Bulletin of the Seismological Society of America, Vol. 94, No. 4, pp. 1528–1545, August 2004

Ⓔ

Collection of a Reference Event Set for Regional and Teleseismic Location Calibration

by István Bondár, E. Robert Engdahl, Xiaoping Yang, Hafidh A. A. Ghalib, Abraham Hofstetter, Victor Kirichenko, Robert Wagner, Indra Gupta, Göran Ekström, Eric Bergman, Hans Israelsson, and Keith McLaughlin

Abstract A three-year consortium project, with members of Science Applications International Corp., University of Colorado at Boulder, Harvard University, Multimax Inc., Geophysical Institute of Israel, Western Services, and University of California at San Diego, was initiated in 2000 to improve locations and reduce uncertainties in the Middle East, North Africa, Europe, and Western Eurasia. The consortium developed high-resolution three-dimensional models of the Earth's mantle to generate accurate travel-time predictions for regional and teleseismic *P* phases. Since the approach was purely model-based, a large set of high-quality reference events was needed to validate the model predictions. The consortium has spent considerable effort to collect, vet, and validate reference events located with 5-km accuracy or better by local networks and "promoted" reference events located with an accuracy of 7 km or better by application of multiple-event location techniques. Consortium members built an extensive network of contacts to solicit candidate reference events from local, regional, and national network operators. Strict methodologies were developed to identify candidate reference events in earthquake bulletins, and to validate and quality control the selected candidate reference events. The outcome of the consortium effort was a quality-controlled reference event list with nearly 2000 events and over 200,000 arrivals. The Reference Event List is provided as an electronic supplement to this article.

Online material: Reference event database.

Introduction

In recognition of the importance of location calibration of sparse global seismic networks, such as the International Monitoring System network, a consortium was formed in 2000 to improve locations and reduce uncertainties in the Middle East, North Africa, Europe, and Western Eurasia. The consortium (named the Group 2 Consortium) included scientists from both academia and industry. The Group 2 Consortium was led by Science Applications International Corp. (SAIC), with University of Colorado at Boulder (CUB), Harvard University, Multimax Inc., Geophysical Institute of Israel (GII), Western Services, Russia, and University of California at San Diego as consortium members.

The Group 2 Consortium followed a model-based approach to generate path-dependent travel-time correction surfaces relative to the IASPEI (Kennett and Engdahl, 1991) travel-time tables. Part of the Consortium effort constituted the development of a global upper mantle model by CUB (Shapiro and Ritzwoller, 2002, 2004) and a global whole mantle model by Harvard University (Antolik *et al.*, 2003). Both models incorporate the global CRUST2.0 model (Bassin *et al.*, 2000) of the Earth's crust. The CUB model was used to calculate travel-time corrections for regional phases (*Pn, Sn*) while the Harvard J362 model was employed to generate correction surfaces for teleseismic *P* in the distance range between 25 and 97°.

It was recognized early that in order to test and validate models and corrections, a high-quality ground truth data set is needed. Ground truth ideally represents events with exactly known location, depth, and origin time. However, there are only a handful of events that would satisfy these requirements and be large enough to be recorded at regional distances. Ground truth (GT) information in this context is therefore defined as epicenters with known accuracy. We adopted the "GTX" classification of Bondár *et al.* (2001), which indicates that the true epicenter lies within "X" km of the estimated epicenter at high confidence level. Since the expected improvements due to location calibration are in the order of 10 km, we aimed for GT0–5 events. Furthermore, as for nuclear monitoring purposes deep events are of less interest, we concentrated on collecting shallow-depth (crustal) ground truth events.

The consortium put special emphasis on establishing selection criteria to identify candidate GT5 events at high confidence level by searching earthquake bulletins. Arrival data from various bulletins were merged to create a comprehensive bulletin for every event in the reference event data set. We relied greatly on data extracted from the bulletins of the International Seismological Centre (ISC; www.isc.ac.uk), the National Earthquake Information Center (NEIC; http://wwwneic.cr.usgs.gov), the Prototype International Data Center (PIDC; www.cmr.gov), and the European–Mediterranean Seismological Centre (EMSC; www.emsc-csem.org), as well as the groomed ISC data set (EHB; Engdahl *et al.*, 1998). The EHB was also used to identify candidate GT5 events. We have also shared location and arrival data of reference events with the two other consortia (Armbruster *et al.*, 2002; Murphy *et al.*, 2002) formed for the location calibration of Eastern Asia.

During the consortium project, nearly half of the resources were devoted to the reference event collection effort. The task proved to be more difficult than initially anticipated. The motivation of this article is to document this effort; discuss the methodologies we developed to identify, vet, and validate candidate GT events; and present the final products. A related paper by Yang *et al.* (2004) discusses the validation of the CUB and Harvard model travel-time predictions by the relocation of the reference events presented here.

Methodologies and Criteria for Assigning GTX

Selection Criteria to Identify GT5 Candidates

Since location uncertainties cited in earthquake bulletins are in reality only measures of precision that are often unrealistic, researchers in the location calibration community set out to establish criteria to identify candidate events at various GT levels without relying on reported nominal errors. Most of these criteria exploit parameters readily available or easily extracted from earthquake bulletins, and attempt to describe favorable network geometries using the number of stations, primary and secondary azimuthal gaps (e.g., Sweeney, 1998; Dewey and Kork, 2000; Myers and Schultz, 2000a). The Group 2 Consortium helped develop selection criteria for GT5 candidate events (McLaughlin *et al.*, 2001, 2002). A detailed overview of the development of GT selection criteria is given in Bondár *et al.* (2004).

Other approaches include the analysis of satellite imagery (Gupta and Pabian, 1996; Gupta and Rich, 1996; Albright *et al.*, 1998; Barker *et al.*, 1999; Richards, 2000; Bhattacharyya *et al.*, 2002; Fisk, 2002; Skov *et al.*, 2002) to obtain accurate locations for underground nuclear explosions. Analysis of Interferometric Satellite Radar (InSAR) images is a promising technique to obtain well-defined locations for large, shallow earthquakes (e.g., Saikia *et al.*, 2001, 2002). For instance, InSAR analysis (Martínez-Díaz *et al.*, 2002) of the February 1999 Murcia, Spain, earthquake sequence confirmed the GT5 locations identified in the Instituto Geografico Nacional (IGN) bulletin. While InSAR has great promise for generating future GT events, our unsuccessful InSAR analysis of three earthquakes in Algeria taught us that InSAR coverage is still sparse and moderate-sized events must be shallow to form unambiguous InSAR signatures.

We applied the local network selection criteria developed by Bondár *et al.* (2004) on the bulletins to identify candidate GT5 events. The selection criteria require that an event be located by a local velocity model with at least 10 stations, all within 250 km, with an azimuthal gap less than 110° and a secondary azimuthal gap less than 160°, as well as with at least one station within 30 km of the epicenter. The latter constraint provides confidence for the depth of the event. The criteria identify GT5 events at the 95% confidence level and GT10 events at the 99% confidence level. Each earthquake was required to pass these selection criteria. Because their locations are obtained from independent, nonseismic sources, explosions may not pass the GT5 selection criteria.

Secondary azimuthal gap is defined as the largest azimuthal gap filled by a single station. Figure 1 illustrates the idea of the secondary azimuthal gap for an underground nuclear explosion carried out in Novaya Zemlya on 25 October 1984. Although the primary azimuthal gap of about 90° suggests reasonable station coverage, CMAR closes a 135° azimuthal gap, thus having the largest influence on the location. A constraint on the secondary azimuthal gap therefore reduces the risk of mislocation due to picking errors at stations with large relative importance in the location process.

Validation of GT5 Candidate Events and Generation of Promoted GT7 Reference Events via Multiple Event Location Techniques

Candidate reference events are validated if multiple-event location of clustered events, using phase arrival times at regional and teleseismic distance ranges can be shown to be consistent with the corresponding local network solutions (Engdahl *et al.*, 2002). Events within a cluster may also be “promoted” to reference event status if the confidence ellipses for those events meet set criteria. Since multiple-event location techniques use more than one arrival along similar paths, these techniques also can be used to identify outliers in the phase arrival times.

To generate promoted reference events from selected GT5 reference events Joint Hypocenter Determination (JHD) analysis (Douglas, 1967; Dewey, 1972) was applied to 11 event clusters located in the Mediterranean and Central Europe. The analysis was based on first arrival-time data retrieved from ISC and NEIC bulletins. Only arrivals recorded at a distance greater than 5 times the event cluster diameter (typically 50 km) and less than 90° were used. Most of the

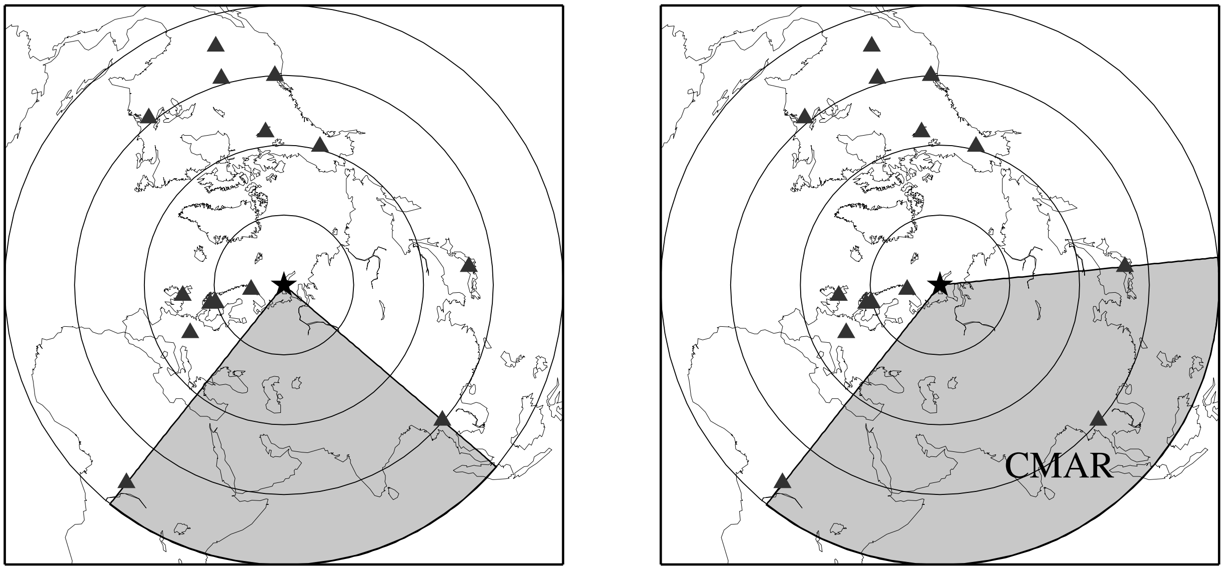

Figure 1. Illustration of secondary azimuthal gap for the 25 October 1984 underground nuclear explosion in Novaya Zemlya. While the primary azimuthal gap (gray area in the left panel) suggests reasonable station coverage, the secondary azimuthal gap (gray area in the right panel) indicates that CMAR is closing about a 135° azimuthal gap, thus having the largest influence on the location.

clusters included events recorded at local and regional distances. The lower distance limit on station data was therefore imposed to avoid violating the JHD assumption that the path correction for a given station and cluster is the same for all events in the cluster. Each event cluster was constrained by one or more reference events (GT5 or better) already in the reference event list. Events inside the cluster located with JHD were accepted as GT7 events if the estimated semimajor axis of the 90% error ellipse was less than 5 km and the distance to the nearest constraining event was less than 5 km. These two criteria for promoted GT7 (or better) reference events were satisfied by 28 events in 4 of the 11 clusters.

