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

Xiaoping Yang[1], István Bondár[1], Joydeep Bhattacharyya[1], Michael Ritzwoller[2], Nikolai Shapiro[2], Michael Antolik[3], Göran Ekström[3], Hans Israelsson[1], and Keith McLaughlin[1]

[1] Science Applications International Corporation

[2] University of Colorado at Boulder

[3] Harvard University

SUMMARY

Recently, significant improvements in global seismic velocity models have Rate that significant improvements can be achieved using 3D city band models for 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 Mediste generated from a "constrained bootstrapping" technique. The advantages of validate event location improvements using this new data set by applying the regional and teleseismic model-

demional (3D) Earth structures.

locating small events with sparse network data. Besides relocating events using all station arrivals, a subset of the GT events was also relocated using controlled station geom

this approach include: (1) generating simulated sparse networks (Simulated Sparse Network Bulletin or SSNB), (2) increasing the statistical power of the tests, (3) reducing the effect of correlated errors to ensure valid 90% error ellipse coverage statistics, and (4) measuring location bias due to un-modeled three-dim

With respect to the GT events, we compared event relocations, with and without t time calibrations, considering statistics of mislocation, error ellipse area, 90% coverage, origin time bias, origin time errors, and misfit. Relocations of more than 1000 GT0-GT10 reference events show significant reductions in location bias and uncertainty. Pn and/or P calibration reduces mislocation for between 60% and 70% of the events. Joint regional Pn and teleseismic P travel-time calibration provided the largest location improvements, and approximately achieved GT5 accuracy levels. Due to correlated

based travel-time calibrations (independently and jointly) to

rranean, North Africa, the Middle East, and Western Eurasia. In this study, we

errors, calibrated event locations using large numbers of stations have deficient 90% error ellipse coverage. However, the coverages derived from the model errors are appropriate for sparse regional and teleseismic networks. This validation effort demonstrates that the global model-based travel-time calibrations of Pn and teleseismic P travel-time reduce both location bias and uncertainty over wide areas.

Key words: traveltime, seismic velocities, epicenter, Europe, the Mediterranean, and Western Eurasia

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 the calibration is to improve location accuracy and reduce the uncertainty while maintaining targets for 90% coverage, i.e. 90% of the true locations are found inside the respective error ellipses. To demonstrate these goals, reliable ground truth (GT) events are required with high confidence epicentral estimates and high quality arrival times.

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 one-

1991). They amount to two- (2D) and three-dimens

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

given for each observed arrival and both errors are used to weight the arrivals 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. for Eastern Asia (Asrbrust et al., 2002). Th

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 observations has been used in developing calibrated travel-

dimensional seismic velocity models (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 mostly focused on calibrating regional Pn phases since regional travel times are sensitive to small-scale sensitivity and deviate from IASP91 travel-time tables more severely

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 one-second bias in a teleseismic travel time is

uncertainty of the underlying model. Another error, measurement error, is

more severe than that caused by the same bias in a regional travel time. Teleseismic phases comprise an important fraction of IMS arrival data motivating joint calibration of regional and teleseismic phases. Global 3D velocity models have been developed (J362 model and its predecessors, e.g. SP12 model) and demonstrated to improve event locations (Antolik et al., 2001; 2003) with teleseismic phases alone.

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 goodness, including comparisons between predicted versus empirical path corrections (e.g. Ritzwoller et al., 2003), cross-

and Vincent, 2002), and evaluations of location bias using randomly selected subsets of

three ways (e.g. Antolik et al., 2001). Effective testing depends on the accuracy, the amount, and the 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, ndef) 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 the model development should be used in validation testing. This requirement is particularly difficult to fulfill for

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

validation of RMS residual reductions (e.g. Johnson   5

kriged calibrations. Since many events are clustered, more objective testing would

require leave-one-path-

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.

Group-2

set of reference GT events in Europe, Africa, and Eurasia (Bondár et al., 2003b). The

Consortium database includes nearly 2000 GT0-GT10 origins with known or well-

estimated location accuracy. This database provides an excellent set of arrivals and

empirical path corrections for model validation. Comparisons between predicted

empirical path corrections are given in Ritzwoller et al. (2002a) and Bhatacharryya et al.

(2003). This paper 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.

Many GT events used in the previous relocation studies (e.g. Ritzwoller, et al.,

2002a; Antolik et al., 2001) and in the Group-2 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 calibration

but instead Calibrations Consortium recently collected data set high quality

improvements/deteriorations. This paper 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 previously been used 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 Bulletin (SSNB) 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. It has recently been argued that the size of the so-called "secondary azimuthal gap" (sgap), i.e. 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-regional and teleseismic travel-time calibrations. In addition to relocating GT events using all available quality arrival, 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-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°x2° 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 sc 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 $dln(Vp)/dln(Vs) \sim 0.5$, in which perturbations are taken relative to the S and P velocities (Vs and Vp, 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 (2002). In essence, given the mineralogical composition, the bases for the anelastic

8

his 3D global model.

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 Vs model is converted to temperature, and then converted to Vp. In CUB2 this transformation has not yet been regionally tuned; mineralogical composition is homogeneous across the region of study and 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) show event location improvements using regional calibrations derived from t

To calculate Pn calibration tables (SSSCs) based on CUB1 or CUB2, we used a 2D raytracer that handles refracted and reflected compressional waves in 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 second. After subtracting the predicted IASP91 travel times, the resulting travel-time corrections were interpolated to a 1°×1° 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 CUB Pn SSSCs, as shown in Figure 2, using travel-time misfits obtained from the events selected from the 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 3-D 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 (Figure 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.

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 GT0-10 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 ray tracer. Because the perturbation ray-tracer is not valid in regions of triplication or diffraction, we limited our study to travel-time corrections between 25° and 95°, bounded by West longitude -80° to North latitude and -40°E to 100° E, for 5° N and 80° N. 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 $2°$ by $2°$ rectangular grid

2000 stations. Figure 1 shows an example of teleseismic P SSSCs for ABKT. modeling errors, based on our experience with the regional calibrations we adopted a scaled version of the baseline IASP91 teleseismic P error model (see Figure 2).

Baseline Differences between Regional and Teleseismic Travel-Times

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 if biases exists between the regional and teleseismic calibrations and whether any baseline shifts are warranted. To assess the baselines of each model, comparisons were made between regional and teleseismic phases for travel-time residuals of over 700 GT events 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 5 or more Pn phases and 5 or more P phases). For each GT event, residuals of Pn and P phases were calculated using observed and predicted travel times from each model for fixed 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 the smallest standard deviation, but an adjustment can 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 uncalibrated IASP91 Pn may be a bad match.

A baseline correction can be applied to any of the models to reduce the mean differences, but the standard deviations reveal event-by-event baseline variations that cannot be remedied by simple bulk adjustments to the travel-time tables. For our purposes we chose to reduce the CUB1 baseline shift by simply applying a bulk correction of 0.75 sec. CUB2 requires no baseline shifts. We have examined the geographic patterns of these biases between models and we found that they are systematic and reflect un-modeled differences between the broad tectonic provinces of shield and platform regions versus tectonically active regions. These systematic deviations suggest that future 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. 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. We evaluated location improvements by comparing the calibrated results with uncalibrated results (IASP91 baseline) using a set of well-statistical metrics. In all our relocation tests, depth was fixed to zero to simulate a nuclear monitoring scenario. In our data set all events are either near the surface (explosions, mine collapses, etc.) or at crustal depths (earthquakes).

Data Selection for Event Relocations

We selected all GT0-GT10 events in the Group-2 datab

relocated (1) using only regional phases, (2) using only teleseismic phases, and (3) using

both regional and teleseismic phases. To minimize ambiguity in arrival data, we only

used Pn arrivals within 15º and P arrivals between 25º and 97º. We selected events in

each of the three groups of events with a secondary azimuthal gap (sgap) less than 160º.

The sgap is the largest azimuthal gap that would result from deleting one station. This

guarantees the azimuthal gap (azgap) is less than 160º and guarantees a minimum of 5

stations. This selection criterion produces locations less susceptible to single station

outliers (Bondár et al., 2003a).

In the IMS event location algorithm, reproduce events are used to weigh and are reduce if (and number of defining phas

calculate error ellipses. The errors are partitioned into model errors 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 and a measurement error of 1.0 sec was therefore assigned to those arrivals

by default.

Pn Data Set

Figure 3 shows 526 GT0-GT10 events with sgap < 160º for Pn arrivals within 15º.

This data set provides a fairly good geographic coverage across the region. They are

well-

ndef of 6), azgap of 76º, and sgap of 99º. Most of the events are GT5 (87%), and 64% of

the events were 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

13

ase that can be reliably

cluster analyses (Engdahl and Bergman, 2001). The rest (10%) of the GT events are nuclear explosions, chemical explosions, or mining events (GT0-GT2). Pn SSSCs were calculated for 1098 stations (CUB1 and CUB2).

Teleseismic P Data Set

Figure 4 shows 793 GT0-GT10 events with sgap < 160º for P arrivals between 25º and 97º. These events were very well recorded, with a median ndef of 100, azgap of 68º and sgap of 85º. Teleseismic P SSSCs (J362) were calculated for 2821 stations (sources 25º to 97º from each station). 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 eve are nuclear explosions located at historic test sites. Of the remaining events, about 17% are so-called peaceful nuclear explosions (PNE's) scattered across the former Soviet Union.

Pn and P Data Set

There are 1234 GT0-GT10 events with sgap < 160º with Pn arrivals within 15º and P arrivals between 25º and 97º in our dataset. There are 28 GT0, 328 GT1 (27%), 14 GT2, 650 GT5 (53%), and 214 GT10 (17%) events with 111,498 P phases from 2823 stations and 39,017 Pn phases from 1166 stations. The events were also well recorded, with a median ndef of 84, azgap of 58º and sgap of 76º. Most of the events (80%) have more teleseismic P than Pn arrivals with a median ratio of P to Pn ndef about 3. The maximum azimuthal gap is 154º and 87% of the events have an azimuth gap less than 100º. A large number of the events (24%) are explosions at historic test sites or PNE explosions scattered across the former Soviet Union (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 with sgap < 160º using Pn and P together compared to P alone.

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 over emphasizing a single key station/arrival. We chose to simulate sparse network geometries with 10 stations and secondary azimuthal gaps less than 160, called Simulated Sparse Network Bulletins (SSNBs). We find acceptable SSNB geometries by building the 'sgap tree', illustrated by the fictitious example in Figure 5. The sgap tree is a hierarchical binary tree where each node represents a station that splits the remaining azimuthal gap in a way 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 sgap 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 above. 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. Such that statistical certainties it also provides 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 localization tests, we selected well-recorded GT0 to GT1 representative sub-networks 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 mislo

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 gene

events to provide good geographical coverage. We limited the number of events from the same event cluster to 10 to avoid over-representing clusters with a large number of events. Sampling a cluster by several events allowed us to examine the consistency of location bias estimates and 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 events in total, provided the basis for the evaluation of jointly regional and teleseismic calibration travel times.

For direct comparisons, SSNB seed events were also relocated using all stations, denoted as

paper 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 (GTX) 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 GTX. Furthermore, we split the category of events with both calibrated and uncalibrated mislocation greater than 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 uncertainty, 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) defined above, we compare the overall percentages of events "Improved" (III+IV), "Deteriorated" (II+V), and "Indecisive" (I). We also compare the median mislocations with and without

17

"all-station (seed)". Note that in general the all-station cases discussed in this

travel-time calibrations, and the median of the improved mislocations (better) versus the median of the deteriorated mislocations (worse).   It is important to verify the improvements are generally larger than the instability of locations. Performance is decisive when

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 w.r.t. 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, i.e. normalized improvement and deterioration. Different from the categories defined as in: Table 2 that show how well the events are located relative to GTX, these metrics measure by how much the events have improved or deteriorated. Events may have im

the normalized improvement/deterioration is within unity.

The coverage parameter, E, is defined as the mislocation normalized to the 90% error ellipse in the direction of mislocation. If the GT location lies within the error ellipse then 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

$$E = x^2/(smajax^2+GTX^2)+y^2/(sminax^2+GTX^2),$$ (1)

where x and y are location components in the coordinate system defined by the semi-major (S_{majax}) and semi-minor (S_{minax}) axes of the ellipse. Under the assumptions of Gaussian and independent errors, E follows a χ^2 distribution with two-

freedom. The 90th percentile should correspond to E = 1.0 if the *a-priori* error models are properly calibrated. An event is "covered" if E <= 1.0. We define "coverage" for a test data set as the percentage of events with E <= 1.

