

A hybrid multiple event location technique to obtain ground truth event locations

István Bondár,^{1,*} Eric Bergman,² E. Robert Engdahl,³ Ben Kohl,¹ Yu-Long Kung¹ and Keith McLaughlin¹

¹Science Applications International Corporation, La Jolla, CA, USA. E-mail: istvan@isc.ac.uk

²Global Seismological Services, Golden, CO, USA

³University of Colorado at Boulder, CO, USA

Accepted 2008 June 3. Received 2008 May 29; in original form 2007 January 22

SUMMARY

We present a multiple event location technique that is able to produce minimally biased (location accuracy of 5 km or better) events from a cluster of earthquakes, without relying on prior ground truth information or dense local seismograph networks. In the hybrid hypocentroidal decomposition and reciprocal cluster analysis (HDC-RCA) method, HDC determines accurate relative locations in an event cluster (but a biased hypocentroid) using regional and/or teleseismic data. RCA uses local phases and invokes the reciprocity principle by relocating the local station network centroid with the events fixed (by HDC) as fictitious stations. Since the location of the local station network centroid is exactly known, the mislocation vector between the true and relocated centroids represents the offset with which the entire event cluster must be shifted to obtain unbiased absolute locations. We have developed applicability criteria to help determine under what conditions HDC-RCA is able to determine minimally-biased locations in an event cluster. The performance of the HDC-RCA method is validated on several event clusters containing previously determined ground truth events and applied to clusters for which ground truth events were not previously determined. Events promoted to GT5 level are provided in an electronic supplement.

Key words: Computational seismology.

INTRODUCTION

A set of accurate seismic event locations are necessary to develop and validate velocity models and perform location calibration studies. If the event locations are biased, the products (velocity models, traveltimes predictions) will also be biased. Since the main objective of earthquake catalogues is to achieve completeness down to the lowest magnitude levels, published bulletins contain a mixture of accurate, good and poor locations. Thus, event locations in earthquake catalogues should always be treated with caution. To identify accurately located events at a high confidence level in published bulletins, Bondár *et al.* (2004a) developed selection criteria, which are based on the network geometry. However, for events at the GT5 level, these criteria require a dense local network with good azimuthal coverage. Therefore, the scarcity of dense local networks limits our ability to obtain wide coverage by GT events, with this approach.

Multiple event location methods are proven techniques to obtain improved accuracy for relative locations in an event cluster (e.g. Douglas 1967; Dewey 1972; Jordan & Sverdrup 1981; Pavlis & Booker 1983; Got *et al.* 1994; Waldhauser & Ellsworth 2000; Rodi *et al.* 2003; Myers *et al.* 2005, 2007). Although all methods can provide improved relative locations, none of them can resolve absolute event locations without dense local station coverage (Lin & Shearer 2005; Richards *et al.* 2006).

To obtain unbiased absolute locations, modern multiple event location techniques utilize independent GT information such as existing reference events (e.g. Ritzwoller *et al.* 2003; Bondár *et al.* 2004b), seafloor bathymetry (Pan *et al.* 2002), satellite imagery (e.g. Fisk 2002), InSAR interferometry (e.g. Biggs *et al.* 2006; Parsons *et al.* 2006) and active fault lines (Waldhauser & Richards 2004) to estimate the mislocation vector between the true and apparent cluster hypocentroid. Hence, the availability of prior accurate independent GT information limits the applicability of multiple event location methods, for generating new GT events.

Our motivation in this study is to develop a multiple event location technique that provides accurate relative and absolute locations,

*Now at: International Seismological Centre, Pipers Lane, Thatcham, Berkshire, RG19 4NS, UK.

without reliance upon independent GT information or upon dense local networks.

HDC-RCA METHOD

Hypocentroidal decomposition and reciprocal cluster analysis (HDC-RCA) is a hybrid location technique that combines a traditional multiple event location method, using teleseismic and regional data, with local data to obtain more accurate absolute event locations by using local seismic stations as GT0 constraints. In the following, we present a summary of the HDC-RCA method, describe the procedures to form event clusters for the HDC-RCA analysis, present the HDC and RCA algorithms and identify the conditions under which HDC-RCA can be expected to produce locations that are accurate (5 km or better) at a high confidence level. The RCA algorithm should not be confused with the method of Shearer (2001), which used reciprocity to constrain traveltime corrections for events in region A and stations in region B to the reverse scenario of stations in region A to events in region B.

Summary of the HDC-RCA method

The HDC-RCA method combines a traditional multiple event location method with an innovative approach for removing location bias from a cluster of events, when local seismograph stations satisfy certain geometric criteria with respect to the event cluster. In principle the method can work with a single local seismic station. The choice of multiple event relocation method is not critical; any method that can constrain relative locations accurately would be suitable. It would be certainly possible to consolidate the entire procedure into a single piece of software, instead of implementing it as two separate steps. However, because of the natural separation of the problem in the regional/teleseismic and local domains, we decided to employ the already well-tested HDC for the regional/teleseismic relative location problem and concentrate our efforts on the novel RCA method to obtain absolute locations using local data.

The role of HDC analysis in the HDC-RCA method is to provide the best possible estimate of the relative locations, including origin time, of the events in a cluster. The accuracy of the absolute locations, that is, location bias of the cluster as a whole, obtained in the HDC analysis is of less concern because that will be corrected in the RCA analysis. It is only required that the HDC-derived absolute locations, represented by the accuracy of the hypocentroid of the cluster, be close enough to the true location, so that the RCA analysis can converge to the global error minimum in parameter space. This is very much analogous to the issue of having an acceptable starting location in a single event location procedure. In analyses of nearly a hundred different clusters in a variety of tectonic situations, we have never found a case in which the HDC hypocentroid is too far from the global error minimum for the RCA analysis to succeed.

The RCA analysis begins by taking the relative locations determined in the HDC analysis as a fixed constellation of locations, with uncertainties defined by the covariance matrix of the cluster vectors (i.e. relative locations with respect to the HDC hypocentroid). The relative locations will not be changed by the RCA analysis. The starting absolute locations of the cluster events are established by the hypocentroid of the HDC analysis, but, as noted above, the RCA analysis is not very sensitive to this choice.

Given a set of event locations from the HDC analysis and a set of local stations that observed at least some of the cluster events, RCA

proceeds by treating the events as stations and the stations as events. Since we can constrain the relative locations of the local stations from their known locations, RCA locates only the centroid of the local station network, using all the readings at all the stations to any events in the cluster, with appropriate corrections for traveltimes and derivatives. Thus, RCA is very much the same as a single event local earthquake location procedure. The resulting location of the station network centroid is biased (relative to the known centroid of the local station network) according to the bias of the hypocentroid of the HDC process that was used to fix the locations of the fictitious station network. This bias vector is used to correct the absolute locations of the fictitious stations (i.e. the actual seismic events), yielding a set of unbiased locations that can be treated as ground truth or reference events for other purposes. The RCA analysis is illustrated in Fig. 1.