The Hypocentroidal Decomposition (HDC) multiple-event relocation method (Jordan and Sverdrup, 1981) was used to validate candidate GT5 events in 26 earthquake and explosion sequences across Eurasia and North Africa. Situations were sought where a number of moderate-sized earthquakes are clustered (within about 50–100 km of each other) and where at least one or more of the events has been very well located by a local network so that they meet the GT5 selection criteria for candidate reference events. The events in the cluster may be widely distributed in time, as long as arrival time data at common stations are available. The clusters typically are 50–100 km across and comprise up to 100 events of magnitude 3.5 or greater 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 the 1D model ak135 (Kennett *et al.*, 1995). The HDC analyses produce new locations that are defined by cluster vectors in space and origin time relative to the hypocentroid, which is then located in the traditional manner to yield absolute locations and origin times. Obviously, this process is subject to bias. To remove the bias, we shift the hypocentroid in space and time to provide an optimal match on average to one or more reference events that are included in the cluster. This brings all events in the cluster into close alignment (<5 km) with the ground truth reference events. In addition, since the absolute locations and times of all the events in the cluster are now estimated with increased accuracy, many of the events in the cluster may now be promoted to reference event status at an appropriate GT level (GT7 or better).

The degree of consistency between the relative locations as determined by global arrival time data and the relative shifted locations specified by the reference event data are two of the tests we use to validate candidate reference events. Shifts in epicenter and origin time (to best match the reference event locations) are typically in the range of 5–15 km and ±2 sec, respectively. We validate candidate reference events by requiring that the relative shift patterns (or mislocation vectors) of candidate reference events be consistent with the pattern of the corresponding cluster vectors from the HDC analysis. Discrepancies may be resolved by determining that the cluster vector is biased for some reason, or by rejecting the candidate reference event. For this reason, most clusters contributing to the database for this study are calibrated by several reference events. Reference events that could not be validated at the GT5 level by the HDC method, because the estimated semimajor axis of the 90% error ellipse was greater than 5 km, were classified as GT10 if the estimate was less then 10 km and rejected if greater than 10 km.

Engdahl and Bergman (2001) compiled a comprehensive data set of well-located earthquakes and explosions that have been validated by cluster analysis. There are 26 clusters in this data set: 6 explosion clusters, most with source locations known to 2 km or better; 17 earthquake clusters, 12 of which are believed to be accurate to 5 km or better; and the remainder to 10 km or better. For each event, the database contains associated phase-arrival times that were either reported to the ISC or NEIC or contributed by other sources.

Cross-Validation of JHD and HDC Analyses

We compared HDC and JHD results (epicenters and station path corrections) for event clusters near Azgir at the Caspian Sea (underground nuclear explosions), at Racha in the Western Caucasus, and in the Gulf of Aqaba (shallow earthquakes). The two cluster-analysis methodologies were applied to identical data for the Azgir cluster as well as independently compiled and only partly overlapping data for the Racha and Aqaba clusters. The IASPEI travel-time tables were used by both HDC and JHD, which both also used only first arriving *P* times.

As Figure 2 indicates, some minor systematic differences were noted in the size of the error ellipses. The differences can be attributed to the algorithmic differences between HDC and JHD in the weighting of input data and application of ellipticity and elevation corrections. Although these differences might be worthy of further investigation, they do not impact the independent use of the two methodologies for estimating station path corrections for clusters of well-located events with and without ground truth information.

Figure 3 shows the empirical path corrections obtained from HDC (*x*-axis) and JHD (*y*-axis) analyses. In summary, the comparisons did not reveal any statistically significant differences in estimates, which were generally consistent within estimated uncertainties. The standard deviation of the differences in estimated path corrections ranged from 0.1 sec, for Azgir with identical data, to 0.35 sec, for Aqaba and Racha with partly overlapping data.

Outlier Analysis

Outlier Rejection in Event Cluster Analysis

For each cluster, the JHD algorithm was applied three times with a two-step outlier rejection procedure between runs. Prior to the first JHD run, arrivals with residuals larger than 15 sec in the ISC and NEIC bulletins were removed. After the initial JHD run arrival rejection was, in a first step, based on distance-dependent cut-off values ranging from about 1.5 sec to 4 sec (the largest cut-off limit for arrivals around 15°). A standard kurtosis test (at 1% level) was applied on a station-by-station basis as a second outlier rejection step; arrivals failing the kurtosis test but with residuals less than 1 sec were, however, retained. Stations for which more than one-third of the arrivals were rejected in the two steps were omitted entirely from the subsequent processing. The JHD was then run with the groomed data and the two-step outlier rejection was applied once again. On average, about 5% of the original arrivals were rejected in this multistep outlier rejection procedure. A third and final JHD run completed the analysis.

HDC analysis is also useful in grooming the arrival-time data to identify and remove outliers much more precisely than is possible in single-event location. This is possible because HDC analysis estimates the average path anomaly for each seismic station that observes more than one of the events in the cluster. The residuals from this path anomaly are dominated by reading errors. Very large cluster residuals can be rejected as spurious readings, even if their absolute travel-time residual is small. Removing outliers in this way improves the resolution of relative locations and reduces the size of confidence ellipses.

In the HDC method reading uncertainties for each station are estimated directly from the residuals of that station in the cluster being analyzed. In general this uncertainty is smaller than the standard assumptions used in single-event location work, because in the HDC analysis we have removed the contribution of correlated travel-time errors. As a result, confidence ellipses are smaller and our confidence in the statistical validity of the results is raised. For the purposes of generating new promoted ground truth events from HDC analysis, however, this convenient approximation is not appropriate, and it is necessary to be able to combine these two aspects of location uncertainty to arrive at a single measure of uncertainty for an event that has been localized by HDC. In the clusters we have so far examined for ground truth event-validation purposes, the semi-axes of the hypocentroid's 90% confidence ellipse are usually 1–3 km in length. This is not insignificant at GT5 levels of accuracy or even at GT10. In future, as part of a general program to tighten up the statistical underpinnings of HDC for validation work, this issue must be addressed. Hence, the confidence ellipses estimated by either the JHD or HDC methods reflect almost entirely the reading error, as the model error is absorbed in the station path anomalies. As a result, the size of the estimated 90% confidence ellipse estimated by either method is an accurate representation of uncertainties in relative locations due to reading errors.

Analysis of Bulletin Information

Several studies (e.g., Waldhauser and Ellsworth, 2000; Fisk, 2002) have shown that waveform correlation techniques can significantly increase the internal consistency and accuracy of phase picks of similar ray paths. Waveform correlation can also easily identify outliers. However, given the size of the region of interest, the collection and analysis of waveforms would have been a gargantuan task clearly exceeding the resources and scope of the consortium project. Therefore we resorted to the analysis of available bulletin data.

In order to detect possible phase misidentifications, clock errors, and other outliers in the bulletins, we performed

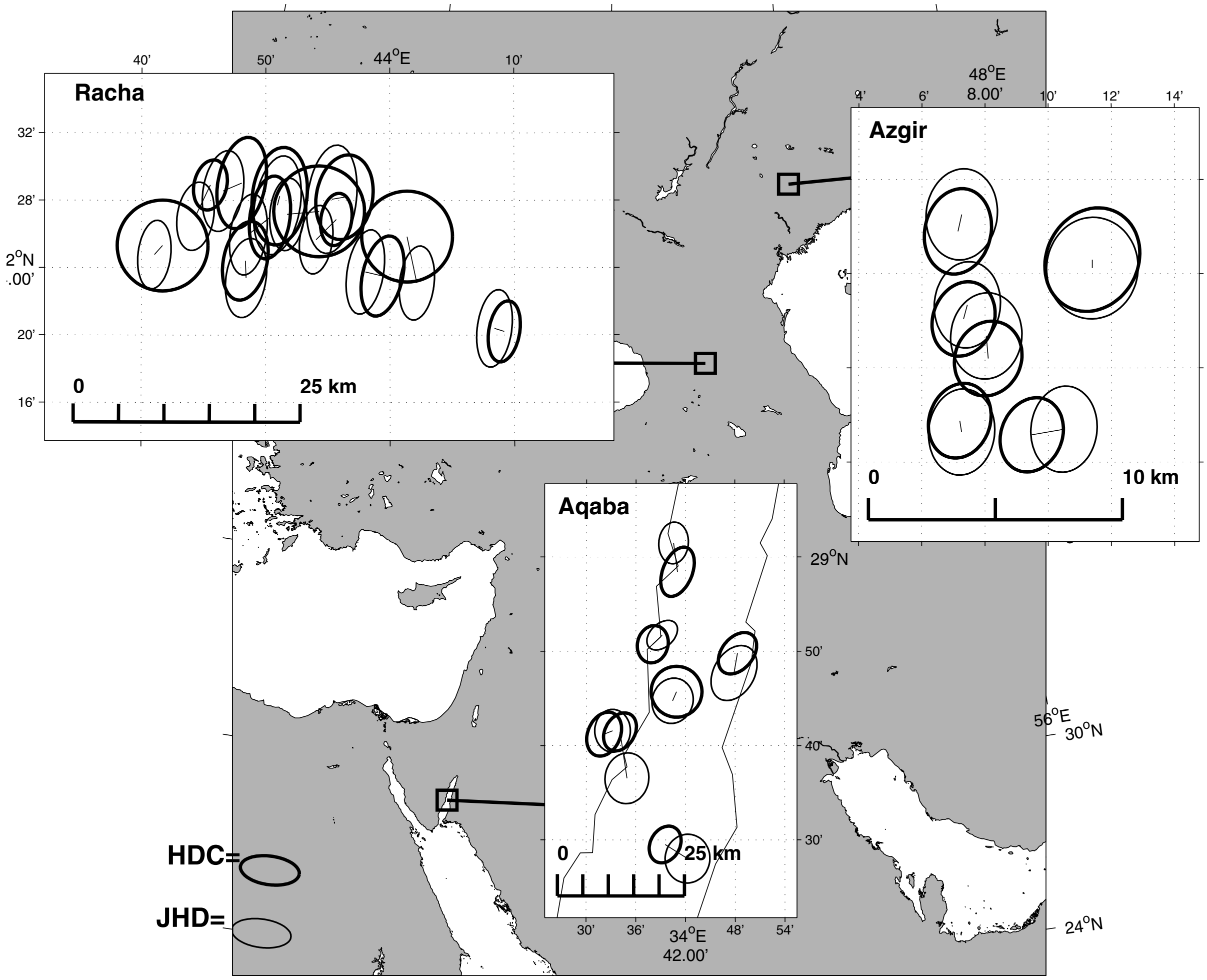

Figure 2. Comparison of locations with error ellipses obtained from HDC (thick lines) and JHD (thin lines) analyses of three event clusters in the Gulf of Aqaba, Azgir, and Racha used in the cross-validation of the two techniques.

an analysis of time residuals with respect to IASPEI predictions. Figure 4a illustrates the effect of phase-identification errors on predicted travel-time residuals. The regions indicated by the ovals are possible phase misidentifications in the near and far regional *Pn* distance ranges. These arrivals are considered outliers, and if individual inspection of data warrants, they are removed from the data set as shown in Figure 4b. Similar analysis is carried out for other phases.

Variogram analysis of arrival data is another powerful technique to identify outliers and clock errors. A variogram summarizes the relationship between differences in pairs of measurements and the distance of the corresponding points from each other (Isaaks and Srivastava, 1989), and is routinely used in geostatistical applications, such as kriging of travel-time surfaces (Myers and Schultz, 2000b). Figure 5a shows the variogram of IASPEI-predicted *Pn* travel-time residuals. Note the unexpected increase in predicted travel-time residual differences at very closely located stations. This is due to scatter in picks at colocated stations. Data from some stations are processed by multiple agencies and reported under different station codes (e.g., the PIDC reported GERES, while Bochum University reports GEC2). When various bulletins are fused into one, this problem becomes profound. To remedy the problem, we vet the merged bulletin in a way that it will contain only one pick for a given phase by taking the average arrival time from the co-located stations, assigning it to the most frequently reporting station code, and removing the other picks. Inspection of variograms also reveals stations that are consistently out of "sync" with nearby stations. These stations produce most of the points outside the 3-sigma range on the variogram and most likely suffer from clock or phase-identification errors, and are therefore flagged as outliers. The variogram after the removal of the outliers is shown in Figure 5b.

