

Validation of Regional and Teleseismic Travel-Time Models by Relocating Ground-Truth Events

by Xiaoping Yang, István Bondár, Joydeep Bhattacharyya, Michael Ritzwoller, Nikolai Shapiro, Michael Antolik, Göran Ekström, Hans Israelsson, and Keith McLaughlin

Abstract Recently developed 3D global seismic velocity models have demonstrated location improvements through independent regional and teleseismic travel-time calibration. Concurrently, a large set of high-quality ground-truth (GT) events with location accuracies 10 km or better (GT0–GT10) has been collected for Europe, the Mediterranean, North Africa, the Middle East, and western Eurasia. In this study we demonstrate event location improvements using this new data set by applying the regional and teleseismic model-based travel-time calibrations (independently and jointly). Besides relocating events using all arrivals, a subset of the GT events was also relocated using controlled station geometries generated from a constrained bootstrapping technique. This approach simulates sparse networks and reduces the effect of correlated errors to ensure valid 90% error ellipse coverage statistics. With respect to the GT events, we compared event relocations, with and without travel-time calibrations, considering statistics of mislocation, error ellipse area, 90% coverage, origin time bias, origin time errors, and misfit. Relocations of over 1200 GT events show that Pn and/or P calibration reduced mislocation for 60%–70% of the events. Joint regional Pn and teleseismic P travel-time calibration provided the largest location improvements and achieved approximately GT5 accuracy levels. Due to correlated errors, event locations using large numbers of stations have deficient 90% error ellipse coverage. However, the coverages derived from the model errors are appropriate for the simulated sparse regional and teleseismic networks. Our validation effort demonstrates that the global model-based calibrations of Pn and teleseismic P travel times reduce both location bias and uncertainty over wide areas.

Introduction

In recent years, significant progress has been made calibrating seismic event location for the sparse seismic network of the International Monitoring System (IMS). The goal of travel-time calibration is to improve location accuracy and reduce the uncertainty while retaining true 90% coverage, that is, 90% of the true locations are found inside the respective error ellipses. To demonstrate these goals, reliable ground-truth (GT) events having high-confidence epicentral estimates and high-quality arrival times are required.

The IMS event-location process (Bratt and Bache, 1988) uses a hierarchy of travel-time corrections (Yang *et al.*, 2001a,b). The corrections include ellipticity, station elevation, optional station bulk corrections, and source-specific station corrections (SSSCs). SSSCs are travel-time corrections (calibrated travel times) relative to the baseline global 1D IASP91 travel-time tables (Kennett and Engdahl, 1991). They amount to 2D and 3D travel-time tables for a specific station as a function of source latitude, longitude, and depth. Associated modeling errors are also specified in the same

manner for each station to reflect the travel-time uncertainty of the underlying model. Another error, measurement error, is given for each observed arrival, and both errors are used to weight the arrival in the location inversion process. As the last step of the location procedure, 90% confidence error ellipses are estimated from the total *a priori* errors (modeling and measurement), assuming Gaussian distributed independent errors.

Both model-based and empirical approaches have been used to improve regional travel-time predictions. One-dimensional models were initially employed in developing the SSSCs for IMS stations in Fennoscandia and North America (Yang *et al.*, 2001a,b). Recently, tomographic inversion was applied to derive high-resolution 3D models in eastern and central Asia (Murphy *et al.*, 2002). GT data were kriged to develop empirical travel times for events in the Caucasus Mountains (Myers and Shultz, 2000a). In regions where both velocity models and GT data are available, a hybrid approach combining both model predictions and travel-time observa-

tions has been used in developing calibrated travel times for eastern Asia (Armbruster *et al.*, 2002). Three-dimensional seismic velocity models (the CUB1 and CUB2 models) with raytracing have most recently demonstrated significant improvements in locating events (Shapiro and Ritzwoller, 2002a,b; Ritzwoller *et al.*, 2003) with regional phases alone.

Previous efforts have focused primarily on calibrating regional P_n phases, since regional travel times are sensitive to small-scale structures and deviate from IASP91 travel-time tables more severely than teleseismic travel times. As a result of these regional calibration efforts, regional travel-time model errors are approaching the levels of teleseismic model errors. Because of the slowness differences between teleseismic and regional phases, mislocation due to a 1-sec bias in a teleseismic travel time is more severe than that caused by the same bias in a regional travel time. Teleseismic phases constitute an important fraction of IMS arrival data, motivating joint calibration of regional and teleseismic phases. Global 3D velocity models have been developed (the J362 model and its predecessors, e.g., the SP12 model) and have been demonstrated to improve event locations with teleseismic phases alone (Antolik *et al.*, 2001, 2003).

Model validation requires robust statistical analyses with a large amount of quality GT origins and arrival data. Various techniques have been used to assess model performance, including comparisons between predicted versus empirical path corrections (e.g., Ritzwoller *et al.*, 2003), cross-validation of root mean square (rms) residual reductions (e.g., Johnson and Vincent, 2002), and evaluations of location bias using randomly selected subsets of arrivals (e.g., Antolik *et al.*, 2001). Effective testing depends on the accuracy, amount, and geographic coverage of GT data. Locations can be severely affected by outliers, particularly for stations that fill large azimuthal gaps (e.g., Johnson and Vincent, 2002). Outlier analysis is therefore useful in providing more robust results (e.g., Ritzwoller *et al.*, 2003). Requirements on minimum azimuthal gap and the number of observations (referred to as the number of defining phases, n_{def}) help eliminate poor events that are unreliable for evaluating model-based location improvements (e.g., Ritzwoller *et al.*, 2003). However, the number and geographic coverage of high-quality GT events have been limited up to now. In particular, only data independent from model development should be used in validation testing. This requirement is particularly difficult to fulfill for the empirical approach of model development, such as kriging and body-wave tomography, in which the best travel-time data are often already included in the model construction. Typically a leave-one-event-out strategy is used in validating kriged calibrations. Since many events are clustered, more objective testing would require leave-one-path-out instead, which significantly reduces the test data set. In contrast, the model-based approach, particularly the surface-wave-based models such as CUB1 and CUB2 considered in this work, has the advantage of exploiting data independent of the body-wave arrivals.

The Group-2 Location Calibration Consortium recently collected a large high-quality set of reference GT events in Europe, Africa, and Eurasia (Bondár *et al.*, 2004b). The consortium database includes nearly 2000 GT0–GT10 origins with known or well-estimated location accuracy. It provides an excellent set of arrivals and empirical path corrections for model validation. This article describes event relocation tests utilizing this new data set to validate travel-time calibrations derived from two 3D global regional models and one 3D teleseismic velocity model. Comparisons between predicted and empirical path corrections are given in Ritzwoller *et al.* (2002) and Bhattacharyya *et al.* (2003).

Many GT events used in the previous relocation studies (e.g., Antolik *et al.*, 2001; Ritzwoller *et al.*, 2002) and in the consortium database have large numbers of observations. In addition, because of the good azimuthal coverage, these events are generally well located, mostly within 10-km uncertainties. Regardless of whether calibrated travel times are applied or not, the locations will not change dramatically. Furthermore, arrivals tend to be clustered on the focal sphere, and these clustered observations are not statistically independent. Therefore, tests that simply relocate such events with all available arrivals have poor statistical power to demonstrate improvements due to travel-time calibrations. This article presents a new approach to increase the statistical power by simulating sparse networks from these large data sets.

Calibrated travel times have the largest impact on sparsely recorded events. Monte Carlo techniques have been used previously to examine the statistics of sparsely recorded events using large events with many arrivals (e.g., Antolik *et al.*, 2001). In this work, we generated simulated sparse network bulletins (SSNBs) from well-recorded events by taking subsets of stations that satisfy constraints on the number of stations and the secondary azimuthal gap. The subsets of stations meeting these constraints thus simulate sparse networks (e.g., IMS) and yet provide reasonable azimuthal coverage to avoid extremely poor locations caused by degenerate network geometries. Bondár *et al.* (2004a) has recently argued that the size of the so-called secondary azimuthal gap (s_{gap}), that is, the largest azimuthal gap filled by a single station, is a good indicator of the sensitivity of the location to individual arrival-time outliers. Therefore, since our goal is to test the travel-time calibrations, it is desirable to design our tests so that they are less sensitive to measurement errors in the test data set.

In this study we demonstrate location improvements in the Group-2 consortium region of interest: Europe, the Mediterranean, North Africa, the Middle East, and western Eurasia. We validate regional and teleseismic calibrations (both independently and jointly) computed from 3D global velocity models: CUB1, CUB2, and J362. Testing is conducted by relocating a large data set of GT0–GT10 events with model-based regional and teleseismic travel-time calibrations. In addition to relocating GT events using all available

quality arrivals, we demonstrate a new procedure whereby we simulate sparse networks.

Three-Dimensional Global Regional and Teleseismic Models

Regional Models: CUB1 and CUB2

The global 3D models CUB1 and CUB2 were constructed using a Monte Carlo inversion method (Shapiro and Ritzwoller, 2002a,b) applied to group (Ritzwoller and Levshin, 1998) and phase velocity dispersion curves (Trampert and Woodhouse, 1995; Ekström and Dziewonski, 1998). Both models are given on a $2^\circ \times 2^\circ$ grid to a depth of 400 km. Below 400 km, both models revert to the global Harvard 3D model S20a (Ekström and Dziewonski, 1998). There are three principal differences between CUB1 and CUB2. First, the crustal reference for CUB1 is CRUST5.1 of Mooney *et al.* (1998), whereas CUB2 uses CRUST2.0 (Bassin *et al.*, 2000). Second, CUB1 was derived using Gaussian tomography that is based on geometrical ray theory with intuitive Gaussian smoothing constraints to simulate surface wave sensitivities (Barmin *et al.*, 2001), while CUB2 is based on diffraction tomography that uses a simplified version of the scattering sensitivity kernels that emerge from the Born or Rytov approximations (Ritzwoller *et al.*, 2002). Diffraction tomography accounts for path-length-dependent sensitivity, wavefront healing, and associated diffraction effects and provides a more accurate assessment of spatially variable resolution than traditional tomographic methods. Third, CUB1 uses a simple, empirical, logarithmic scaling relation $d\ln(V_p)/d\ln(V_s) \sim 0.5$, in which perturbations are taken relative to the S and P velocities (V_s and V_p , respectively) in ak135 (Kennett *et al.*, 1995). CUB2 uses a theoretical conversion based on mineralogical partial derivatives for a hypothetical composition of the upper mantle. The method is based on the work of Goes *et al.* (2000), as described in detail by Shapiro and Ritzwoller (2002a,b). In essence, given the mineralogical composition, the bases for the anelastic corrections are partial derivatives of the elastic moduli with respect to the independent variables at infinite frequency, a mixing law, and a relation between temperatures and shear Q . The V_s model is converted to temperature and then converted to V_p . In CUB2 this transformation has not yet been regionally tuned; mineralogical composition is assumed to be homogeneous across the region of study, there has been no account for the possible effects of fluids in the mantle beneath tectonically deformed regions, and shear Q is purely a function of temperature. Ritzwoller *et al.* (2003) showed event location improvements using regional calibrations derived from this 3D global model.

To calculate Pn calibrated travel-time tables (SSSCs) based on CUB1 or CUB2, we used a 2D raytracer that handles refracted and reflected compressional waves in a 3D laterally inhomogeneous medium along a 2D cross section of a spherical Earth. Regional travel times were computed

by shooting rays along profiles radiating from each station up to distances of 20° for a crustal source depth of 10 km. We chose an azimuthal spacing of 3° between the profiles and computed travel times at distance intervals of 25 km. Extensive tests verified that numerical errors due to model sampling and interpolation were bounded by less than 0.1 sec. After subtracting the predicted IASP91 travel times, the resulting travel-time corrections were interpolated to a $1^\circ \times 1^\circ$ rectangular geographic grid centered on each station. Figure 1 shows an example of Pn SSSCs (calibration correction surfaces) for station ABKT (Alibek, Turkmenistan).

We derived an empirical modeling error for the calibrated CUB Pn travel times, as shown in Figure 2, using travel-time misfits obtained from the events selected from the Engdahl–van der Hilst–Buland (EHB) bulletin (Engdahl *et al.*, 1998). Modeling errors were estimated from the standard deviations of the misfits as a function of epicentral distance. In lieu of more detailed empirical error maps or error surfaces produced by a robust theory of error propagation from uncertainties in the 3D model, we chose this simple and conservative approach to estimate model errors (azimuthally invariant). It is similar in form to the baseline error model currently used in routine IMS location calculations (Fig. 2). Typically, the model variances for the CUB models were taken to be about half of the corresponding values for baseline IASP91 travel-time tables. Note that the latter model errors are unrealistically small at distances within 1° .

Teleseismic Model: J362

Recently Antolik *et al.* (2003) have developed a global joint compressional and shear velocity 3D model of the Earth's mantle, J362. It is a spherical harmonic degree-18 model with a horizontal length scale of 1000 km based on absolute and differential body-wave travel times as well as surface wave dispersion measurements. Using a data set of GT0–G10 earthquakes and explosions, Antolik *et al.* (2003) have shown that J362 achieves about a 10% improvement in rms mislocation for explosions, relative to SP12. J362 also decreases the origin time error by an average of 0.05 sec over the SP12 model.

Teleseismic P -wave SSSCs were computed with a perturbation raytracer. Because the perturbation raytracer approach is not valid in regions of triplication or diffraction, we limited our study to travel-time corrections between 25° and 97° . We included crustal structure by combining J362 with the CRUST2.0 model of Bassin *et al.* (2000), assuming a source depth of 10 km. SSSCs were generated for the study region on a $2^\circ \times 2^\circ$ rectangular grid bounded by -15° N to 80° N and -40° E to 100° E, for a global set of over 2000 stations. Figure 1 also shows an example of teleseismic P SSSCs for ABKT. For modeling errors, based on our experience with the regional calibrations, we adopted a scaled version of the baseline IASP91 teleseismic P error model (see Fig. 2).

