Calibrated Earthquake Locations from a Multiple Event Relocation Analysis

November 1, 2011

Dr. Eric Bergman
Global Seismological Services
http://www.seismo.com
bergman@seismo.com

Abstract

Two methods for determining bias-free ("calibrated") locations for clusters of earthquakes are briefly described. They are both based on a multiple event relocation analysis (Hypocentroidal Decomposition, or HDC) using standard seismic bulletin arrival times, in which arrival time differences are used to constrain the relative locations and origin times of events in a cluster. The key to calibrated locations, however, is the exclusive use of near-source data to establish the location of the cluster in absolute spatial and temporal coordinates. One method of calibration is based on external information on the location of one or more cluster events. Such information can be derived from seismological, geological, or remote sensing data. The second method of calibration is based on analysis a subset of the arrival time information in the cluster at small epicentral distances. The statistical power of both methods is greatly enhanced by a careful analysis of empirical readings errors derived from the arrival time data, which are used both for weighting in the inversion for location and also for identifying outlier readings. Calibrated locations with uncertainties of 2-3 km are achievable with either method. An example is given of the indirect method of calibration, using a cluster of earthquakes in Kyrgyzstan.

Applicability

The methods of location calibration described here can be used at a great range of scales, from a handful of small events recorded locally to several hundred large events recorded globally. The algorithm is best-suited for data sets of about 200 events or less, and it is desirable to restrict the geographic extent of a cluster to approximately 100 km or less. There must be a minimal level of "connectivity" between events in a cluster, meaning repeated observations by the same seismic station. Since calibration depends on data at short range from the source, it is not possible to calibrate the locations of deep earthquakes, although the relative locations can be improved.

Location Calibration Methodology

The location calibration analysis is based on the Hypocentroidal Decomposition (HDC) method for multiple event relocation introduced and by (Jordan and Sverdrup, 1981). The basic algorithm is completely described in that reference. The essence of the HDC algorithm is the use of orthogonal projection operators to separate the relocation problem into two parts:

- The *cluster vectors*, which describe the relative locations in space and time of each event in the cluster. They are defined in kilometers and seconds, relative to the current position of the *hypocentroid*.
- The *hypocentroid*, which is defined as the centroid of the current locations of the cluster events. It is defined in geographic coordinates and Coordinated Universal Time (UTC).

The cluster vectors are defined only in relation to the hypocentroid. The hypocentroid can be thought of as a virtual event with geographic coordinates and origin time in UTC. The orthogonal projection operators act on the data set of arrival times to produce a data set that includes only data that actually bears on the relative location of cluster events, i.e., multiple reports of a given seismic phase at the same station for two or more events in the cluster.

The hypocentroid is located very much as an earthquake would be, except that the data are drawn from all the cluster events. Thus it is typical for the hypocentroid to be determined by many thousands of readings. Nevertheless, the hypocentroid is subject to unknown bias because the theoretical travel times (typically ak135) do not fully account for the three-dimensional velocity structure of the Earth. Geographic locations for the cluster events are found by adding the cluster vectors to the hypocentroid.

The HDC method works iteratively. At each iteration, two inversions are performed, first for the cluster vectors relative to the current hypocentroid, then for an improved hypocentroid. The cluster vectors are added to the new hypocentroid to obtain updated absolute coordinates for each event. The convergence criteria are based on the change in relative location of each event (0.5 km) and the change in the hypocentroid (0.005°). The convergence limits for origin time and depth, for cluster vectors and hypocentroid, are 0.1 s and 0.5 km, respectively. Convergence is normally reached in 2 or 3 iterations.

The data sets used for the two problems need not be (and usually are not) the same. Because the inverse problem for changes in cluster vectors is based solely on arrival time differences, baseline errors in the theoretical travel times drop out and it is desirable to use all available phases at all distances outside the immediate source region. For the hypocentroid, baseline errors in theoretical travel times are more important and one may wish to limit the data set to a phase set, e.g., teleseismic P arrivals in the range 30-90°, to achieve a more stable result. The choice of data set for determining the hypocentroid has great importance in the “direct” calibration method described below.