As Figures 4b and 5b indicate, the remaining scatter in predicted residuals after the removal of outliers is still fairly large. We removed only those arrivals that were unquestionable outliers according to both the IASPEI and CUB (J362 in

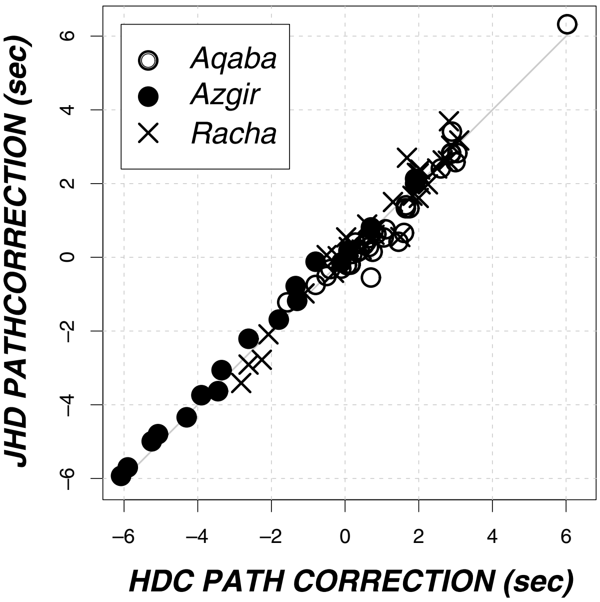

Figure 3. Comparison of empirical station path corrections derived from JHD and HDC analyses of the same event clusters.

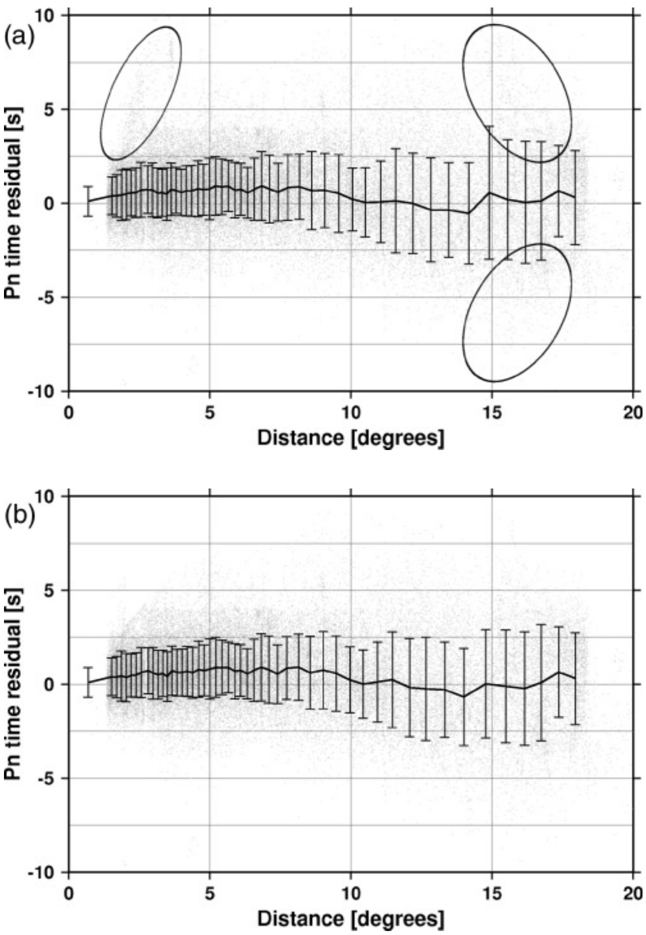

Figure 4. IASPEI-predicted *Pn* travel-time residuals (dots) for events in the Group 2 Reference Event Lists before (a) and after (b) removal of outliers. The thick solid line connects the medians of each 2 percentile points of data. Ovals indicate possible phase-identification errors in the *Pg–Pn* and *Pn–P* crossover distance ranges.

the teleseismic case) model predictions in order to avoid creating self-fulfilling prophecies in favor of a particular model.

The outlier analysis described previously resulted in removing some 1% of the arrival data from the Group 2 Consortium Reference Event List. However, outliers contaminated 767 events, some 40% of the reference events in the database. Therefore, outlier analysis is extremely important.

Reference Event List

The final Group 2 Consortium Reference Event List consists of 1963 GT0–10 events, of which 1852 events have arrival data. The 3-year consortium effort resulted in a more than twofold increase of quality-controlled ground truth events in the region. Every event in the list was vetted and is documented by metadata. (Ⓔ The Group 2 Reference Event List is available online at the SSA Web site.)

The locations of the ground truth events and their distribution by source type and accuracy as well as the number of teleseismic *P* and regional phases are shown in Figure 6. Figure 7 shows the regional ray coverage provided by some 50,000 *Pn* phases. Despite the focused search for candidate GT events, Africa and the Middle East remain poorly covered with reference events, primarily due to the lack of dense local networks. To identify GT5 candidates in the Middle East and Africa we could rely only on temporary networks deployed to study aftershock sequences.

Although nuclear and chemical explosions dominate reference events in northern Eurasia, there are hardly any explosions south of latitude 40°. Tectonic regions are almost exclusively represented by GT5–10 earthquakes. Because explosions typically fall into the GT0–2 categories (with some exceptions for the Peaceful Nuclear Explosions) measuring improvements due to location calibration should be relatively straightforward in Northern Eurasia, while in the south demonstrating improvements is expected to be a more difficult task owing to the more complicated velocity structures in tectonic regions and the somewhat lower accuracy of available GT information. Oceanic seismicity is represented by a few dozen mid-ocean ridge and transform fault events of at best GT10 accuracy, which provides insufficient statistical power to validate travel-time predictions for oceanic paths.

Documentation of ground truth information is critical to understanding the provenance of GT data. We made a considerable effort to collect and maintain information that documents the sources of location (origin) and arrival data. In the Group 2 Reference Event List every single event is documented with metadata.

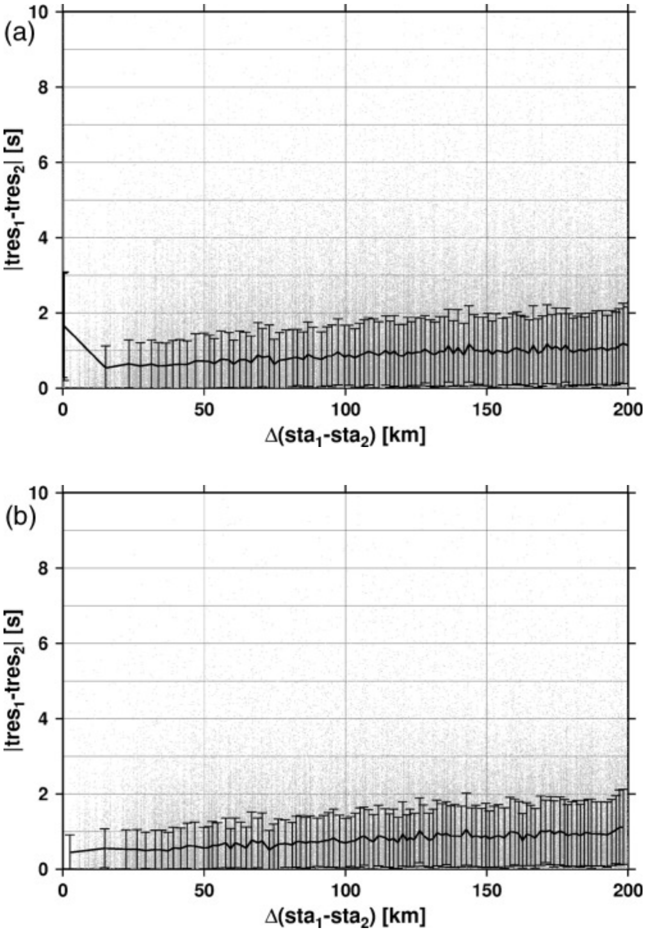

Figure 5. Variogram of IASPEI-predicted *Pn* travel-time residuals for events in the Group 2 Reference Event List before (a) and after (b) removal of outliers. The solid line connects the median values of every 2 percentile points of data. The median residual difference increases almost linearly with increasing distances between station pairs. The sudden increase for closely spaced stations in the bottom panel is due to the scatter of arrival time picks at colocated stations reported by multiple agencies. After removing the outliers the nugget at zero separation is to be estimated of ~0.3 sec.

Table 1 summarizes the sources of information for the events in the Reference Event List. Below we present examples to illustrate our methodological approach in collecting high-quality ground truth events. Since we feel important to highlight the limitations of using bulletin data in identifying ground truth events, examples are given for cases when candidate ground truth events could not be verified.

Existing GT Data prior to the Consortium Effort

One of our major sources for ground truth events was the Center for Monitoring Research (CMR). We obtained two data sets of GT0–GT5 reference events in our region of interest from the CMR databases (Bondár *et al.*, 2001). The first set consisted of confirmed or presumed nuclear explosions carried out in the region by the former USSR, France, China, India, and Pakistan between 1955 and 1998 from the CMR Nuclear Explosion Database (Yang *et al.*, 2004). Arrival data for the nuclear explosions since 1964 were extracted from the ISC/EHB bulletins (Engdahl *et al.*, 1998).

The second set included documented chemical explosions, mining events, and other events collected in the CMR Ground Truth Database (Yang *et al.*, 2000a). These events are located mostly in Fennoscandia and were used in previous location calibration studies (Yang *et al.*, 2001). We selected only those events that are large enough to be recorded at regional distances and well documented with metadata. Reviewed Event Bulletin (REB; www.cmr.gov) arrival data for events in the CMR Ground Truth Database were complemented with arrivals from the CMR Reference Event Database (Yang *et al.*, 2000b) which includes well-recorded events with special emphasis on local and regional seismic network bulletins of various National Data Centers (NDCs) provided to the Prototype International Data Center during the Group of Scientific Experts Technical Test 3.

The Lawrence Livermore National Laboratory (LLNL) made available a set of location and arrival data from the LLNL collection of ground truth events. The combined location and arrival data of GT0–5 events obtained from the CMR and LLNL databases constituted the Consortium's initial Reference Event List of some 850 events.

Local Networks and Bulletins

Unfortunately, no institutional framework exists to routinely identify and collect reference events or to encourage local, regional, and national networks in doing this. While networks in Europe normally follow an open-access policy and report to international data centers, many networks in the Middle East and Africa do not regularly report to international data centers, and only a few share data with neighboring countries. During the past decade the RELEMR (Reducing Earthquake Losses in the Eastern Mediterranean Regions) project, cofunded by the UNESCO (United Nations Educational, Scientific and Cultural Organization) and the USGS (U.S. Geological Survey), spent considerable effort on promoting data exchange between countries in the Mediterranean through the Joint Seismic Observation Period (JSOP) projects coordinated by EMSC (European–Mediterranean Seismological Centre). However, only a few networks contributed regularly to the JSOP project. There are only a few local networks in the Middle East and Africa that are dense enough to locate events with sufficient azimuthal coverage. Thus, data exchange between networks in the region is essential to obtain more accurate locations and to identify GT candidates. The organizers of the annual RELEMR meetings also gave us the opportunity to present our research objectives and results to the participants.