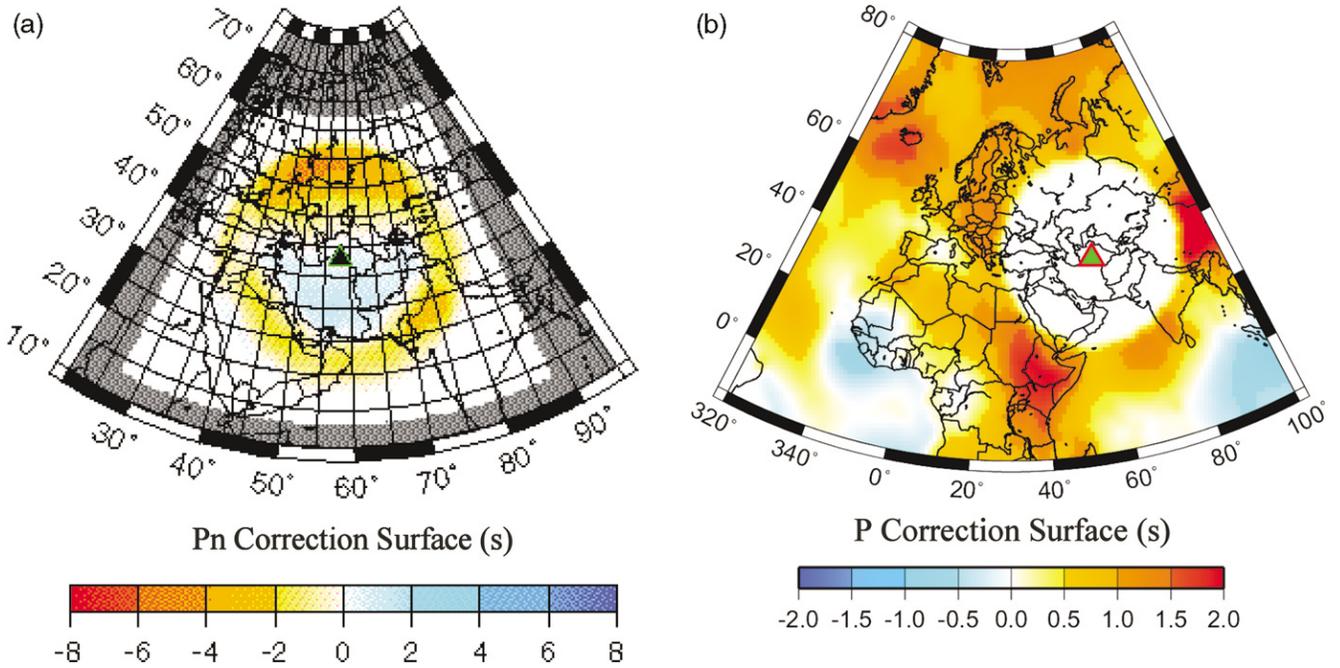


Figure 1. An example of travel-time calibration surfaces (SSSCs) for station ABKT (Alibek, Turkmenistan), calculated using raytracing through 3D global regional and teleseismic models, respectively. (a) Regional Pn SSSCs from CUB1. (b) Teleseismic P SSSCs from J362. Triangles show the station location.

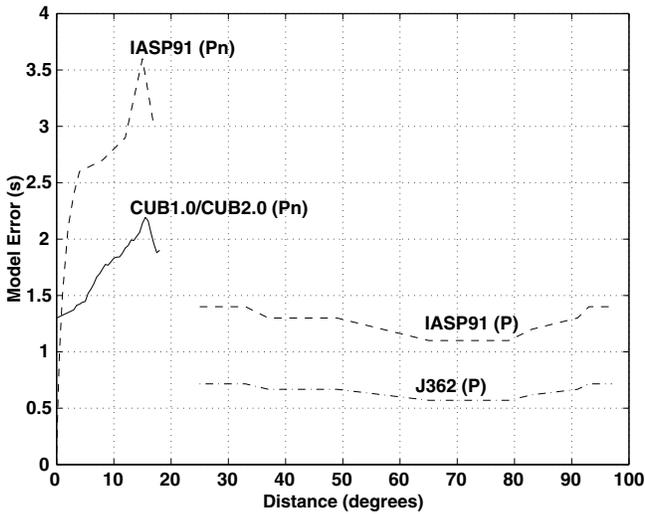


Figure 2. Model errors (standard deviations of travel-time misfits) for regional Pn and teleseismic P phases for the baseline IASP91 (dashed line) travel-time tables, CUB (solid line) Pn calibrations, and J362 (dot-dashed line) teleseismic P calibrations. The CUB and J362 model errors exhibit $\sim 50\%$ variance reduction compared to the baseline model errors. *A priori* model errors are used jointly with measurement (picking) errors to weigh the observations and to predict 90% error ellipses.

Baseline Differences between Regional and Teleseismic Travel Times

In this work, we use these three newly developed models and their raytracers to calibrate regional and teleseismic travel times, respectively. At this time, fully integrated 3D global models are not yet available with high-resolution upper mantle (like CUB1 or CUB2) suitable for combining self-consistent Pn and teleseismic P . Since the regional and teleseismic models were developed independently, it is possible that baseline differences exist between models. Therefore, before relocating events using both Pn and teleseismic P , we examined whether biases exist between the regional and teleseismic calibrations and whether any baseline shifts are warranted.

To assess the baseline of the models, we compared regional and teleseismic phases for travel-time residuals of over 700 GT events (mostly GT5 or better) in the Group-2 database using IASP91, CUB1, CUB2, and J362. We only included reasonable Pn and P arrivals (with absolute residuals less than 5 sec for all models) and reliable GT data (with at least five Pn and five P phases). For each GT event, residuals of Pn and P phases were calculated using observed and predicted travel times from each model for the GT locations and origin times. We then computed the mean and standard deviation of Pn and P time residuals for each GT event and examined the differences between the models. These differences can be interpreted as differences in the origin times estimated by each model for the fixed GT locations. Table 1 reveals that significant baseline differences

exist between several pairs of Pn and P travel-time tables, and the standard deviations reveal which pairs of models are poorly matched. In particular, IASP91 regional Pn and IASP91 teleseismic P phases show a statistically significant baseline offset of -0.23 sec with a standard deviation of 1.12 sec. The combination of CUB2 and J362 has the smallest baseline difference and standard deviation, but an adjustment needs to be made to CUB1 of about 0.75 sec to bring it into alignment with J362. It is noteworthy that combinations of CUB1 and CUB2 with J362 or IASP91 reduce the standard deviation compared to combinations of IASP91 Pn jointly with IASP91 teleseismic P . The results suggest that J362 teleseismic P calibrations jointly with IASP91 Pn may be a poor match.

A baseline correction can be applied to any of the models to reduce the mean differences. For our purposes, we chose to reduce the CUB1 baseline shift by simply applying a bulk correction of 0.75 sec to the travel-time tables. CUB2 requires no baseline shifts. The standard deviations reveal event-by-event baseline variations that cannot be remedied by simple bulk adjustments to the travel-time tables. We have examined the geographic patterns of these biases between models, and we found that they are systematic and reflect unmodeled differences between the broad tectonic provinces of shield and platform regions versus tectonically active regions. These systematic deviations suggest that future global models may benefit from regional tuning.

Methodologies and Data Sets for Validation Testing

We validated the three global models, CUB1, CUB2, and J362, by relocating a large set of GT events that were not involved in constructing the 3D models. High-quality GT events were selected to avoid ambiguity between model and data uncertainties. Events were relocated using all available stations as well as limited sets of stations to simulate sparse networks using regional (Pn -phase) and teleseismic (P -phase) calibrations. We examined the Pn and teleseismic calibrations both independently and jointly. Location improvements were evaluated by comparing the calibrated results with uncalibrated results (IASP91 baseline) from the same location algorithm using a set of well-defined statistical

metrics. In all our relocation tests, the depth was fixed to zero to simulate a nuclear monitoring scenario, as all events in our data set are either near the surface (explosions, mine collapses, etc.) or at shallow crustal depths (earthquakes). Furthermore, for these distance ranges (>150 km), the differences in the travel times for Pn and P between 0- and 10-km source depth are common across the network, and depth and origin time trade off directly.

Data Selection for Event Relocations

We selected all GT0–GT10 events in the Group-2 database that can be reliably relocated (1) using only regional phases, (2) using only teleseismic phases, and (3) using both regional and teleseismic phases. Each of the three data sets provides fairly good geographic coverage across the region. To minimize ambiguity in arrival data, we only used Pn arrivals within 15° and P arrivals between 25° and 97° . We selected events in each of the three groups with a secondary azimuthal gap (s_{gap}) less than 160° . The s_{gap} is the largest azimuthal gap that would result from deleting one station. This guarantees the azimuthal gap (az_{gap}) is less than 160° and guarantees a minimum of five stations. This selection criterion yields a data set whose event locations are less susceptible to single-station outliers (Bondár *et al.*, 2004a).

Figure 3 shows 526 GT0–GT10 events in the Pn -only data set. Most of the events are GT5 (87%). About 64% of the events resulted from the Group-2 GT5 selection criteria at the 95% confidence level. Another 26% of the events are earthquake clusters (mostly GT5) generated from cluster analyses (Engdahl and Bergman, 2001). The rest (10%) of the GT events are nuclear explosions, chemical explosions, or mining events (GT0–GT2).

Figure 4 shows 793 GT0–GT10 events in the P -only data set. A large number of the events are GT1 (40%) and GT5 (35%). About 70% of all events are in clusters, and 30% of these clustered events are nuclear explosions located at historic test sites. Most of the remaining events are peaceful nuclear explosions (PNEs) scattered across the former Soviet Union (17%).

In the Pn and P joint data set, there are 1234 GT0–GT10 events. There are 28 GT0, 328 GT1 (27%), 14 GT2, 650 GT5 (53%), and 214 GT10 (17%) events with 111,498 P

Table 1
Estimates of Baseline Differences between Regional Pn and Teleseismic P -Wave Models

Travel-Time Model		$Pn - P$ Residual Mean (sec)	$Pn - P$ Standard Deviation (sec)	Number of Events
Pn	P			
IASP91	IASP91	-0.23	1.12	730
CUB2	IASP91	-0.48	0.95	768
CUB2	J362	-0.08	0.88	769
IASP91	J362	0.18	1.21	765
CUB1	J362	0.79	1.02	768
CUB1	IASP91	0.38	1.08	765

The values are based on $\text{mean}(Pn) - \text{mean}(P)$ travel-time residual differences for more than 700 GT0–GT10 events with at least five Pn and five P arrivals.

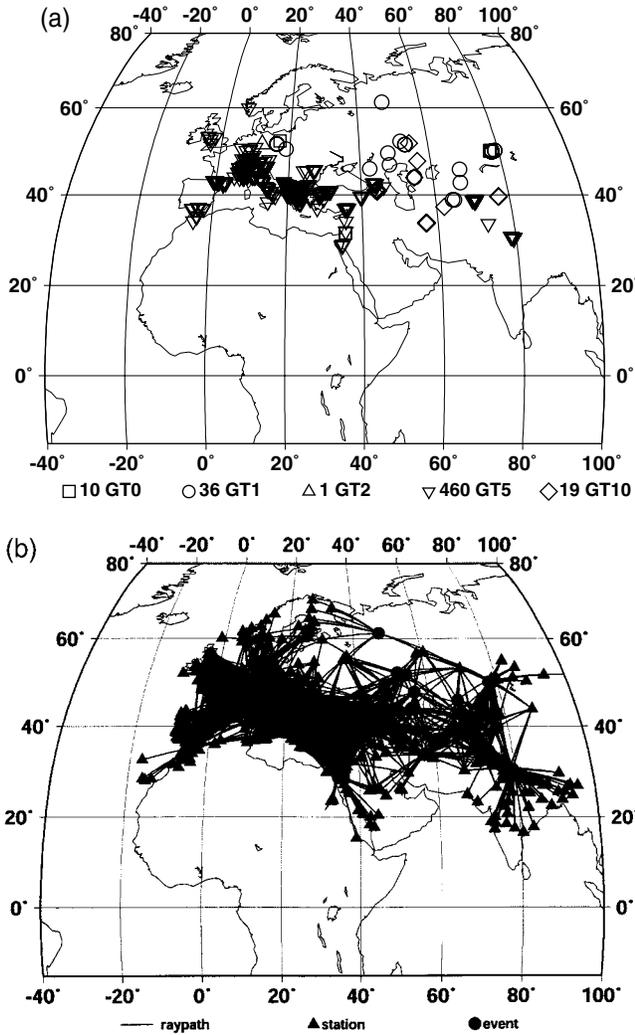


Figure 3. (a) The 526 GT0–GT10 events selected for relocation testing of regional Pn calibration using all stations: 10 GT0 (squares), 36 GT1 (circles), 1 GT2 (triangles), 460 GT5 (inverted triangles), and 19 GT10 (diamonds). (b) Event–station paths with 1098 stations (35,161 arrivals). Triangles show stations and circles show events. These are well-recorded events, with a median n_{def} (number of defining phases) of 51 (minimum n_{def} of 6), az_{gap} of 76°, and s_{gap} of 99°.

phases from 2823 stations and 39,017 Pn phases from 1166 stations. Most of the events (80%) have more teleseismic P than Pn arrivals, with a median ratio of P to Pn n_{def} of about 3. The maximum azimuthal gap is 154°, and 87% of the events have an azimuthal gap less than 100°. A large number of the events (24%) are explosions at historic test sites of PNEs (7%). The majority of the remaining events are GT5 earthquakes or mining-related events (GT1–GT5). Note that there are 441 more events (over 50%) that can be located using Pn and P together, compared to P alone. Most of the events in our three data sets have magnitudes of 4–6.