The level of ground truth (e.g. GT5) depends on the uncertainties (covariance matrix) of the relative locations in the HDC analysis and on the uncertainty of the estimate of the local station network centroid in the RCA analysis. The uncertainty of the HDC hypocentroid

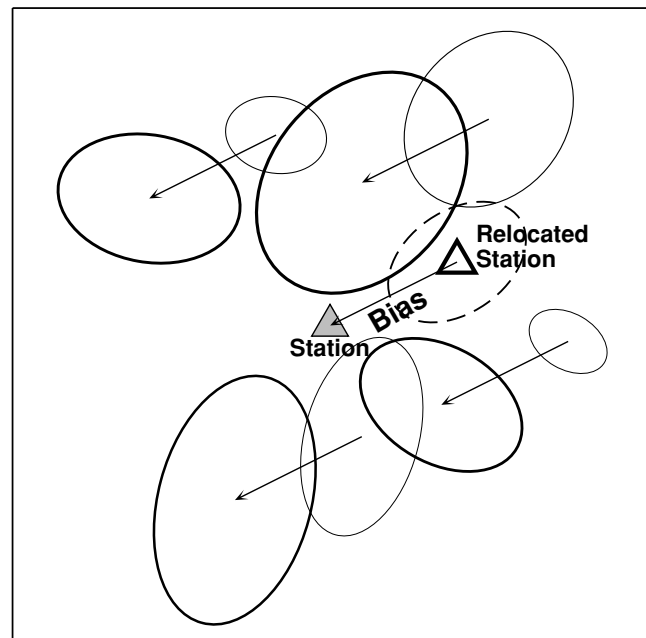


Figure 1. Cartoon of the Reciprocal Cluster Analysis (RCA) analysis, using a cluster of four seismic events and a single local seismic station (large filled triangle). This illustrates the process for the epicentre; focal depth and origin time are treated similarly. Hypocentroidal Decomposition (HDC) is used with phase arrivals at regional and teleseismic distances to constrain the relative locations of the four events, with confidence ellipses for relative location shown as thin lines. The location of the cluster as a whole is biased by some unknown amount due to the use of a global average traveltime model. With these event locations fixed as fictitious stations, phase arrival times recorded at the local station are used to locate the station as if it were an earthquake (large open triangle) with an associated confidence ellipse (dashed), representing the uncertainty of the RCA estimation process. The vector from the relocated station to the known station location is the calibration vector that is applied to the cluster to correct for teleseismic location bias, yielding the unbiased locations. The confidence ellipses for the calibrated locations (heavy lines) are larger than the original HDC confidence ellipses because of the added uncertainty of the RCA estimation process. If there are more than one local stations, the RCA location process operates on the centroid of the local station network, with the same number of free parameters as the single station case.

is not any more relevant than the uncertainty of the starting location in a normal location procedure. Obviously, the level of ground truth is different for each event in the cluster.

HDC-RCA data selection

We typically extract seismic event clusters from an updated EHB (Engdahl *et al.* 1998) bulletin, which spans four decades of groomed International Seismological Centre (ISC 2001) bulletins. Although in many cases the clusters are formed from aftershock sequences, the events in the cluster may actually be widely distributed in time, as long as arrival time data at common stations are available. We seek localities where a number of moderate-size, shallow (crustal) earthquakes are spatially clustered (with about 50–100 km cluster diameter) and a subset of events has been recorded by one or more local stations. This last criterion is absolutely essential to the RCA method. It is not necessary that the local stations be numerous enough or distributed so as to be able to locate any of the events in a traditional manner. In many regions of the world, local seismograph networks have been deployed only recently. Hence, although we can use the entire span of the EHB bulletin to build an HDC cluster, the subset of locally recorded events in these regions is often restricted to the most recent events. It is desirable to include events that are well recorded at regional and teleseismic distances, even if they occurred prior to the installation of local seismograph stations, because they add resolving power for the relative locations of all events in the HDC analysis.

Multiple event location techniques implicitly rely on the assumption that the events in a cluster and the stations used in the inversion are well connected, that is, all events are well observed by stations, which also observed many other events in the cluster. This assumption is not always valid, especially for clusters that span a long time period. We require that all events are linked to all other events through two or more stations, and all stations are linked to all stations through two or more events (biconnected).

To build a cluster for HDC analysis, we extract the largest biconnected cluster from the initial EHB event cluster, using stations in the regional and teleseismic distance range (usually, beyond 3° epicentral distance). During the HDC analysis, individual arrivals or entire events may be rejected on the basis of outlier analysis. For the RCA analysis, we again extract the largest biconnected cluster using stations only in the 0°–1.5° distance range. Thus, the connectivity requirements are enforced for the HDC and the RCA steps independently. The choice for the RCA distance range is motivated by the desire to select truly local phases (*Pg*, *Sg*, *Pb*, *Sb*) and avoid phase identification problems in the *Pg/Pn* crossover distance range. We prefer to choose stations in an even closer distance range (up to 100 km); stations further away (up to 1.5°) are selected only if they are deemed necessary to close a large azimuthal gap.

Hypocentroidal decomposition

The basic HDC algorithm is described in detail in Jordan & Sverdrup (1981). The HDC method separates the multiple event location problem into the estimation of the relative event locations defined by cluster vectors (relative to the hypocentroid) and the location of the hypocentroid itself, which establishes the absolute locations and origin times for all cluster events. Estimation of the cluster vectors and the hypocentroid are done as two separate inversions, using different

data sets and weighting schemes. The two-step process is repeated iteratively until convergence is reached.

For the estimation of cluster vectors, we normally use all phase arrivals (first arriving and secondary) beyond 3° epicentral distance. An analysis of observed residuals is used to make a robust estimate of empirical reading errors for each station-phase pair, which includes traditional picking errors as well as variability in arrival times due to station equipment changes, small changes in station location, timing errors, unmodelled velocity variations across the cluster and departures from the assumption of common ray paths and correlated path errors. The data are weighted inversely to these reading errors, which are also used in an analysis to identify and remove outliers from the data set.

For this study, estimation of the hypocentroid is normally done with only the teleseismic *P* phase between 30° and 90° epicentral distance. This provides a stable solution that is usually less biased than an epicentre that includes secondary phases or *P* phases at regional distances because the predicted traveltimes for these phases from the Ak135 model (Kennett *et al.* 1995) are less reliable. In some cases, regional phases might be used to provide adequate azimuthal coverage. Data weighting for the hypocentroid includes the empirical reading errors discussed above, as well as a weight term that reflects the scatter of specific phases with respect to the Ak135 model (i.e. lateral heterogeneity).

