkereit, C. *et al.* (2000). Preliminary aftershock analysis of the Mw=7.4 Izmit and Mw=7.1 Duezce earthquake in western Turkey, In: The 1999 Izmit and Duezce earthquakes: preliminary results (Editors: A. Barka, Oe. Kozaci, S. Akyuez and E. Altunel), 179--187, Istanbul Technical University.

Sultanov, D.D., J.R. Murphy and K.D. Rubenstein (1999), A seismic source summary for Soviet Peaceful Nuclear Explosions, Bull. Seis. Soc. Amer., 89, 640-647.

Sweeney, J.J. (1998), Criteria for Selecting Accurate Event Locations from NEIC and ISC Bulletins, Lawrence Livermore National Laboratory, Technical Report UCRL-JC-130655.

Thurber, C.H., H.R. Quin, and P.G. Richards (1993), Accurate locations of nuclear explosions in Balapan, Kazakhstan, 1987 to 1989, Geoph. Res. Letters, 20, 399-402.

Thurber, C.H., H.R. Quin and P.G. Richards (1994), Catalog of locations of nuclear explosions at Balapan, Kazakhstan, 1965 to 1985, Bull. Seis. Soc. Amer., 84, 458-461.

Tibi, R. *et al.* (2001), Rupture process of the 1999 August 17 Izmit and November 12 Duezce (Turkey) earthquakes, Geophys. J. Int., 144, F1--F7.

Xu, S. (2001), The Study of the XIUYAN Earthquake on Nv. 29, 1999, as an Example of an Event for Local Calibration (presentation), Third Location Workshop of the IDC Technical Experts Group on Seismic Event Location, Oslo.