In the IMS event-location algorithm, *a priori* errors are used to weigh arrivals and calculate error ellipses. The errors

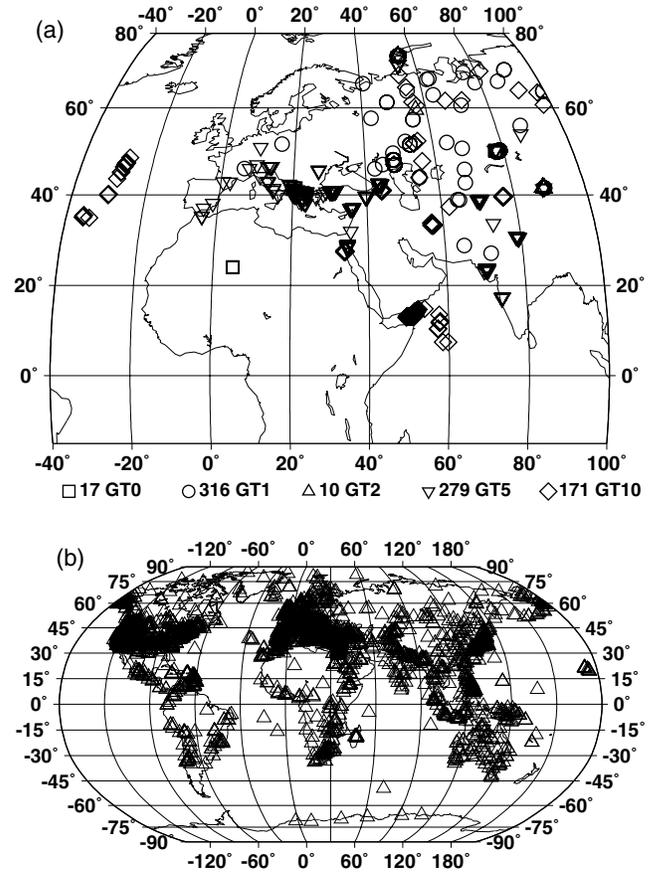


Figure 4. (a) The 793 GT0–GT10 events selected for use in relocation testing of teleseismic P calibration using all stations: 17 GT0 (squares), 316 GT1 (circles), 10 GT2 (triangles), 279 GT5 (inverted triangles), and 171 GT10 (diamonds). These events were very well recorded, with a median n_{def} of 100, az_{gap} of 68°, and s_{gap} of 85°. (b) The 2821 stations that contributed to 108,172 arrivals for the events shown in panel (a). Triangles show station locations.

are partitioned into model errors (Fig. 2) and measurement (picking) errors. Measurement errors are generally considered to be a function of phase type and signal-to-noise ratio. However, signal-to-noise ratios are not available from bulletin data; a measurement error of 1.0 sec was therefore assigned to those arrivals by default.

Constrained Bootstrapping

In order to maximize the discriminatory power of the relocation test, we devised a technique that, for a fixed number of arrivals, (1) samples azimuths as uniformly as possible to avoid statistical dependence on correlated ray paths, (2) minimizes the largest azimuthal gap at each realization for stable locations, and (3) avoids overemphasizing a single key station/arrival. We chose to simulate sparse network geometries with 10 stations and secondary azimuthal gaps less than 160°, in obtaining SSNBs. We find acceptable

SSNB geometries by building the s_{gap} tree, illustrated by the fictitious example in Figure 5. The s_{gap} tree is a hierarchical binary tree where each node represents a station that splits the remaining azimuthal gap in a way such that the secondary gap closed by the station is maximized. If there are several stations situated at the same azimuth, we pick one randomly. The s_{gap} tree stemming from a particular station is deterministic and unique. Cutting the tree at the predefined number of stations defines the minimal spanning tree providing the smallest secondary azimuthal gap for the given number of stations. If this secondary gap is smaller than the prescribed secondary gap limit, we accept the geometry. The procedure is repeated for each station in the network, which yields all network geometries meeting the constraints defined earlier. However, because of possible symmetries, different root stations may generate the same station configurations. To avoid repeating geometries, we retain only the unique station configurations.

We select from this set of sparse networks a balanced subset that does not overly emphasize any single station. The sets of acceptable station geometries are typically large. Some stations, situated at strategic azimuths, are sampled multiple times, while less important stations appear only once or twice. The individual station configurations can therefore be considered as basis vectors spanning the network geometry space. We orthogonalize this space using an algorithm analogous to the Gram–Schmitt orthogonalization of a matrix and select the 20 most representative network geometries to provide a reasonable statistical sample. The selected station configurations will then represent the SSNB realizations of an event. Since we select the most representative subnetworks, as opposed to just blindly taking random subsets of stations (as in traditional bootstrapping), we call this approach “constrained bootstrapping.” This procedure provides a controlled relocation experiment and allows better estimates of calibrated versus uncalibrated mislocations and their statistical uncertainties. It also provides some quality control over the test data set, as some events are found to contain outlier arrivals that produce unstable populations of SSNB locations.

Simulated Sparse Network Bulletins

To generate SSNBs for relocation tests, we selected well-recorded GT0–GT10 events to provide good geographical coverage. We limited the number of events to 10 from the same event cluster to avoid overrepresenting clusters with a large number of events. Sampling a cluster by several events allowed us to examine the consistency of location bias estimates and to identify outliers. The SSNB seed events were selected for the validation of regional and teleseismic calibrated travel times, respectively, from the events shown in Figures 3–4. The common events, 116 in total, provided the basis for evaluating regional and teleseismic calibration travel times jointly.

For direct comparisons, SSNB seed events were also relocated using all stations, denoted as “all-station (seed).”

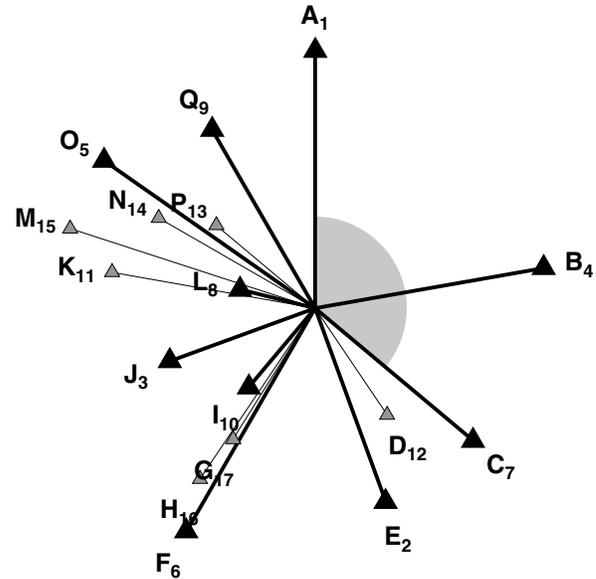


Figure 5. Illustration of a fictitious network and its SSNB realization stemming from station A. Each subsequent station is selected in a way that it splits the remaining azimuthal gap as evenly as possible. The selection order of stations is indicated as a subscript to the station names. In this example, a 10-station network is selected (large triangles) and station B closes the largest, 130° secondary azimuthal gap.

Note that in general the all-station cases discussed in this article refer to the entire data sets. The all-station (seed) results are for the SSNB seed events only, which are subsets of the entire data sets (282 versus 526 events for Pn data sets, 359 versus 793 events for P data sets, and 116 versus 1234 events for Pn and P data sets).

Location Evaluation Metrics

For location improvement comparisons, we evaluated statistics of mislocation, 90% error ellipse area, error ellipse coverage, origin time bias, origin time error, and standard deviation of observations. GT uncertainties (GTXs) are included when evaluating location improvement and coverage statistics. For one performance metric, categories are defined based on whether calibrated and uncalibrated mislocation is either less than or greater than the GTX. Furthermore, we split the category of events with both calibrated and uncalibrated mislocation greater than the GTX into two categories as to whether calibration moved the location toward or away from the GT epicenter. Events are considered indecisive when both the calibrated and uncalibrated locations are within the GTX or when they are identical. The numbers of events in these five categories, as defined in Table 2, are tabulated for different sets of models.

In addition to the five relocation categories (I–V given in Table 2), we compare the overall percentages of events improved (III + IV), degraded (II + V), and indecisive (I). We also compare the median mislocations with and without

Table 2
Five Categories (I–V) of Events Based on Epicentral Mislocation Comparisons with Respect to GTX

Mislocation Categories	Calibrated (CAL)	
	CAL Mislocation \leq GTX	CAL Mislocation $>$ GTX
Uncalibrated (UNCAL) UNCAL Mislocation \leq GTX	I. CAL and UNCAL inside GTX (indecisive)	II. CAL moved outside GTX
UNCAL Mislocation $>$ GTX	III. CAL moved inside GTX	IV. CAL moved toward GT V. CAL moved away from GT

There are two sets of relocation results (i.e., using calibrated and uncalibrated travel times).

travel-time calibrations and the median of the improved mislocations (better) versus the median of the degraded mislocations (worse). It is important to verify that the improvements are generally larger than the inevitable degradations.

Comparisons using the aforementioned categories provide direct assessment of mislocation relative to GTX with and without calibration. Note that these categories are described as relative to the IASP91 location (Table 2), since our major goal is to compare location improvement with respect to this baseline. Similar analysis can also be done for direct comparisons between any two calibrated models (e.g., CUB1 and CUB2).

We also directly evaluate the median mislocation differences with respect to GTX, that is, normalized improvement and degradation. The difference is insignificant when the normalized improvement/degradation is within unity. These metrics measure the degree to which the events have improved or degraded, as opposed to the categories defined in Table 2 that show how well the events are located relative to GTX.

The coverage parameter, E , is defined as the mislocation normalized to the 90% error ellipse in the direction of mislocation (Fig. 6). If the GT location lies within the error ellipse, the coverage parameter is less than unity. A realistic error model is one that covers the GT location with the error ellipse for 90% of all events. The adjusted coverage parameter, E , is then defined as

$$E = x^2/(s_{\text{majax}}^2 + \text{GTX}^2) + y^2/(s_{\text{minax}}^2 + \text{GTX}^2),$$

where x and y are mislocation components in the coordinate system defined by the semimajor (s_{majax}) and semiminor (s_{minax}) axes of the ellipse. Under the assumptions of Gaussian and independent errors, E follows a χ^2 distribution with 2 degrees of freedom. The ninetieth percentile should correspond to $E = 1.0$ if the *a priori* error models are properly calibrated. We define coverage for a test data set as the percentage of events with $E \leq 1$.

Location bias can be estimated from a number of realizations provided by SSNB relocations. The distance between the GT location and the centroid of all the realizations

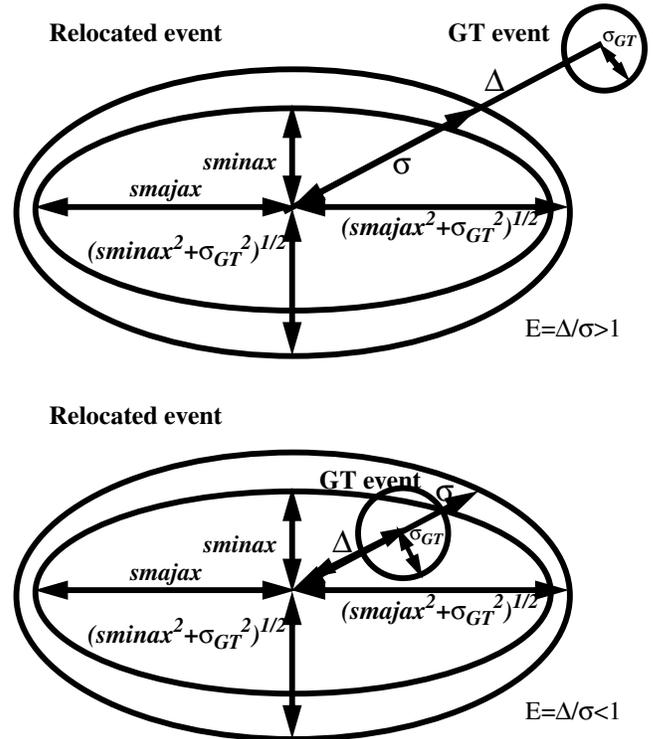


Figure 6. Coverage parameter $E = \Delta/\sigma$ where Δ is the mislocation and σ is the ellipse axis in the same direction. The event is not covered when $E > 1$ (top), and is covered when $E < 1$ (bottom). $E = 1.0$ corresponds to the 90% coverage. E should be χ^2 distributions with 2 degrees of freedom.

serves as an estimate of the scalar location bias (with estimated uncertainty) due to unmodeled lateral heterogeneities.

In summary, the evaluation metrics were designed to test and document that travel-time calibrations

- Reduce median mislocation
- Improve more events than degrade events
- Provide median improvement larger than median degradation
- Reduce median error ellipse area
- Provide coverage (percent of events with $E \leq 1$) of 90%

- Reduce misfit (standard deviation of residuals)
- Reduce origin time errors

Event Relocation Using Regional P_n and Teleseismic P Calibrated Travel Times

We performed two sets of location experiments using the data sets described in the previous section. We first present relocation tests using bulletins of all available arrivals and then describe the results using SSNBs.