Location bias can be estimated from a number of realizations provided by SSNB relocations for an event. In such a case the distance between the GT location and the centroid of all the realizations serves as an estimate of the location bias (with estimated uncertainty) due to un-modeled lateral heterogeneities.

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

- Reduce median mislocation,

- Improve more events than deteriorate events,

- Provide median improvement larger than median deterioration,

- Reduce median error ellipse area,

- Provide coverage (% of events with E <= 1) of 90%,

- Reduce misfit (standard deviation of residuals), and

- Reduce origin time errors.

EVENT RELOCATIONS USING REGIONAL Pn 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 Simulated Sparse Network Bulletins (SSNBs).

All-Station Relocation

We relocated all 526 GT events in the Pn data set (Figure 3), 793 GT events in the P data set (Figure 4), and 1234 GT events in the Pn and P data set using all avail stations. The GT events in all three data sets (Pn, P, and Pn & P) are generally well located ensuring good azimuthal distributions (azgaps between 58° and 76°, sgaps 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 deterioration for Pn only (CUB1), P (J362) only, and jointly Pn 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 deteriorated for CUB1 and J362 models. There are between 20 and 40% more events improved than deteriorated. For example, CUB1 improves 48% of the events versus deteriorating 29%. Between 10% and 20% of the events are located within GTX uncertainty 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 uncertainty is too large to discern a significant mislocation. This underscores the importance of test events with location accuracies better than 5 km for relocation testing since events with large uncertainties contribute little to the test except to verify that calibrations do no harm to these already well-located events.

Table 3 also shows the details of each mislocation category. For both cases that mislocations are inside and outside GTX, more events are improved than deteriorated. This is particularly striking for J392 with more than twice amount of events improved than deteriorated (9% versus 4% inside GTX, and 56% versus 22% outside GTX).

able

Calibration moves events inside than outside GTX uncertainty of the GT location ("Improved at GTX versus Deteriorated at GTX"). Calibration improves more events that lie outside GTX uncertainty ("Improved Outside GTX versus Deteriorated Outside GTX").

Also shown in Table 3 are percentages of events with improvements and deteriorations by more than GTX; the remaining events are indecisive. CUB1 and J362 locations are closer to the GT locations than IASP91 by more than the GTX for significantly more events than they are further away from the GT location by the same uncertainty. For J362 38% of the events improved by more than GTX compared to 6% of the events deteriorated by more that GTX.

Summary statistics of median mislocation, median error ellipse area, 90% coverage, origin time error, and standard deviation of observations (sdobs) are summarized in Table 4. The median mislocations all range between 6 and 8 km, with and without calibration. Overall, calibration reduced median mislocation by 10% to 20%. By far, teleseismic calibration has delivered the largest fractional improvement with a 27% in the median mislocation. Using both regional and teleseismic calibrations the median mislocation has been reduced to nearly the level of GT5 uncertainty (5.7 km). As shown in the Appendix, there is large mislocation improvement using CUB1 but CUB2 does not show overall significant improvement. Unlike regional calibration alone, results with J362+CUB1 are only somewhat better than those with J362+CUB2 since teleseismic phases play a dominant role for most events in this dataset when teleseismic and regional phases are jointly. CUB1 location results are slightly better with the baseline shift. A

more direct comparison of CUB1 versus CUB2 and the baseline correction to CUB1 are discussed in the next Section.

Figures 7a-c

each data set. For all 526 GT0-GT10 events in the Pn data set, the median CUB1 mislocation is 6.6 km. The largest CUB1 improvement occurred at the 95th percentile, with mislocations reduced by 27% and demonstrates that Pn calibration made the grea 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 30th 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 (MORT), did not significantly improve using J362. For all 1234 events in the Pn and P data set, there is significant location improvement using CUB1 and J362 as those in independent calibration with a median mislocation of only 5.7 km. As in the P only case, the largest improvement with joint calibration occurred at the 25th percentile, as the most poorly located events did not improve. The cumulative distributions of the improvements/deteriorations normalized to GTX are shown in Figures 7d-f and illustrate that improvements (positive) are generally larger than deteriorations (negative), as given in Table 3 (last two rows).

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. As expected, there are more events improved than deteriorated with jointly calibrations, compared with events improved using regional or teleseismic calibration alone. Median mislocation is

shows the cumulative mislocations with and without calibration for

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% to 16%).

Table 4 also shows that calibration significantly reduces median error ellipse ar origin time error, and the standard deviation of observations (sdobs). 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 calibration misfit is reduced for the majority of events and the median misfit (sdobs) is reduced. The median origin time bias is generally increased (mostly positive) indicating that these models are biased slightly fast. 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 (most better than 5 km), most origin times (including most explosions) are inferred from seismic data and are not as reliable. In all cases the estimated origin time error is reduced as a result of a reduction in misfit and smaller a-priori errors.

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 large ndef events this is not too surprising. Figure 8 shows the coverage parameter, E, versus number of defining

ea.

address the problem jointly with sufficient Pn and P coverage for events. That likelihood that coverage is insufficient increases with increasing ndef. The uncalibrated 90% coverage is prope station relocation tests and the calibrated 90% coverage is maintained for ndef = 27. It was 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.

Simulated Sparse Network Bulletin 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

number of arrivals as well as to improve the statistical power of the relocations since a large number of the events in the all-

shows examples of 10 station SSNB relocations for a GT5 earthquake in Pakistan on 02/14/1977 and a GT1 PNE in Russia on 08/20/1972. 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 above.

Table 5 summarizes the overall percentages of events improved, deteriorated or remaining indecisive w.r.t. IASP91 due to the application of calibrated travel times; More detailed statistics are given in the Appendix. The statistics of the SSNB centroids represent the reduction in location bias. The SSNB 10-station locations represent the

24

reduction in mislocation obtained from the constrained bootstrapping. For comparison, the results from the all-station (seed) locations of the same SSNB seeds are also compared. More events are improved than deteriorated. Calibrated travel-times reduce both bias and mislocation. When mixing calibrated travel-times with uncalibrated ones, the improvements tend to fade (Appendix).Thus it is important to calibrate both regional and teleseismic phases whenever possible. We find, however, that mixing calibrated and uncalibrated Pn and teleseismic P arrivals is generally favorable to ignoring the

SSNB bias estimate provided that suitable a priori model clusters are available to properly weight the calibrated and uncalibrated data.

Figure 10 shows the mean mislocation from the all-station (seed) locations (top) compared to the mean location improvements from the 10-station sparse network solutions (middle) 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). 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 using calibrated or uncalibrated travel times. 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 significant effect on our results. While using all-stations (seed) 57% of

travel-times, the

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 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 s

striking difference between the median area 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. Since more than 100 stations typically record the SSNB seed events, many arrivals sample similar ray paths. Thus the assumption of independent errors made by the location algorithm is violated, leading to unrealistically small er ellipses. We further discuss the issue of correlated errors in a following Discussion section.

We compared the percentage of events with improved/deteriorated locations due to calibrated travel-times for the regional (CUB1, CUB2), teleseismic (J362), and the jointly regional and teleseismic (CUB1+J362, CUB2+J362) cases. In all cases calibration improved roughly 2/3's of the events (60-70%). CUB1 improved somewhat more events than CUB2 when only regional phases were used or when jointly with J362 teleseismic travel-

reduction. The statistics for the area of error ellipse and coverage do not significantly differ between CUB1 and CUB2 either alone, or jointly with J362.

tmalleprediectiornsixir@nutheallbbhetehaand CCalTB2atteiddasvselienthtcheslarNotdothattion bias

Compar

all combinations of location (Pn, P, and Pn & P) shows that in general calibrated travel-times reduce location bias. Median bias reduction (Figure 11a) and percentages of events impro

consequences of using calibration and relocation using regional Pn alone, teleseismic P alone, and joint Pn & 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. Pn & P location improves between 61% and 74% of events with between 21% and 28% reduction in the median bias. Calibrated 10-station spares network Pn and P median mislocations approach the GT5 accuracy of the majority of the test events. Adding calibration and combining Pn 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 deterioration 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.

DISCUSSIONS

Three topics touched upon in the Section above deserve some additional discussion. In this section we examine the results of the baseline 0.75-sec shift applied to CUB1, directly compare CUB1 versus CUB2, and particularly examine the issue of correlated errors in more detail.

isechs/ Figure constrained bootstratech for esults info the distforre matrix o that compares the

Baseline Shift between Regional and Teleseismic Calibrations

As discussed above, we applied a bulk shift of 0.75 sec to the CUB1 regional Pn calibrations in order to correct the baseline between the CUB1 regional Pn and J362 teleseismic P travel-

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. Slightly more events are improved (27%) than deteriorated (24%), but about half of the events (49%) do not resolve the difference. This confirms that applying the baseline shift is beneficial.

Direct Comparisons between CUB1 and CUB2

It appears that the CUB1 model provides better location calibration for our target region but that CUB2 may perform somewhat better in Asia. Detailed statistics are given in the Appendix. More events are improved by CUB1 than CUB2 (50% versus 28% for regional only, and 36% versus 28% for regional and teleseismic jointly). These events are mostly west of 40° E, and are generally located in the Middle East and the Mediterranean (Figure 12). Direct comparisons show besides CUB2, J362 also performs poorly, and CUB2+J362 calibration deteriorates more events than it improves in this region (Appendix).

The majority of the GT5 earthquakes in the Pn 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 not vetted using cluster analysis. Initially we might suspect that because these events are less well constrained, that the better performance of CUB1 might be illusory. However, it is also

time models. To assess the effect of this practice, we also relocated

plausible that, upper mantle mineralogies differ systematically east-to-west, such that the different 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 Pn 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 many of lesser GTX quality. They 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 CUB1 and CUB2 predicted path effects (calibrations) are the poorest in Greece highlighting where these two models are the most different. In this area the CUB1 model is faster than IASP91 while the CUB2 model is slower than IASP91.

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 would predict. The fact that J362 actually performs worse for this subset of events than would be predicted by random chance suggests that the root cause is not due to random errors in the GT event locations but rather un-modeled structure. Also since Bhattacharya et al. (2003) finds more clusters with statistically significant correlations between empirical path corrections and the CUB1 model-based path corrections than with those derived from CUB2 suggest 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 calibration. While the problem may be due to underestimated *a priori* errors, the SSBN bootstrapping results reveal that this is due to the breakdown in the assumption of uncorrelated errors.

Herrin and Taggart (1968) have shown by analyzing the Long Shot nuclear explosion that a large number of arrivals sampling similar ray paths along un-modeled three-di

large number of ray paths travel through a subducted oceanic slab with 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 assumed 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 principal of kriging; similar paths through the Earth have similar path effects.

Although in the past few years several non-linear hypocenter location methods were developed to account for non-linearity and non-Gaussian error distributions (e.g. Billings, 1994; Billings *et al*., 1994; Sambridge and Gallagher, 1993; 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 developing the constrained bootstrapping technique was to minimize the effect of correlated errors on the validation test.

To demonstrate how correlated errors may deteriorate coverage, intr

location bias and produce unrealistic error ellipses, we performed an SSNB study on a

GT1 Chinese underground nuclear explosion on 10/07/1994. A total of 562 stations at

distances between 25º and 97º recorded this event. The station distribution is far from

uniform and dominated by stations in Europe, Japan and California.   We generated

SSNBs with the number of defining phases (ndef) ranging from 6 to 400 stations and

with each ndef 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 (sgap) as a function of ndef.   The location bias

increases (almost monotonically) with increasing number of stations regardless of

calibrat

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 ignores the fact that travel-time predictions are correlated

along similar ray paths. Note that even if the J362 model precisely accounted for all

three-dimensional 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. After this point adding further stations simply increases

data redundancy.   However, the area of the error ellipse shrinks indefinitely with

increasing number of arrivals because the covariance matrix is calculated incorrectly by

conduc(Figure 13). As more and more stations are added to the solution, the location

assuming 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 offer any clue 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 it may not always be the case. Location bias may or may not increase as more arrivals are added to the solution. The actual behavior depends on how the net weights of closely clustered stations are balanced with respect to the un-modeled 3D Earth structures.

Correlated errors will always produce unrealistic error ellipses for events located by 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-stations (seed) locations and the 10-station sparse network 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 90th percentile of coverage is still low for the 10-station SSNB tests for the calibrated teleseismic case. A conservative re-scaling 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 90th 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.