We realized that personal contacts and correspondence are more likely to succeed in obtaining arrival data than official inquiries at institutional levels. Consortium members have devoted a considerable amount of time to establish per-

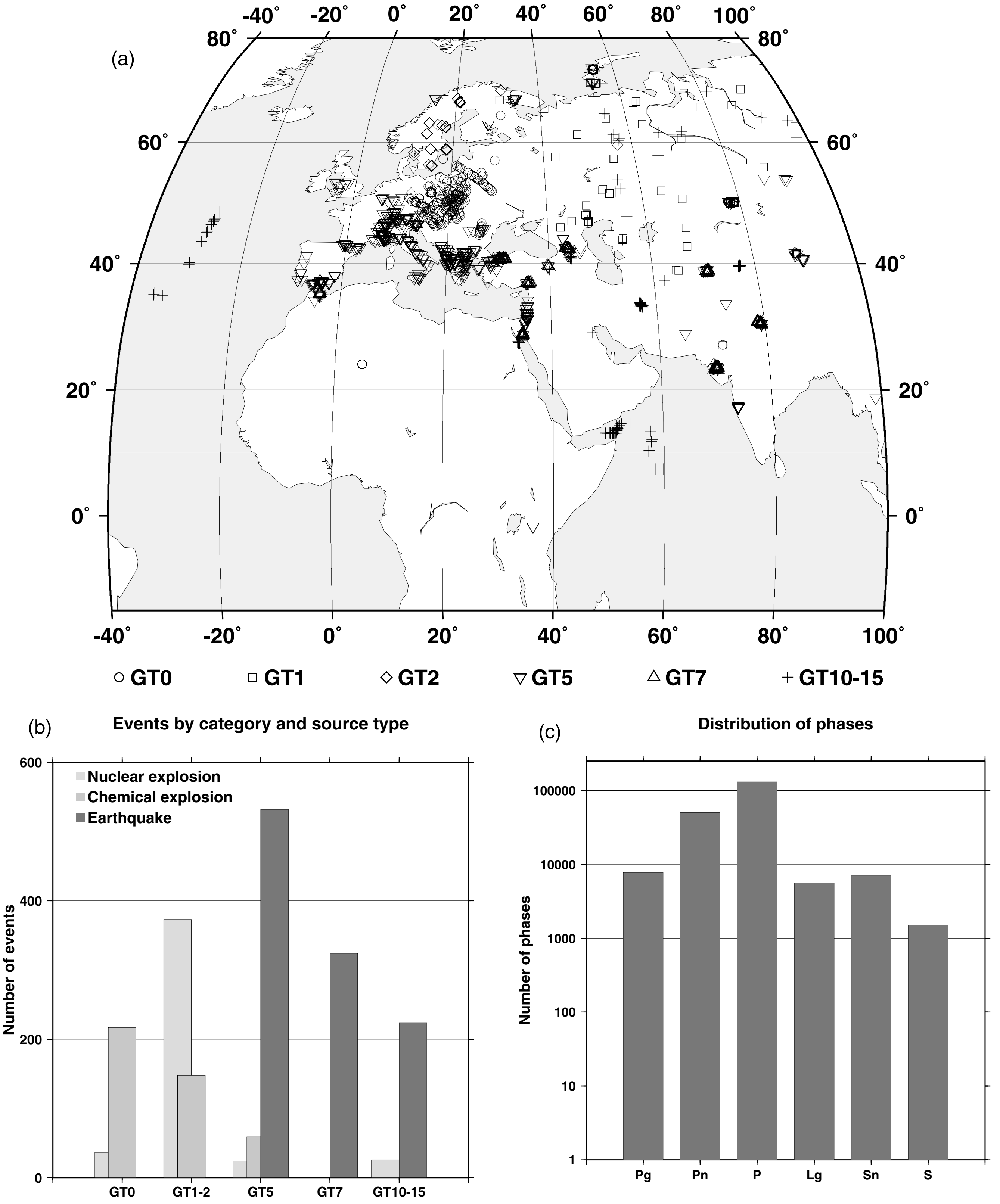

Figure 6. Location of ground truth (GT) events and number of teleseismic *P* and regional phases. (a) Location of GT0 (circles), GT1 (squares), GT2 (diamonds), GT5 (inverted triangles), GT7 (triangles), and GT10–15 (crosses) events in the Group 2 Reference Event List. While Europe is well covered with GT events, Africa and the Middle East remains poorly covered with GT events. (b) Distribution of events by GT category and source type. The overwhelming majority of the 1963 reference events are earthquakes. (c) Number of *Pg*, *Pn*, *P*, *Lg*, *Sn*, and *S* phases in the Reference Event List. The number of *P* and *Pn* phases is an order of magnitude larger than the number of other phases.

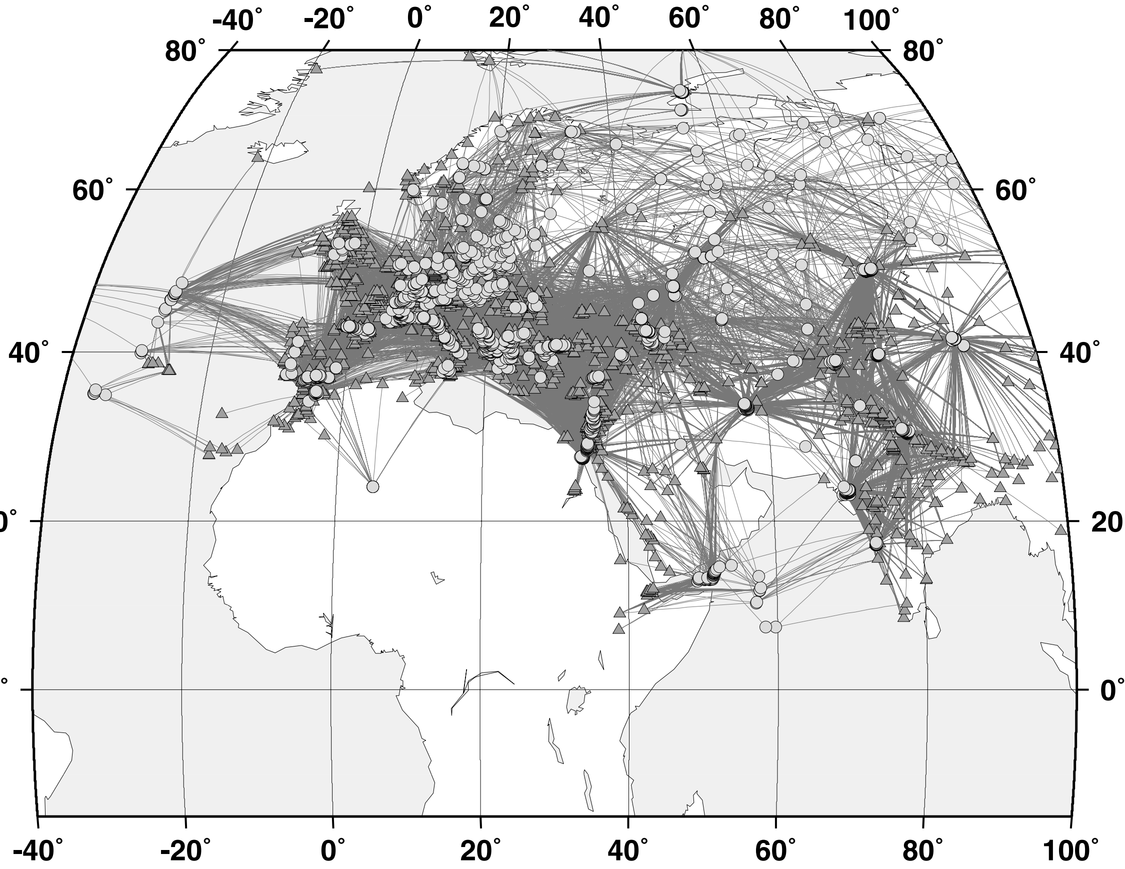

Figure 7. Ray coverage by some 50,000 *Pn* phases in the Group 2 Reference Event List. While most of the Group 2 region of interest is well covered, Africa and the Persian Gulf remain poorly covered.

sonal contacts at meetings and by e-mail inquiries with researchers in the Group 2 region of interest.

Waveforms at some future IMS stations are not regularly analyzed. The Geophysical Institute of Israel made an effort to collect waveforms from these stations or their surrogates and pick arrivals for events in the Reference Event List: AAE, Ethiopia; ATD, Djibouti; IDI, Greece; KEG, Egypt; MDT, Morocco; RAYN, Saudi Arabia; and EIL and MRNI, Israel.

Multimax obtained bulletins of the Minagish, Kuwait, aftershock sequence (Bou-Rabee, 2000) from the Kuwait National Seismic Network. The local bulletins were complemented with ISC readings. Additional arrivals were picked from station RAYN, Saudi Arabia, waveforms. As the 18 September 1997 main shock was the only event in Kuwait ever recorded instrumentally at regional distances, we were not able to perform a cluster analysis. Since the event almost satisfied our GT5 criteria, we accepted it as GT10.

We obtained local bulletins in Morocco between 1992 and 1999 via personal communication with researchers at the Université Mohamed V, Dept. de Physique du Globe (DPG), and at the Centre National Coordination et de Planification de la Recherche Scientifique et Technique (CNRM). Both centers operate seismic networks in Morocco of about 14 and 24 stations, respectively. We selected the 12 best reference event candidates for local network relocation by merging arrivals from the local networks and ISC or EHB bulletins. Four events passed our GT5 selection criteria, and were validated by HDC analysis.

The Geology Department of the University of Nairobi provided us local bulletins from the 15-station temporary deployment around Lake Magadi, Kenya. The locations of the swarm events occurred during the lifetime of the network were refined using local tomography (Ibs-von Seht *et al.*, 2001). Of the 12 best-recorded events only 2 met the GT5 selection criteria. Although we set out to specifically look for arrivals for these events in the waveforms of stations at regional distances from the Kenya Rift (e.g., KMBO, ATD, BGCA, FURI, RAYN, AAE) we could not detect any arrivals.

Using data from a temporary deployment to study the Bhuj aftershock sequence (Bodin *et al.*, 2001), HDC analysis gave very consistent results with the local network locations. The Bhuj sequence produced 79 reference events altogether. Similarly, HDC analysis of the Chamoli sequence provided

Table 1
Metadata for Events in the Group 2 Reference Event List

Event Type	Event Origin Sources	Arrival Sources*	No. of Events†	GT
Nuclear Explosions			459	0–10
Balapan, Kazakhstan	N. N. Belyashova, personal comm., 1999; Bocharov *et al.*, 1989; Lilwall *et al.*, 1990; Murphy and Jenab, 1992	HDC, IDG, LDEO, EHB, ISC	99	1
Degelen, Kazakhstan	Bocharov *et al.*, 1989; Lilwall *et al.*, 1990	EHB, HDC, IDG, LDEO, ISC	151	1
India	Barker *et al.*, 1998; Gupta and Pabian, 1996	EHB, REB	2	0–5
Lop Nor, China	Bhattacharyya *et al.*, 2002; Douglas *et al.*, 1993; Gupta, 1995; Gupta and Rich, 1996; Fisk, 2002	EHB, HDC, ISC	29	0–5
Novaya Zemlya, Russia	Khristoforov, 1996; Marshall *et al.*, 1994; Richards, 2000	EHB, IDG, ISC, HDC	42	1–5
Peaceful Nuclear Explosions, USSR	Sultanov *et al.*, 1999	EHB; Murphy *et al.*, 2002; IDG; ISC; LDEO; HDC	98	1–10
Pakistan	Barker *et al.*, 1998; Albright *et al.*, 1998	EHB, GII, REB	1	5
Sahara, Algeria	Bolt, 1976; Duclaux and Michaud 1970	HDC	13	0
Semipalatinsk, Kazakhstan	Bocharov *et al.*, 1989; Lilwall *et al.*, 1990	EHB, IDG, LDEO, HDC	24	0–5
Calibration Shots and Seismic Profiles			216	0
Balapan shots, Kazakhstan	Jih and Wagner, 2000; Kazakhstan	Multimax, REB	7	0
CELEBRATION2000	Guterch *et al.*, 2001	Austria, Hungary, Slovenia, NEIC, REB	147	0
Dead Sea Shots	Gitterman and Shapira, 2001	GII, REB	3	0
European GeoTraverse	Aichroth *et al.*, 1990; Buness 1990; Egger *et al.*, 1988	EHB, ISC	6	0
EUROBRIDGE	EUROBRIDGE Seismic Working Group, 1999; Kvaerna *et al.*, 2000	REB	25	0
Kola calibration shots, Russia	Russia	Finland, Norway, REB	3	0
Negev calibration shot, Israel	GII	GII	1	0
POLONAISE97	Guterch *et al.*, 1999	REB	15	0
VRANCEA99	M. Popa, personal comm., 2001	GII; M. Popa, personal comm., 2001	9	0
Mining Events, Accidents			202	0–5
Abakan and Kuzbass mining blasts, Russia	Emanov *et al.*, 1999	REB	2	5
Ammunition explosion, Switzerland	Switzerland	Czech Republic, France, Germany, Spain, Switzerland, UK	1	0
Factory explosion, Thailand	REB	REB	1	5
Mine explosions, Fennoscandia	H. Israelsson, personal comm., 1999; Sweden	REB	176	2–5
Mine tremors, Poland	G. Gibowicz, personal comm., 2000	REB	13	1–2
Quarry blasts, Kola peninsula, Russia	NORSAR, 2000	NORSAR, 2000	5	1–2
Quarry blasts, Israel	GII	Israel	2	1–2
Solikamsk mine collapse, Ural, Russia	ISC	ISC	1	2
Teutschenthal salt mine collapse, Germany	Germany	Croatia, Finland, France, Germany, Hungary, Italy, Netherlands, Norway, REB, Slovenia, Switzerland, UK	1	2
Earthquakes			499	5–15
EHB	Engdahl *et al.*, 1998	EHB, GII, LLNL	432	5
IGN bulletin, Spain	Chan *et al.*, 2000	EHB, GII, Spain	9	5
Fennoscandia	Grant *et al.*, 1993	Grant *et al.*, 1993	10	5
Revda, Norway	NORSAR, 2000	NORSAR, 2000	3	5
Caucasus	Kirichenko *et al.*, 2001	Kirichenko *et al.*, 2001	6	5