All-Station Relocation

We relocated all 526 GT events in the P_n data set (Fig. 3), 793 GT events in the P data set (Fig. 4), and 1234 GT events in the P_n and P data set using all available stations. The GT events in all three data sets (P_n , P , and P_n and P) are generally well located, ensuring good azimuthal distributions (az_{gap} 's between 58° and 76° , s_{gap} 's between 76° and 99°) and large numbers of observations (median number of defining phases between 51 and 100). Table 3 summarizes the percentages of improvements versus degradation for P_n only (CUB1), P only (J362), and jointly P_n and P (CUB1 + J362); more detailed statistics on a case-by-case basis are given in the Appendix. The overall statistics show that more events are significantly improved than degraded for the CUB1 and J362 models. Between 20% and 40% more events were improved than were degraded. For example, CUB1 reduced mislocations for 49% of the events but increased mislocations for 29% of the events. Between 10% and 20% of the events are located within GTX, with or without calibration, and therefore are categorized as “indecisive” in Table 3. This fraction of the population is well located with or without calibration (within ~ 5 km), or the GTX is too large to discern a significant mislocation. This underscores the importance of using test events having location accuracies better than 5 km for relocation testing. Events with large uncertainties contribute little to the test except to verify that the calibrations do no harm to these already well-located events.

Table 3 also shows the details of each mislocation category. Regardless of whether mislocations are inside or outside the GTX, more events are improved than degraded. It is particularly striking for J362, with more than twice the number of events improved than degraded (9% versus 4% for the case of mislocation inside the GTX and 56% versus 22% for the case of mislocation outside the GTX). Calibration moves more events inside than outside the GTX of the GT location and also improves more events that lie outside the GTX.

Also shown in Table 3 are percentages of events with improvements and degradations normalized by the GTX; the changes for the remaining events are insignificant. Locations were improved for significantly more events than those degraded by the same uncertainty. For J362, 38% of the events improved compared to 6% of the events degraded by the same uncertainty.

Table 3

Summary of All-Station Relocation Statistics for Location Improvement Using Models CUB1 and J362, Both Independently and Jointly (P_n and P)

Epicentral Mislocation	P_n	P	P_n and P
	CUB1	J362	CUB1 + J362
Number of events	526	793	1234
III + IV: reduced (%)	49	65	52
II + V: increased (%)	29	26	31
I: indecisive (%)	22	9	17
III: moved inside GTX (%)	11	9	11
II: moved outside GTX (%)	9	4	10
IV: moved toward GTX (%)	38	56	41
V: moved away from GTX (%)	20	22	21
<i>Normalized Improvement/Degradation</i>			
Improvement \geq GTX (%)	15	38	26
Degradation \geq GTX (%)	4	6	5

Values are given by percentages of events using CUB1, J362, and CUB1 + J362.

Summary statistics of median mislocation, median error ellipse area, 90% coverage, origin time error, and standard deviation of observations (sd_{obs}) are given in Table 4; more details are included in the Appendix. The median mislocations all range between 6 and 8 km, with and without calibration. Overall, calibration reduced median mislocation by 10%–20%. By far, teleseismic calibration has delivered the largest fractional improvement with a 27% improvement in the median mislocation. Using both regional and teleseismic calibrations, the median mislocation has been reduced to nearly the level of the GT5 uncertainty (5.7 km).

Figure 7a–c shows the cumulative mislocations with and without calibration for each data set. For all 526 GT0–GT10 events in the P_n data set, the median CUB1 mislocation is 6.6 km. That the largest CUB1 improvement occurred at the ninety-fifth percentile, with mislocations reduced by 27%, demonstrates that P_n calibration made the greatest improvements to some of the events with the largest mislocations. For all 793 GT0–GT10 events in the P -only data, the J362 median mislocation is 6.1 km. The largest improvements occurred at the thirtieth percentile, with J362 mislocation reduced by 29%, demonstrating that some of the larger uncalibrated mislocations are not improved. The poorly located events, mostly GT10 events along the mid-ocean ridge and transforms, did not significantly improve using J362. As in the independent calibration, for all 1234 events in the P_n and P data set there is also significant location improvement, with a median mislocation of only 5.7 km. As in the P -only case, the largest improvement with joint calibration occurred at the twenty-fifth percentile, as the most poorly located events did not improve. The cumulative distributions of the improvements/degradations normalized to the GTX are shown in Figure 7d–f and illustrate that improvements (positive) are generally larger than degradations (negative), as given in Table 3 (the last two rows).

Table 4

Summary of All-Station Relocation Statistics for Mislocations, Error Ellipse Area, Origin Times (OT), and Standard Deviation of Observations (sd_{obs}) using CUB1, J362, and CUB1 + J36

Metrics	Pn		P		Pn and P	
	IASP	CUB1	IASP	J362	IASP	CUB1 + J362
Mislocation (km)	7.1	6.6	8.3	6.1	7.1	5.7
Error ellipse area (km ²)	448	253	464	322	394	233
Coverage (%)	83	76	65	75	75	77
Median OT from GT (sec)	0.21	-0.52	0.73	0.97	0.67	0.90
Median OT error (sec)	0.79	0.55	0.43	0.32	0.50	0.36
Median sd_{obs} (sec)	1.19	1.11	0.84	0.77	1.04	0.92

We also compared 227 events for which Pn -only, P -only, and joint Pn -and- P locations can all be directly compared, with and without calibration (details are not tabulated here). As expected, there are more events that were improved than degraded with joint calibrations, compared with events improved using regional or teleseismic calibration alone. Median mislocation is the smallest with joint regional and teleseismic calibration (7.2 km for Pn only, 8.1 km for P only, and 5.4 km for Pn and P jointly). All combinations of calibration location are generally better than the uncalibrated locations. The calibrated median mislocation reduction is similar in each case (13%–16%).

Table 4 also shows that calibration significantly reduces median error ellipse area, origin time error, and the standard deviation of observations (sd_{obs}). The reduction of the error ellipse areas is a direct consequence of the reduced *a priori* model variances used to compute the error ellipse. The misfit (standard deviation of observations, sd_{obs}) is reduced for the majority of events, and its median is reduced. The estimated origin time error is also reduced as a result of a reduction in misfit and smaller *a priori* errors. The median origin time bias is generally increased (mostly positive), indicating that these models are biased slightly fast compared to the IASP91 baseline. This may indicate that some baseline biases may still remain in the models. However, these results should be considered with caution. While most epicenters are fairly accurate (better than 5 km), most origin times (including most explosions) are inferred from seismic data and are not as reliable.

As seen in Table 4, regardless of whether uncalibrated or calibrated travel times are utilized, the error ellipse coverages are significantly lower than the desired 90%. For example, the Pn location 90% error ellipses adjusted for GTX contain only 83% (uncalibrated) and 76% (calibrated) of the GT locations. Likewise, the coverage percentages are only 65% (uncalibrated) and 75% (calibrated) for teleseismic P relocations. While the calibrated coverages are generally better than the uncalibrated coverages, the error ellipses are clearly too small. However, for events with large numbers of observations, this is not too surprising. Figure 8 shows the coverage parameter, E , versus the number of defining phases (n_{def}) for the joint regional Pn and teleseismic P case.

The likelihood that coverage is insufficient increases with increasing n_{def} . The uncalibrated 90% coverage is properly maintained for $n_{\text{def}} \leq 50$, and the calibrated 90% coverage is maintained for $n_{\text{def}} \leq 27$. To address this problem and several other issues, we designed a new set of relocation tests that attempt to simulate the effects of calibration on events located by sparse regional and teleseismic networks.

SSNB Relocation

The SSNB realizations were relocated using both uncalibrated (IASP91) and calibrated (CUB1 and CUB2) regional travel times. SSNB location experiments were designed in part to address the problems with insufficient error ellipse coverage for events with a large number of arrivals. They also improve the statistical power of the relocations since a large number of the events in the all-station relocation tests fall into the indecisive category. Figure 9 shows examples of 10-station SSNB relocations for a GT5 earthquake in Pakistan on 14 February 1977 and a GT1 PNE in Russia on 20 August 1972. As expected, calibrated travel times reduce location bias for these two events. The plotted results of 20 SSNB realizations show the estimated uncalibrated and calibrated vector biases (with scatter) from 20 different well-distributed 10-station networks. This procedure was repeated for Pn -only locations (282 events), P -only locations (359 events), and joint Pn -and- P locations (116 events) using the database described earlier.

Table 5 summarizes the overall percentages of events improved, degraded, or remaining indecisive with respect to IASP91 due to the application of calibrated travel times; more details are given in the Appendix. As the statistics of the SSNB centroids represent the reduction in location bias, the SSNB 10-station locations represent the reduction in mislocation obtained from the constrained bootstrapping. The results from the all-station (seed) locations of the same SSNB seeds are also compared in the Appendix. More events are improved than degraded, and calibrated travel times reduce both bias and mislocation. When mixing calibrated travel times with uncalibrated ones, the improvements tend to fade (see the Appendix), so it is important to calibrate both regional and teleseismic phases whenever possible. We find, however, that mixing calibrated and uncalibrated Pn

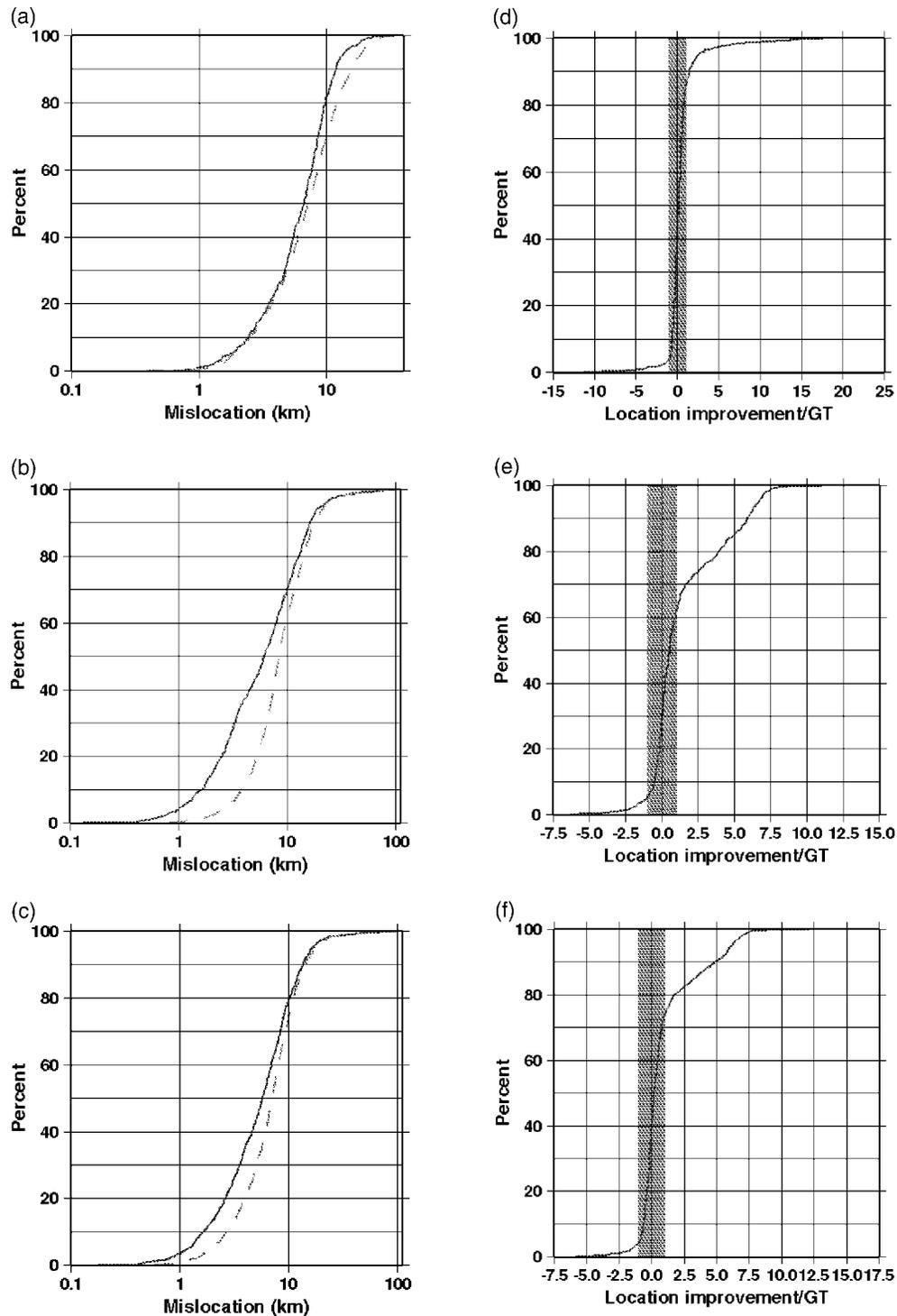


Figure 7. Cumulative mislocations calibrated (solid) and uncalibrated (dashed) are shown for (a) CUB1 P_n , (b) J362 P , and (c) CUB1 P_n and J362 P jointly. In all cases the calibrated distribution (solid) is shifted to the left, indicating a reduction of mislocation at all percentiles. The regional calibration has the biggest gain above the median (>7 km), while teleseismic calibration has the biggest gain below the median (<7 km). The cumulative distribution of mislocations for the joint P_n and P location case (c) is strongly influenced by the teleseismic distribution. Cumulative distributions of mislocation improvement normalized to GTX are shown for (d) CUB1, (e) J362, and (f) CUB1 + J362. Improvement is positive; degradation is negative. Changes in mislocation less than GTX are indecisive (shaded). In all cases the positive tail (improvement \geq GTX) of the distribution is larger than the negative tail (degradation \geq GTX), indicating that improvements are larger than degradations. The three percentage fractions (improvement \geq GTX, degradation \geq GTX, and insignificant) and their median values (median improvement, median degradation) are tabulated in Tables 3–4 and in the Appendix.