Similarly, weighting schemes are different for the two inversions, reflecting the different natures of the two problems. Empirical reading errors for each station-phase pair are used in weighting data for estimating both the hypocentroid and cluster vectors, but the uncertainty of the theoretical travel times, which are estimated empirically for each phase from the residuals of previous runs, is relevant only to the hypocentroid.

Until this point, the HDC algorithm is used only to obtain improved relative locations for the cluster events, with a geographic location for the cluster (the hypocentroid) that is biased to an unknown degree by unmodeled Earth structure convolved with the unbalanced distribution of reporting seismic stations. The calibration process attempts to remove this bias.

Calibration of a cluster is done in two ways, which we refer to as "indirect" and "direct" calibration.

Indirect Calibration

If the location and origin time of one or more of the cluster events can be specified with high accuracy from independent information, we can calibrate the entire cluster by shifting it in space and time to optimally match the known location of the calibration event(s). The approach is illustrated in Figure 1.

The most common source of such independent information are temporary seismic network deployments that capture an event with a large number of stations at very short epicentral distances, and which is also large enough to be well recorded at regional and teleseismic distances. Aftershock studies are a frequent source of such data. We normally obtain the temporary network arrival time data and relocate calibration events ourselves, using a local velocity model when available, to ensure reliable locations. We have also used InSAR data for this purpose, but this requires the use of at least some seismic data at short distance to calibrate origin time. In a few cases, mapped faulting from large events can be used to help constrain the location of calibration events.

When we use the direct calibration approach we take into account the uncertainty of the calibration data, and when there is more than one calibration event we also include a contribution to uncertainty to reflect any discrepancy between the relative locations of the calibration events and the cluster vectors of the corresponding events.