(*continued*)

Table 1

Continued

Event Type	Event Origin Sources	Arrival Sources*	No. of Events†	GT
Eastern Pyrenees	A. Roca, personal comm., 2000	EHB, A. Roca, personal comm., 2000	2	5
Gilad and Golan, Israel	GII	GII, Israel	2	5
Lake Magadi, Kenya	Ibs-von Seht *et al.*, 2001	Ibs-von Seht *et al.*, 2001	2	5
Mid-ocean ridges (Carlsberg ridge, Mid-Atlantic ridge)	Pan *et al.*, 2000	EHB, GII, LLNL	29	15
Minagish, Kuwait	Bou-Rabee, 2000	Bou-Rabee, 2000	1	10
Siberia, Russia	Emanov *et al.*, 1999	REB	1	5
Umbria-Marche, Italy	Amato *et al.*, 1998	GII, Italy	1	5
Valentine day earthquake, Pakistan	Seeber and Armbruster, 1979	ISC	1	5
Event Clusters			587	5–10
Adana, Turkey	GII, Engdahl and Bergman, 2001; Engdahl *et al.*, 2002	HDC, GII	4 (16)	5–7
Annecy, France	Thouvenot *et al.*, 1998; H. Israelsson, personal comm., 2001	Croatia, Czech Republic, France, Germany, Hungary, JHD, Italy, Netherlands, REB, UK, Spain, Switzerland	1 (9)	5–7
Bhuj, India	Bodin *et al.*, 2001; Engdahl and Bergman, 2001; Engdahl *et al.*, 2002	HDC	6 (73)	5–7
Chamoli, India	Saikia *et al.*, 2001; Engdahl and Bergman, 2001; Engdahl *et al.*, 2002	HDC	8 (58)	5–7
Duzce, Turkey	GII; Engdahl and Bergman, 2001; Engdahl *et al.*, 2002	HDC, GII	3 (21)	5–7
Erzincan, Turkey	Fuenzalida *et al.*, 1997a; Engdahl and Bergman, 2001; Engdahl *et al.*, 2002	HDC, GII, LLNL	3 (6)	5–7
Garm, Tajikistan	G. Pavlis, personal comm., 2000; Engdahl and Bergman, 2001; Engdahl *et al.*, 2002	HDC	4 (22)	5–7
Gubal Island, Egypt	Hurukawa *et al.*, 2001; Engdahl and Bergman, 2001; Engdahl *et al.*, 2002	HDC; Hurukawa *et al.*, 2001	1 (25)	10
Gulf of Aden	Pan *et al.*, 2000; Engdahl and Bergman, 2001; Engdahl *et al.*, 2002	HDC, GII, LLNL	5 (50)	10
Gulf of Aqaba	GII; Engdahl and Bergman; 2001, Engdahl *et al.*, 2002	HDC, GII	1 (32)	5–7
Hoceima, Morocco	Gupta and Wagner, 2001; Engdahl and Bergman, 2001; Engdahl *et al.*, 2002	HDC, GII	3 (19)	5–7
Izmit, Turkey	GII, Engdahl and Bergman, 2001; Engdahl *et al.*, 2002	HDC, GII	5 (19)	5–7
Jiashi, China	Xu, 2000; Engdahl and Bergman, 2001; Engdahl *et al.*, 2002	HDC	1 (68)	10
Koyna, India	Gupta *et al.*, 2002; Engdahl and Bergman, 2001; Engdahl *et al.*, 2002	HDC	9 (11)	5–7
Krn, Slovenia	M. Zivcic, personal comm., 2000, H. Israelsson, personal comm., 2001	EHB, JHD, REB	2 (9)	5–7
Loja, Spain	Chan *et al.*, 2000; H. Israelsson, personal comm., 2001	EHB, JHD, Spain	4 (10)	5–7
Racha, Georgia	Fuenzalida *et al.*, 1997b; Engdahl and Bergman, 2001; Engdahl *et al.*, 2002	HDC, GII	5 (30)	5–7
Spitak, Armenia	Dorbath *et al.*, 1992; Engdahl and Bergman, 2001; Engdahl *et al.*, 2002	HDC, GII	2 (9)	10
Tabas, Iran	Berberian, 1982; Engdahl and Bergman, 2001; Engdahl *et al.*, 2002	HDC, LLNL	2 (31)	10

*EHB: Engdahl *et al.*, 1998; GII: Geophysical Institute of Israel; HDC: HypoCentroidal Decomposition; IDG: Institute for the Dynamics of the Geospheres, Moscow; ISC: International Seismological Centre; JHD: Joint Hypocentre Determination; LDEO: Lamont-Doherty Earth Observatory; LLNL: Lawrence Livermore National Laboratory; NEIC: National Earthquake Information Center; REB: Reviewed Event Bulletin of the Prototype International Data Center.

†The number in parentheses indicates the number of events promoted to GT level during the HDC/JHD analysis. All other events in the event clusters as well as single earthquakes were assigned a GT level based on the criteria of Bondár et al. (2003).

very consistent results with the local network solutions of a temporary deployment (Saikia *et al.*, 2001) and generated 66 more GT7 events.

Although the Spitak, Armenia, aftershocks of the December 1988 main shock (Dorbath *et al.*, 1992) represent an event cluster that is very well recorded at regional and teleseismic distances, the relative locations from the HDC analysis were inconsistent with the local network locations. Therefore we could only accept the two candidate Spitak reference events as only GT10. Nine promoted events were identified by the HDC analysis and also assigned to the GT10 category.

A temporary network was used to study a massive swarm in Jiashi beginning in 1997 (Xu, 2000). Only one event was well located by the local network, and this event was used as a master event to locate the remaining events. However, when performing HDC analysis, we found that the master event solutions are inconsistent with the relative location pattern obtained from HDC. We also found a large, ~3-sec origin-time discrepancy between the local network and the HDC solution for the reference event. Therefore we could not validate the event as GT5, but accepted it (along with a further 67 events) as GT10.

We obtained the local bulletins of the Hurghada Seismological Network (Hurukawa *et al.*, 2001), jointly operated by the International Institute of Seismology and Earthquake Engineering (IISEE; Japan) and the National Research Institute of Astronomy and Geophysics (NRIAG; Egypt) in the Southern Gulf of Suez. Although none of the events satisfied our GT5 criteria, HDC analysis confirmed that the best candidate event on 17 December 1996 could be accepted as GT10. HDC analysis has also produced further 25 GT10 events, mostly from the 1969 Shadwan aftershock sequence.

Calibration Shots and Refraction Profiles

The Geophysical Institute of Israel made an effort to collect all available station readings for the Dead Sea calibration explosions (Gitterman and Shapira, 2001). This included arrival data from the networks in Cyprus, Jordan, and Turkey. During the RELEMR meeting in 2001, we were provided with digital waveforms of the Dead Sea shots from the Saudi National Seismic Network.

Large shots from long refraction profiles may provide invaluable GT0 events. An effort was made to canvass the region for well-recorded explosions. Unfortunately, publications describing the seismic experiments often do not list the exact locations and origin times of the shots.

We obtained GT0 information via personal correspondence on the European Geotraverse (1983–1986) Northern and Southern segments; the CELEBRATION2000 shots in Central Europe; the Saudi Arabia refraction profile carried out in 1978; the VRANCEA99 experiment, Romania; the EUROBRIDGE95 shots in Eastern Europe, as well as on refraction profiles near the Spitsbergen islands. The ISC, PDE, and EHB bulletins were searched for arrival data.

Unfortunately, most of the shots were too small to be recorded beyond local distances, and in the case of the older profiles, even if the shots were sizeable (such as the shots in the Red Sea and offshore Tunisia) they were poorly recorded due to the lack of regional station coverage. However, we included the most recent GT0 shots carried out in Europe (EUROBRIDGE95, POLONAISE97, VRANCEA99, and CELEBRATION2000) in the Reference Event List as they may be valuable sources for the location calibration of local networks in the region.

Mid-Ocean Ridge and Transform Fault Events

Pan *et al.* (2000) devised a method to identify ground truth event candidates at mid-oceanic ridges and transform faults. Depending on the source mechanism from Harvard Centroid Moment Tensor (CMT) solutions events are identified as transform fault (strike-slip) and ridge (normal fault) events, and associated with the closest corresponding feature (with the correct strike) in bathymetry maps. The events are then relocated with the constraint that they must occur along the associated ridge or transform fault. Finally, event clusters are built around these master events, and a JHD analysis is carried out to improve locations of smaller events that do not have CMT solutions. The technique was successfully applied on the Mid-Atlantic ridge, the Carlsberg ridge, and in the Gulf of Aden. However, because of the resolution of the bathymetry maps, and the occasional violation of the assumption that events occur on faults and ridges, the method can produce no better than GT10 candidates, and most of them were accepted as only GT15 events. An HDC analysis of the Gulf of Aden events validated 5 candidate events as GT10, and promoted a further 50 events to GT10 status.

GT5 Selection from the EHB Bulletin

Myers and Schultz (2001) assessed the location accuracies of events in the EHB bulletin as GT15 at the 95% confidence level (excluding subduction zone events) for events with azimuthal gap less than 90°. Therefore the EHB bulletin serves as a good starting point for identifying candidate GT5 events.

Multimax performed local network relocations of 596 EHB GT5 candidate events covering 40 geographic regions in the consortium's area of interest (Gupta and Wagner, 2001). Events were chosen from a prioritized list of over 6000 potential reference event candidates on the basis of geographic distribution, magnitude, number of stations and azimuthal coverage. In general, we tried to achieve uniform geographic coverage by selecting the largest, best-recorded events with depths less than 40 km. In the relocations we used only stations within 300 km from the epicenter thus simulating local network solutions, albeit we used the IASPEI travel times for lack of local velocity models.

Relocations were carried out uniformly using first arriving *P* phases only, as well as using secondary phases in order to identify possible phase misidentifications. In the latter case we used *Pn*, *Pb*, *Pg*, *Sn*, *Sb*, *Sg*, and *Lg* phases recorded at over 700 stations. In the review process, most phase types remained unchanged, but many were renamed for consistency with epicentral distance or if the time residual indicated that a change may improve the hypocentral location. For example, at very close distances, *Pn* or *Pb* phases were often changed to *Pg*.