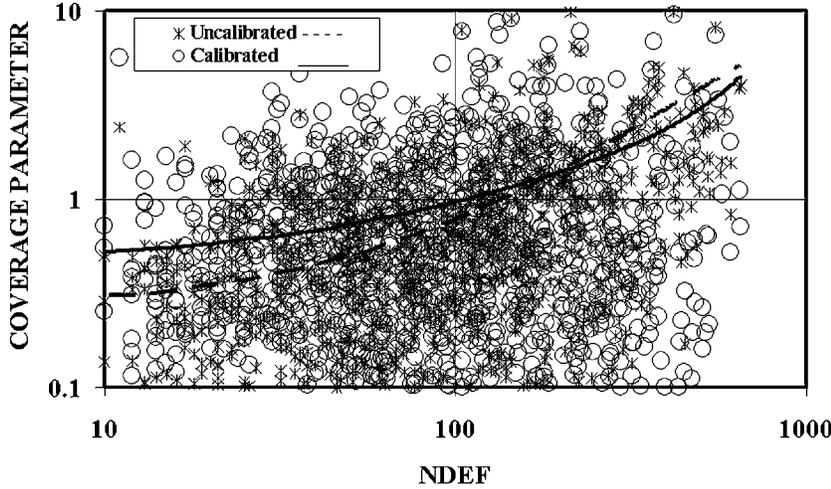


Figure 8. Coverage parameter, E , versus number of defining phases (n_{def}) using joint CUB1 P_n and J362 P calibrations (triangles) and uncalibrated travel times (diamonds). Median trends are shown as solid lines. The likelihood of insufficient coverage increases with increasing number of defining phases (n_{def}). Coverage parameters, E , are less than unity for 90% of the events with $n_{\text{def}} \leq 50$ (uncalibrated) and $n_{\text{def}} \leq 27$ (calibrated).

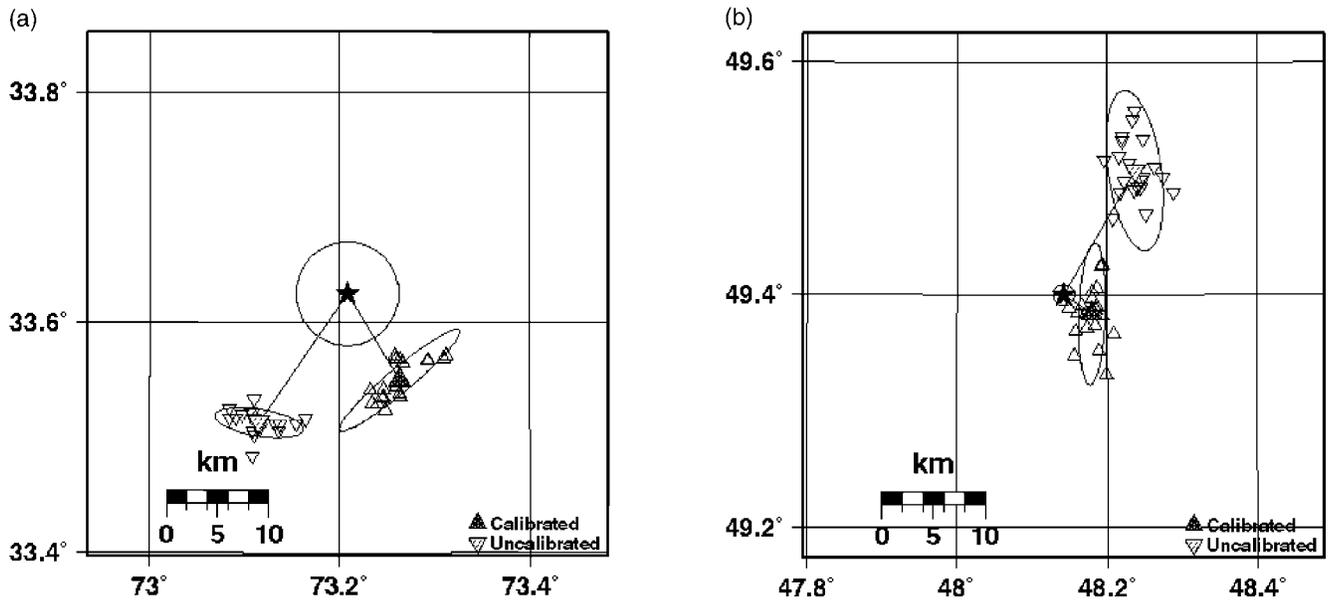


Figure 9. SSNB example for (a) the 14 February 1977 00:22:37 earthquake in Pakistan, and (b) the 20 August 1972 03:00:00 PNE. Open triangles represent the relocations with the 20 most characteristic SSNB 10-station subnetworks with calibrated (CUB1, triangles) and uncalibrated (IASPEI, inverted triangles) regional P_n travel times. The vector location bias estimate (line) connects the GT location (star) with the centroid of the SSNB realizations (solid triangle). The uncertainty in the location bias estimate (ellipses) is derived from the individual SSNB locations.

and teleseismic P arrivals is generally favorable to ignoring the uncalibrated arrivals provided that suitable *a priori* model errors are available to properly weight the calibrated and uncalibrated data.

Figure 10 shows the mean mislocation from the all-station (seed) locations compared to the mean location bias from the 10-station sparse network solutions using regional and teleseismic calibrated (CUB1 + J362) and uncalibrated (IASP91) travel times for event clusters with multiple SSNB seeds. The event clusters are sorted by their GT accuracy (thick line). Both the mean mislocation and bias across all

the clusters are about 5–7 km. Therefore, for clusters of GT10 quality, most of the events are located within the GT accuracy, regardless of whether calibrated or uncalibrated travel times are used. This result demonstrates that GT10 events have very limited utility for location calibration testing and validation. We did not separate the GT10 events from our overall statistics since the small number of such events did not have a significant effect on our results. While using all stations (seed), 57% of the clusters show location improvements due to calibrated travel times and the SSNB bias estimate exhibits improvement for 78% of the clusters.

Table 5
Summary of SSNB Relocation Statistics for Location Improvement

Epicentral Mislocation	<i>P_n</i> CUB1	<i>P</i> J362	<i>P_n</i> and <i>P</i> CUB1 + J362
Reduced (%)	51	62	58
Increased (%)	30	28	25
Indecisive (%)	19	10	17
<i>Normalized Improvement/Degradation</i>			
Improvement \geq GTX (%)	17	25	25
Degradation \geq GTX (%)	8	9	7

Values are given as percentages of events using CUB1, J362, and CUB1 + J362.

In general, the measurable improvements are larger and more consistent for SSNBs, indicating that the constrained bootstrapping provides a more sensitive experiment to evaluate calibrated travel times and indeed to increase the statistical power of the relocation tests.

The principal metrics on location bias, mislocation, coverage, and the area of the 90% error ellipse are summarized in Table 6 for the cases of regional, teleseismic, and jointly regional and teleseismic calibration; more details are given in the Appendix. The location bias and mislocation are reduced in all cases, but the improvements are significantly smaller when mixing uncalibrated and calibrated travel times. There are striking differences between the median areas of error ellipses obtained from the SSNB and all-station locations. The ellipses are too small when using all stations, resulting in much poorer coverage statistics. More than 100 stations typically record the SSNB seed events, and many arrivals sample similar ray paths. Thus the assumption made by the location algorithm of independent errors is violated, leading to unrealistically small error ellipses. We will further address the issue of correlated errors in the Discussion.

We compared the percentage of events with improved/degraded locations due to calibrated travel times for the regional (CUB1, CUB2), teleseismic (J362), and joint regional and teleseismic (CUB1 + J362, CUB2 + J362) cases. In all cases calibration improved most of the events. CUB1 improved somewhat more events than CUB2 when only regional phases were used or when used jointly with J362 teleseismic travel-time predictions. On the other hand, CUB2 yields slightly greater reduction in location bias. The statistics for the area of error ellipse and coverage do not significantly differ between CUB1 and CUB2, either alone or jointly with J362.

Comparisons of constrained bootstrapping results for the 116 events common to all combinations of location (*P_n*, *P*, and *P_n* and *P*) show that, in general, calibrated travel times reduce location bias. Figure 11 illustrates median bias reduction (Fig. 11a) and percentages of events improved (Fig. 11b) for a simple decision matrix that compares the consequences of using calibration and relocation using regional *P_n* alone, teleseismic *P* alone, and joint *P_n* and *P*

location based on 10-station network SSNB simulations. Between 65% and 70% of events are improved by calibration with between 20% and 24% reduction in the median bias. *P_n* and *P* location improves between 61% and 74% of events, with between 21% and 28% reduction in the median bias. Calibrated 10-station sparse network *P_n* and *P* median mislocations approach the GT5 accuracy of the majority of the test events. Adding calibration and combining *P_n* and *P* is always desirable. The most profound reduction in bias was achieved when both regional and teleseismic calibrated travel times were used. The median area of the error ellipse is also significantly reduced in all cases, dropping below or approaching 1000 km² for the 10-station SSNB sparse networks. The reduction in the area of error ellipse is achieved without significant degradation in coverage. Since constrained bootstrapping increases the statistical power of the test, it shows more significant improvements between uncalibrated and calibrated travel times than the all-station relocations.

Discussion

Two topics touched upon in the previous section deserve some additional discussion. In this section we directly compare CUB1 versus CUB2 and particularly examine the issue of correlated errors in more detail.

Direct Comparisons between CUB1 and CUB2

As shown in the Appendix, there is large mislocation improvement using CUB1, but CUB2 does not show overall significant improvement. Significantly more events are improved by CUB1 than CUB2 using regional phases (50% of events improved versus 28% of events degraded, as shown in Fig. 12). Unlike regional calibration alone, results with J362 + CUB1 are only somewhat better than those with J362 + CUB2 (36% of events improved versus 28% of events degraded), since teleseismic phases play a dominant role for most events in this data set when teleseismic and regional phases are used jointly.

It appears that the CUB1 model provides better location calibration for the western part of the region but that CUB2 may perform somewhat better in Asia. Of the events that were improved more by CUB1 than CUB2, most are west of 40° E and are generally located in the Middle East and the Mediterranean (Fig. 12; Appendix). Besides CUB2, J362 also performs poorly in this western region, resulting in poor performance in joint calibration.

The majority of the GT5 earthquakes in the *P_n* data set are located in this western region. Most of them (91%) are GT5, and 66% of all events in this region were selected from the EHB bulletin based on the Group-2 GT5 event selection criteria and were not vetted using cluster analysis. Initially we might suspect that, because these events are less well constrained, the better performance of CUB1 might be illusory. However, it is also plausible that upper mantle mineralogies differ systematically east to west, such that the dif-

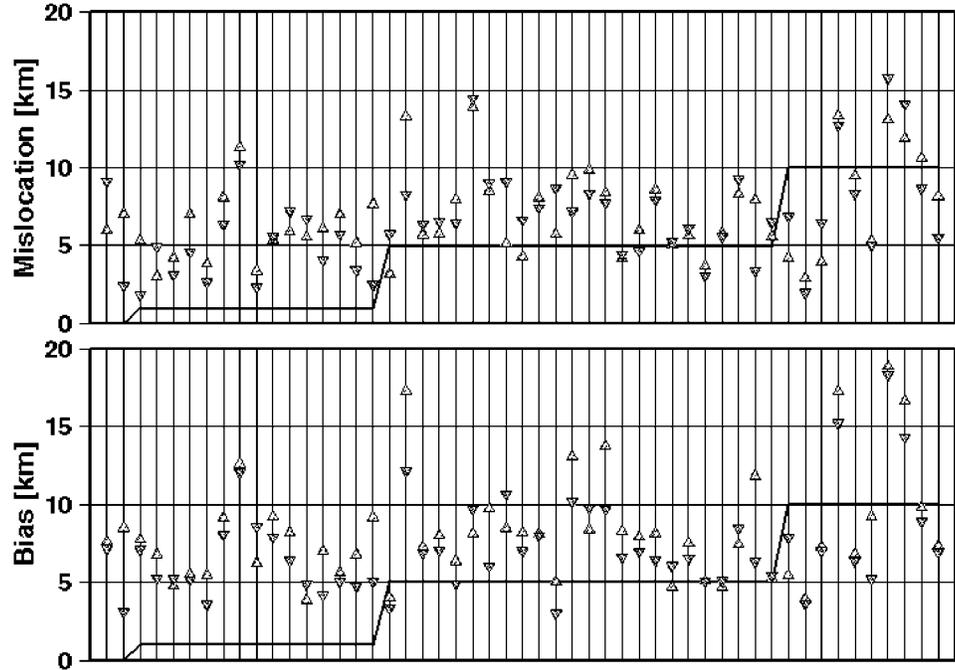


Figure 10. Top: Mean mislocations using all stations (seed). Results shown for relocations calculated with (inverted triangles) and without (triangles) calibration. Bottom: Mean bias from SSNB locations with (inverted triangles) and without (triangles) calibration. Event clusters are sorted by GT category. The thick line indicates the GT accuracy of the reference events in the clusters.

Table 6

Summary of SSNB Relocation Statistics for Location Bias, Mislocation, Error Ellipse Area, and Coverage using CUB1, J362, and CUB1 + J362

Metrics	P_n		P		P_n and P	
	IASP	CUB1	IASP	J362	IASP	CUB1 + J362
Mean bias (km)	8.4	7.2	9.2	7.9	7.7	6.6
Mislocation (km)	8.8	7.6	11.1	9.9	9.9	8.7
Error ellipse area (km ²)	1522	745	1953	1133	1815	971
Coverage (%)	99	96	92	85	97	92

ferent S -to- P mappings used for CUB1 and CUB2 each work better in different regions.