The cluster hypocentroid is located in an absolute sense, as if all the data were from a single event, using the Ak135 model (Kennett *et al.* 1995). Because of the use of a 1-D earth model, the HDC cluster hypocentroid is biased due to unmodelled velocity heterogeneities in the Earth and their interaction with the particular set of ray paths from the cluster events to observing stations. Even for well-constrained events, the bias often exceeds 10 km in continental regions and may reach 50 km in the vicinity of subduction zones. For poorly constrained events or events located primarily with regional arrivals, the level of bias is usually worse. The goal of the subsequent RCA analysis is to minimize this location bias.

Reciprocal cluster analysis

RCA utilizes stations up to 1.5° epicentral distance (preferably within 100 km), a completely separate data set from the one used in the HDC analysis and one which is especially suited to estimating absolute event locations because the accumulated errors in predicted traveltimes are automatically limited by the short ray paths. We use a local velocity model, whenever available, to obtain traveltime predictions for local phases. We apply a distance-dependent uncertainty to predicted traveltimes (Fig. 2), for two reasons:

- (1) Uncertainty in the predicted traveltimes should be propagated into the uncertainty of the station network centroid, which establishes the location bias of the HDC cluster.
- (2) To help account for lateral heterogeneity within the region defined by the event cluster and local seismograph stations.

Clearly, the possibility exists that there could be strong local velocity heterogeneity, whose biasing effect will not be captured in this error term.

In the RCA inversion step, we fix the ‘pattern’ of relative hypocentres and origin times obtained from the HDC step and then locate the centroid of the local seismic station network, using the HDC event locations as fixed fictitious stations. The RCA concept envisions the local station network and the event cluster as two rigid meshes. One mesh is fixed; the other mesh is allowed to move but only as

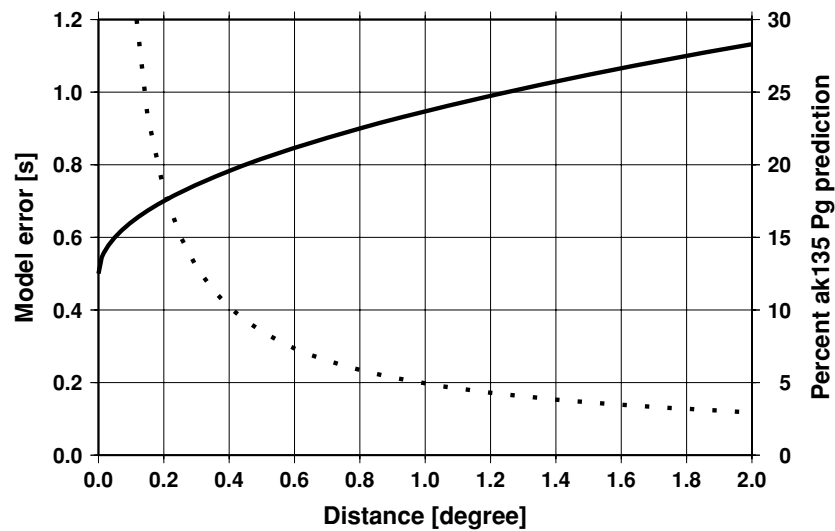


Figure 2. Distance-dependent model error curve (thick line), employed to account for local traveltime prediction errors. The dotted line represents the model errors as the percentage of the Ak135 *Pg* traveltime for a surface-focus event. The model errors allow for large deviations from the Ak135 velocity model for the typically short ray paths ($<1.5^\circ$) used in the RCA processing.

a semi-rigid entity. The amount by which the centroid of the free mesh moves is the RCA correction to the HDC centroid bias.

Because the event pattern of locations and origin times is fixed by HDC, the RCA inversion solves only for an over-determined shift in the centroid of the local station network required by the local data. Note that in this representation, it is irrelevant which mesh (HDC relative event locations or known pattern of local stations) is fixed and which one is free. Therefore, locating the local station network centroid is equivalent to relocating the hypocentroid of the HDC cluster. RCA solves for four or three unknowns, depending on whether depth is a free parameter or not. The centroid of the local

station is defined as the mean latitude and longitude of the station coordinates and zero depth and origin time. The origin times of the individual events are adjusted by the mean of time residuals.

Therefore, when the rigid body constraints are added, the multiple event location system of equations with local data, which would typically pose an underdetermined and unstable inversion problem due to the low number of observations relative to the number of model parameters, simplifies to solving for the overdetermined station centroid. Any algorithm for non-linear minimization of the weighted sum of the squares of residuals (chi-squared statistic) may be utilized. An existing standard earthquake

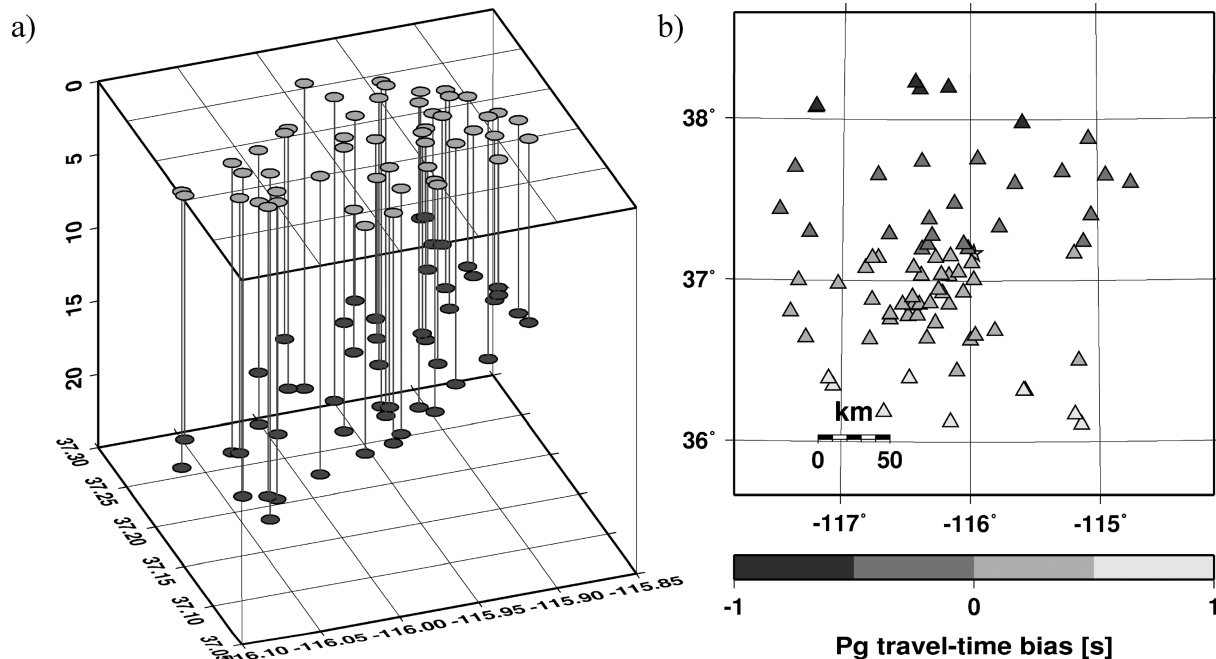


Figure 3. (a) Synthetic cluster generated from Yucca Flat, Nevada Test Site nuclear explosions. Events were projected to a ‘fault’ plane (strike = 225° , dip = 45°), and arrival times were calculated using IASP91 predictions and distance- and azimuth-dependent delays. (b) Local network configuration for the synthetic cluster. Traveltime biases were introduced, as shown by the shading of the symbols, such that *Pg* arrival times at stations to the south are increasingly late and those to the north are increasingly early.

location programme can be readily modified and is usually stable and fast.