Finally, from the set of the relocations using first arriving *P* phases only, we accepted the events that satisfied our GT5 selection criteria. This resulted in 432 GT5 events from the EHB bulletin.

Merged Data Sets

Our objective was to create a comprehensive bulletin for each individual event in the Reference Event List. Merging bulletins of local, regional, and teleseismic networks have several advantages, but in doing so one has to be aware of some possible pitfalls. Running the GT5 selection criteria on bulletins of dense local networks or temporary deployments identified most of the candidate GT5 events. These bulletins typically do not contain stations from regional and teleseismic distances. To facilitate event cluster analysis, which uses regional and teleseismic data to validate GT5 candidates, we searched global bulletins, such as the ISC, EHB, and REB, to add arrival data from stations beyond the local distance range. We disregarded events that were not recorded at regional distances.

Some local networks do not report to international agencies; therefore, the station codes used by the network operators may be in conflict with registered station codes at the NEIC. To resolve this conflict we assigned unique station codes to those stations that reused already registered codes. For registered stations, possible conflicts between station coordinates maintained by the NEIC and those reported by the local networks were resolved by personal communications with the network operators.

Often various agencies operate networks in the same region. A typical example is the Caucasus region, where at least eight different local networks coexist. As a study, performed by consortium member Western Services in cooperation with the Geophysical Survey of the Russian Academy of Sciences (Kirichenko *et al.*, 2001; Gabsatarova *et al.*, 2002), pointed out, none of these local networks could meet the GT5 selection criteria, but combining the various bulletins made it possible to identify GT5 events.

Another example for merging arrival data is the Peaceful Nuclear Explosions (PNEs) conducted by the former USSR. We used several data sources for the PNE data. The first data set was from the Institute for the Dynamics of the Geospheres (IDG); these readings were obtained and made available by consortium member Harvard University. It consisted of 2665 phases from 168 stations for 83 PNEs, with epicentral distances from 1.6 to 150°. The second data source was the EHB bulletin with arrivals from 72 PNEs recorded at 1046 stations. The IDG and EHB readings were merged with quality-controlled BRVK (Borovoye, Russia) readings from the Lamont–Doherty Earth Observatory (Kim *et al.*, 2001). The combined data set included 14,515 *P* and *S* arrivals from 1208 stations for 83 PNEs. Murphy *et al.* (2002) made available quality-controlled *Pn* arrivals for 72 PNEs, with 814 *Pn* arrivals from 73 former USSR stations with epicentral distances between 1.5 and 22°. They consistently reviewed and removed outliers from this data set. We combined the PNE data sets with priority given to the Murphy *et al.* (2002) picks, followed by the Kim *et al.* (2001) readings, and the Harvard data set. The final merged PNE bulletins represent the most comprehensive PNE bulletins to date.

Summary and Discussion

The seismic velocity models that motivated this effort will ultimately be replaced by better models, but this testing and validation data set will help develop and/or test future models. Multiple-event location techniques, such as JHD and HDC analyses, were not only employed to validate candidate GT5 events, but also to provide empirical station path corrections relative to the IASPEI travel-time tables, together with their corresponding uncertainties. The empirical path corrections are derived from a number of observations along repeated ray paths between a station and the cluster centroid, thus representing travel-time bias due to lateral heterogeneities along the ray path unaccounted for by the underlying velocity model. Three-dimensional model predictions should correlate well with the empirical path corrections. Path corrections have proven to be an invaluable asset to validate models (Ritzwoller *et al.*, 2003).

Bergman and Engdahl (2001) give a detailed discussion on the event clusters processed by CUB using the HDC approach. Figure 8 shows the ray coverage provided by empirical path corrections in the regional distance range, obtained from JHD and HDC analyses of event clusters. JHD clusters are located in the Mediterranean and Europe, and HDC clusters are mostly located in the Middle East and Asia. HDC analysis was also performed for events at known test sites, such as Balapan, Degelen, Lop Nor, Novaya Zemlya, and the historical French test site in Algeria, in order to obtain empirical path corrections. Note that many of the JHD clusters are constrained by only GT25 events. These JHD clusters were specifically formed to generate path corrections in the Mediterranean. Although the location accuracy of the centroid is somewhat poorer quality, the pattern of path corrections can still be relevant in validating model predictions.

Figure 9 shows the distribution of path corrections and their standard deviations as a function of epicentral distance. Although the corrections are not brought to a common origin time baseline, the trend indicated by the medians (solid lines) shows that path corrections approach zero-mean only beyond 40° epicentral distances. The median standard deviation steadily increases at regional distances, reaching the maximum around 18°, it then decreases and finally stabilizes at around 0.3 sec (the typical reading error for WWSSN seismograms) at teleseismic distances. Since path corrections represent the path effects due to lateral heterogeneities, the

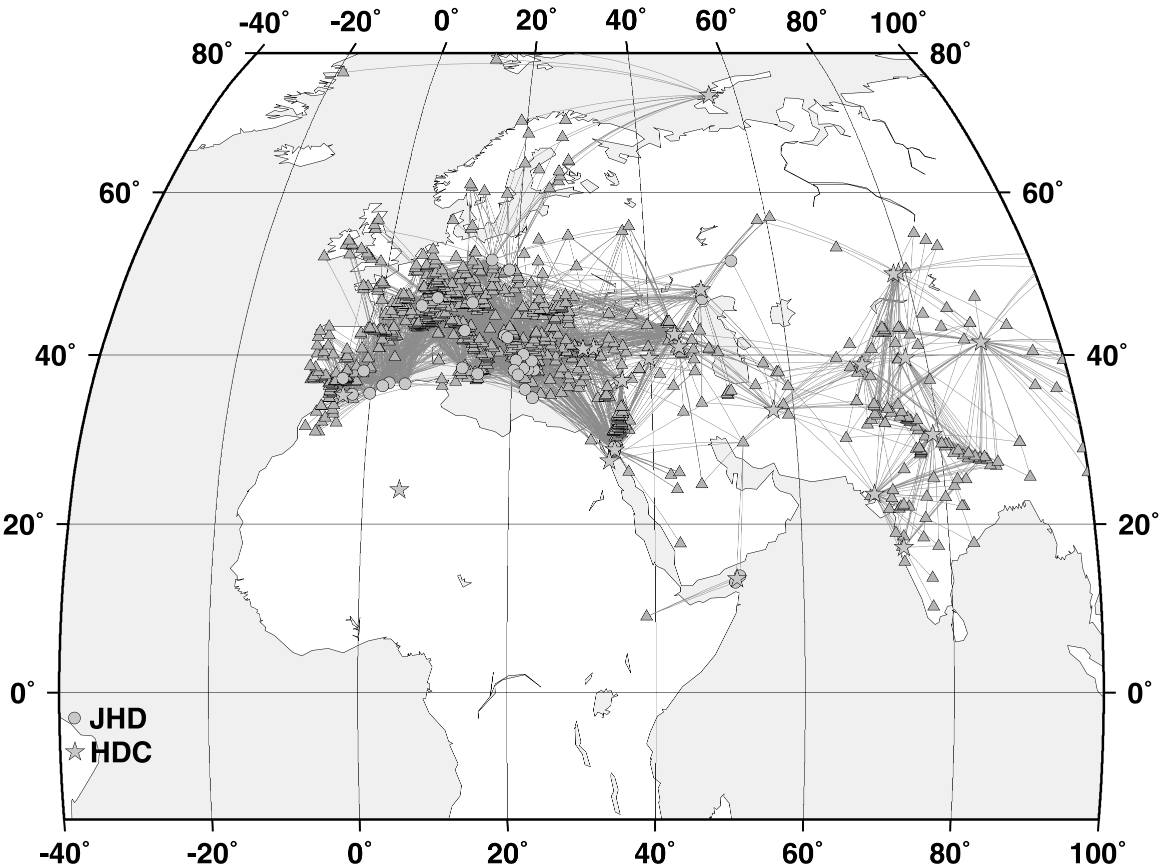

Figure 8. Regional ray coverage provided by the 4014 regional empirical *Pn* path corrections from HDC and JHD analysis of event clusters.

median standard deviation curve can be considered as a general approximation of the combined model and measurement (picking) errors.

The Group 2 Consortium reference event data collection effort demonstrated the power of a focused team approach to collecting data from a wide variety of sources, developing and performing quality control, and merging the data into a test and validation database. The effort was motivated by the need for well-recorded GT5 or better events distributed over a wide area to verify that calibrated travel times do indeed decrease mislocation and reduce uncertainty. In the course of the 3-year effort, consortium members and associated colleagues have developed new criteria for selecting candidate GT events from local, regional, and teleseismic bulletins, and they have demonstrated that some criteria previously used for these purposes were probably inadequate (Bondár *et al.* 2003) to ensure high confidence of a GTX origin. Cluster analysis (both JHD and HDC) was used in new novel ways to both test and validate candidate GT5 or better events, to promote new GT7 or better events, and to reject both origin and arrival outliers from the database.

The provenance of each origin is documented in the database along with the criteria stated for their selection. Reference events are most useful if they are well recorded at regional and teleseismic distances. Many networks do not report arrivals or may fail to measure arrivals from earthquakes outside their networks. For example, explosions in the Mediterranean as part of the European Geotraverse were not systematically read at regional distances. Also, some stations are not closely affiliated with local networks and therefore are not systematically read for regional or teleseismic arrivals. In order to provide regional phases from GT events, adjoining network bulletins were collected and fused or waveforms were recovered from selected stations. Despite these efforts, many candidate reference events failed to achieve GT5 status, or regional arrivals could not be detected to make them useful for testing and validation efforts. On paper several countries are operating extensive networks. Unfortunately, several networks have suffered from lack of maintenance or political strife, and some countries are just beginning to recover from a hiatus in seismological monitoring. In some Mediterranean countries, seismic phase data is considered a military secret. Political unrest in the region has shut down at least one network that on paper should have had sufficiently dense coverage and seismicity to produce valuable GT5 events. Seismology has always relied on international cooperation. Efforts such as the RELEMR conferences have been an important conduit for establishing the

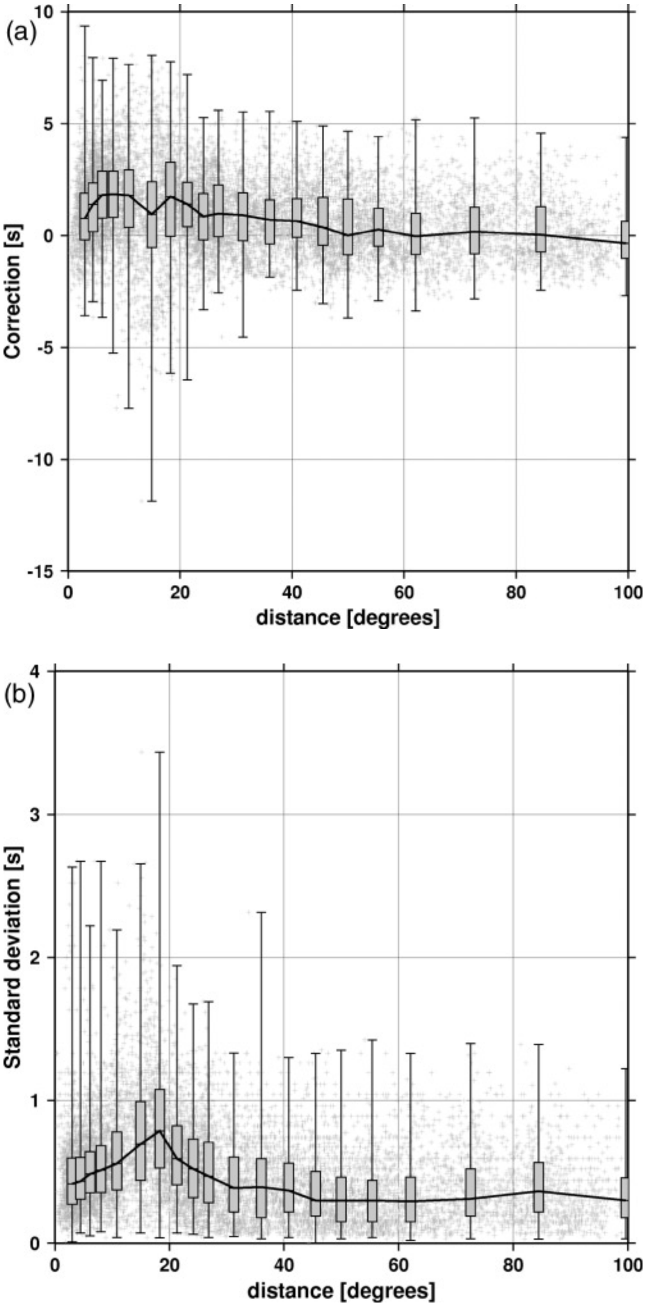

Figure 9. Empirical path corrections (dots) as a function of epicentral distance (a) and their standard deviations versus epicentral distance (b). The solid line connects the medians of every 5 percentile points of data, the bars indicate the range, and the boxes indicate the 25% and 75% quartiles.

all-important personal relationships between seismologists that make data exchange possible.