A fundamental approach in addressing the 90% coverage problem is to construct the covariance matrix in a way that it takes into account the correlation between similar ray paths. Chang et al. (1983) proposed a scheme to estimate the covariance matrix in the presence of inter-correlated errors. Bayesian kriging (Myers and Schultz, 200 routinely deals with this problem. A revised location algorithm should take into consideration non-linearity, as well as non-

independent 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 large number of stations to locate events, such as the ISC (International Seismological Centre) and NEIC (National Earthquake Information Center), are likely to be underestimated.

CONCLUSIONS

We have validated location improvement from model-based calibration using the 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 or both regional Pn (CUB) and teleseismic P (J362) calibration. Our result is consistent with variance reductions estimated from cluster analysis (Bhattacharyya et al. 2003) as well as other previous relocation studies (Ritzwoller et al. 2003; Antolik et al. 2003) that demonstrated significant improvement using either regional or teleseismic model-based calibration alone with smaller datasets.

Gaussian, non-zero

0a)

Since in practice regional and teleseismic phases are used jointly, we tested Pn and P calibrations with more than 1200 GT0-GT10. Our relocation tests show that the largest location improvement is achieved by jointly regional and teleseismic calibration, and has nearly reached 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 rely on converting an S model to a P model, we find that the S to P mappings may be regional dependent and we suggest that future regional tuning of these models may provide additional benefits.

We introduced constrained bootstrapping that not only provides a controlled experiment designed to increase the statistical power of the relocation test, but also provides location bias estimates due to un-modeled 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 sub-networks 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 minimizes the effect of correlated errors, the ultimate remedy would be to improve the location algorithms, e.g. using non-Gaussian/non-linear methods, and accounting for off-diagonal elements in the covariance matrix.

Using our unique data set and new sub-network sampling techniques, we have demonstrated that event locations for small events recorded by sparse networks can be significantly improved using 3D global models. As shown in Figure 11, the median 10-station location bias for the jointly calibrated regional and teleseismic case is about 6.0 km (7.0 km for Pn only and 8.0 km for P only), approaching the GT5 uncertainty of the majority of GT event locations. In general, mislocation reductions for model-calibration are over 20%. About 2 out of 3 (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.

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.

based

Antolik, M., Y.J. Gu, G. Ekström, A.M. Dziewonski (2003), J362D28: a new joint model of compressional and shear velocity in the Earth's mantle, Geophysical Journal International, 153:2, 443 - 466, 2003.

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, Proceedings of 24th Seismic Research Review- Nuclear Explosion Monitoring: Innovation and Integration, Ponte Vedra Beach, Florida, September 17-19.

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 Trans AGU, 81.

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, in preparation.

Billings (1994), Simulated annealing for earthquake location, Geophys. J. Int., 118, 680-692.

Billings, S.D., M.S. Sambridge, B.L.N. Kennett (1994), Errors in hypercentral location: picking, model, and magnitude dependence, Bull. Seism. Soc. Am. 84, 1978-1990.

Bondár, I., S. Myers, E.R. Engdahl, and E. Bergman (2003a), Epicenter accuracy base on seismic network criteria, Geophy. J. Int. in press.

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 (2003b), Collection of a reference event set for regional and telseismic location calibration. Submitted to Bull. Seism. Soc. Am.

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.

Dziewonski, A.M., and D.L. Anderson (1981), Preliminary reference earth model, Phys. Earth Plannet. Inter., 25, 297-356.

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.

Engdahl, E.R. and E.A. Bergman (2001), Validation and generation of reference events by cluster analysis, Proceedings of 23rd Seismic Research Review, Jackson Hole, WY, October 15.

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.

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, Proceedings of 24th Seismic Research Review- N Explosion Monitoring: Innovation and Integration, Ponte Vedra Beach, Florida, September 17-19.

Myers, S.C. and C. Shultz (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 Trans. AGU, 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 models to locate regional events, J. Geophys. Res., in press.

Rodi, W., C. Shultz, W. Hanley, S. Sarkar, and H. Kuleli (2002), Grid-search location methods for

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. Kennet (2001), Seismic event location: non- 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 u mantle, Geophys. J. Int., 151, 88-105.

Shapiro, N.M. and M.H. Ritzwoller (2002b), Thermodynamic constraints on seismic inversions, submitted to Geophys. J. Int.

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.

shear inversion

ground-truth collection from location and regional seismic networks.

Table 1. *Estimates of baseline differences between regional Pn and teleseismic P-wave models based on mean(Pn) - mean(P) travel-time residual differences for more than 700 GT0-10 events with at least 5 Pn and 52 P arrivals.*

Travel-Time Model		Pn-P Residual Mean (sec)	Pn-P Standard Deviation (sec)	Number of Events
Pn	P			
IASP91	IASP91	0.48	0.95	762
CUB2	IASP91	-0.23		
CUB2	J362			
IASP91				
CUB1				
CUB1				

-0.08

Table 2. *Five categories (I-V) of events based on mislocation comparisons w.r.t. GTX for two sets of relocation results (e.g. using calibrated, CAL, and uncalibrated, UNCAL, travel times).*

Mislocation Categories		Calibrated (CAL)		
		CAL Mislocation = GTX	CAL Mislocation > GTX	
Uncalibrated (UNCAL)	UNCAL Mislocation = GTX	I. CAL & UNCAL inside GTX (Indecisive)	II. CAL Moved outside GTX (Deteriorated at GTX)	
	UNCAL Mislocation > GTX	III. CAL Moved inside GTX (Improved at GTX)	IV. CAL Moved toward GT (Improved outside GTX)	V. CAL Moved away from GT (Deteriorated outside GTX)

Table 3. *All-station relocation statistics summary on location improvement for models CUB1 and J362, separate and jointly (Pn & P), given by percentages of events using CUB1, J362, and CUB1+J362.*

Improvement/Deterioration Percentages	Pn CUB1	P J362	Pn & P CUB1+J362
Number Events	526	793	
Improved (%)	48	65	52
Deteriorated (%)	29	26	31
Indecisive (%)	22	9	17
Mislocation Categories			
Improved at GTX (%)	11	9	11
Deteriorated at GTX (%)	9	4	10
Improved outside GTX (%)	38	56	41
Deteriorated outside GTX (%)	20	22	21
Normalized Improvement/Deterioration			
Improvement = GTX (%)	15	38	26
Deterioration = GTX (%)	4	6	5

234

IASP91   0.38   1.08   765

Table 4. 0.41 *All-station* 0.62 *relocation statistic summary* 0.97 *on mislocatios, error ellipss, origin times (OT), and standard deviation of observations (sdobs) using CUB1, J362, and* **CUB1+J36.** 0.79   0.55   0.43   0.32   0.50   0.36   0.89

Metrics	Pn		P		Pn & 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 (km²)	448	253	464	322	394	233

(overlapping values: 1.11 0.84 0.77 1.04 0.92)

relocation statistic summary on location bias by percentages of 89 76 75

Median OT from "GT" (sec)
Median OT error
Median sdobs (sec)

Table 5. *SSNB improvement versus deterioration using CUB1, J362, and CUB1+J362.*

Improvement/Deterioration Percentages	Pn CUB1	P J362	Pn & P CUB1+J362
Improved (%)	51	62	58
Deteriorated (%)	30	28	25
Indecisive (%)	19	10	17
Improvement = GTX (%)	17	25	25
Deterioration = GTX (%)	8	9	7

Table 6. *SSNB relocation statistic summary on location bias, mislocation, error ellipse size, and coverage using CUB1, J362, and CUB1+J362.*

Metrics	Pn		P		Pn & P		
	IASP	CUB1	IASP	J362	IASP	CUB1+J362	8.7
Mean Bias (km)	8.4	7.2	9.2	7.9	7.7	6.6	
Mislocation (km)	8.8	7.6					
Error ellipse area (km²)	1522	745	1953	1133	1815	971	
Coverage (%)	99	96	92	85	97	92	

FIGURE CAPTIONS

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

Figure 2. Model errors for regional Pn and teleseismic P phases for the baseline IASP91 (dashed line) travel-time tables, CUB (solid line) Pn calibrations, and J362 (dot-line) teleseismic P calibrations. The CUB and J362 model errors amount to about a 50% variance reduction over the baseline model errors. A-priori model errors jointly with measurement (picking) errors are used to weigh the observations and to predict 90% error ellipses.

Figure 3. (a) 526 GT0-GT10 selected events for relocation testing of regional Pn calibration using all stations: 10 GT0 (square), 36 GT1 (circle), 1 GT2 (triangle), 450 GT5 (invert triangle), and 19 GT10 (diamond). (b) Event-station paths with 1098 stations (35161 arrivals). Triangles show stations and circles show events.

Figure 4. (a) 793 GT0-GT10 events selected for use in relocation testing of teleseismic P calibration using all stations: 17 GT0 (square), 316 GT1 (circle), 10 GT2 (triangle), 279 GT5 (invert triangle), and 171 GT10 (diamond). (b) 2821 stations (108,172 arrivals). Triangles show station locations.

Figure 5. Illustration of sgap tree for station "A" of the fictitious network shown in the inset. The sgap tree is a hierarchical binary tree where each subsequent station is selected in a way that it splits the remaining azimuthal as evenly as possible. The selection order of stations is shown on the right axis, and the numbers at the left and right of the n

dashed

indicate the remaining azimuthal gaps. Cutting the sgap tree at any level of the hierarchy (dashed line) provides the sub-network with the most uniform azimuthal coverage and the minimal secondary azimuthal gap (in this example 130°, provided by stat station sub-network). Different root stations may generate different sgap trees.

Figure 6. Coverage parameter E. E should be χ^2 distribution with two-freedom. E=1.0 corresponds to the 90% coverage. In this diagram the event, E > 1, is not covered.

Figure 7. Cumulative mislocations calibrated (red) and uncalibrated (green) are shown for (a) CUB1 Pn, (b) J362 P, and (c) CUB1 Pn and J362 P jointly. In all cases the calibrated distribution (red) is shifted to the right indicating the 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 joint Pn & P location case 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; Deterioration is negative. Changes in mislocation less than GTX are indecisive (shaded green). In all cases the positive tail (improvement = GTX) of the distribution is larger than the negative tail (deterioration = GTX) indicating that improvements are larger than deteriorations. The three percentage fractions ("Improvement = GTX, Deterioration = GTX , and Indecisive) and their median values (Median Improvement, Median Deterioration) are tabulated in Tables 3-4.

Figure 8. Coverage parameter, E, versus number of defining phases (ndef) using joint CUB1 Pn and J362 P calibrations (red triangles) versus the uncalibrated travel-times

degrees of   43

ion B for a 10-

station sub-networks with calibrated (CUB1, red) and

(blue diamonds). Trend lines are shown as solid lines. The likelihood of insufficient coverage increases with increasing number of defining phases (ndef). Coverage parameters, E, are less than unity for 90% of the events with ndef = 50 (uncalibrated) and ndef = 27

Figure 9. SSNB example for the 1997/02/14 00:22:37 earthquake in Pakistan (a), and the 1972/08/20 03:00:00 PNE Region-3 (b). Open triangles represent the relocations with the 20 most characteristic SSNB 10-

uncalibrated (IASPEI, blue) regional Pn travel-times. The vector location bias estimate (solid 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.

Figure 10. (top) Mean mislocations using all stations (seed) with (red inverted triangles) and without (blue triangles) calibration. (middle) Mean bias from SSNB locations with (red inverted triangles) and without (blue triangles) calibration. (bottom) Mean improvement in bias (red circles) from SSNB locations and mean improvement in location (blue squares) from all-station (seed) locations due to calibrated regional and teleseismic (CUB1+J362) travel-times for event clusters sampled by more than 3 SSNB seed events. Event clusters are sorted by GT category. The thick green line indicates GT accuracy of the reference events in the clusters.

Figure 11. SSNB location improvements for 116 common events due to calibrated travel times (horizontal arrows) and the combination of regional and teleseismic phases (vertical arrows). (a) Median bias (mislocation) reductions by model-based calibration.
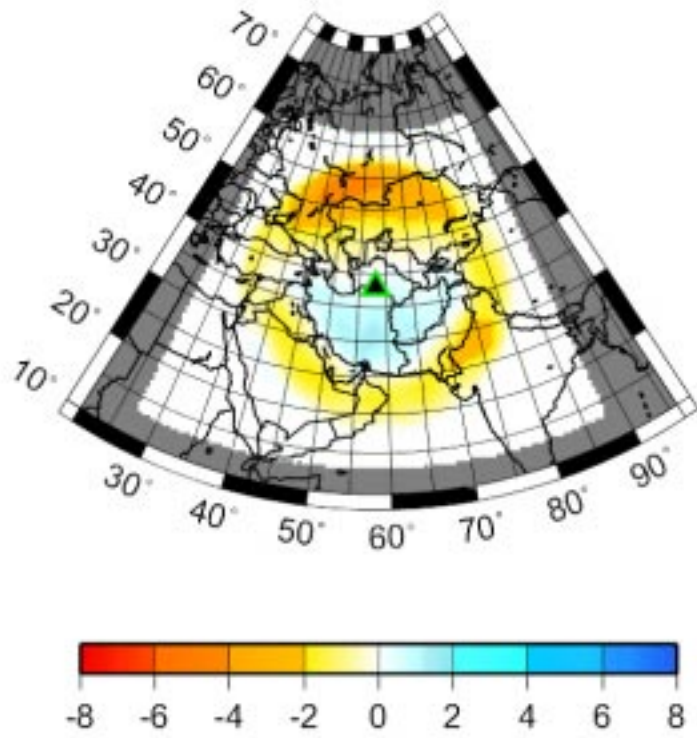
(calibrated).