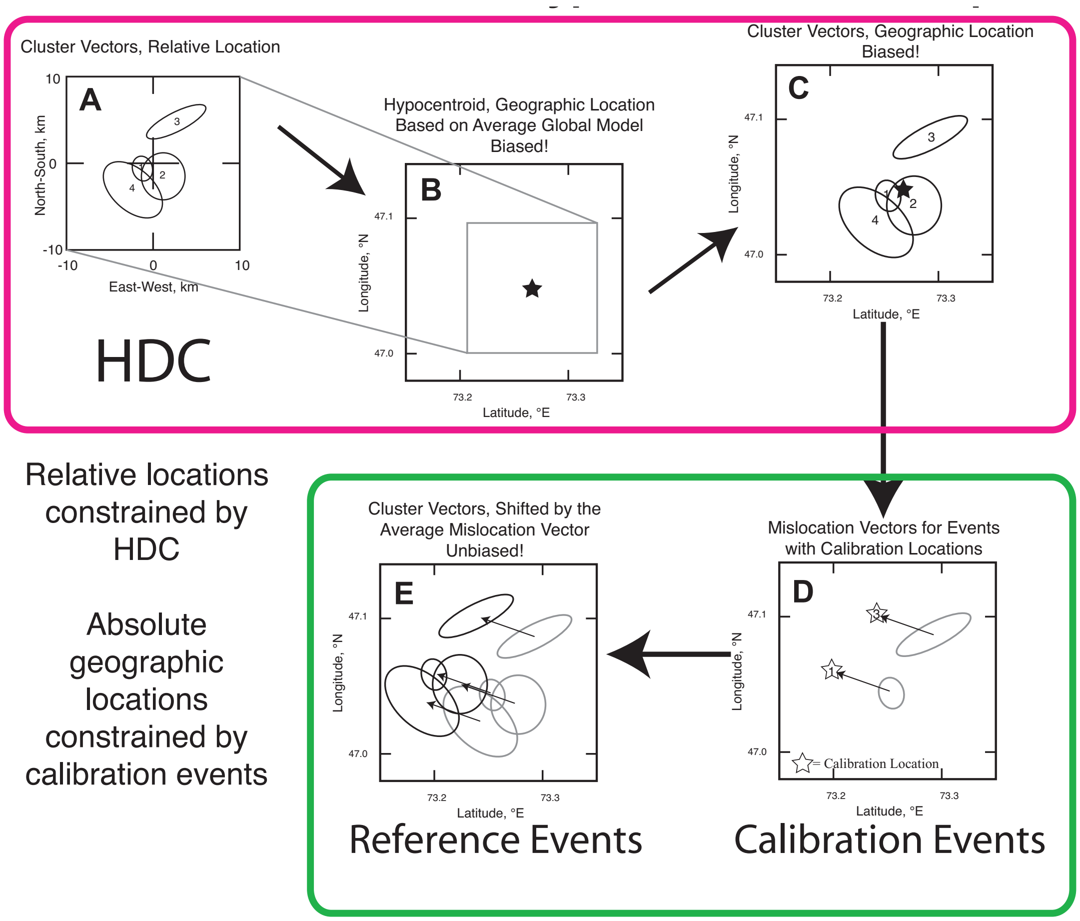

Figure 1. Indirect Calibration

Direct Calibration

It is sometimes the case that there are a few permanent seismic stations close to a cluster, but that no single event is well-enough recorded to reach the level of accuracy necessary to serve as a calibration event. On the other hand, the handful of local seismic stations may have recorded many events in the cluster, so that the number of “short distance” readings is rather large, and well-enough distributed to allow the hypocentroid to be located using only these data. We have found that good results can be obtained as long as the epicentral distance is kept less than about 150-200 km. At greater distances we often see rapidly increasing scatter in residuals at different azimuths. The process is illustrated in Figure 2, which can be contrasted with Figure 1.

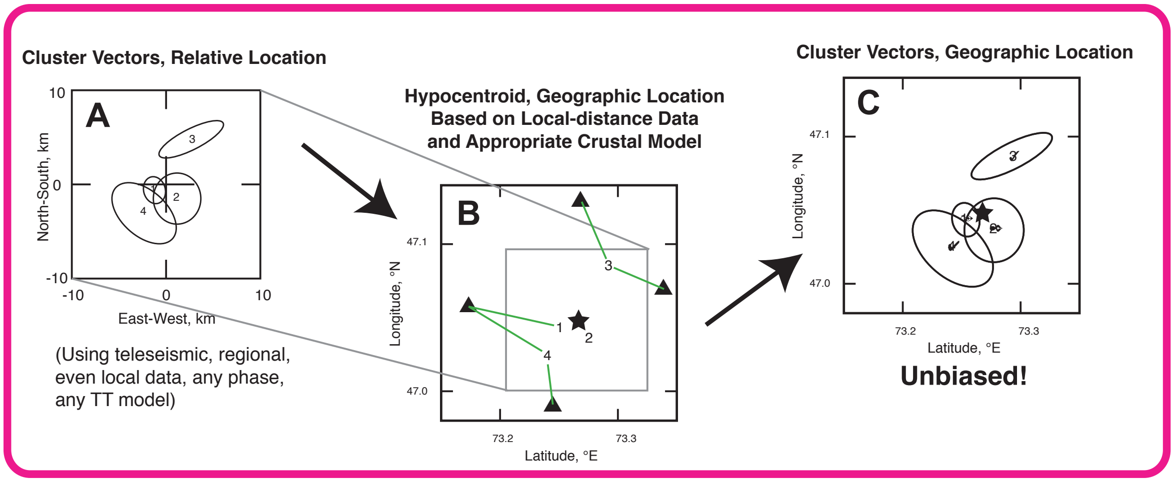

Figure 2. Direct Calibration

In any case where we have the data to locate one or more calibration events for the indirect method, we also have the option to use those same data in the direct calibration method. The decision on which to use is made on a case-by-case basis, because characteristics of the data sets may lead to a better result with one method than the other. Of course, if we are using InSAR data or other remote sensing or geological information to constrain the calibration, we must use the indirect method.