Several times, seemingly precise origins (location and time) from local networks were found to be inconsistent with either other local networks or with precise relative locations derived from cluster analysis of regional and teleseismic data. These events were either rejected or downgraded. The high confidence required for reference events requires that additional tests must be applied to ensure accurate locations. Quality-control procedures revealed that origin times for some mining explosion data sets were reported to the incorrect minute. Merging data sets is time consuming and requires patience and attention to detail. Many networks duplicate station names. The same arrivals are sometimes duplicated with different station names. Arrivals may be read and reported at the same station by multiple agencies and reported using either the same or a different station name. Phase association and naming (*Pn* versus *P*, *Pg* versus *Pn*, etc.) strongly depends upon the underlying set of travel-time tables used by the analyst. Observers use inconsistent nomenclature for *P* and *Pn* from 10 to 20° and for *P*, *P**, *Pb*, *Pg*, *Sg*, *Sn*, and *Lg* at shorter distances. Although many institutions still use the Jeffreys-Bullen (JB) travel times, they do not publish their travel-time tables with their bulletins. While travel-time perturbations are second order for small changes in models, distances of triplications and crossover distances are not second order.

In our searches of the literature, we examined several studies of aftershock sequences. Most recent studies failed to tabulate the best aftershock locations, and few if any relocated the main-shock or even large aftershocks based on corrections derived from close-in aftershock sequences. These are important fundamental data and need to be tabulated. Furthermore, these valuable data need to be collected in larger data repositories.

Unfortunately, the seismological community lacks a proactive institutional solution for the systematic collection and categorization of quality GT locations. The International Seismological Centre and European—Mediterranean Seismological Centre are valuable repositories of multiple seismic locations, but they rely on voluntary submissions, and there are no accepted quality-control criteria to categorize the location accuracy of the locations at high confidence levels. The IASPEI Working Group on Reference Events (http://lemond.colorado.edu/~copgte/) is soliciting such information and beginning to recommend guidelines for quality control and categorization methods. This work needs encouragement and broad dissemination. Just as the international geodetic community developed a global network of benchmarks and observatories to measure the shape of the Earth, seismologists need to develop a global network of well-recorded benchmark reference events. Collection of reference event and GT data requires a higher level of recognition in the seismological community and a broader understanding that this very basic data make fundamental contributions to our science.

Acknowledgments

The Group 2 Consortium effort was supported by the Defense Threat Reduction Agency contract DTRA01-00-C-0013. We are greatly indebted to the institutions and individual researchers who provided us ground truth information on refraction profiles, calibration shots, local bulletins and sta-

tion readings, and helped to establish personal correspondence with researchers in the region. Without their voluntary contributions, this work could not have been succeeded.

References

Aichroth, B., S. Ye, J. Feddersen, M. Maistrello, and R. Pedone (1990). A compilation of data from the 1986 European Geotraverse Experiment (main line) from Genova to Kiel, Geophysical Institute, University of Karlsruhe, Open-File Report 90-1.

Albright, D., C. Gay, and F. Pabian (1998). New details emerge on Pakistan's nuclear test site, *Earth Observation Magazine* December 1998/ January 1999, 18–21.

Amato, A., A. Azzara, C. Chiarabba, G. B. Cimini, M. Cocco, M. di Bona, L. Margheriti, S. Mazza, F. Mele, G. Selvaggi, A. Basili, E. Boschi, F. Courboulex, A. Deschamps, S. Gaffet, G. Bittarelli, L. Chiaraluce, D. Piccinini, and M. Ripepe (1998). The 1997 Umbria–Marche, Italy, earthquake sequence: a first look at the main shocks and aftershocks, *Geophys. Res. Lett.* **25,** 2861–2864.

Antolik, M., Y. J. Gu, G. Ekström, and A. M. Dziewonski (2003). J362D28: a new joint model of compressional and shear velocity in the Earth's mantle, *Geophys. J. Int.* **153,** 443–466.

Armbruster, J., V. Burlacu, M. Fisk, V. Khalturin, W. Kim, I. Morozov, E. Morozova, P. Richards, D. Schaff, and F. Waldhauser (2002). Seismic location calibration for thirty International Monitoring System stations in Eastern Asia, in *Proc. 24th Seism. Res. Rev.*, Ponte Vedra Beach Florida, October 2002, 231–241.

Barker, B., M. Clark, P. Davis, M. Fisk, M. Hedlin, H. Israelsson, V. Khalturin, W.-Y. Kim, K. McLaughlin, C. Meade, J. Murphy, R. North, J. Orcutt, C. Powell, P. Richards, R. Stead, J. Stevens, F. Vernon, and T. Wallace (1998). Monitoring nuclear tests, *Science* **281,** 1967–1968.

Bassin, C., G. Laske, and G. Masters (2000). The current limits of resolution for surface wave tomography in North America, *EOS* **81,** F897.

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.

Bergman, E. A., and E. R. Engdahl (2001). Validation and generation of reference events by cluster analysis, *Seism. Res. Lett.* **72,** 3.

Bhattacharyya, J., B. Kohl, H. Israelsson, and B. Zuzolo (2002). Lop Nor underground nuclear explosions: summary of preferred locations and origin times, Center for Monitoring Research Technical Report, CMR-02/01.

Bocharov, V. S., S. A. Selentov, and V. N. Michailov (1989). Characteristics of 92 underground nuclear explosions at the Semipalatinsk Test Site, *Atomaya Energia* **87,** no. 3 (in Russian).

Bodin, P., S. Horton, A. Johnston, G. Patterson, J. Bollwerk, P. Rydelek, G. Steiner, C. McGoldrick, K. P. Budhbhatti, R. Shah, and N. Macwan (2001). Aftershocks of the India Republic Day earthquake: the MAEC/ISTAR temporary seismograph network, *EOS* **82,** S22a-07, 2001.

Bolt, B. A. (1976). *Nuclear Explosions and Earthquakes: The Parted Veil*, W. H. Freeman, New York.

Bondár, I., S. C. Myers, E. R. Engdahl, and E. A. Bergman (2003). Epicenter accuracy based on seismic network criteria, *Geophys. J. Int.* **156,** 483–496, doi 10.1111/j.1365-246X.2004.02070.x.

Bondár, I., X. Yang, R. North, and C. Romney (2001). Location calibration data for CTBT monitoring at the Prototype International Data Center, *Pure Appl. Geophys.* **158,** 19–34.

Bou-Rabee, F. (2000). Seismotectonics and earthquake activity in Kuwait, *J. Seism.* **4,** 133–141.

Buness, H. (1990). A compilation of data from the 1983 European Geotraverse Experiment from the Ligurian Sea to the Southern Alps, Institute of Geophysics, Free University Berlin.

Chan, W., W. Rivers, I. Gupta, and R. Wagner (2000). Selection of reference events in Spain from IGN bulletins, Multimax Report, Largo, Maryland.

Dewey, J. (1972). Seismicity and tectonics in western Venezuela, *Bull. Seism. Soc. Am.* **62,** 1711–1751.

Dewey, J. W., and J. O. Kork (2000). Selection of earthquakes for calibration events in the United States, in *2nd Workshop on Location Calibration*, Oslo, Norway.

Dorbath, L., C. Dorbath, L. Rivera, A. Fuenzalida, A. Cisternas, R. Tatevossian, J. Aptekman, and S. Arefiev (1992). Geometry, segmentation and stress regime of the Spitak (Armenia) earthquake from the analysis of the aftershock sequence, *Geophys. J. Int.* **108,** 309–328.

Douglas, A. (1967). Joint epicentre determination, *Nature* **251,** 47–48.

Douglas, A., P. D. Marshall, and K. H. Jones (1993). Body-wave magnitude and locations of explosions at the Chinese test site: 1967–1989, Atomic Weapons Establishment, AWE Report No. O 12/93.

Duclaux, F., and L. Michaud (1970). Conditions experimentales des tirs nucleaires souterrains Français au Sahara, 1961–1966, *C. R. Acad. Sci. Paris* **270,** Series B, 189–192.

Egger, A., M. Demartin, J. Ansorge, E. Banda, and M. Maistrello (1988). The gross structure of the crust under Corsica and Sardinia, *Tectonophysics* **150,** 363–389.

Emanov, A. F., A. G. Filina, V. I. Khalturin, W.-Y. Kim, and P. G. Richards (1999). Earthquakes and large mining blasts in Southwestern Siberia, Russia, in *1st Workshop on IMS Location Calibration*, Oslo, Norway, 12–14 January 1999.

Engdahl, E., E. Bergman, M. Ritzwoller, N. Shapiro, and A. Levshin (2002). A reference event data set for validating 3-D models, in *Proc. 24th Seism. Res. Rev.*, Ponte Vedra Beach, Florida, October 2002.

Engdahl, E. R., and E. A. Bergman (2001). Validation and generation of reference events by cluster analysis, in *Proc. 23rd Seism. Res. Rev.*, Jackson Hole, Wyoming, 2–5 October 2001, 205–214.

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.

EUROBRIDGE Seismic Working Group (1999). Seismic velocity structure across the Fennoscandia–Sarmatia suture of the East European craton beneath the EUROBRIDGE profile through Lithuania and Belarus, *Tectonophysics* **314,** 193–217.

Fisk, M. D. (2002). Accurate locations of nuclear explosions at the Lop Nor test site using cross-correlated seismograms and Ikonos satellite imagery, *Bull. Seism. Soc. Am.* **92,** 2911–2925.

Fuenzalida, H., L. Dorbath, H. Ediyogan, A. Barka, L. Rivera, H. Haessler, H. Philip, and N. Lyberis (1997a). Mechanism of the 1992 Erzincan earthquake and its aftershocks, tectonics of the Erzincan basin and decoupling on the North Anatolian fault, *Geophys. J. Int.* **129,** 1–28.

Fuenzalida, H., L. Rivera, H. Haessler, D. Legrand, H. Philip, L. Dorbath, D. McCormack, S. Arefiev, C. Langer, and A. Cisternas (1997b). 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,** 29–64.

Gabsatarova, I. P., O. Y. Starovoit, Y. A. Kraev, and V. V. Kirichenko (2002). Selection of candidate reference events in the Northern Caucasus region, in *4th Workshop on IMS Location Calibration*, 22–26 April 2002, Oslo, Norway.

Gitterman, Y., and A. Shapira (2001). Dead Sea seismic calibration experiment contributes to CTBT monitoring, *Seism. Res. Lett.* **72,** 159–170.

Grant, L., J. Coyne, and F. Ryall (1993). CSS ground-truth database: version 1 handbook, CSS Technical Report C93-05.

Gupta, H. K., P. Mandal, and R. K. Rastogi (2002). How long will triggered earthquakes at Koyna, India continue?, *Curr. Sci. India* **82,** 202–210.