Both Ritzwoller *et al.* (2002) and Bhattacharya *et al.* (2003) compared P_n travel times predicted by CUB1 and CUB2 to empirical path corrections derived from clusters. Ritzwoller *et al.* examined only a few clusters west of 40° E. Bhattacharya *et al.* considered clusters throughout the region but used many of lesser GTX quality. Their study found that the general spatial distribution of anomalies is similar for both models, but there are some important regional variations. With respect to empirical path corrections, CUB1 is generally better correlated and shows greater variance reduction than CUB2 in the Mediterranean. Correlations between predicted path effects (calibrations) for CUB1 and CUB2 are the poorest in Greece, highlighting where these two models differ the most. In this area the CUB1 model is

faster than IASP91, while the CUB2 model is slower than IASP91.

As shown in the Appendix, the CUB1 model does perform better than would be expected from random chance in this region, while the CUB2 model performs about as well as random chance (two equal values) would predict. The fact that the independent model J362 actually performs worse for this subset of events than would be predicted by random chance suggests that the problems do not arise from random errors in the GT event locations, but rather from unmodeled structure. Moreover, Bhattacharya *et al.* (2003) found more clusters with statistically significant correlations between empirical path corrections and the CUB1 model-based path corrections than with those derived from CUB2. This suggests that the underlying differences are structural and not due to the GT event locations.

Correlated Errors versus 90% Coverage

As shown in Figure 8, the 90% coverage is very low using all stations for each case with independent or joint regional and teleseismic calibration. While the problem may be due to underestimated *a priori* errors, the SSNB bootstrapping results reveal that this is due to a breakdown in the assumption of uncorrelated errors.

Herrin and Taggart (1968) have shown in an analysis of the LONGSHOT nuclear explosion (29 October 1965, Amchitka) that a large number of arrivals sampling similar ray paths along unmodeled 3D Earth structure may introduce

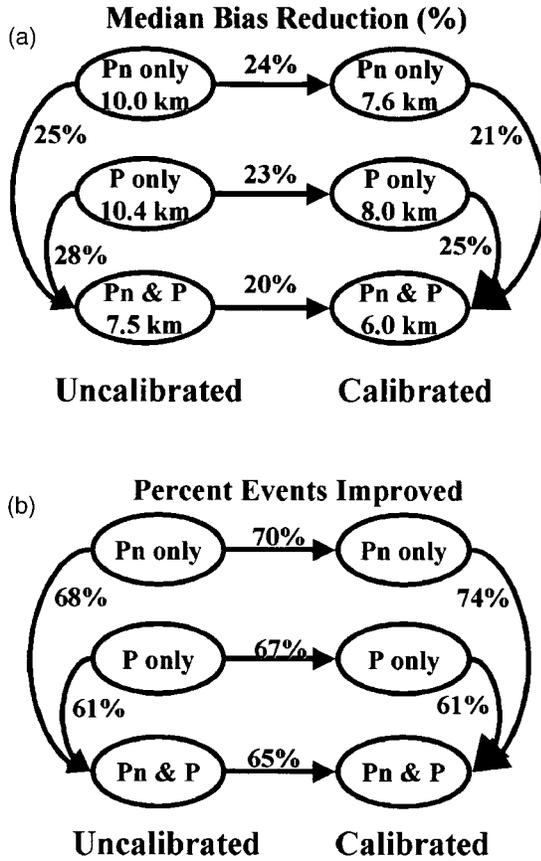


Figure 11. SSNB location improvements for 116 common events resulted from using calibrated travel times (horizontal arrows) and the combination of regional and teleseismic phases (vertical arrows). (a) Median bias (mislocation) reductions by model-based calibration. (b) Percentages of event locations improved by model-based calibration. Joint Pn and P travel-time calibration improves two out of three events and reduces bias 20% or more.

location bias. In the LONGSHOT case, a large number of ray paths travel through subducted oceanic slab having a high seismic velocity, thus arriving systematically earlier than the predicted arrival times. The systematic travel-time prediction bias resulted in a 26-km mislocation that was far outside the error ellipse (139 km^2), which was calculated assuming uncorrelated errors. More recently, Myers and Schultz (2000b) have pointed out that travel-time prediction errors are typically correlated for similar ray paths. In essence, this is the underlying principle of kriging: similar paths through the Earth have similar path effects.

Although in the past few years several nonlinear hypocenter location methods have been developed to account for nonlinearity and non-Gaussian error distributions (e.g., Sambridge and Gallagher, 1993; Billings, 1994; Billings *et al.*, 1994; Lomax *et al.*, 2000; Sambridge and Kennett, 2001; Rodi *et al.*, 2002), most routinely used location algorithms (including the one we used in this study) assume Gaussian, independent errors. One of our major motivations for de-

veloping the constrained bootstrapping technique was to minimize the effect of correlated errors on the validation test.

To demonstrate how correlated errors may degrade coverage, introduce location bias, and produce unrealistic error ellipses, we performed an SSNB study on a GT1 Chinese underground nuclear explosion that occurred on 7 October 1994. A total of 562 stations at distances between 25° and 97° recorded this event. The station distribution is far from uniform and is dominated by stations in Europe, Japan, and California. We generated SSNBs having the number of defining phases (n_{def}) ranging from 6 to 400 stations and having each n_{def} value represented by 20 SSNB realizations. Figure 13 shows the SSNB J362 versus IASP91 location bias, misfit, 90% coverage, error ellipse area, azimuthal gap, and secondary azimuthal gap (s_{gap}) as a function of n_{def} . The location bias increases (almost monotonically) with increasing number of stations, regardless of calibration (Fig. 13). As more and more stations are added to the solution, the location is driven away from the GT1 location. Although J362 decreases location bias, the pattern of the location changes between the two is quite similar, indicating that the relative importance of some station clusters (possibly the Californian network) steadily increases because the location algorithm does not account for correlated travel times along similar ray paths. Note that even if the J362 model precisely accounted for all 3D heterogeneities in the Earth, the false assumption of independent errors would still have resulted in a biased location and error ellipse.

As shown in Figure 13, the information carried by the network geometry is exhausted relatively early. The azimuthal gap stabilizes after about 20 stations, and the secondary gap after 100 stations. Then adding further stations merely increases data redundancy. However, the area of the error ellipse shrinks indefinitely with increasing number of arrivals because the covariance matrix is calculated incorrectly, based on the assumption that the errors are uncorrelated. This results in an approximately linear increase in the coverage parameter, meaning that it is guaranteed that the true epicenter will lie outside the error ellipse, given a sufficient number of arrivals. On the other hand, the median misfit (rms residual) does not provide any information about the location quality: the misfit remains basically constant once the secondary gap information is exhausted. In this particular example, the location bias due to correlated errors increases with the number of stations, but this may not always be the case. Location bias may or may not increase as more arrivals are added to a solution. The actual behavior depends on how the net weights of closely clustered stations are balanced with respect to the unmodeled 3D Earth structures.

Correlated errors will always produce unrealistic error ellipses for events located using a large number of arrivals. The constrained bootstrapping was developed to minimize the effect of correlated errors. Figure 14 shows the cumulative histograms of the coverage parameter for the all-station (seed) locations and the 10-station sparse network

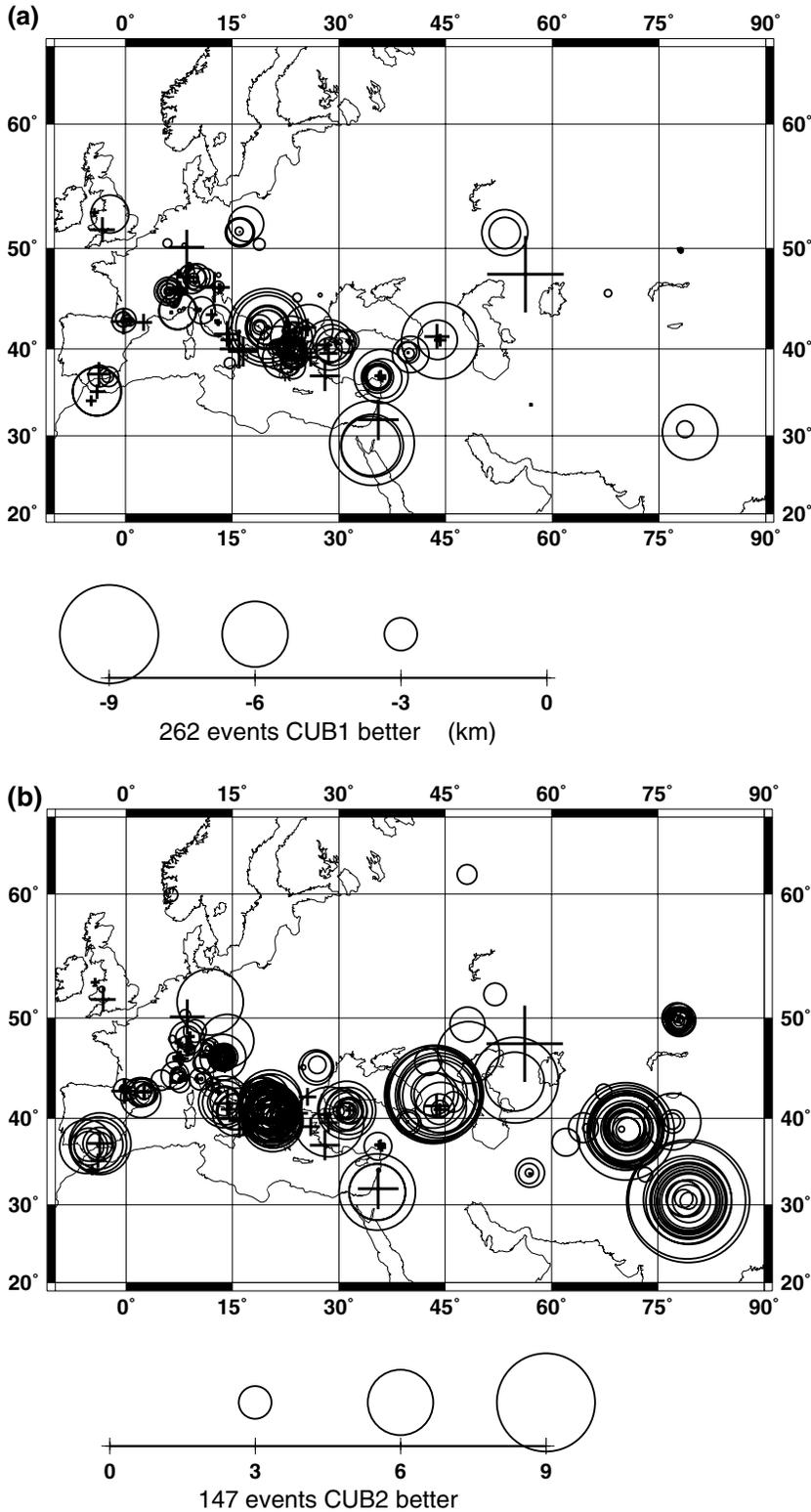


Figure 12. Mislocation differences between CUB1 and CUB2 in kilometers. Circles show events where (a) CUB1 is better or (b) CUB2 is better, and crosses show events that are indecisive. CUB1 improves 262 (50%) events (mostly in the Mediterranean), and degrades 147 (28%) events within GT uncertainty with respect to CUB2 (see Appendix). There are 117 (22%) indecisive events that cannot discriminate between CUB1 and CUB2.

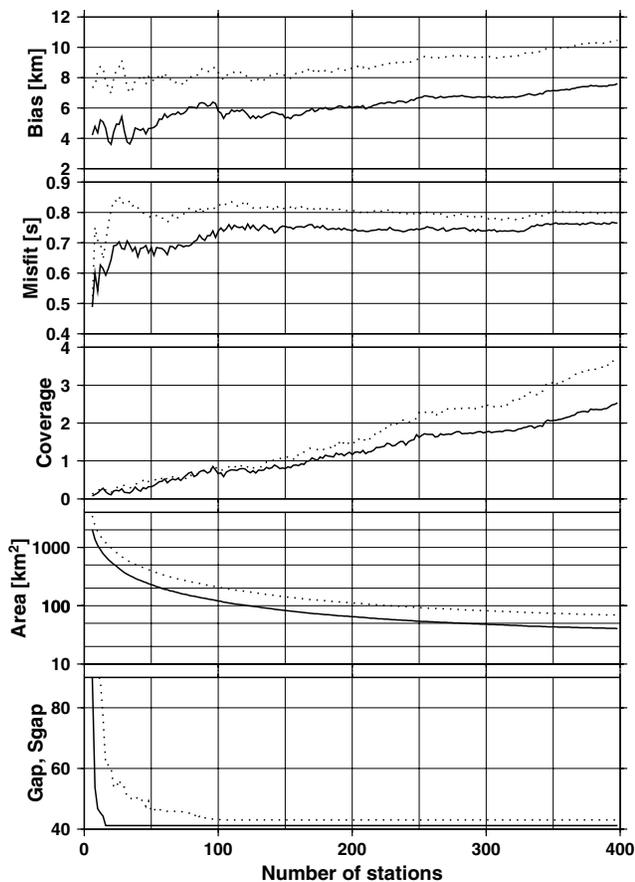


Figure 13. Example of the GT1 Chinese underground nuclear explosion detonated on 7 October 1994. Location bias, misfit, coverage parameter, error ellipse area, median gap (solid), and s_{gap} (dashed) are plotted as a function of number of phases used in the SSNB locations with calibrated (solid line; J362) and uncalibrated (dashed line; IASP91) travel times.

locations for the same events using calibrated and uncalibrated regional and teleseismic travel times. While the error ellipses of the all-station (seed) locations cover only 70% of the true epicenters, about 90% of the SSNB calibrated locations are covered by their 90% coverage ellipses. The error models appear to be correct for sparse station networks.