(b) Percentages of event locations improved by model-based calibration. Joint Pn and P travel time calibration improves 2 out of 3 events and reduces bias 20% or more.

Figure 12. Mislocation differences between CUB1 and CUB2 in km. Green circles show events where CUB1 is better, red circles show events where CUB2 is better, and blue circles show events that are indecisive. CUB1 improves 262 events (mostly in the Mediterranean), and deteriorates 147 events within GT uncertainty w.r.t. CUB2 (Appendix). 113 events are indecisive and cannot discriminate between CUB1 and CUB2.

Figure 13. Location bias, misfit, coverage parameter, error ellipse area, median gap (orange) and sgap (green) are plotted as a function of number of phases used in the SSNB locations with calibrated (red; J362) and uncalibrated (blue; IASP91) travel-times for a GT1 Chinese underground nuclear explosion on 10/07/1994.

Figure 14. Cumulative histograms of the coverage parameter with all-station seed (a) and the SSNB 10-station sparse networks (b) with calibrated (red) and uncalibrated (blue) regional and teleseismic travel-times. The theoretical χ^2 distribution (assuming Gaussian and independent errors) of the coverage parameter is shown as a black line.
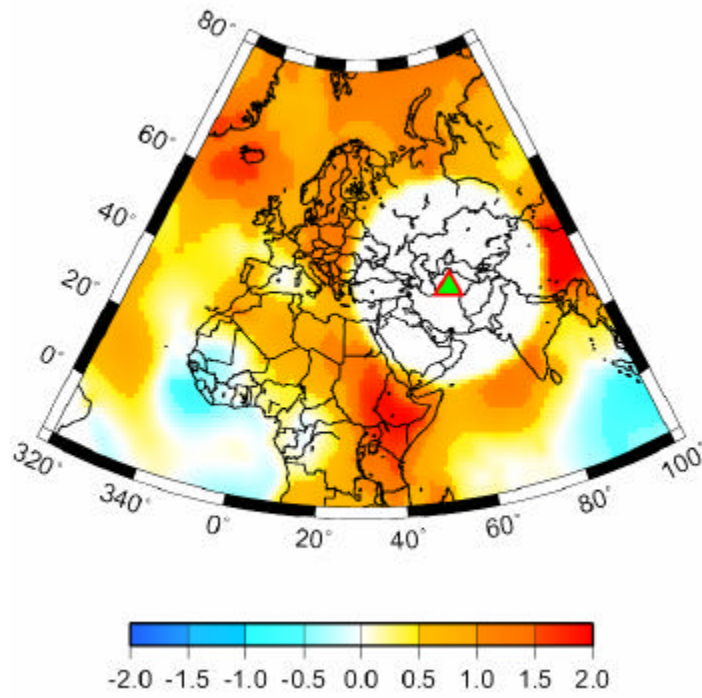
Figure 1.

Figure 2.

(a)

□ 10 GT0   ○ 36 GT1   △ 1 GT2   ▽ 460 GT5   ◇ 19 GT10

——— raypath   ▲ station   ○ event

(b)

Figure 3.

(a)

□ 17 GT0   ○ 316 GT1   △ 10 GT2   ▽ 279 GT5   ◇ 171 GT10

(b)

Figure 4.

Figure 5.

Relocated event

GT event

σ_{GT}

Δ

σ

sminax

smajax

$(smajax^2+\sigma_{GT}^2)^{1/2}$

$(sminax^2+\sigma_{GT}^2)^{1/2}$

Figure 6.

(a)

(d)

(b)

(e)

(c)

(f)
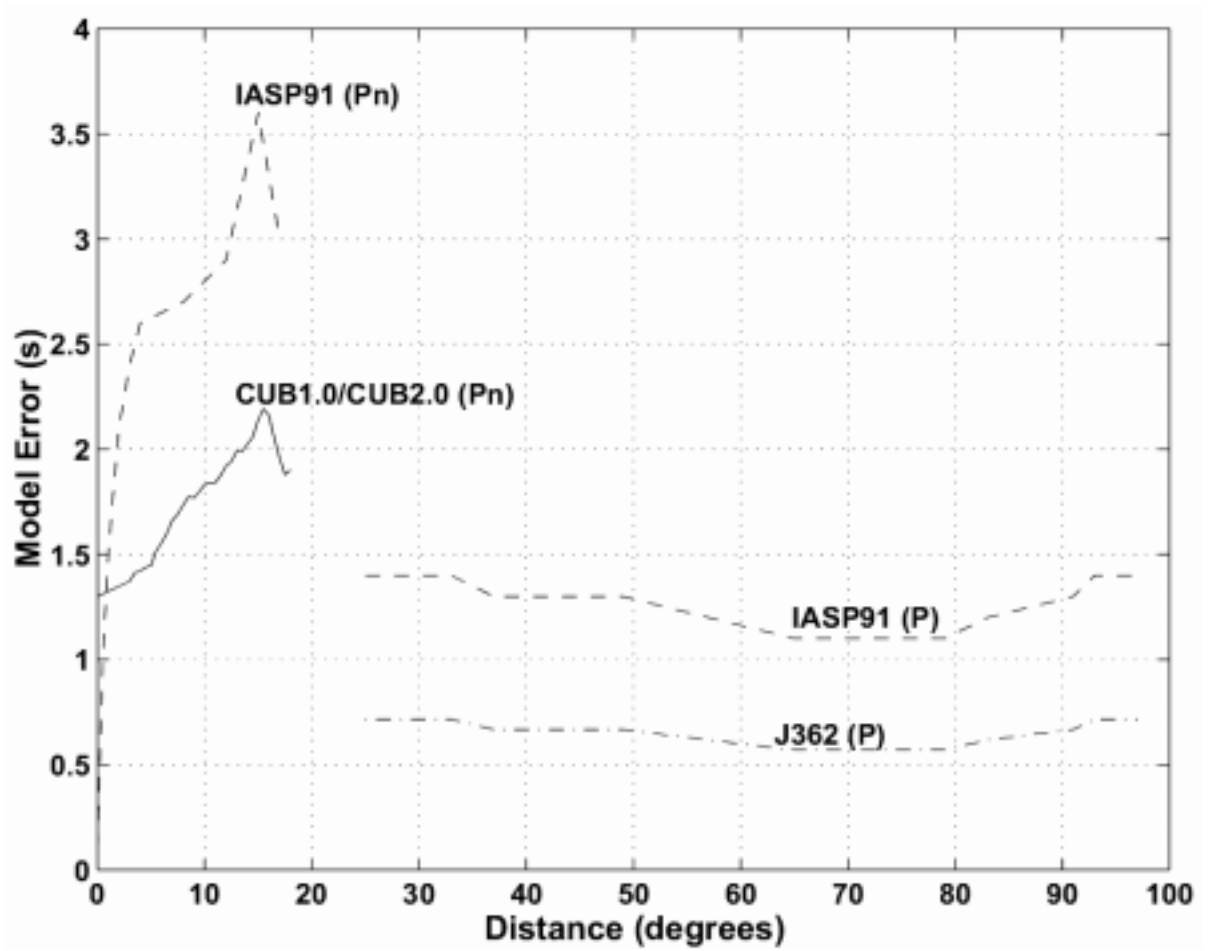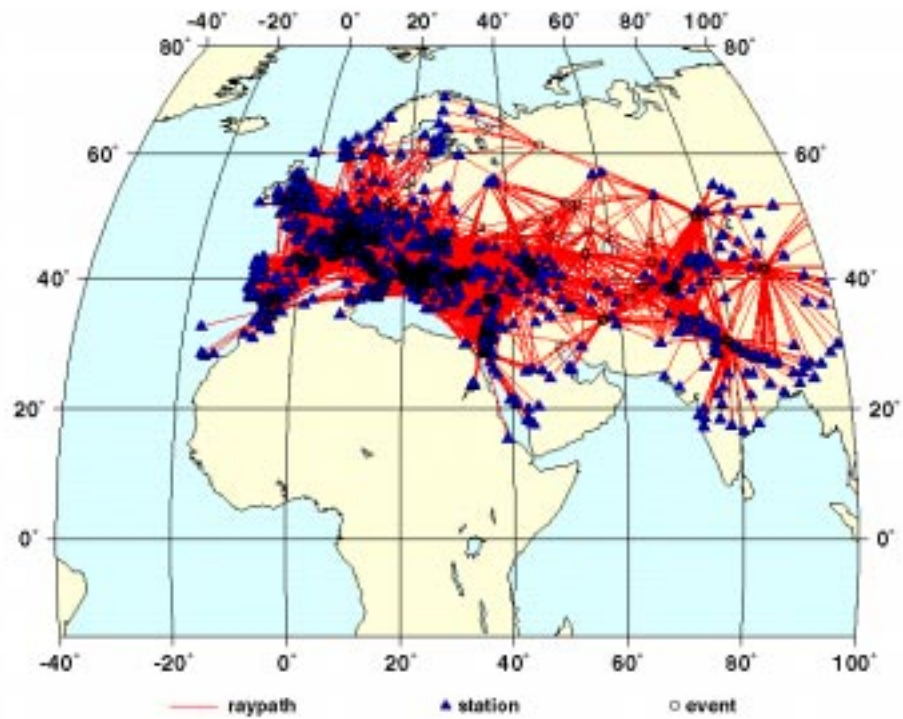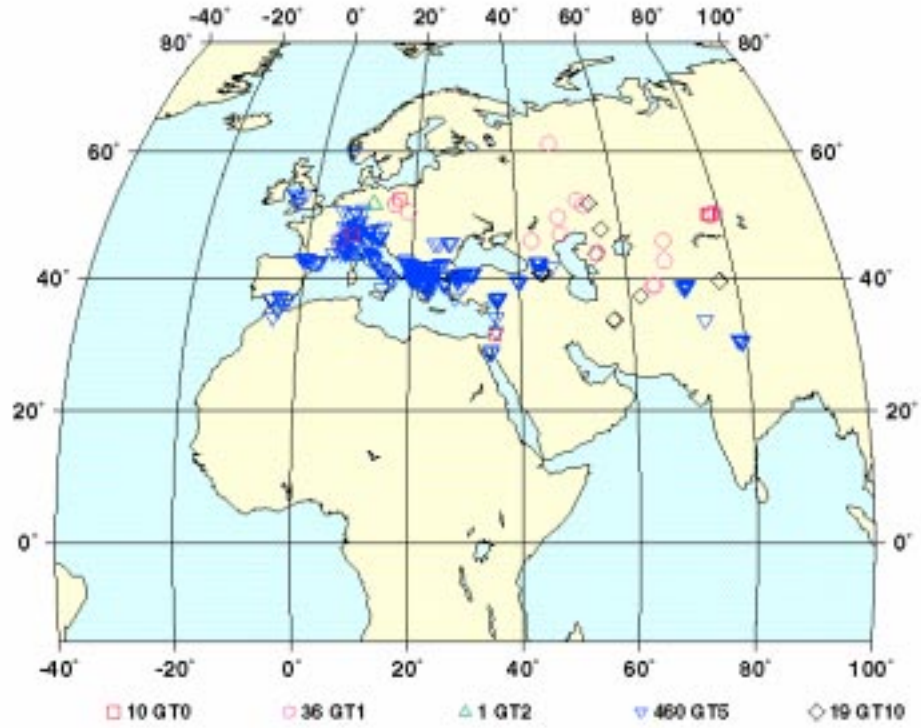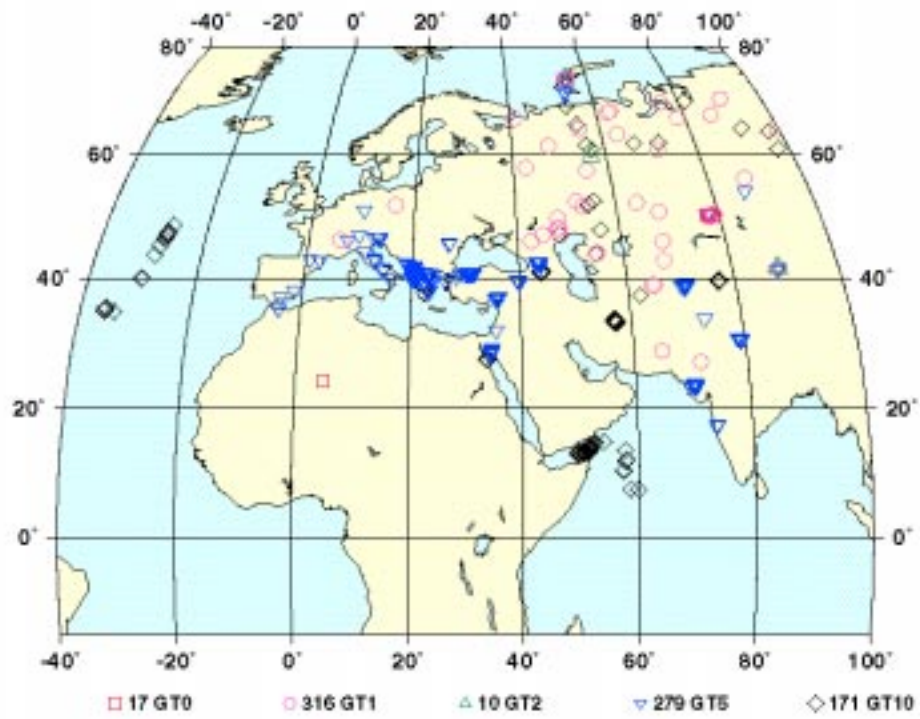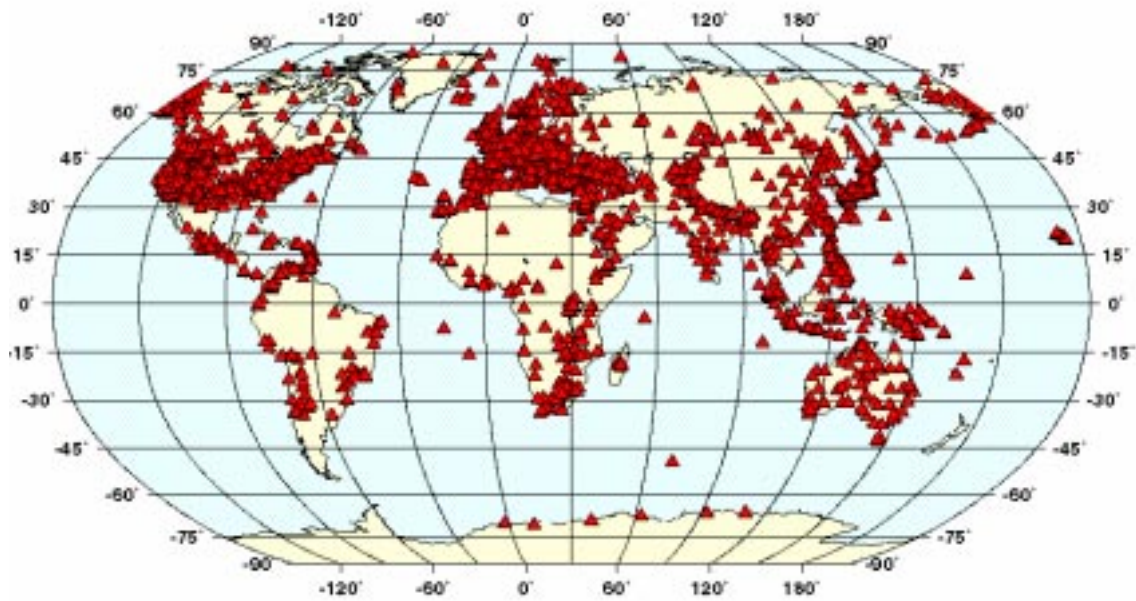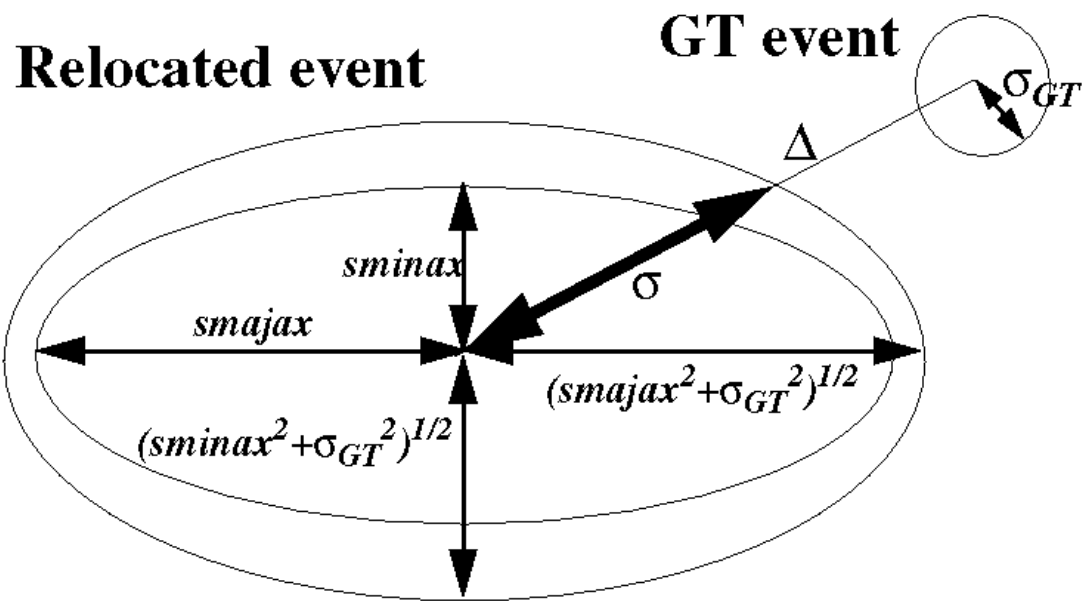
Figure 7.