Assume that there are $k = 1 \dots K$ local stations and $i = 1 \dots N$ events in a cluster of events. Each event has M_i local observations, which constitute $M = \sum_{i=1}^N M_i$ total number of observations. We allow the use of secondary phases in the RCA inversion, and a station does not necessarily record every event in the cluster. The residual of the j th observation at the k th station for the i th event is then written as the difference between the observed and predicted arrival time $d_{kji} = (T_{kji}^{\text{obs}} - T_{kji}^{\text{pred}})$. The model parameter perturbations are the horizontal and vertical coordinates of the station centroid shift and a single origin time-shift: $[\Delta\lambda_C, \Delta\varphi_C, \Delta z_C, \Delta\tau_C]^T$. We solve the system of linearized equations shown below by an iterative weighted least-squares method.

$$\begin{bmatrix} \frac{\partial T_{1,1}}{\partial \lambda_1} & \frac{\partial T_{1,1}}{\partial \varphi_1} & \frac{\partial T_{1,1}}{\partial z_1} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial T_{1,N}}{\partial \lambda_1} & \frac{\partial T_{1,N}}{\partial \varphi_1} & \frac{\partial T_{1,N}}{\partial z_1} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial T_{k,i}}{\partial \lambda_k} & \frac{\partial T_{k,i}}{\partial \varphi_k} & \frac{\partial T_{k,i}}{\partial z_k} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial T_{K,N}}{\partial \lambda_K} & \frac{\partial T_{K,N}}{\partial \varphi_K} & \frac{\partial T_{K,N}}{\partial z_K} & 1 \end{bmatrix} \begin{bmatrix} \Delta\lambda_C \\ \Delta\varphi_C \\ \Delta z_C \\ \Delta\tau_C \end{bmatrix} = \begin{bmatrix} d_{1,1} \\ \vdots \\ d_{1,N} \\ \vdots \\ d_{k,i} \\ \vdots \\ d_{K,N} \end{bmatrix}$$

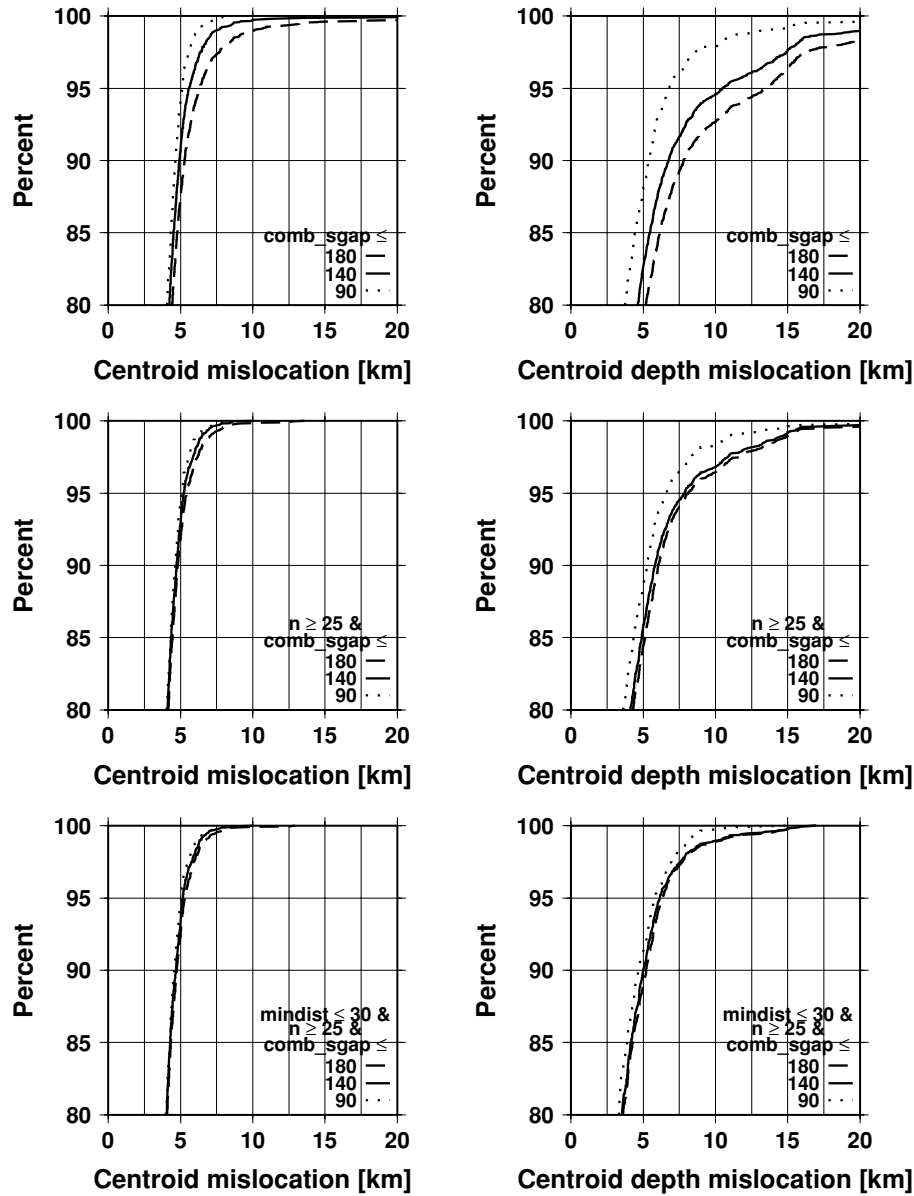


Figure 4. Cumulative distributions of the RCA event hypocentroid mislocations from the Monte Carlo experiment with a synthetic cluster. (a) horizontal mislocation for combined secondary azimuth gaps less than 90°, 140° and 180°. With a combined secondary azimuthal gap less than 90°, the event centroid is recovered within 5 km, 95 per cent of the time. (b) vertical mislocation for combined secondary gaps less than 90°, 140° and 180°. (c) as in (a) and at least 25 observations. (d) as in (b) and at least 25 observations. (e) as in (c) and at least one station closer than 30 km. (f) as in (d) and at least one station closer than 30 km. Depth is now recovered within 6–8 km at the same, 95 per cent confidence level.