Groomed Readings

Our relocation and calibration methodology departs from standard practice in dealing with outlier arrival time readings by emphasizing consistency between repeated readings (of a given phase from different events in a cluster to the same station) rather than the size of residual against some reference travel time model. A considerable amount of work goes into estimation of what we refer to as "empirical reading errors" from the specific arrival time data set. This estimate is based on a robust estimator of spread (Croux and Rousseeuw, 1992) applied to the travel time residuals for a specific station and phase. The estimator S_n is defined as

$$S_n = c_n 1.1926 \operatorname*{lomed}_{i=1,\ldots,n} \operatorname*{lomed}_{j \neq i} |x_i - x_j|$$

where "lomed" is the low median and the factor 1.1926 establishes consistency at a normal distribution. The term c_n is a correction term for small sample sizes, which has been tabulated. S_n makes no assumption about the underlying distribution and requires no estimate of central location. A minimum value of 0.15 s for S_n is normally enforced to avoid numerical instability.

In addition to their use in weighting the arrival time data for inversion, we also use empirical reading errors to detect outliers in the data, which are flagged. Because outlier readings can cause good readings to have large residuals, the process needs to be done incrementally, starting with the largest residuals, to avoid rejection of good data. We gradually remove the largest outliers, followed by relocation, until the normalized, de-meaned distribution of each station-phase approaches a Gaussian distribution. It is continued out until the residuals satisfy a 3σ criterion, using the current empirical reading error as the estimate of σ. This “cleaning” process is crucial in providing a self-consistent statistical framework for estimating location uncertainties. The process is illustrated in Figure 3.

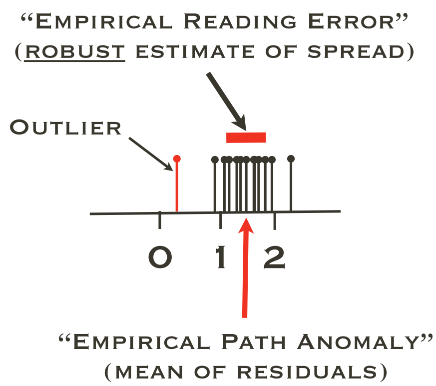

Figure 3. Outlier Removal

The effect of systematic and comprehensive (all phases and stations with 2 or more samples) estimation of empirical reading errors and associated outlier rejection is that the arrival time data set is reduced to one that more closely approaches the assumption of zero-mean, normally-distributed data which underlies the estimation of improved locations and their uncertainties. Gross errors are still possible from, for example, mis-identified phases, mis-associated readings, incorrect station locations, and temporary equipment or operational problems at stations. The distributions of empirical reading errors for the most common phases, measured from a calibrated event data set, and after cleaning, are shown in Figure 4. Note that the time scales differ significantly for different phases.