Figure 8.

(a)

(b)

Figure 9.

Figure 10.

(a)

Median Bias Reduction (%)

Uncalibrated   Calibrated

Pn only 10.0 km → 24% → Pn only 7.6 km

25%

P only 10.4 km → 23% → P only 8.0 km

21%

28%

Pn & P 7.5 km → 20% → Pn & P 6.0 km

25%

(b)

Percent Events Improved

Uncalibrated   Calibrated

Pn only → 70% → Pn only

68%

P only → 67% → P only

74%

61%

Pn & P → 65% → Pn & P

61%
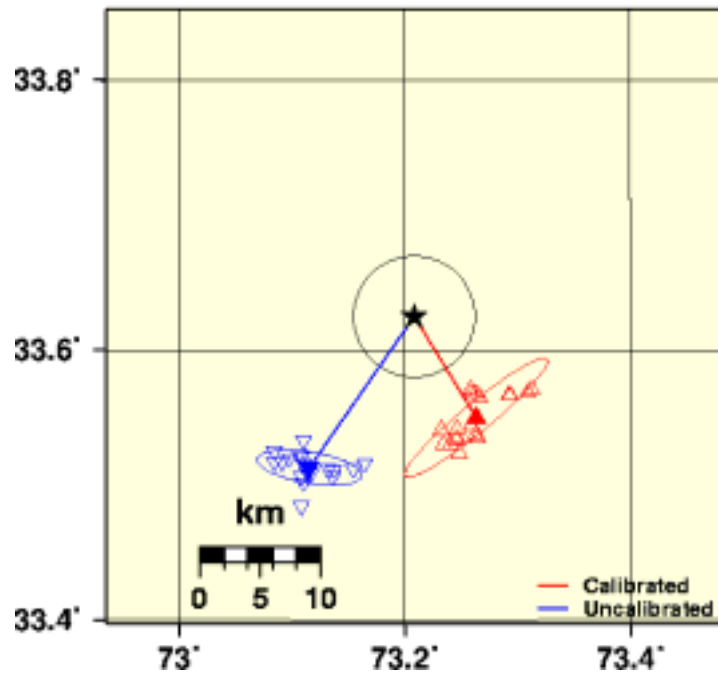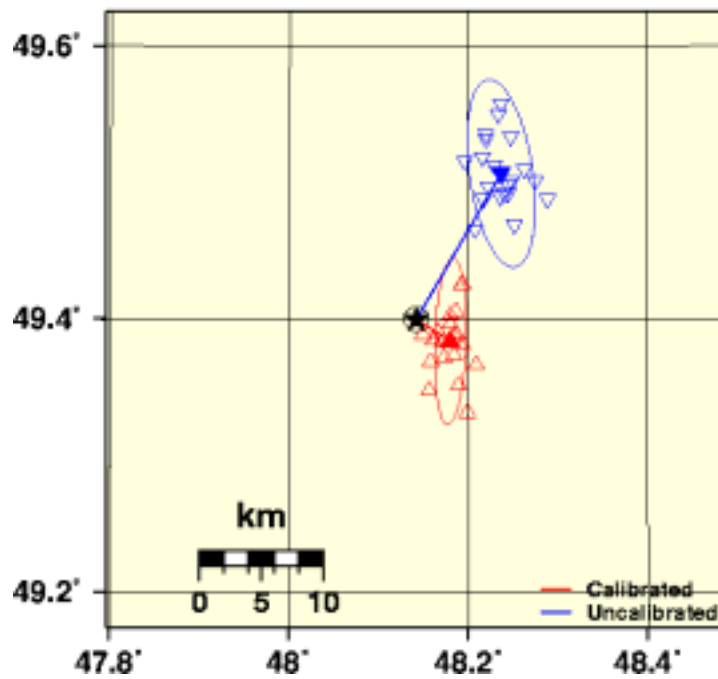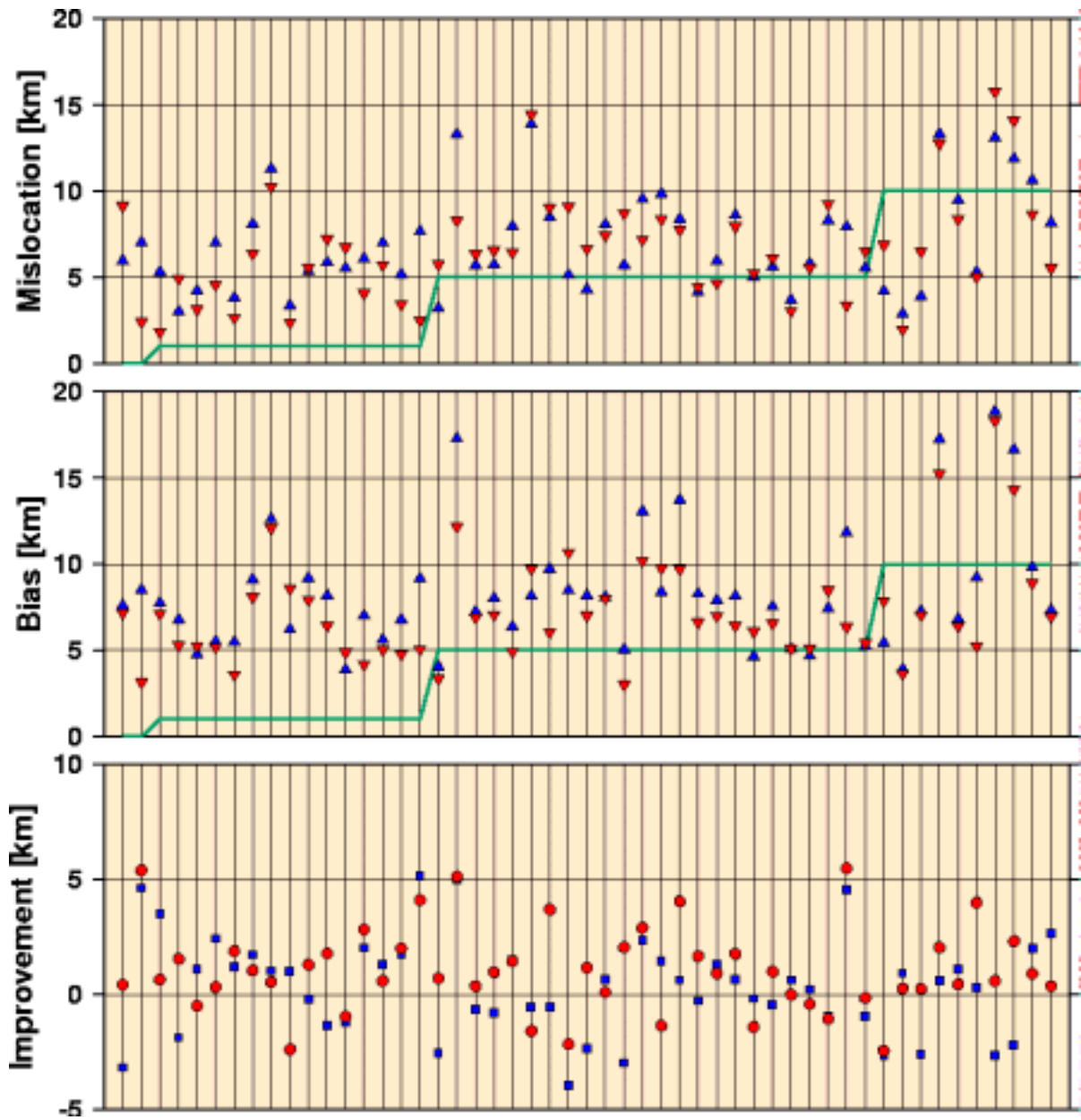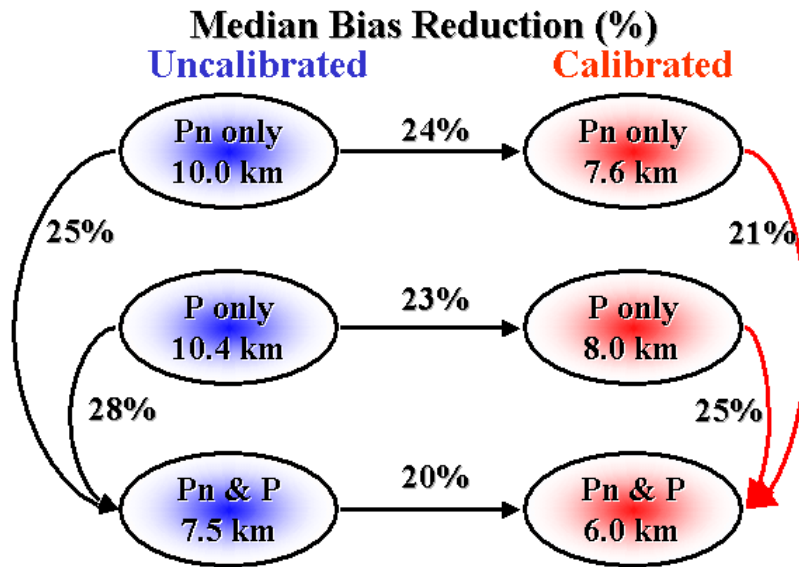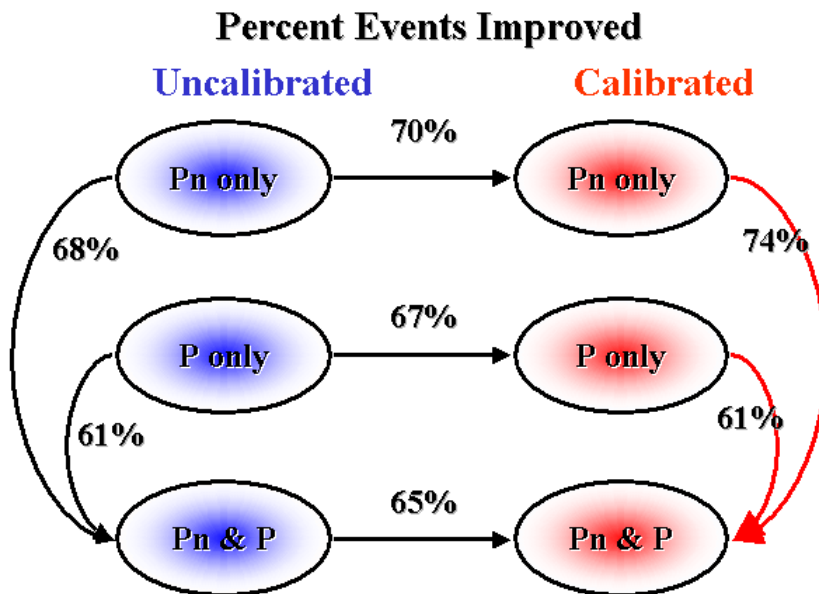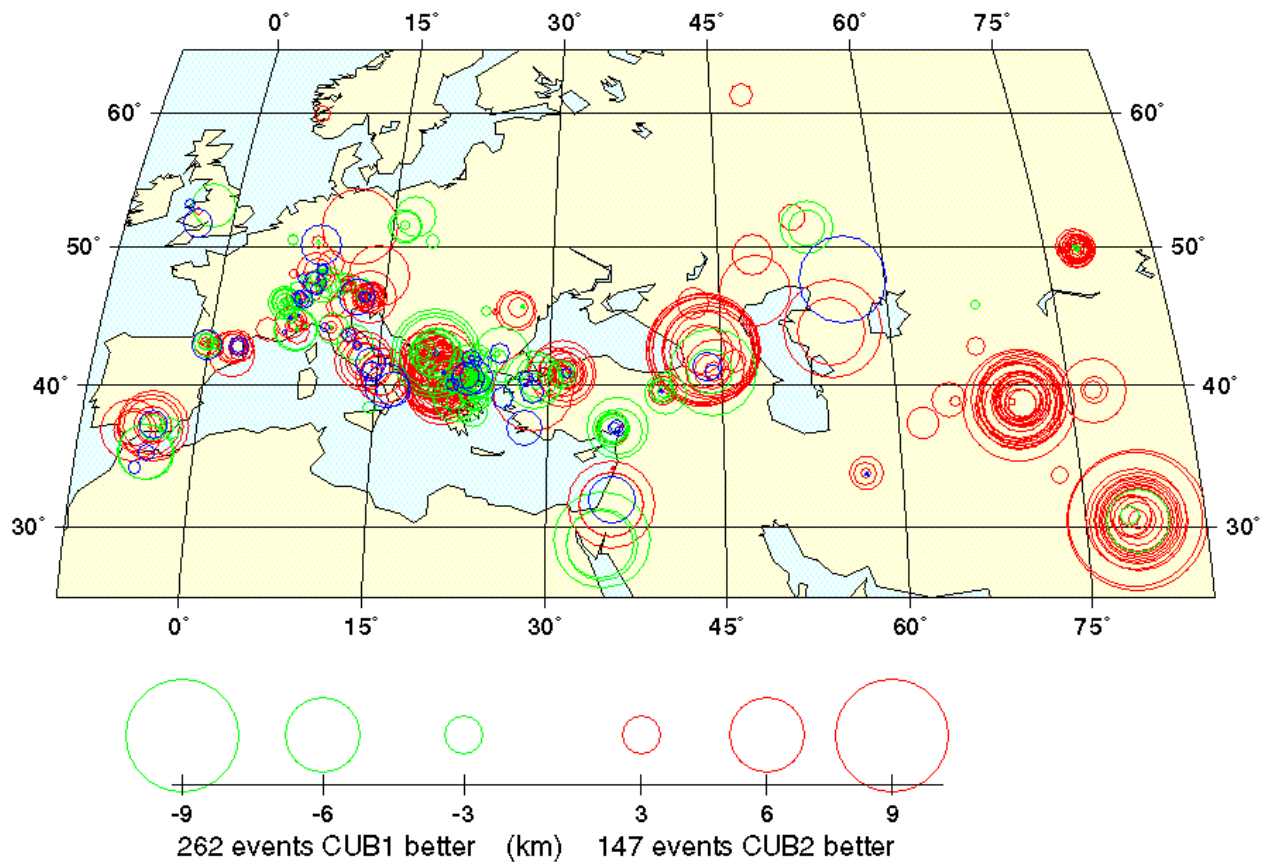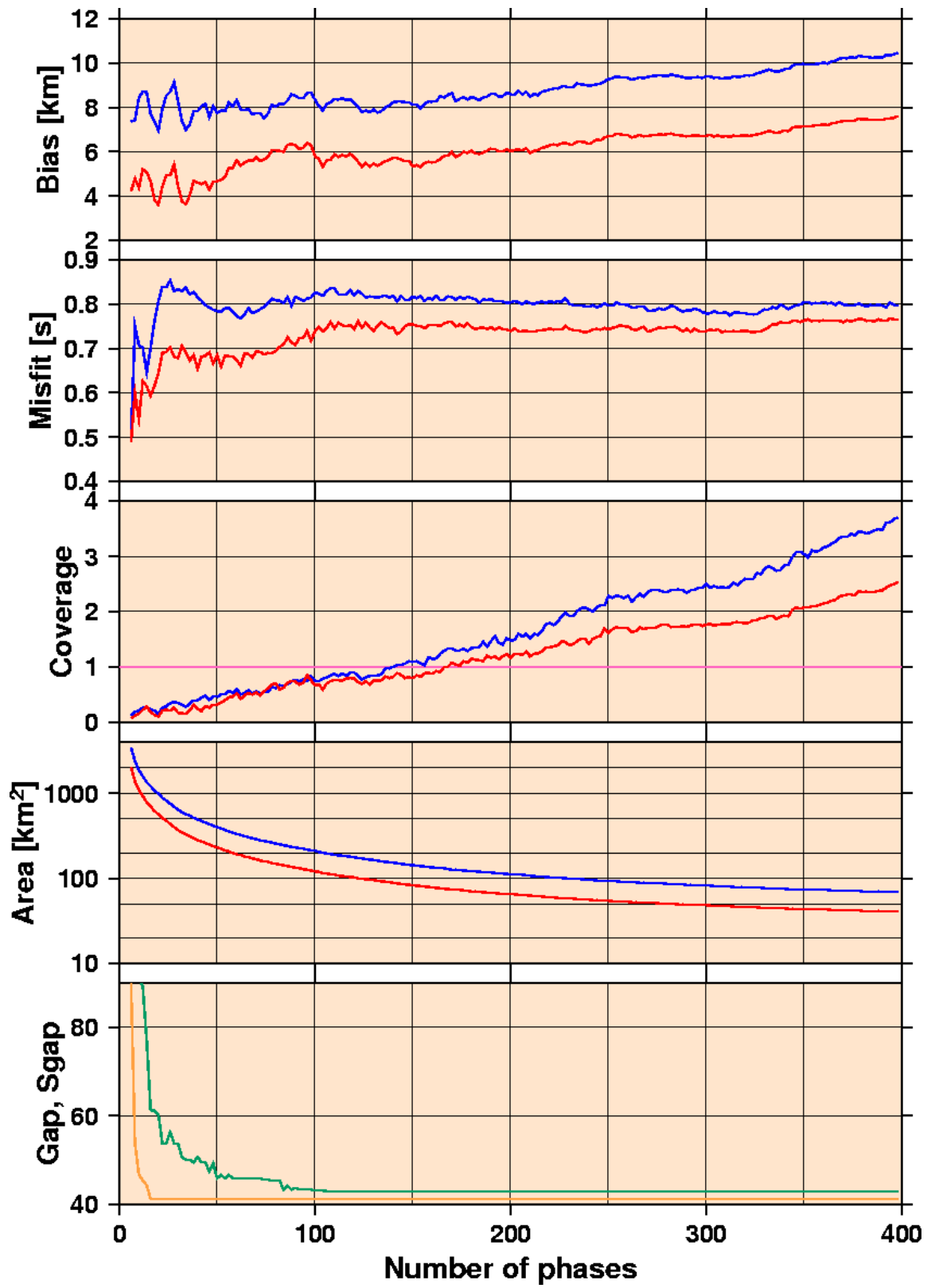
Figure 11.

Figure 12.

Figure 13.