The size of the matrix of partial derivatives is $M \times 4$. Writing the system of equations in this way represents the underlying RCA concept of treating the stations and events as two separate rigid meshes that are only allowed to move relative to each other, without changing the relative patterns of events or stations. The reciprocity principle is reflected in the fact that we evaluate the partial derivatives of the predicted traveltimes at the station locations (fictitious events).

In the iterative least-squares process, the residuals are weighted by their *a priori* estimates of uncertainty in the standard manner. Applying the reciprocity principle implies that we do not know exactly where our fictitious stations are, due to uncertainties in the relative event locations (covariance matrix of the cluster vectors in HDC). To account for this extra error term, we propagate the relative event location uncertainties into the local traveltime uncertainties. This is done by multiplying the HDC model covariance matrices (Σ_i) describing the spatial and origin time uncertainties in the event locations (with respect to the hypocentroid) by the predicted slowness vector (s_{kji}) for the i th event at the k th station arrival time observation: $\sigma_{kji,rel}^2 = s_{kji}^T \Sigma_i s_{kji}$. Thus, readings for events with large relative uncertainties (with respect to the hypocentroid) are down-weighted in the RCA inversion. This is equivalent to allowing the event pattern to deform, consistent with the individual HDC output error ellipses. The total *a priori* uncertainty in an observation is then expressed by the sum of the measurement, model and relative location errors: $\sigma_{kji,total}^2 = \sigma_{kji,meas}^2 + \sigma_{kji,model}^2 + \sigma_{kji,rel}^2$.

Convergence is evaluated at the end of each iteration in both model and data space. Convergence is achieved if in three consecutive iterations, the norm of the model adjustment vector decreases and

is less than a threshold or the expression $\frac{\|\mathbf{G}^T \mathbf{d}\|}{\|\mathbf{G}\| \|\mathbf{d}\|} < \varepsilon$ (where \mathbf{G} is the weighted derivative matrix and \mathbf{d} is weighted data residual vector) is satisfied (IDC Documentation 1999).

Once a convergent solution is reached, the mislocation vector between the known and estimated station centroids represents the bias correction to the HDC locations. The entire cluster (including events that did not take part in the RCA inversion) is then shifted so that the apparent and true station centroids coincide, thus yielding unbiased absolute event locations and origin times. Finally, the model covariance matrix of the station centroid shift is added to the covariance matrix of the relative event locations to obtain the absolute location uncertainties for each event in the cluster. The covariance matrices are scaled to the 95 per cent confidence level and the 95 per cent coverage error ellipses, depth and origin time errors are calculated. Events with a semi-major axis of less than 5 km are promoted to GT5 status.

Selection criteria

It is useful to have a way to evaluate the likelihood that a data set of interest, comprising some distribution of earthquakes and local seismic stations, could be analysed by the HDC-RCA method and be expected to yield ground truth events at GT5 or better. Our objective is to derive applicability criteria, similar to those developed by Bondár *et al.* (2004a), to identify prospective GT5 event locations in published bulletins.

To answer this question, we performed a Monte Carlo analysis on a synthetic event cluster created by projecting Yucca Flat,

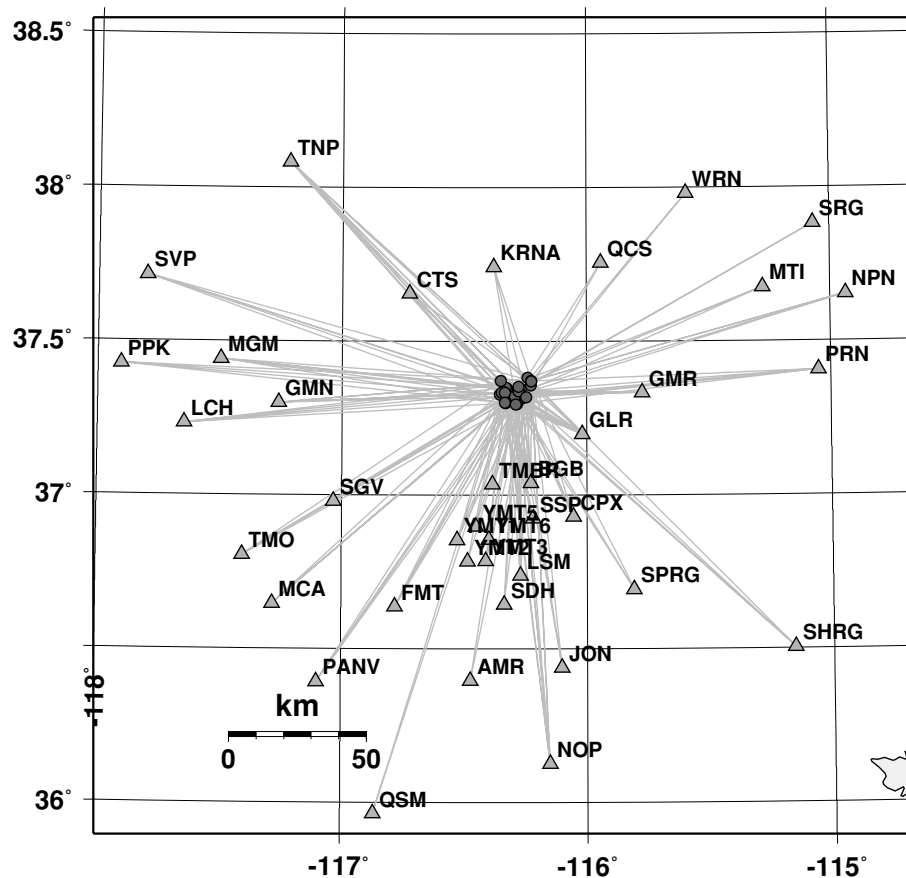


Figure 5. The Pahute Mesa cluster used in the validation tests. The cluster consists of 52 GT0 events (circles), each recorded by at least two stations from the local network of 38 stations (triangles). Each station recorded at least two events.

Nevada Test Site, GT0 nuclear explosions, along a hypothetical fault plane (Fig. 3a). The arrival times are generated as IASP91 (Kennett & Engdahl 1991) traveltimes predictions, at local (P_g), regional (P_n) and teleseismic (P) distances. We added distance- and azimuth-dependent delays to the arrival times to simulate separate local, regional and teleseismic source–receiver path anomalies. We then relocated the events with HDC, keeping the depths fixed and using P_n and P phases only, which resulted in a 12 km location bias. This constituted the synthetic input cluster for the RCA Monte Carlo experiment, where we used P_g arrivals only. The P_g travel-time bias is about ± 1 s over 150 km, which is roughly equivalent to a ± 5 per cent velocity perturbation with respect to IASP91. The travel-time biases are such that even a perfectly balanced network would produce a biased location, with stations to the south being increasingly slow and those to the north being increasingly fast (Fig. 3b).