As shown in Table 6, the 90% coverages are improved using calibration, but the ninetieth percentile of coverage is still low for the 10-station SSNB tests in the calibrated teleseismic case. A conservative rescaling of the IASP91 model errors may be estimated from the square root of the coverage ratio as a function of cumulative percentage. To ensure 90% coverage at the ninetieth percentile, the model error needs to be increased to approximately 70% of the IASP91 model error (standard deviation). This still corresponds to a net 50% variance reduction with respect to the original baseline IASP91 model errors.

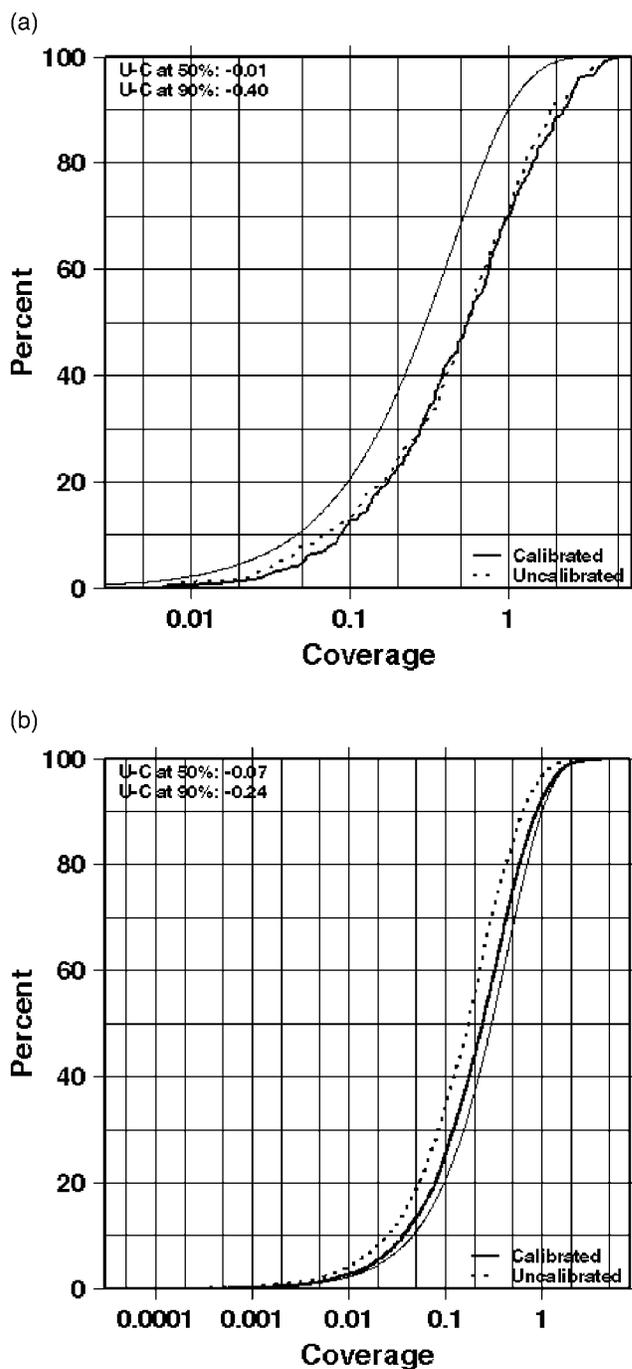


Figure 14. Cumulative histograms of the coverage parameter with (a) all-station seed and (b) the SSNB 10-station sparse networks with calibrated (solid) and uncalibrated (dashed) regional and teleseismic travel times. The theoretical χ^2 distribution (assuming Gaussian and independent errors) of the coverage parameter is shown as a thin solid line. For both figures, the differences in the coverage parameters between uncalibrated and calibrated cases are also given for the fiftieth and ninetieth percentiles.

A fundamental approach in addressing the 90% coverage problem is to construct the covariance matrix in such a way that it accounts for the correlation between similar ray paths. Chang *et al.* (1983) proposed a scheme to estimate the covariance matrix in the presence of intercorrelated errors. Bayesian kriging (Myers and Schultz, 2000a) routinely deals with this problem. A revised location algorithm should take into consideration nonlinearity, as well as non-Gaussian, nonzero mean and nonindependent errors. Sparse networks, such as the IMS network, are less vulnerable to correlated errors. However, the formal location uncertainties published by international agencies that use a large number of stations to locate events, such as the International Seismological Centre and National Earthquake Information Center, are likely to be underestimated.

Conclusions

We have validated location improvement from model-based calibration based on 3D global CUB and J362 models using a large set of high-quality GT events. Relocations using all-station and SSNB bootstraps show that there are significant location improvements using either regional P_n (CUB) and teleseismic P (J362) calibration or both. Our result is consistent with variance reductions estimated from cluster analysis (Bhattacharyya *et al.*, 2003), as well as with other previous relocation studies (Antolik *et al.*, 2003; Ritzwoller *et al.*, 2003) that demonstrated significant improvement using either regional or teleseismic model-based calibration alone with smaller data sets. Since in practice regional and teleseismic phases are used jointly, we tested P_n and P calibrations with more than 1200 GT0–GT10 events. Our relocation tests show that the largest location improvement is achieved by joint regional and teleseismic calibration and has reached nearly the GT5 accuracy level for well-distributed sparse networks. In combining regional and teleseismic calibrations, any baseline differences need to be adjusted and appropriate weights applied. In evaluating the two upper mantle models (CUB1 and CUB2), which are derived by converting an S model to a P model, we find that the S -to- P mappings may be regionally dependent and future regional tuning of these models may provide additional benefits.

We introduced a method of constrained bootstrapping, which not only provides a controlled experiment designed to increase the statistical power of the relocation test but also provides estimates of location bias due to unmodeled 3D Earth structures. In general, the documented improvements are larger and more consistent for SSNBs than those for the all-station locations. A corollary of the SSNB approach is that for any well-recorded event, a set of subnetworks exists that locates the event better than the entire network.

The 90% coverage error ellipses obtained from the all-station locations are unrealistically small and do not cover 90% of the GT locations. This is due to the false assumptions of Gaussian and independent errors that are built into most

routinely used location algorithms. While constrained bootstrapping reduces the effect of correlated errors, the ultimate remedy would be to improve the location algorithms, that is, by using non-Gaussian/nonlinear methods and accounting for off-diagonal elements in the covariance matrix.

Using our unique data set and new subnetwork sampling techniques, we have demonstrated that locations for events recorded by sparse networks can be significantly improved using 3D global models. The median 10-station location bias for the jointly calibrated regional and teleseismic case is 6.6 km (7.2 km for P_n only and 7.9 km for P only), approaching the GT5 uncertainty of the majority of GT event locations. In general, mislocation reductions for model-based calibration are over 20%. About two out of three (60%–70%) event locations improved with calibration. Improvements due to the combination of regional and teleseismic phases are comparable to those for either regional or teleseismic improvements alone. Model-based calibrated travel times provide improvements in all cases, and combining regional and teleseismic phases further improves event locations, using either uncalibrated or calibrated travel times.

Acknowledgments

This work was sponsored by the Defense Threat Reduction Agency under Contract Number DTRA01-00-C-0013. We are greatly indebted to the institutions and individual researchers who provided ground-truth information and station readings. Without their voluntary contributions, this work could not have succeeded. We thank the editors and reviewers for their help and suggestions.

References

- Antolik, M., G. Ekström, and A. Dziewonski (2001). Global event location with full and sparse data sets using three-dimensional models of mantle P -wave velocity, *Pure Appl. Geophys.* **158**, 291–317.
- 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.-Y. 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. of 24th Seismic Research Review: Nuclear Explosion Monitoring—Innovation and Integration*, Ponte Vedra Beach, Florida, 17–19 September.
- Barmin, M. P., M. H. Ritzwoller, and A. L. Levshin (2001). A fast and reliable method for surface wave tomography, *Pure Appl. Geophys.* **158**, 1351–1375.
- Bassin, C., G. Laske, and G. Masters (2000). The current limits of resolution for surface wave tomography in North America, *EOS* **81**, F897.
- Bhattacharyya, J., H. Israelsson, K. McLaughlin, E. R. Engdahl, and E. Bergman (2003). Validation of a global and a regional model using empirical path corrections derived from event clusters, Group-2 Consortium final report, Arlington, Virginia.
- Billings, S. D. (1994). Simulated annealing for earthquake location, *Geophys. J. Int.* **118**, 680–692.
- Billings, S. D., M. S. Sambridge, and B. L. N. Kennett (1994). Errors in hypocentral location: picking, model, and magnitude dependence, *Bull. Seism. Soc. Am.* **84**, 1978–1990.
- Bondár, I., S. Myers, E. R. Engdahl, and E. Bergman (2004a). Epicenter

- accuracy based on seismic network criteria, *Geophys. J. Int.* **156**, no. 3, 483–496.
- Bondár, I., E. R. Engdahl, X. Yang, H. Ghalib, A. Hofstetter, V. Kirichenko, R. Wagner, I. Gupta, G. Ekström, E. Bergman, H. Israelsson, and K. McLaughlin (2004b). Collection of a reference event set for regional and teleseismic location calibration, *Bull. Seism. Soc. Am.* **94** (in press).
- Bratt, S., and T. Bache (1988). Locating events with a sparse network of regional arrays, *Bull. Seism. Soc. Am.* **78**, 780–798.
- Chang, A., R. Shumway, R. Blandford, and B. Barker (1983). Two methods to improve location estimates: preliminary results, *Bull. Seism. Soc. Am.* **73**, 281–295.
- Ekstrom, G., and A. Dziewonski (1998). The unique anisotropy of the Pacific upper mantle, *Nature* **394**, 168–172.
- Engdahl, E. R., and E. A. Bergman (2001). Validation and generation of reference events by cluster analysis, in *Proc. of 23rd Seismic Research Review*, Jackson Hole, Wyoming, 15 October.
- Engdahl, E. R., R. van der Hilst, and R. Buland (1998). Global teleseismic earthquake relocation with improved travel times and procedures for depth determination, *Bull. Seism. Soc. Am.* **88**, 727–743.
- Goes, S., R. Govers, and R. Vacher (2000). Shallow mantle temperatures under Europe from *P* and *S* wave tomography, *J. Geophys. Res.* **105**, 11,153–11,169.
- Herrin, E., and J. Taggart (1968). Source bias in epicenter determinations, *Bull. Seism. Soc. Am.* **58**, 1791–1796.
- Johnson, M., and C. Vincent (2002). Development and testing of a 3D velocity model for improved event location: a case study for the India–Pakistan region, *Bull. Seism. Soc. Am.* **92**, 2893–2910.
- 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. Buland (1995). Constraints on seismic velocities in the Earth from travel times, *Geophys. J. Int.* **122**, 108–124.

Table A1

Statistics on Relocation Improvement Using All Stations for Models CUB1, CUB2, and J362, Independently and Jointly (*Pn* and *P*)

Metrics	<i>Pn</i> CUB1	<i>Pn</i> CUB2	<i>P</i> J362	<i>Pn</i> and <i>P</i> CUB1' + J362	<i>Pn</i> and <i>P</i> CUB1 + J362	<i>Pn</i> and <i>P</i> CUB2 + J362
<i>Mislocation Given by Percentage of Events</i>						
Number of events	526	526	793	1234	1234	1234
III + IV: reduced (%)	49	41	65	52	52	50
II + V: increased (%)	29	39	26	30	31	32
Indecisive (%)	22	20	9	18	17	17
<i>Mislocation Categories Given by Percentage of Events</i>						
III: Moved inside GTX (%)	11	10	9	11	11	10
II: Moved outside GTX (%)	9	11	4	9	10	9
IV: Moved toward GTX (%)	38	31	56	41	41	40
V: Moved away from GTX (%)	20	28	22	21	21	23
I: Indecisive (%)	22	20	9	18	17	17
<i>Normalized Improvement/Degradation Given as Percentage of Events</i>						
Improvement \geq GTX (%)	15	10	38	26	26	26
Degradation \geq GTX (%)	4	8	6	7	5	7
Insignificant (%)	81	82	56	67	69	67
<i>Mislocation Given by Kilometer</i>						
Number of events	526	526	793	1234	1234	1234
Uncalibrated median mislocation (km)	7.1	7.1	8.3	7.1	7.1	7.1
Calibrated median mislocation (km)	6.6	7.2	6.1	5.9	5.7	5.9
Median improvement (km)	2.6	2.7	3.8	3.1	2.9	3.0
Median degradation (km)	1.7	2.2	1.8	2.0	1.9	2.1
<i>Ellipse Area and Coverage</i>						
Events ellipse area reduced (%)	100	100	100	100	100	100
Uncalibrated median ellipse area (km ²)	448	448	464	394	394	394
Calibrated median ellipse area (km ²)	253	253	322	233	233	232
Uncalibrated 90% coverage (%)	83	83	65	75	75	75
Calibrated 90% coverage (%)	76	70	75	76	77	75
<i>Origin Time (OT)</i>						
Events OT improved (%)	44	41	27	37	34	33
Uncalibrated median OT error (sec)	0.21	0.21	0.73	0.67	0.67	0.67
Calibrated median OT error (sec)	−0.52	0.74	0.97	0.81	0.90	0.97
Events OT error improved (%)	100	100	100	100	100	100
Uncalibrated median OT error (sec)	0.79	0.79	0.43	0.50	0.50	0.50
Calibrated median OT error (sec)	0.55	0.55	0.32	0.36	0.36	0.36
<i>Misfit (Standard Deviation of Observation, sd_{obs})</i>						
Events with sd_{obs} reduced (%)	65	61	84	67	79	79
Uncalibrated median sd_{obs} (sec)	1.19	1.19	0.84	1.04	1.04	1.04
Calibrated median sd_{obs} (sec)	1.11	1.13	0.77	0.96	0.92	0.94

CUB1 without the 0.75-sec baseline correction is denoted CUB1'.