(a)
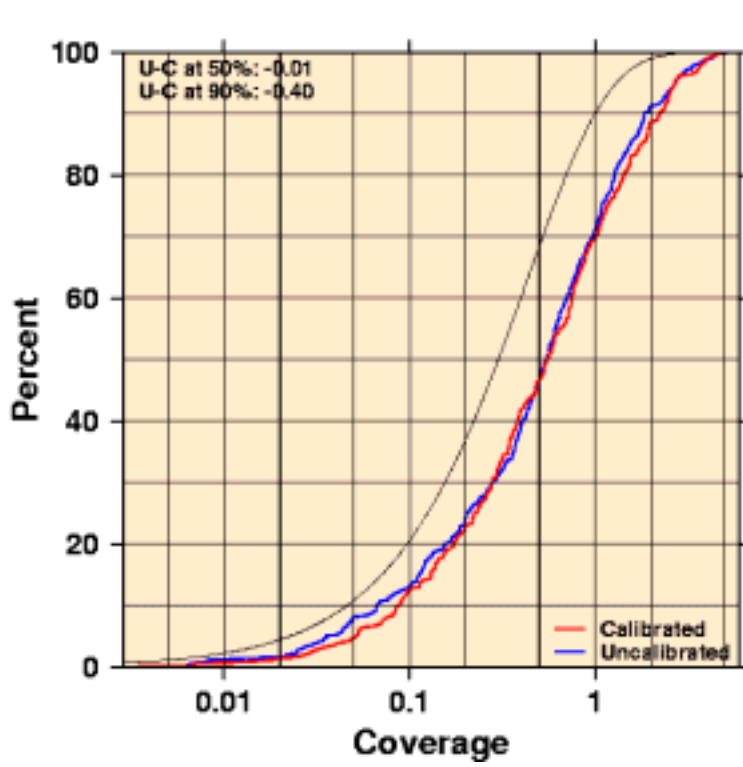
(b)
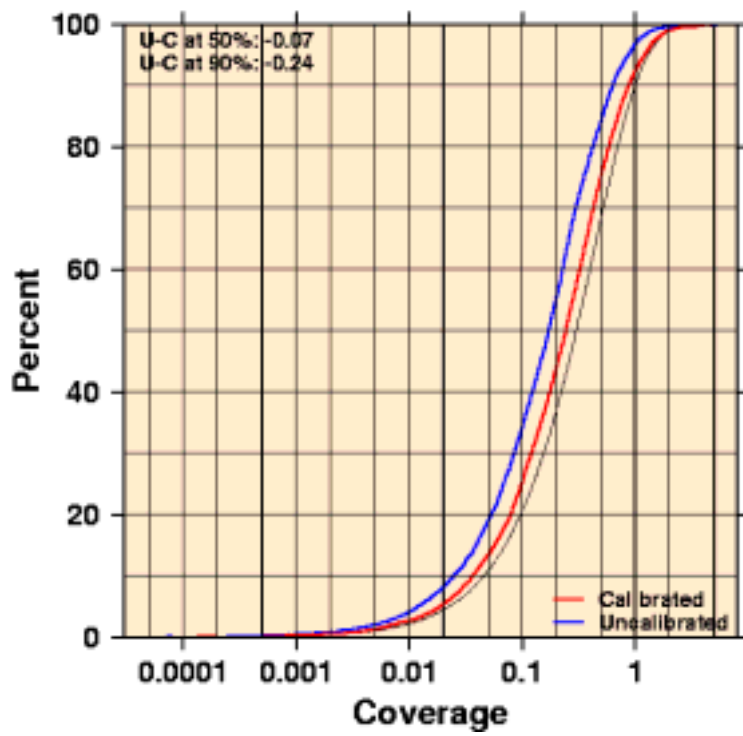
Figure 14.

APPENDIX: RELOCATION STATISTICS

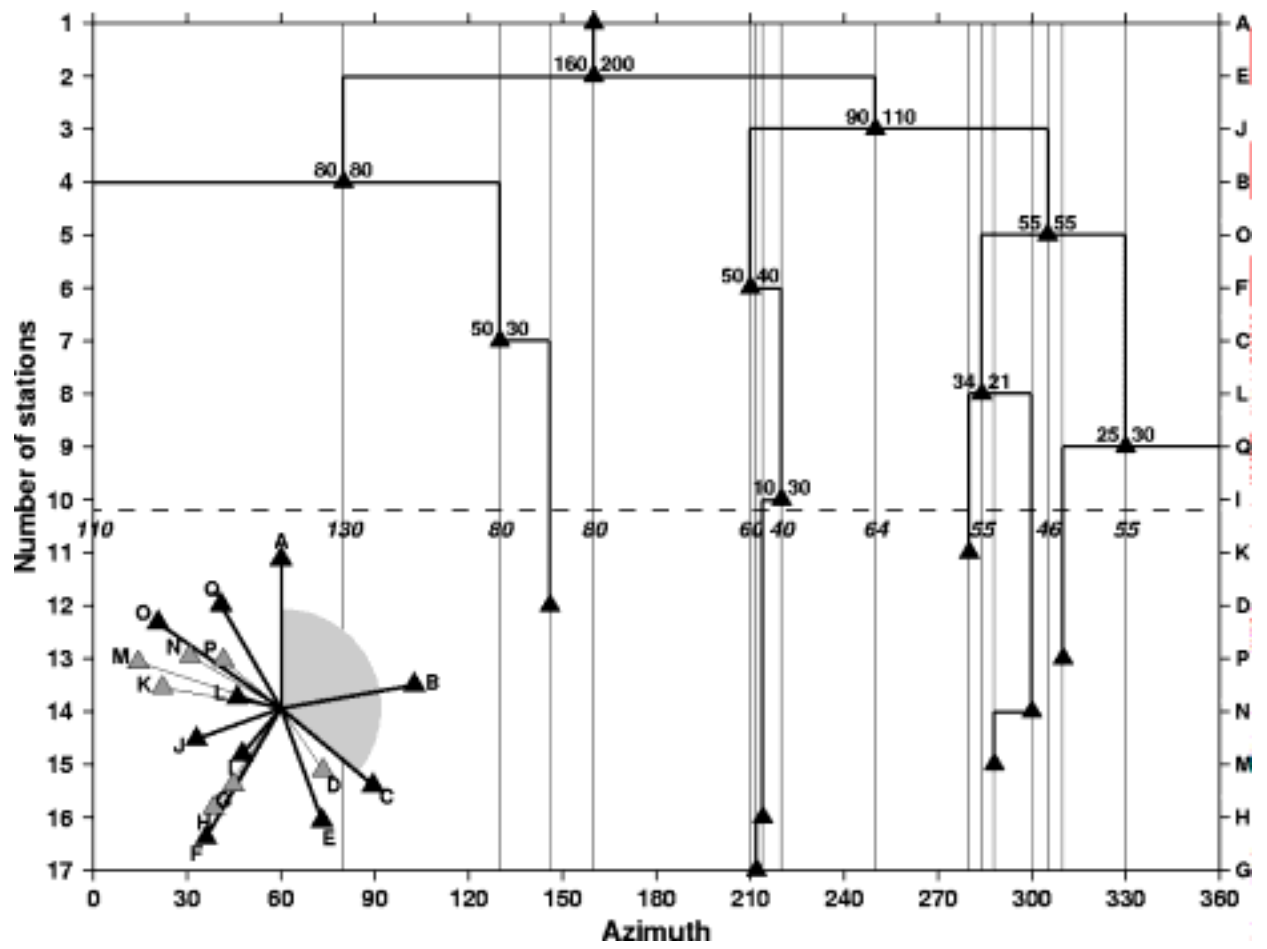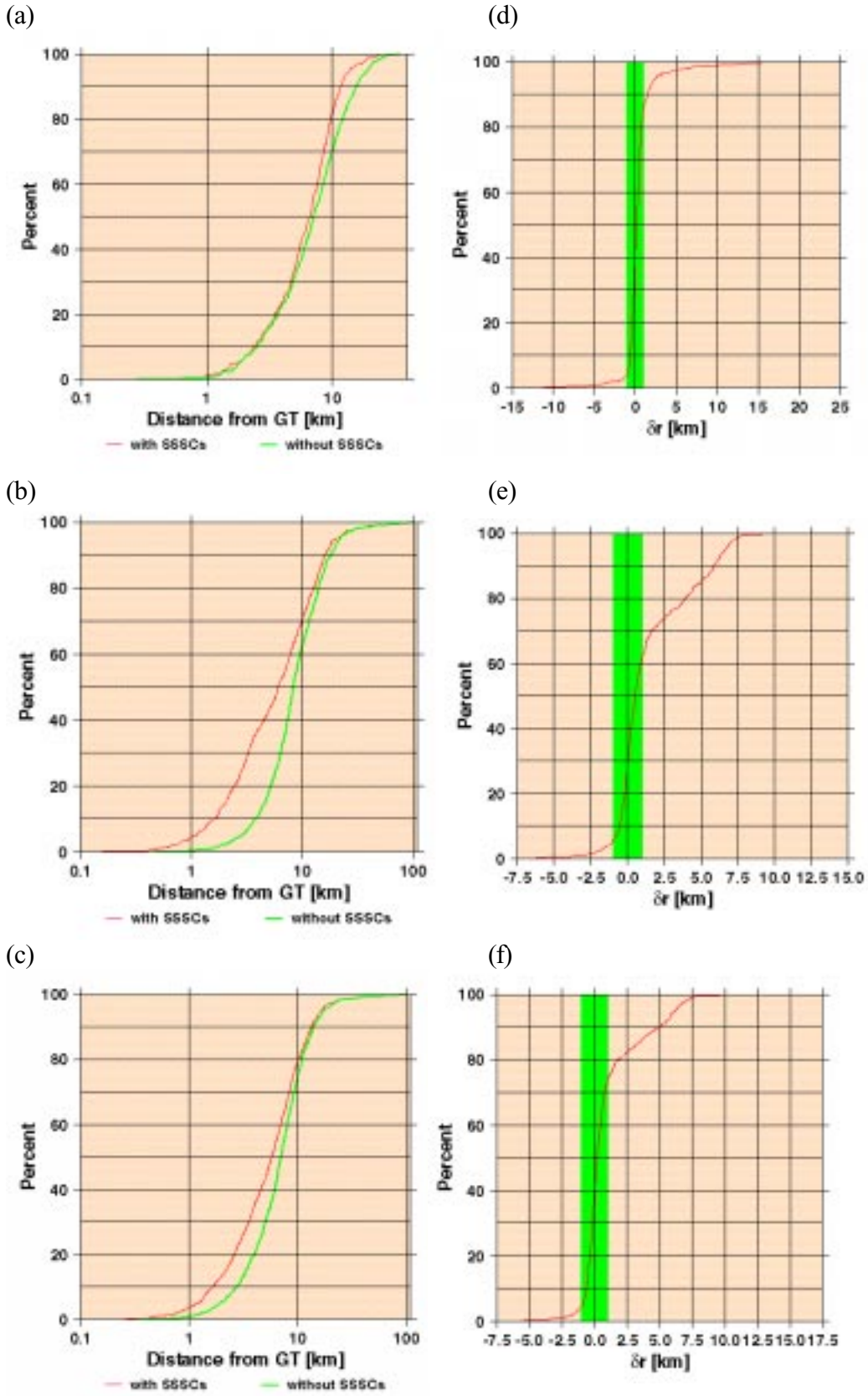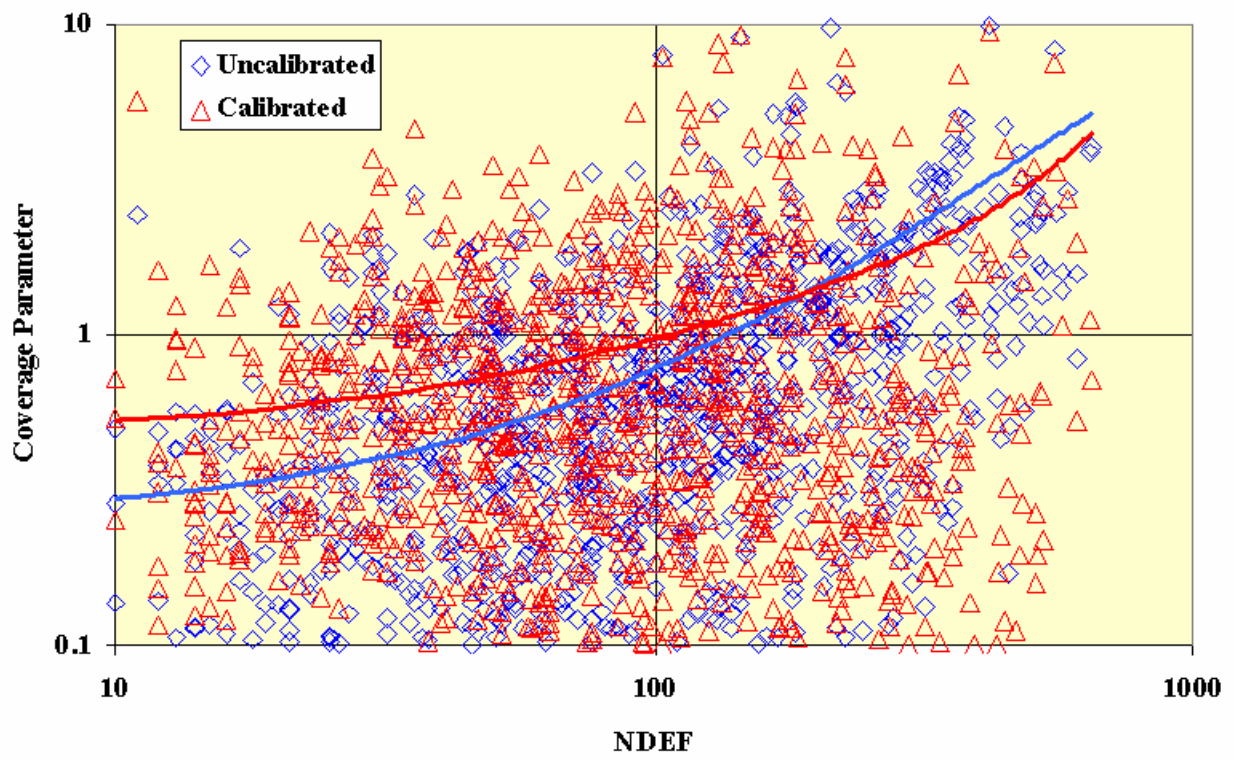In this Appendix we present detailed relocation statistics on improvements versus deterioration on a case-by-case basis for Pn calibrations only (CUB1 and CUB2, respectively), P (J362) calibration only, and the jointly Pn and P calibrations (CUB1+J362 and CUB2+J362, respectively). Tables 7-8 show the 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 deviation of observations (sdobs), and the SSNB centroids (location bias). 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 this Appendix we also include the detailed statistics on GT events west of 40°E (Table 9). These events are shown in Figure 12, mostly in the Middle East and the Mediterranean. Unlike the joint CUB1+J362 calibration, CUB2+J362 calibration deteriorates more events than it improves in this region.