For the Monte Carlo experiment, we selected 5000 realizations of random, biconnected clusters (subsets of events and stations) from the synthetic cluster. For each realization, we performed RCA and generated various metrics to assess the quality of the RCA results.

We found that the combined secondary azimuthal gap, defined as the largest secondary azimuthal gap when considering the azimuths of all event–station pairs, provides a robust metric to predict the location accuracy of the event cluster centroid.

Fig. 4(a) shows the cumulative distributions of the event hypocentroid horizontal mislocation for combined secondary gaps less than a specific threshold. RCA is not very sensitive to the combined secondary azimuth below a confidence level of about 95 per cent, but the figure suggests that we are able to recover the true cluster centroid within 5 km at the 90 per cent confidence level, when the combined secondary gap is less than 140° . However, as Fig. 4(b) shows, the centroid depth can only be recovered within 5 km at about 83 per cent confidence, with the same combined secondary azimuth gap (or within about 7 km at 90 per cent confidence).

The results shown in Figs 4(a) and (b) are based on subsets of stations and events of variable size; some are quite small. It is easy to show that the minimal (given our biconnectivity constraint) RCA graph is defined by three events and two stations, where both stations record all events, thus defining six event–station pairs. In a perfectly symmetric arrangement, this geometry would provide a 120°

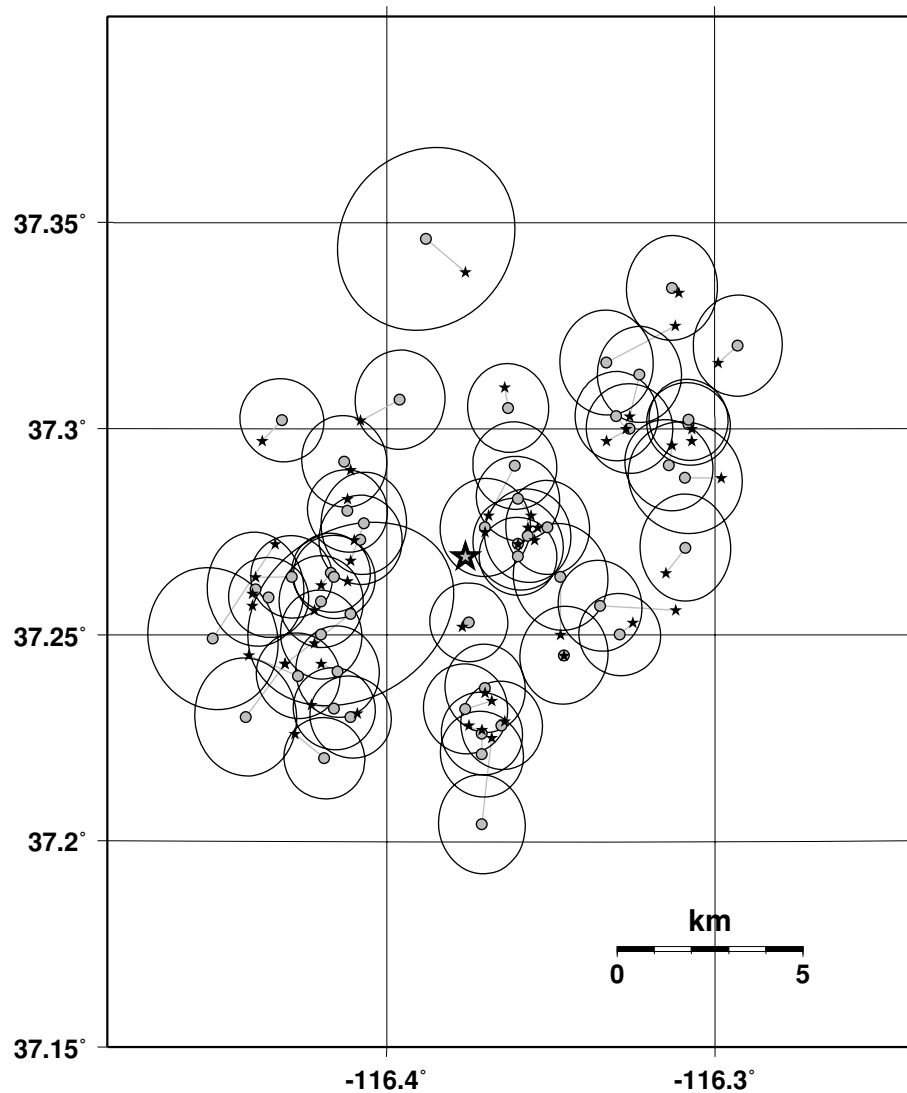


Figure 6. Pahute Mesa, NTS GT0 cluster. When the HDC location bias is removed, the HDC (circles) relative error ellipses (calculated at 90 per cent confidence) cover 83 per cent of the true (stars) locations.

combined secondary azimuthal gap. However, this would hardly constitute an overdetermined inversion problem. Fig. 4(c) is equivalent to Fig. 4(a), with the requirement that there be at least 25 event–station pairs. In this case, a combined secondary gap of less than 140° recovers the event centroid within slightly more than 5 km at the 95 per cent confidence level, and the result is largely insensitive to combined secondary azimuth gap. Although it appears that we could relax the constraint on the combined secondary gap, perhaps to as much as 180° , we feel that the 140° threshold on the combined secondary gap is a reasonable conservative choice.

Even with 25 or more event–station pairs, the accuracy with which we can determine the event hypocentroid depth is less than that for the epicentre (Fig. 4d). Our tests indicate, as expected, that resolution of the centroid depth is improved if there is a station close to the centroid. This is illustrated in Figs 4(e) and (f), where we have added the requirement that there must be at least one station within 30 km of the centroid. This makes almost no difference on our ability to recover the epicentre (compare Figs 4c and e), but it improves the ability to recover the centroid depth by several kilometres at a high confidence level (compare Figs 4d and f).

We summarize these results below to provide GT5₉₅ (i.e. at 95 per cent confidence) applicability criteria for the event cluster epicentroid, analogous to the GT5 selection criteria of Bondár *et al.* (2004a) for single event locations.

- (1) The combined secondary azimuthal gap is less than 140° .
- (2) There are at least 25 event–station pairs.

The centroid depth can only be resolved at about 6–8 km at the same confidence level, with better resolution, when there is a station at a distance less than about 30 km. To call these conditions ‘necessary’ would be incorrect in a strict mathematical sense. It is not impossible that a cluster failing to meet these criteria could generate GT5₉₅ locations using the HDC-RCA method, but we consider it unlikely. We can say, with greater certainty, that these conditions are not sufficient to generate GT5₉₅ locations. We have encountered more than a few clusters, which meet the above criteria yet fail to yield good results due to various problems with the data sets and network geometry. Hence, we consider the above conditions as rejection criteria, rather GT5 classification metrics. If a cluster fails these criteria, we reject all events in the cluster as non-GT5; if a cluster meets the criteria, we inspect every event before promoting it to GT5. We prefer a conservative approach, as we would rather lose some GT5 events than promote questionable events to GT5.