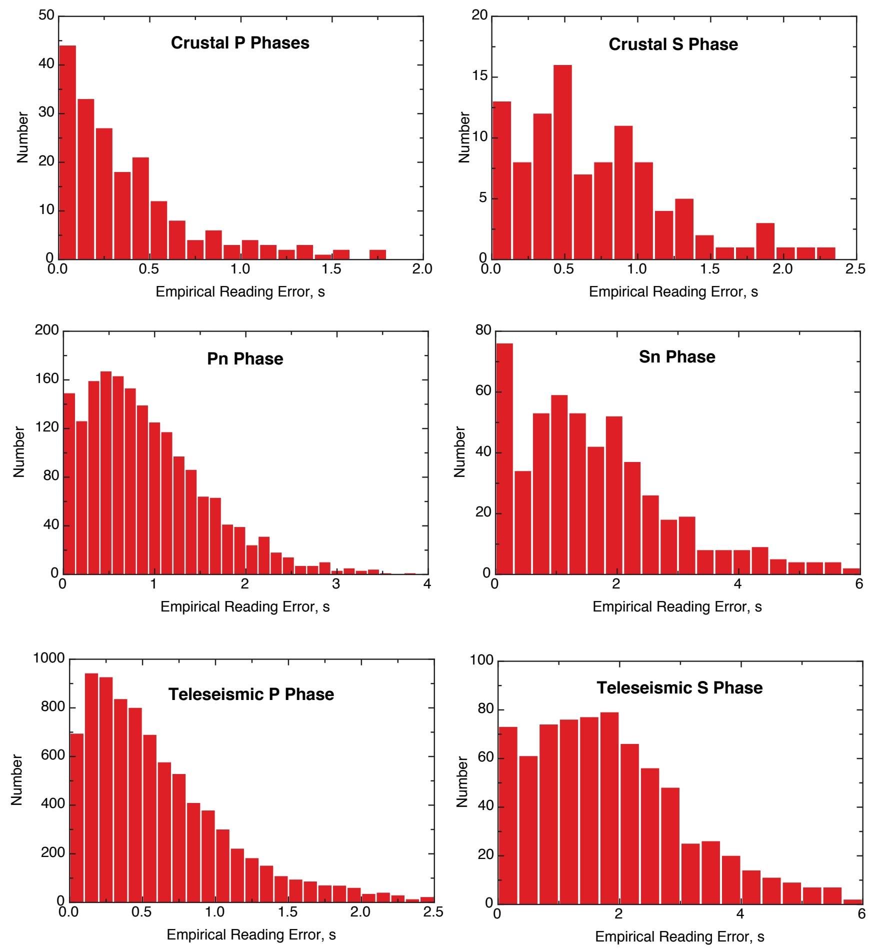

Figure 4. Histograms of Empirical Reading Errors

Calibration Levels

It has become common to use the GTX formulation ("Ground Truth to within X km") to describe the location accuracy of seismic events, although actual ground truth information is rarely available. As Bondár et al. (2004) noted, it is also necessary to specify a level of confidence for such a designation. In our study confidence ellipses are calculated at the 90% level. We always use the confidence level subscript when making a quantitative statement or when referring to a specific calibrated data set, but it may be dropped in a more general context.

Moreover, it is necessary to be clear about what metric is used for the GT classification. We take the nearest integer of the length of the semi-major axis of the 90% confidence ellipse that includes both the uncertainty of relative location and the uncertainty of the calibration process. In our experience it is possible to achieve GT_{90} levels of 2-3 km in the most favorable circumstances.

Depending on the purpose to which one might want to apply a set of reference events, the desired level of accuracy will vary. GT5 is the most commonly discussed level for calibrated earthquake locations, and GT5 (or better) locations certainly represent a significant improvement in accuracy over what is available in standard catalogs. In our calibration studies we have generally discarded events whose relative location within the cluster cannot be determined to better than 10 km (semi-major axis length).

Focal Depth

The HDC algorithm is poorly suited to including focal depth as a free parameter in relocation because it would require “connectivity” between all events with observations constraining depth, such as stations within 1-2 focal depths or teleseismic depth phases. If even one event in a cluster has no near-source readings or teleseismic depth phases that are shared with other events, the inversion for cluster vectors becomes singular.

For this reason most calibration location studies are performed with fixed focal depths and considerable effort goes into obtaining the best possible constraints on focal depth for these events. We make a careful review of observed depth phases and body waveform studies for this purpose. During relocation we often experiment with a range of depths and discover that there is strong sensitivity to changes of more than about 5 km in average assumed depth, relative to the preferred depth. In addition, when we have data at short epicentral distance for direct calibration, it is common to observe curvature of the travel-time vs. distance curve at short distances, which is diagnostic of source depth.

Our general philosophy has been to establish a default depth for each cluster, based on the best-constrained events in the cluster. Tests have shown that errors of less than about 15 km in depth have negligible effect on the estimated epicenter.

An Example

As an example of the application of the calibrated location procedures described above, we consider a cluster of earthquakes in Kyrgyzstan based mainly on a large (M_S 7.5), damaging earthquake on August 19, 1992, called the “Susamyr” event and it’s aftershocks. Calibration is done with the indirect method, using the closely-controlled location and origin time of a large engineering explosion conducted in the region on December 22, 1999, the so-called “Kambarata” shot.

The cluster was formed by searching the ISC Bulletin for events in the vicinity of the Kambarata explosion, yielding 38 earthquakes between 1966 and 2008, 25 of which belong to the Susamyr sequence in 1992. The first relocation was done without the Kambarata explosion in the data set, locating the hypocentroid of the cluster by fitting 2628 teleseismic P readings (30°-90° epicentral distance) to the ak135 travel time model (Figure 5). Depths were fixed at the values reported in the ISC Bulletin, ranging from 0 to 70 km.