- Lomax, A., J. Virieux, P. Volant, and C. Berge (2000). Probabilistic earthquake location in 3D and layered models: introduction of a Metropolis–Gibbs method and comparison with linear locations, in *Advances in Seismic Event Location*, C. H. Thurber and N. Rabinowitz (Editors), Kluwer, Hingham, Massachusetts, 101–134.
- Mooney, W. D., G. Laske, and G. Masters (1998). CRUST5.1: a global model at 5 degrees by 5 degrees, *J. Geophys. Res.* **102**, 727–748.
- Murphy, J., W. Rodi, M. Johnson, J. Sultanov, T. Bennett, M. N. Toksoz, C. Vincent, V. Ovtchinnikov, 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. of 24th Seismic Research Review: Nuclear Explosion Monitoring—Innovation and Integration*, Ponte Vedra Beach, Florida, 17–19 September.
- Myers, S. C., and C. Schultz (2000a). 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. Schultz (2000b). Calibration of seismic travel time using events with seismically determined locations and origin times, *EOS* **81**, F845.
- Ritzwoller, M. H., and A. L. Levshin (1998). Eurasian surface wave tomography: group velocities, *J. Geophys. Res.* **103**, 4839–4878.
- Ritzwoller, M. H., N. M. Shapiro, M. P. Barmin, and A. Levshin (2002). Global surface wave diffraction tomography, *J. Geophys. Res.* **107**, ESE4.
- Ritzwoller, M. H., N. M. Shapiro, A. Levshin, E. A. Bergman, and E. R. Engdahl (2003). The ability of a global 3-D model to locate regional events, *J. Geophys. Res.* **108**, no. 2353, ESE 9-1–ESE 9-24.
- Rodi, W., C. Shultz, W. Hanley, S. Sarkar, and H. Kuleli (2002). Grid-search location methods for ground-truth collection from location and regional seismic networks, in *Proc. of 24th Seismic Research Review: Nuclear Explosion Monitoring—Innovation and Integration*, Ponte Vedra Beach, Florida, 17–19 September.
- Sambridge, M., and K. Gallagher (1993). Earthquake hypocenter location using genetic algorithms, *Bull. Seism. Soc. Am.* **83**, 1467–1491.
- Sambridge, M., and B. L. N. Kennett (2001). Seismic event location: non-linear inversion using a neighborhood algorithm, *Pure Appl. Geophys.* **158**, 241–257.
- Shapiro, N. M., and M. H. Ritzwoller (2002a). Monte-Carlo inversion of broad-band surface wave dispersion for a global shear velocity model of the crust and upper mantle, *Geophys. J. Int.* **151**, 88–105.
- Shapiro, N. M., and M. H. Ritzwoller (2002b). Thermodynamic constraints on seismic inversions, Group-2 Consortium report, Arlington, Virginia.
- Trampert, J., and J. Woodhouse (1995). Global phase velocity maps of Love and Rayleigh waves between 40 and 150 seconds, *Geophys. J. Int.* **122**, 675–690.
- Yang, X., I. Bondár, K. McLaughlin, R. North, and W. Nagy (2001a). Path-dependent regional phase travel-time corrections for the International Monitoring System in North America, *Bull. Seism. Soc. Am.* **91**, 1831–1850.
- Yang, X., I. Bondár, K. McLaughlin, and R. North (2001b). Source-specific station corrections for regional phases at Fennoscandian stations, *Pure Appl. Geophys.* **158**, 35–57.

Appendix

Relocation Statistics

In this Appendix, we present detailed relocation statistics on improvements versus degradation on a case-by-case basis for P_n calibrations only (CUB1 and CUB2, respectively), P calibration only (J362), and the joint P_n and P

Table A2

Statistics on Relocation Improvement Given by Percentage of Events Using SSNBs for CUB1, CUB2, and J362

Metrics	P_n	P_n	P_n
	CUB1	CUB2	J362
<i>Mislocation</i>			
SSNB centroid bias:			
Number of events	276	273	331
Reduced (%)	51	50	62
Increased (%)	30	33	28
Indecisive (%)	19	17	10
SSNB 10-station locations:			
Number of events	4856	4856	6836
Reduced (%)	52	52	57
Increased (%)	34	35	34
Indecisive (%)	14	13	9
All-station locations (seed):			
Number of events	276	273	331
Reduced (%)	56	48	58
Increased (%)	27	36	28
Indecisive (%)	17	16	14
<i>Normalized Improvement/Degradation</i>			
SSNB centroid bias:			
Number of events	276	273	331
Improvement \geq GTX (%)	17	18	25
Degradation \geq GTX (%)	8	11	9
Insignificant (%)	75	71	66
SSNB 10-station locations:			
Number of events	4856	4856	6836
Improvement \geq GTX (%)	5	5	3
Degradation \geq GTX (%)	1	1	1
Insignificant (%)	94	94	96
All-station locations (seed):			
Number of events	276	273	331
Improvement \geq GTX (%)	18	17	26
Degradation \geq GTX (%)	8	12	7
Insignificant (%)	74	71	67
<i>Mislocation Categories</i>			
SSNB centroid bias:			
Number of events	276	273	331
Moved inside GTX (%)	11	10	9
Moved outside GTX (%)	6	8	4
Moved toward GTX (%)	40	40	53
Moved away from GTX (%)	24	25	24
Indecisive (%)	19	17	10
SSNB 10-station locations:			
Number of events	4856	4856	6836
Moved inside GTX (%)	11	13	6
Moved outside GTX (%)	7	7	4
Moved toward GTX (%)	41	39	51
Moved away from GTX (%)	27	28	30
Indecisive (%)	14	13	9
All-station locations (seed):			
Number of events	276	273	331
Moved inside GTX (%)	12	12	8
Moved outside GTX (%)	6	8	3
Moved toward GTX (%)	44	36	50
Moved away from GTX (%)	21	28	25
Indecisive (%)	17	16	14

calibrations (CUB1 + J362 and CUB2 + J362, respectively). Tables A1 through A5 show the results for all-station and SSNB relocations, respectively. These correspond to the

summaries given in Tables 3–4 for all-station and in Tables 5–6 for SSNB relocations. Statistics include mislocations, median error ellipse area, origin time error, the standard de-

Table A3
Statistics on Relocation Improvement Given by Percentage of Events Using SSNBs for CUB1 + J362, CUB2 + J362, CUB1 + IASP, CUB2 + IASP, and IASP + J362

Metrics	<i>P_n</i> and <i>P</i> CUB1 + J362	<i>P_n</i> and <i>P</i> CUB2 + J362	<i>P_n</i> and <i>P</i> CUB1 + IASP	<i>P_n</i> and <i>P</i> CUB2 + IASP	<i>P_n</i> and <i>P</i> IASP + J362
<i>Mislocation</i>					
SSNB centroid bias:					
Number of events	355	356	359	357	355
Reduced (%)	58	56	49	45	48
Increased (%)	25	28	35	40	35
Indecisive (%)	17	16	16	15	17
SSNB 10-station locations:					
Number of events	7145	7142	7175	7158	7127
Reduced (%)	57	56	47	45	48
Increased (%)	33	35	34	37	41
Indecisive (%)	10	9	19	18	11
All-station locations (seed):					
Number of events	355	356	359	357	355
Reduced (%)	48	49	46	41	45
Increased (%)	33	34	35	44	35
Indecisive (%)	19	17	19	15	20
<i>Normalized Improvements/Degradation</i>					
SSNB centroid bias:					
Number of events	355	356	359	357	355
Improvement \geq GTX (%)	25	24	17	22	17
Degradation \geq GTX (%)	7	10	6	11	8
Insignificant (%)	68	66	77	67	75
SSNB 10-station locations:					
Number of events	7145	7142	7175	7158	7127
Improvement \geq GTX (%)	3	3	3	3	3
Degradation \geq GTX (%)	1	1	1	1	1
Insignificant (%)	96	96	96	96	96
All-station locations (seed):					
Number of events	355	356	359	357	355
Improvement \geq GTX (%)	22	22	13	14	16
Degradation \geq GTX (%)	7	7	4	8	6
Insignificant (%)	71	71	83	78	78
<i>Mislocation Categories</i>					
SSNB centroid bias:					
Number of events	355	356	359	357	355
Moved inside GTX (%)	9	9	6	8	6
Moved outside GTX (%)	3	5	5	8	4
Moved toward GTX (%)	49	47	43	37	42
Moved away from GTX (%)	22	23	30	34	31
Indecisive (%)	17	16	16	13	17
SSNB 10-station locations:					
Number of events	7145	7142	7175	7158	7127
Moved inside GTX (%)	8	9	7	7	6
Moved outside GTX (%)	5	6	5	7	5
Moved toward GTX (%)	49	47	40	38	42
Moved away from GTX (%)	28	29	38	39	36
Indecisive (%)	10	9	10	9	11
All-station locations (seed):					
Number of events	355	356	359	357	355
Moved inside GTX (%)	10	11	8	9	10
Moved outside GTX (%)	9	10	8	12	7
Moved toward GTX (%)	38	38	38	32	35
Moved away from GTX (%)	24	24	27	32	28
Indecisive (%)	19	17	19	15	20

variation of observations (sd_{obs}), and the SSNB centroids (location bias). Results of partially calibrated, joint regional, and teleseismic relocations are also listed. For SSNB, 10-station locations and all-station locations (seed) are both given. Calibrated and uncalibrated cases are referred to as “cal” and “uncal,” respectively. In general, event locations are significantly improved using travel-time calibration.

In Table A1, CUB1 without the 0.75-sec baseline correction is also given (denoted CUB1'). As discussed previously, we applied a bulk shift of 0.75 sec to the CUB1 regional Pn calibrations in order to correct the baseline difference between the CUB1 regional Pn and J362 teleseismic P travel-time models. To assess the effect of this practice, we also relocated events using joint CUB1 Pn and J362 P calibrations without the bulk correction. Direct comparisons between the results with and without the baseline shift show that the results are somewhat better when using the baseline correction (not tabulated here). Slightly more events are improved (27%) than degraded (24%), but about half of the events (49%) do not resolve the difference. This confirms that applying the baseline shift is beneficial.

In this Appendix, we also include the detailed statistics on GT events west of 40° E (Table A6). These events are shown in Figure 12, mostly in the Middle East and the Mediterranean. In this region, more events are improved by CUB1 than CUB2, and J362 and joint regional and teleseismic calibration also perform poorly.

Science Applications International Corporation
1300 N. 17th Street
Arlington, Virginia 22209
(X.Y., I.B., J.B., H.I., K.M.)

University of Colorado
Department of Physics
Boulder, Colorado 80309
(M.R., N.S.)

Harvard University
Department of Geology
Cambridge, Massachusetts 02138
(M.A., G.E.)

Manuscript received 23 July 2003.

Table A4
Statistics on Location Bias, Mislocation, Area of Error Ellipse, and Coverage of the 90% Error Ellipse Using SSNBs for CUB1, CUB2, and J362

Metrics	CUB1		CUB2		J362	
	uncal	cal	uncal	cal	uncal	cal
SSNB:						
Mean bias (km)	8.4	7.2	8.2	7.0	9.2	7.9
Mislocation (km)	8.8	7.6	8.8	7.6	11.1	9.9
Error ellipse area (km ²)	1522	745	1522	730	1953	1133
Coverage (%)	99	96	99	93	92	85
All stations (seed):						
Mislocation (km)	7.9	6.8	7.8	7.3	8.2	7.1
Error ellipse area (km ²)	518	293	529	288	411	274
Coverage (%)	80	79	81	71	63	64

Table A5
Location Bias, Mislocation, Area of Error Ellipse, and Coverage of the 90% Error Ellipse Using SSNBs for CUB1 + J362, CUB2 + J362, CUB1 + IASP, CUB2 + IASP, and IASP + J362

Metrics	CUB1 + J362		CUB2 + J362		CUB1 + IASP		CUB2 + IASP		IASP + J362	
	uncal	cal	uncal	cal	uncal	cal	uncal	cal	uncal	cal
SSNB:										
Mean bias (km)	7.7	6.6	7.7	6.4	7.8	7.3	7.8	7.5	7.7	7.1
Mislocation (km)	9.9	8.7	9.9	8.6	9.9	9.1	9.9	9.2	9.9	9.5
Error ellipse area (km ²)	1815	971	1815	959	1816	1420	1816	1461	1814	1300
Coverage (%)	97	92	97	92	96	95	97	95	97	93
All station (seed):										
Mislocation (km)	6.6	5.7	6.6	5.9	6.6	6.4	6.6	6.9	6.6	6.0
Error ellipse area (km ²)	250	171	251	167	254	206	254	200	253	203
Coverage (%)	71	69	71	68	70	64	71	60	71	74

Table A6
Relocation Statistics using All Stations for GT Events West of 40° E.

Mislocation	<i>P_n</i>			<i>P</i>	<i>P_n and P</i>		
	CUB1	CUB2	CUB1	J362	CUB1 + J362	CUB2 + J362	CUB1 + J362
Model 2	IASP91	IASP91	CUB2	IASP91	IASP91	IASP91	CUB2 + J362
Number events	415	415	508	179	508	508	508
Reduced (%)	41	35	41	40	35	33	41
Increased (%)	32	41	33	50	39	43	33
Indecisive (%)	27	24	26	10	26	24	26
Moved inside GTX (%)	10	11	10	5	8	11	10
Moved outside GTX (%)	11	14	11	5	12	14	11
Moved toward GTX (%)	31	24	31	35	27	22	31
Moved away from GTX (%)	21	27	22	45	27	28	22
Model 1 median mislocation (km)	6.1	6.1	6.1	11.4	6.5	6.5	6.5
Model 2 median mislocation (km)	6.1	6.2	6.2	11.2	6.4	6.4	6.4