A linear discriminant analysis using the combined secondary azimuthal gap, the distance of the closest station and the number of event–station pairs indicates that the above criteria are on the conservative side, by minimizing type II errors (false GT) at the expense of type I errors (missed GT).

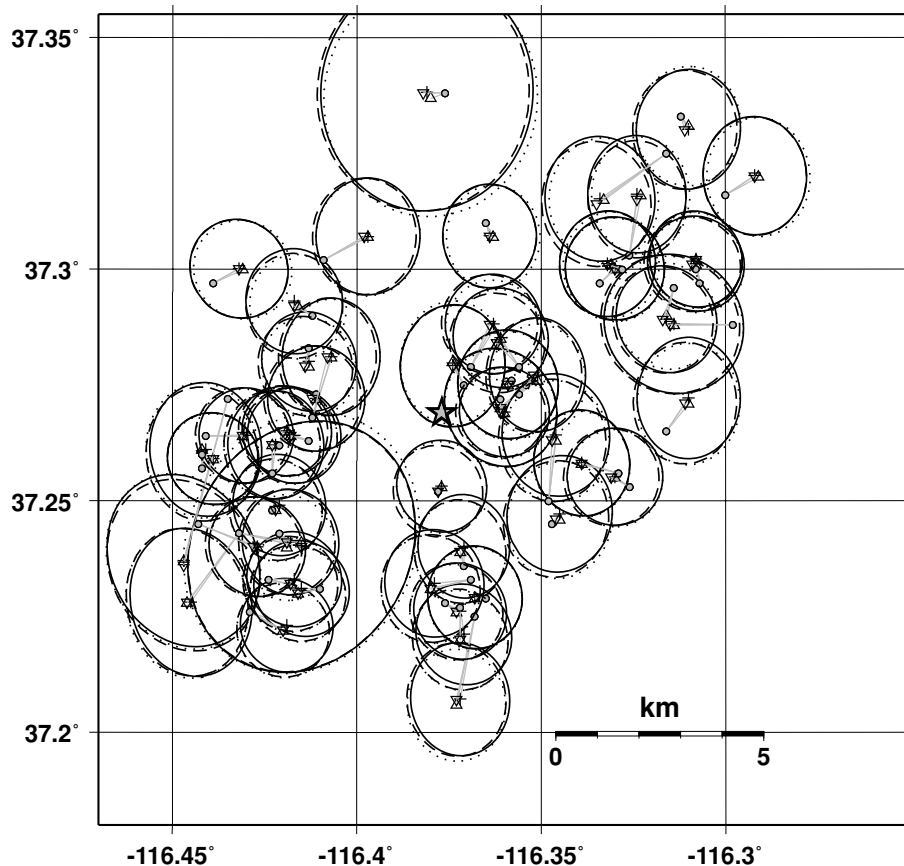


Figure 7. The cluster vectors for the Pahute Mesa cluster from three separate experiments are overlaid, that is, differences in average location have been removed. The known locations of the events are indicated by filled circles. The three experiments use different data sets to estimate the hypocentroid: teleseismic P (30° – 90°), shown by crosses and thin line ellipses; regional P (3° – 20°), shown by triangles and dotted lines and all first-arriving P phases in the distance range 3° – 180° , shown by inverted triangles and dashed lines. There is no significant difference in the estimated cluster vectors (relative locations) for the three cases. The GT locations are connected to the HDC location by thin grey lines; stars represent the true (grey) and the shifted (hollow) centroids (they actually overlap).

VALIDATION TESTS

In this section, we present validation tests using nuclear tests from Pahute Mesa at the Nevada Test Site (Springer *et al.* 2002), which are well recorded by a dense network of 38 local stations (Fig. 5), to show that the underlying assumptions in the RCA algorithm are justified. The Pahute Mesa cluster includes 52 events for which the location, depth and origin time are known with exceptional accuracy (i.e. they can be considered as GT0 events). We also present clusters with previously determined GT information to demonstrate that RCA locations are consistent with prior GT5 locations.

Validation using GT0 events

Recovery of the correct pattern of relative locations

One of the fundamental assumptions in the RCA algorithm is that HDC accurately recovers the event pattern. The HDC analysis of the Pahute Mesa cluster, using regional and teleseismic phases with Ak135 (Kennett *et al.* 1995) traveltimes tables, mislocates the hypocentroid by 12 km. Fig. 6 illustrates that when the HDC bias is removed, using the known event locations, the HDC event pattern (dots) matches the true event pattern (stars) quite well and the 90 per cent confidence ellipses (for relative location) provide 83 per cent coverage. We expect the formal confidence ellipses for relative location to be somewhat underestimated because we have not included error terms to account for several factors:

- (1) Real seismic data are not drawn from a Gaussian distribution.
- (2) We do not account for velocity variations across the cluster.

Also, in making the comparison of coverage shown in Fig. 6, it is necessary to use an estimation process to remove the location bias of the HDC hypocentroid, and this estimation process has its own uncertainty that should be added. It is possible to add an *ad hoc* error term to the HDC algorithm that expands the confidence ellipses for the cluster vectors to better match the GT0 data, but we prefer to avoid such steps in most cases because there is too little data available on the average value and variability of such a correction, especially for earthquake data sets. For the Pahute Mesa cluster, such a term appears to be about 0.5 km of additional uncertainty. When the variance matrix form of this term is added to the formal covariance matrix, the change in dimension of the combined ellipse is, of course, less than the nominal extra uncertainty. That is, an assumed 0.5 km of extra uncertainty causes the combined confidence ellipses to expand by less than 0.5 km.

Dependence of relative event pattern on hypocentroid

For a given cluster, the relative locations (cluster vectors) determined in the HDC analysis should, within reason, be independent of the hypocentroid, which can vary, depending on the specific choice of data set and procedures. We tested this assumption by locating the Pahute Mesa cluster with three different data sets for the hypocentroid:

- (1) first-arriving teleseismic *P* between 30°–90° (our standard approach);
- (2) first-arriving *P* phases at regional distance, 3°–20° and
- (3) all first-arriving *P* phases between 3°–180°.