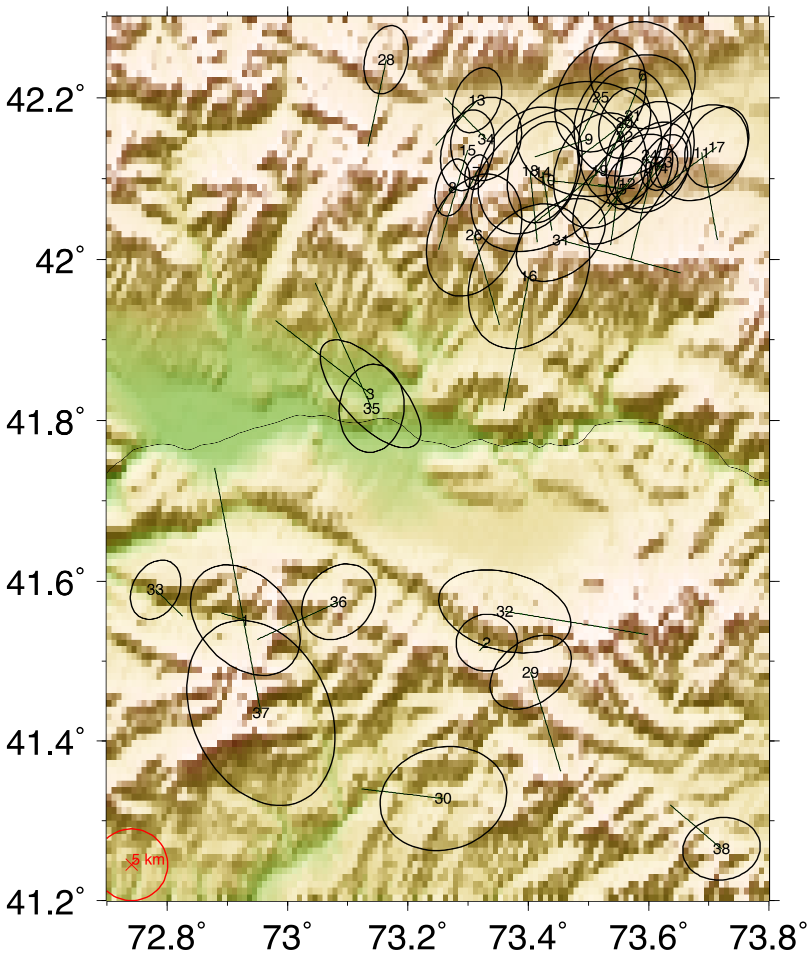

Figure 5. The Susamyr cluster initial relocation, without the calibration event, with default reading errors, and no cleaning of outliers. Black vectors show change in location from the location provided in the input data files (ISC locations). Confidence ellipses are at 90% level of confidence, for relative locations. The red circle (radius 5 km) is shown for scale.

The formal uncertainty of the location of the hypocentroid is 3.4 km (largest semi-axis of the 90% confidence ellipse), but there is an unknown level of bias in the location, which propagates to the location of each event, due to departures of the true Earth from the assumed velocity model. Our goal is to try to minimize this bias, and also to reduce the size of the confidence ellipses for relative location in Figure 5 by estimating empirical reading errors for each station-phase set and using that information to identify and remove outlier readings, and to weight the inversion. The uncertainties of the relative locations range from 2.4 to 13.4 km, measured as the largest semi-axis of the confidence ellipse.

For subsequent relocation runs, the events for which teleseismic depth phases were reported were reviewed and depths were fixed at values that satisfy the depth phase data. The remaining events were held to a default depth for the cluster of 16 km that is based on the average depth of events for which there is depth constraint from teleseismic depth phases.

The cleaning process, removing the largest outliers and relocating, required 16 more runs of the relocation program. The calibrated locations are shown in Figure 6.