Table 7. *Statistics on relocation improvement using all stations for models CUB1, CUB2, and J362, separate and jointly (Pn & P). CUB1 without the 0.75 sec baseline correction is denoted 'CUB1'.*

a. *Improvement versus deterioration percentages:*

Improvement/Deterioration Percentages	Pn CUB1	Pn CUB2	P J362	Pn & P CUB1'+J362	Pn & P CUB1+J362	Pn & P CUB2+J362	1234
Number Events	526	526	793				
Improved (%)	48	41	65	52	52	50	
Deteriorated (%)	29	39	26	30	31	32	
Indecisive (%)	22	20	9	18	17	17	

b. *Improvement/Deterioration/Indecisive percentages:*

Significant Improvement/Deterioration	Pn CUB1	Pn CUB2	P J362	Pn & P CUB1'+J362	Pn & P CUB1+J362	Pn & P CUB2+J362
Improvement = GTX (%)	15	10	38	26	26	26
Deterioration = GTX (%)	4	8	6	7	5	7
Indecisive (%)	81	82	56	67	69	67

c. *Improvement/deterioration categories:*

Improvement/Deteriation Categories	Pn CUB1	Pn CUB2	P J362	Pn & P CUB1'+J362	Pn & P CUB1+J362	Pn & P CUB2+J362
Improved at GTX (%)	11	10	9	11	11	10
Deteriorated at GTX (%)	9	11	4	9	10	9
Improved Outside GTX (%)	38	31	56	41	41	40
Deteriorated Outside GTX (%)	20	28	22	21	21	23
Indecisive (%)	22	20	9	18	17	17

d. *Mislocation:*

Mislocation Statistics	Pn CUB1	Pn CUB2	P J362	Pn & P CUB1'+J362	Pn & P CUB1+J362	Pn & P CUB2+J362
Number Events	526	526	793	1234	1234	1234
Uncalibrated Median Mislocation	7.1	7.1	8.3	7.1	7.1	7.1
Calibrated Median Mislocation	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 Deterioration (km)	1.7	2.2	1.8	2.0	1.9	2.1

e. *Ellipse area and coverage:*

Confidence Ellipse Statistics	Pn CUB1	Pn CUB2	P J362	Pn & P CUB1'+J362	Pn & P CUB1+J362	Pn & P CUB2+J362
% 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

f. *Origin time (OT):* 0.90 0.97

Origin Time Statistics	Pn CUB1	Pn CUB2	P J362	Pn & P CUB1'+J362	Pn & P CUB1+J362	Pn & P CUB2+J362
% 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
*dabis)*ated Median OT Error (Sec)	-0.52	0.74				
% 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

g. *Misfit (standard deviation of observation, s*

Misfit Statistics	Pn CUB1	Pn CUB2	P J362	Pn & P CUB1'+J362	Pn & P CUB1+J362	Pn & P CUB2+J362
% Events with Sdobs Reduced (%)	65	61	84	67	79	79
Uncalibrated Median sdobs (sec)	1.19	1.19	0.84	1.04	1.04	1.04
Calibrated Median sdobs (sec)	1.11	1.13	0.77	0.96	0.92	0.94

Table 8. *Statistics on relocation improvement using SSNB for CUB1, CUB2, J362, CUB1+J362, CUB2+J362, CUB1+IASP, CUB2+IASP, and IASP+J362.*

a. *Location impr*ovements/deteriorations given by percent of events:

Improvement/Deterioration Percentages	Pn CUB1	Pn CUB2	Pn J362
SSNB centroid bias			
Number of events	276	273	331
Improved (%)	51	50	62
Deteriorated (%)	30	33	28
Indecisive (%)	19	17	10
SSNB 10-			
Number of events	4856	4856	6836
Improved (%)	52	52	57
Deteriorated (%)	34	35	34
Indecisive (%)	14	13	9
All-station locations	station locations		
Number of events	276	273	331
Improved (%)	56	48	58
Deteriorated (%)	27	36	28
Indec/Deter(%)	17	16	14

b. *Location improvements/deteriorations given by percent of events:*

Significant Improve	Pn CUB1	Pn CUB2	Pn J362
SSNB centroid bias			
Number of events	276	273	331
Improved >= GTX (%)	17	18	25
Deteriorated >= GTX (%)	8	11	9
Indecisive (%)	83	82	75
SSNB 10-			
Number of events	4856	4856	6836

Improved >=			
Deteriorated >= GTX (%)	1	1	1
Indecisive (%)	95	95	97
All-station locations			
Number of events	276	273	331
Improved >= GTX (%)	18	17	26
Deteriorated >= GTX (%)	8	12	7
Deteriorated Outside GTX (%)	24 82	25 83	24 74

c. *Location bias improvements/deteriorations given by percent of events:*

Improvement/Deteriation Categories	Pn CUB1	Pn CUB2	Pn J362
SSNB centroid bias			
Number of events station locations	276	273	331
Improved at GTX (%)	11	10	9
Deteriorated at GTX (%)	6	8	4
Improved Outside GTX (%)	40	40	53
Deteri			
Indecisive (%)	19	17	10
SSNB 10-			
Number of events	4856	4856	6836
Improved at GTX (%)	11	13	6
Deteriorated at GTX (%)	7	7	4
Improved Outside GTX (%)	41	39	51
Deteriorated Outside GTX (%)	27	28	30
Indecisive (%)	14	13	9
All-station locations			
Number of events	276	273	331
Improved at GTX (%)	12	12	8
Deteriorated at GTX (%)	6	8	3
Improved Outside GTX (%)	44	36	50
Deteriorated Outside GTX (%)	21	28	25
Indecisive (%)	17	16	14

d. *Location improvements/deteriorations given by percent of events for Pn & P relocations:*

Improvements/Deterioration Percentages	Pn & P CUB1+J362	Pn & P CUB2+J362	Pn & P CUB1+IASP	Pn & P CUB2+IASP	Pn & P IASP+J362
SSNB centroid bias					
Number of events	355	356	359	357	355
Improved (%)	58	56	49	45	48
Deteriorated (%)	25	28	35	40	35
Indecisive (%)	17	16	16	15	17
SSNB 10-					
Number of events	7145	7142	7175	7158	7127
Improved (%)	57	56	47	45	48
Deteriorated (%)	33	35	34	37	41
Indecisive (%)	10	9	19	18	11
All-station locations					
Number of events	355	356	359	357	355
Improved (%)	48	49	46	41	45
Deteriorated (%)	33	34	35	44	35
Indecisive (%)	19	17	19	15	20

GTX (%)	5	5	3

e. *Location significant improvements/deteriorations given by percent of events for Pn & P relocations:*

Significant Improvements/Deteriorations	Pn & P CUB1+J362	Pn & P CUB2+J362	Pn & P CUB1+IASP	Pn & P CUB2+IASP	Pn & P IASP+J362
SSNB centroid bias					
Number of events	355	356	359	357	355
Improved >= GTX (%)	25	24	17	22	17
Deteriorated >= GTX (%)	7	10	6	11	8
Indecisive (%)	75	76	83	78	83
SSNB 10-					
Number of events	7145	7142	7175	7158	7127
Improved >= GTX (%)	3	3	3	3	3
Deteriorated >= GTX (%)	1	1	1	1	1
Indecisive (%)	96	96	96	96	96
All-station locations (seed)					
Number of events	355	356	359	357	355
Improved >= GTX (%)	22	22	13	14	16
Deteriorated >= GTX (%)	7	7	4	8	6
Indecisive (%)	78	78	87	86	84

f. *Location improvement/deterioration GTX Categories given by percent of events for Pn & P relocations:*

Improvement/Deteriation Categories	CUB1 +J362	CUB2 +J362	CUB1 +IASP	CUB2 +IASP	IASP +J362
SSNB centroid bias					
Number of events	355	356	359	357	355
Improved at GTX (%)	9	9	6	8	6
Deteriorated at GTX (%)	3	5	5	8	4
Improved Outside GTX (%)	49	47	43	37	42
Deteriorated Outside GTX (%)	22	23	30	34	31
Indecisive (%)	17	16	16	13	17
SSNB 10-					
Number of events	7145	7142	7175	7158	7127
Improved within GTX (%)	8	9	7	7	6
Deteriorated within GTX (%)	5	6	5	7	5
Improved Outside GTX (%)	49	47	40	38	42
Deteriorated Outside GTX (%)	28	29	38	39	36
Indecisive (%)	10	9	10	9	11
All-station locations (seed)					
Number of events	355	356	359	357	355
Improved at GTX (%)	10	11	8	9	10
Deteriorated at GTX (%)	9	10	8	12	7
Improved Outside GTX (%)	38	38	38	32	35
Deteriorated Outside GTX (%)	24	24	27	32	28
Indecisive (%)	19	17	19	15	20

g. *Mean bias, mislocation, area of error ellipse, and coverage of the 90% error ellipse using uncalibrated (uncal) or calibrated (cal) travel-times:*

Metrics	CUB1 uncal	CUB1 cal	CUB2 uncal	CUB2 cal	J362 uncal	J362 cal
SSNB 9.9						
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		

Error ellipse area (km²)	1522	745	1522	730	1953	1133
Coverage (%)	99	96	99	93	92	85
All stations (seed)	*times:*					
Mislocation (km)	7.9	6.8	7.8	7.3	8.2	7.1
Error ellipse area (km²)	518	293	529	288	411	274
Coverage (%)	80	79	81	71	63	64

h. Location bias, mislocation, area of error ellipse, and coverage of the 90% error ellipse using uncalibrated (uncal) and calibrated (cal) travel-

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 (km²)	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 (km²)	250	171	251	167	254	206	254	200	253	203
Coverage (%)	71	69	71	68	70	64	71	60	71	74

Table 9. *Relocation statistics using all-stations for GT events west of 40°E.*

	Pn	Pn	Pn	P	Pn & P	Pn & P	Pn & P
Model 2	CUB1	CUB2	CUB1	J362	CUB1+J362	CUB2+J362	CUB1+J362
Model 1	IASP91	IASP91	CUB2	IASP91	IASP91	IASP91	CUB2+J362
Number Events	415	415	508	179	508	508	508
Improved (%)	41	35	41	40	35	33	41
Deteriorated (%)	32	41	33	50	39	43	33
Indecisive (%)	27	24	26	10	26	24	26
Model1 median mislocation	6.1	6.1	6.1	11.4	6.5	6.5	6.5
Model2 median mislocation outside GTX (%)	6.1	6.2	6.2	11.2	6.4	6.4	6.4
Indecisive (%)	27	24	26	10	26	24	26
Improved at GTX (%)	10	11	10	5	8	11	10
Deteriorated at GTX (%)	11	14	11	5	12	14	11
Improved outside GTX (%)	31	24	31	35	27	22	31
Deteriorated	21	27	22	45	27	28	22