Gupta, I. N., and R. A. Wagner (2001). Local network relocation of selected events from the EHBGT5 database, Multimax Report, 21 June 2001, Largo, Maryland.

Gupta, V. (1995). Locating nuclear explosions at the Chinese test site near Lop Nor, *Sci. Glob. Sec.* **5,** 205–244.

Gupta, V., and F. Pabian (1996). Investigating the allegations of Indian nuclear test preparations in the Rajasthan desert, *Sci. Glob. Sec.* **6,** 101–189.

Gupta, V., and D. Rich (1996). Locating the denotation point of China first nuclear explosive test on 16 October 1964, *Int. J. Remote Sens.* **17,** 1969–1974.

Guterch, A., M. Grad, and A. Keller (2001). Seismologists celebrate the new millenium with an experiment in Central Europe, *EOS* **82,** 45.

Guterch, A., M. Grad, H. Thybo, G. R. Keller, and The POLONAISE Working Group (1999). POLONAISE97: an international seismic experiment between Precambrian and Variscan Europe in Poland, *Tectonophysics* **314,** 101–121.

Hurukawa, N., N. Seto, H. Inoue, K. Nishigami, I. Marzouk, A. Megahed, E. M. Ibrahim, H. Murukami, M. Nakamura, T. Haneda, S. Sugiyama, T. Ohkura, Y. Fujii, H. M. Hussein, A. S. Megahed, H. F. Mohammed, R. Abdel-Fattah, M. Mizoue, S. Hashimoto, and D. Suetsugu (2001). Seismological observations in and around the southern part of the Gulf of Suez, Egypt, *Bull. Seism. Soc. Am.* **91,** 708–717.

Ibs-von Seht, M., S. Blumenstein, R. Wagner, D. Hollnack, and J. Wohlenberg (2001). Seismicity, seismotectonics and crustal structure of the southern Kenya Rift: new data from the Lake Magadi area, *Geophys. J. Int.* **146,** 439–453.

Isaaks, E. H., and R. M. Srivastava (1999). *Applied Geostatistics*, Oxford University Press.

Jih, R.-S., and R. A. Wagner (2000). Location calibration of Kazakhstan based on DTRA (US)–NNC (RoK) joint experiment (1997–1999) in the former Semipalatinsk test site, in *Proc. 22nd Seism. Res. Symp.,* New Orleans, Louisiana, 12–15 September 2000.

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

Kennett, B., and E. R. Engdahl (1991). Travel times for global earthquake location and phase identification, *Geophys. J. Int.* **105,** 429–465.

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

Khristoforov, B. (1996). About the control of the underwater and above water nuclear explosions by hydroacoustic methods, Institute for Dynamics of Geosphere, Russian Academy of Sciences Report, SPC-95-4049, Moscow, October 1996.

Kim, W.-Y., P. G. Richards, V. Adushkin, and V. Ovtchinnikov (2001). Borovoye digital seismogram archive for underground nuclear tests during 1966–1996, Lamont-Doherty Earth Observatory Report, April 2001.

Kirichenko, V. V., Y. A. Kraev, I. P. Gabsatarova, and O. Starovoit (2001). Selection of candidate reference events in the Northern Caucasus region, Western Services Corp. and Geophysical Survey of the Russian Academy of Sciences Report, September 2001.

Kvaerna, T., L. Taylor, and J. Schweitzer (2000). The Eurobridge profile: ground truth observations at the Fennoscandian arrays, in *2nd Workshop on IMS Location Calibration*, Oslo, Norway, 20–24 March 2000.

Lilwall, R. C., and J. Farthing (1990). Joint epicenter determination of Soviet underground nuclear explosions 1973–89 at the Semipalatinsk Test Site, Atomic Weapons Establishment, AWE Report No. O 12/90.

Marshall, P. D., D. Porter, J. B. Young, and P. A. Peachell (1994). Analysis of short period seismograms from explosions at the Novaya Zemlya test site in Russia, Atomic Weapons Establishment, AWE Report No. O 2/94.

Martínez-Díaz, J. J., A. Rigo, L. Louis, R. Capote, E, Hernández, E. Carreño, and M. Tsige (2002). Interferometría diferencial aplicada al caracterización símotectónica del terremoto del Mula (febrero de 1999, mb: 5.0), *Boletín Geológico y Minero* **113,** 23–33.

McLaughlin, K., X. Yang, I. Bondár, J. Bhattacharyya, H. Israelsson, B. North, V. Kirichenko, Y. Kraev, R. Engdahl, M. Ritzwoller, A. Levshin, N. Shapiro, M. Barmin, M. Antolik, A. Dziewonski, G. Ekström, I. Gupta, R. Wagner, W. Chan, W. Rivers, A. Hofstetter, A. Shapira, G. Laske, and N. Cotte (2001). Seismic location calibration in the Mediterranean, North Africa, Middle East and Western Eurasia, in *23rd CTBT Seism. Res. Symp.*, Jackson Hole, Wyoming, 1–5 October 2001.

McLaughlin, K., I. Bondár, X. Yang, J. Bhattacharyya, H. Israelsson, B. North, V. Kirichenko, R. Engdahl, M. Ritzwoller, A. Levshin, N. Shapiro, E. Bergman, M. Antolik, A. Dziewonski, G. Ekström, H. Ghalib, I. Gupta, R. Wagner, W. Chan, W. Rivers, A. Hofstetter, A. Shapira, and G. Laske (2002). Seismic location calibration in the Mediterranean, North Africa, Middle East and Western Eurasia, in *Proc. 24th Seism. Res. Rev.*, Ponte Vedra Beach, Florida, October 2002, 340–350.

Murphy, J., and J. Jenab (1992). Development of a comprehensive seismic yield estimation system for underground nuclear explosions, Maxwell Technical Report, PL-TR-92-2076, SSS-TR-92-13129.

Murphy, J., W. Rodi, M. Johnson, J. Sultanov, T. Bennettt, M. Toksöz, C. Vincent, V. Ovtchinnikow, B. Barker, A. Rosca, and Y. Shchukin (2002). Seismic calibration of Group 1 International Monitoring System (IMS) stations in Eastern Asia for improved event location, in *Proc. 24th Seism. Res. Rev.*, Ponte Vedra Beach, Florida, October 2002, 340–350.

Myers, S. C., and C. A. Schultz (2000a). Calibration of seismic travel time using events with seismically determined locations and origin times, *EOS Trans. AGU* **81,** F845.

Myers, S. C., and C. A. Schultz (2000b). Improving sparse network seismic location with Bayesian kriging and teleseismically constrained calibration events, *Bull. Seism. Soc. Am.* **90,** 199–211.

Myers, S. C., and C. A. Shultz (2001). Statistical characterization of reference event accuracy, *Seism. Res. Lett.* **72,** 244.

NORSAR (2000). Seismic calibration of the European Arctic ground truth events, Special Technical Report No. 5, NORSAR, June 2000.

Pan, J., M. Antolik, and A. Dziewonski (2000). Locations of mid-oceanic earthquakes constrained by sea-floor bathymetry, *EOS* **81,** F868.

Richards, P. G. (2000). Accurate estimates of the absolute location of underground nuclear tests at the northern Novaya Zemlya Test Site, in *2nd Workshop on IMS Location Calibration*, Oslo, Norway, 20–24 March 2000.

Ritzwoller, M. H., N. M. Shapiro, E. A. Levshin, E. A. Bergman, and E. R. Engdahl (2003). Ability of a global three-dimensional model to locate regional events, *J. Geophys. Res.* **108,** no. B7, 2353, doi 10.1029/2002JB002167.

Saikia, C. K., G. Ichinose, and C. Ji (2001). Seismic event location strategy and path calibration in and around Indian subcontinent, in *Proc. 23rd Seism. Res. Rev.*, Jackson Hole, Wyoming, October 2001, 343–354.

Saikia, C. K., R. Lohman, G. Ichinose, D. V. Helmberger, M. Simons, and P. Rosen (2002). Ground truth locations—a synergy of seismic and synthetic aperture radar interferometric methods, in *Proc. 24th Seism. Res. Rev.*, Ponte Vedra Beach, Florida, October 2002.

Seeber, L., and J. Armbruster (1979). Seismicity of the Hazara Arc in northern Pakistan: decollement vs. basement faulting, in *Geodynamics of Pakistan*, A. Farah and K. A. DeJong (Editors), Geological Survey of Pakistan, Karachi, 131–142.

Shapiro, N. M., and M. H. Ritzwoller (2002). Monte Carlo inversion for a global shear velocity model of the crust and upper mantle, *Geophys. J. Int.* **151,** 88–195.

Shapiro, N. M., and M. H. Ritzwoller (2004). Thermodynamic constraints on seismic inversions, *Geophys. J. Int.* **157,** 1175–1188, doi 10.1111/j1365-246X.2004.02254.x.

Skov, M., B. Kohl, and R. L. Woodward (2002). Imagery assets for supporting nuclear explosion monitoring research and development, in *Proc. 24th Seism. Res. Rev.*, Ponte Vedra Beach, Florida, October 2002.

Sultanov, D. D., J. R. Murphy, and Kh. D. Rubinstein (1999). A seismic source summary for Soviet peaceful nuclear explosions, *Bull. Seism. Soc. Am.* **89,** 640–647.

Sweeney, J. J. (1998). Criteria for selecting accurate event locations from NEIC and ISC Bulletins, Lawrence Livermore National Laboratory, UCRL-JC-130655.

Thouvenot, F., J. Frechet, P. Tapponier, J.-C. Thomas, B. Le Brun, G. Menard, R. Lacassin, L. Jenatton, J.-R. Grasso, O. Coutant, A. Paul, and D. Hatzfeld (1998). The ML 5.3 Epagny (French Alps) earthquake of 1996 July 15: a long-awaited event on the Vuache Fault, *Geophys. J. Int.* **135,** 876–892.

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,** 1353–1368.

Xu, S. (2000). A review of information on seismic event location in China, in *2nd Workshop on IMS Location Calibration*, Oslo, Norway, 20–24 March 2000.

Yang, X., I. Bondár, and C. Romney (2000a). PIDC ground truth (GT) database (revision 1), Center for Monitoring Research Technical Report, CMR-00/15.

Yang, X., I. Bondár, and K. McLaughlin (2000b). PIDC reference event database (REDB), Center for Monitoring Research Technical Report, CMR-00/17.

Yang, X., I. Bondár, K. McLaughlin, and R. North (2001). Source specific station corrections for regional phases at Fennoscandian stations, *Pure Appl. Geophys.* **158,** 35–57.

Yang, X., R. North, C. Romney, and P. Richards (2003). Worldwide nuclear explosions, in *International Handbook of Earthquake and Engineering Seismology*, Vol. 81B, W. H. Lee, H. Kanamori, P. Jennings, and C. Kisslinger (Editors), Academic Press, London.

Yang, X., I. Bondár, J. Bhattacharyya, M. Ritzwoller, N. Shapiro, M. Antolik, G. Ekström, H. Israelsson, and K. McLaughlin (2004). Validation of regional and teleseismic travel-time models by relocating ground truth events, *Bull. Seism. Soc. Am.* **94,** 897–919.

Science Applications International Corporation
M/S 2-1 1953 Gallows Road, Suite 260
Vienna, Virginia 22182
(I.B., X.Y., H.I., K.M., V.K.)

University of Colorado at Boulder
Department of Physics, Campus Box 390
Boulder, Colorado 80309
(E.R.E.)

Multimax, Inc.
1141 McCormick Drive
Largo, Maryland 20774
(H.A.A.G., R.W., I.G.)

Geophysical Institute of Israel
P.O.B. 182, LOD 71100, Israel
(A.H.)

Harvard University
Department of Earth and Planetary Sciences
20 Oxford Street
Cambridge, Massachusetts 02138
(G.E.)

Global Seismological Services
601 16th Street #C390
Golden, Colorado 80401
(E.B.)

Manuscript received 23 June 2003.