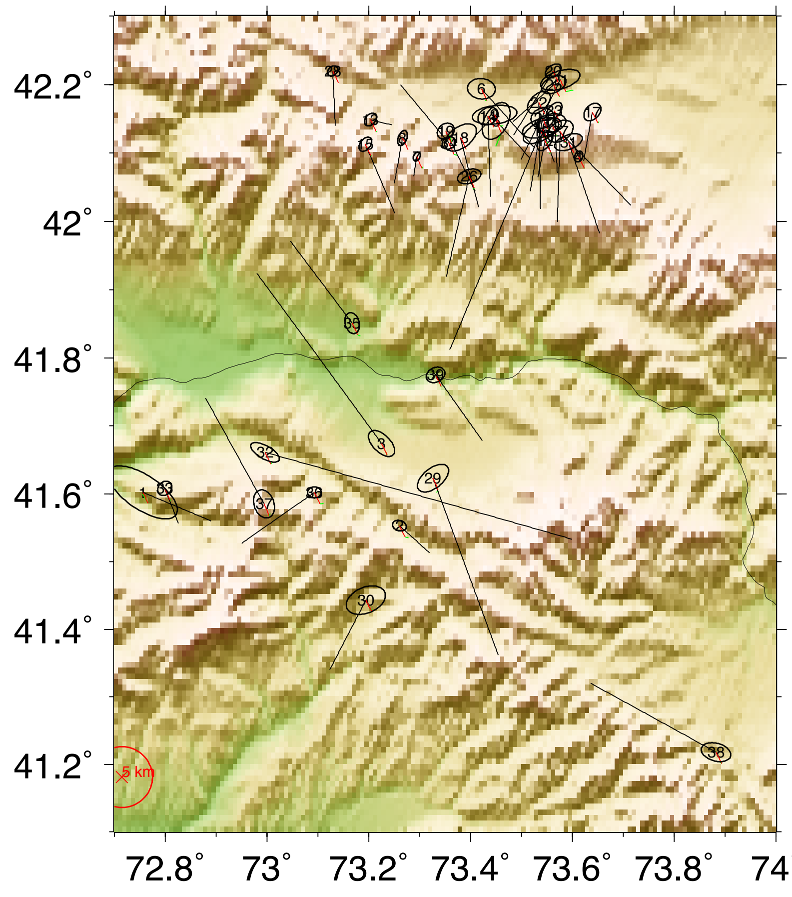

Figure 6. Final calibrated locations of the Susamyr cluster, after outlier removal and weighting according to empirical reading errors, with calibration based on the Kambarata explosion. Red line (barely visible) shows the small shift of the cluster required to fit the

calibration event. Other features as in Figure 5.

In the case of the Susamyr cluster, there was very little location bias (2.1 km) for the cluster as a whole, because of the excellent azimuthal coverage by teleseismic P waves and the fact that the ak135 travel time model performs rather well in this region. Location bias of 10 km or more is not uncommon in many regions. Although the epicentral bias in this case is small, the shift in origin time for the cluster as a whole (-1.1 s) is significant for any application, such as tomography, that would make use of absolute travel times from this cluster.

Moreover, the changes to the locations of individual events in the cluster from their single event locations are tens of km in many cases, and resolution of the pattern of aftershock seismicity for the 1992 Susamyr earthquake is dramatically improved by relocation. After relocation the uncertainty in absolute location ranges from 1.6 to 6.7 km. Because the absolute location of the cluster has been calibrated, these locations can be more usefully interpreted in terms of seismotectonics and seismic hazard.

The Susamyr cluster is typical in terms of the improvement in resolution of relative locations and the level of absolute location accuracy that can be obtained through the procedures described here.

Applications

The methods described here for developing calibrated clusters of earthquakes have been used in a number of studies, including:

Biggs, J., Bergman, E.A., Emmerson, B., Funning, G.J., Jackson, J.A., Parsons, B.E., and Wright, T.J., 2006, Fault identification for buried strike-slip earthquakes using InSAR: The 1994 and 2004 Al Hoceima, Morocco earthquakes: Geophysical Journal International, v. 166, p. 1347–1362.

Bondar, I., Bergman, E.A., Engdahl, E.R., Kohl, B., Kung, Y.-L., and McLaughlin, K., 2008, A hybrid multiple event location technique to obtain ground truth event locations: Geophysical Journal International, v. 175, no. 1, p. 185–201.

Bondar, I., Engdahl, E.R., Yang, X., Ghalib, H.A., Hofstetter, A., Kirichenko, V., Wagner, R., Gupta, I., Ekström, G., Bergman, E.A., Israelsson, H., and McLaughlin, K., 2004, Collection of a reference event set for regional and teleseismic location calibration: Bulletin of the Seismological Society of America, v. 94, no. 4, p. 1528–1545.

Nissen, E., Yamini-Fard, F., Tatar, M., Gholamzadeh, A., Bergman, E.A., Elliott, J.R., Jackson, J.A., and Parsons, B.E., 2010, The vertical separation of mainshock rupture and microseismicity at Qeshm island in the Zagros fold-and-thrust belt, Iran: Earth and Planetary Science Letters, v. 296, no. 3-4, p. 181–194.

Parsons, B.E., Wright, T.J., Rowe, P., Andrews, J., Jackson, J.A., Walker, R.T., Khatib, M., Talebian, M., Bergman, E.A., and Engdahl, E.R., 2006, The 1994 Sefidabeh (eastern Iran) earthquakes revisited: new evidence from satellite radar interferometry and carbonate dating about the growth of an active fold above a blind thrust fault: Geophysical Journal International, v. 164, p. 202–217.

Rastogi, B.K., Bergman, E.A., and Engdahl, E.R., 2005, Improved earthquake locations and estimation of Pn and Sn path anomalies for India, using multiple event relocation and reference events: Current Science, v. 88, no. 10, p. 1586–1591.

Ritzwoller, M.H., Shapiro, N.M., Levshin, A.L., Bergman, E.A., and Engdahl, E.R., 2003, Ability of a global three-dimensional model to locate regional events: Journal of Geophysical Research, v. 108, no. B7, p. 2353–2353.

Roustaei, M., Nissen, E., Abbassi, M.R., Ghorashi, M., Gholamzadeh, A., Tatar, M., Yamini-Fard, F., Bergman, E.A., Jackson, J.A., and Parsons, B.E., 2009, Vertical separation of surface folding, earthquake faulting, and aftershocks in the Zagros Simply Folded Belt (Iran): Geophysical Journal International, v. 142, p. 1–24.

Tatar, M., Jackson, J.A., Hatzfeld, D., and Bergman, E.A., 2007, The 2004 May 28 Baladeh earthquake (Mw 6.2) in the Alborz, Iran: Overthrusting the South Caspian Basin margin, partitioning of oblique convergence and the seismic hazard of Tehran: Geophysical Journal International, v. 170, p. 249–261.

Walker, R.T., Bergman, E.A., Szeliga, W., and Fielding, E.J., 2011, Insights into the 1968-1997 Dasht-e-Bayaz and Zirkuh earthquake sequences, eastern Iran, from calibrated relocations, InSAR and high-resolution satellite imagery: Geophysical Journal International,, p. no–no.

Walker, R.T., Bergman, E.A., Jackson, J.A., Ghorashi, M., and Talebian, M., 2005, The 2002 June 22 Changureh (Avaj) earthquake in Qazvin province, northwest Iran: epicentral location, source parameters, surface deformation and geomorphology: Geophysical Journal International, v. 160, p. 707–720.

References

Bondar, I., Myers, S.C., Engdahl, E.R., and Bergman, E.A., 2004, Epicentre accuracy based on seismic network criteria: Geophysical Journal International, v. 156, p. 483–496.

Croux, C., and Rousseeuw, P.J., 1992, Time-efficient algorithms for two highly robust estimators of scale: Computational Statistics, v. 1, p. 411–428.

Jordan, T.H., and Sverdrup, K.A., 1981, Teleseismic location techniques and their application to earthquake clusters in the South-Central Pacific: Bulletin of the Seismological Society of America, v. 71, no. 4, p. 1105–1130.