The hypocentroids determined in these three cases vary by almost 20 km, but the cluster vectors determined in the three cases are essentially identical (Fig. 7). Furthermore, when using these different

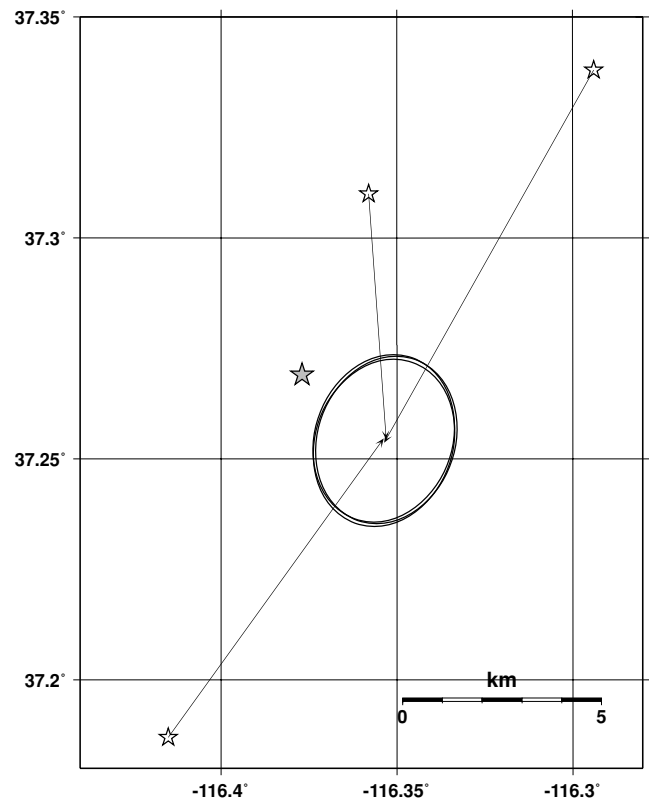


Figure 8. RCA analysis for the Pahute Mesa cluster when three different assumptions are made for the data set used to estimate the hypocentroid in the HDC analysis (Fig. 7). The open stars show the three different starting locations, vectors show the calibration shift from the RCA analysis with 95 per cent confidence ellipses for each case. The true location of the event centroid is shown by the filled star.

solutions as starting points, RCA analysis recovers the same calibration shift in each case (Fig. 8). Unfortunately, Fig. 8 also reveals that the true location of the Pahute Mesa cluster's station centroid lies outside the 95 per cent confidence ellipse of the RCA location, regardless of starting location. This is investigated further in the next section.

Bias in RCA due to lateral heterogeneity

As we noted earlier, RCA eliminates the regional/teleseismic bias in the HDC hypocentroid. However, RCA is still susceptible to bias due to unmodelled local velocity structure. We account for traveltimes prediction errors, inherent in a local velocity model, by introducing generic, distance-dependent model errors (Fig. 2).

We again used the Pahute Mesa cluster to validate that (1) the true event centroid is located within 5 km of the event centroid recovered by RCA and (2) the RCA error ellipse covers the true centroid 95 per cent of the time. Since for this experiment we are only interested in RCA performance and the effect of the local velocity model, we start from the true event pattern with no relative errors (i.e. we used the GT0 locations as the initial locations for RCA). We used the Western US velocity model by Ritsema & Lay (1995) and performed a bootstrapping RCA experiment by selecting all unique subnetworks, ranging from 3 to 30 stations with a secondary azimuthal gap less than 280° from the local network shown in Fig. 5.

The results are summarized in Figs 9–12, considering both the coverage statistic (<1 if the confidence ellipse covers the true

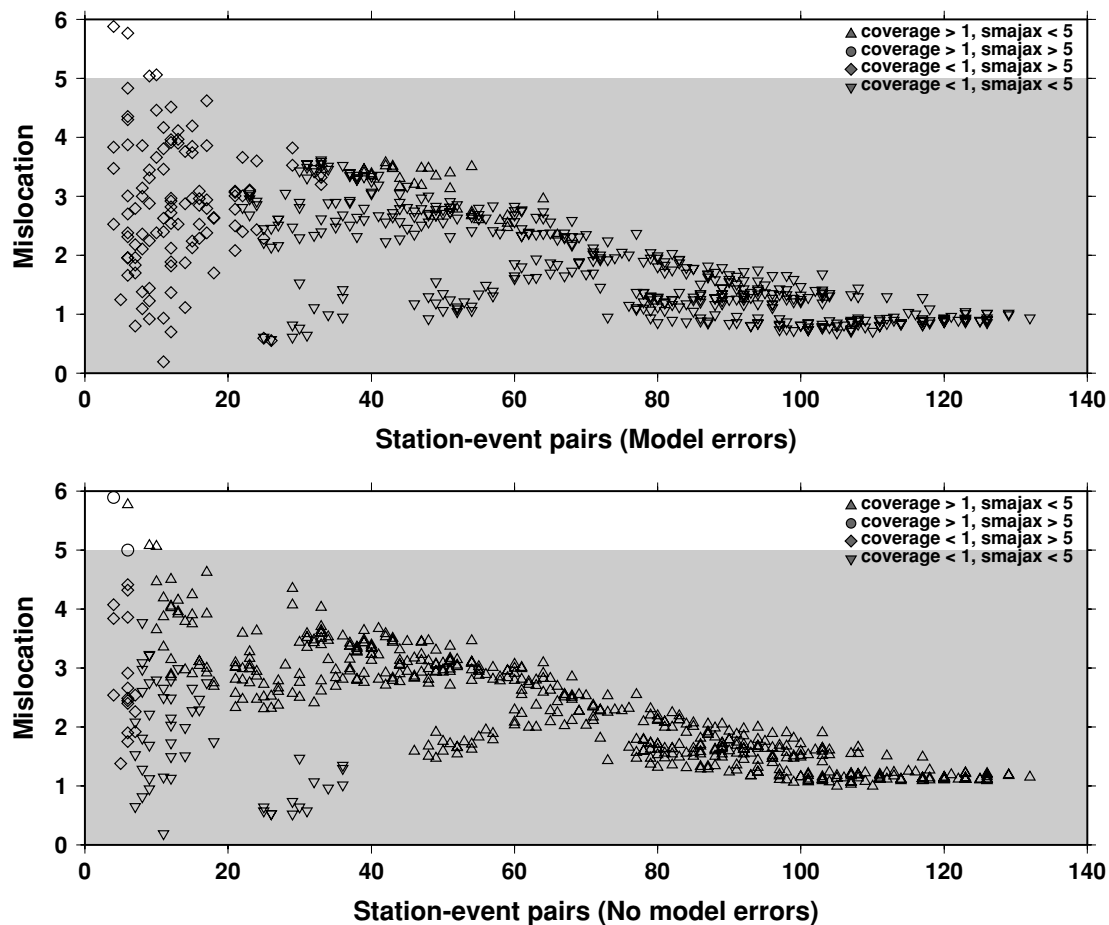


Figure 9. Results of a Monte Carlo test using different numbers of station–event pairs in the RCA analysis of the Pahute Mesa cluster of nuclear tests to investigate the ability of RCA to recover the true event centroid. Cluster event starting locations were the known (GT0) locations. Two cases are shown—with and without distance-dependent traveltime (model) errors (Fig. 2). In each realization, the centroid and 95 per cent confidence ellipse estimated by RCA are compared with the known centroid to calculate the coverage statistic (< 1 if the 95 per cent confidence ellipse covers the true centroid) and mislocation, in km. Symbols are the same in both plots and show four combinations of coverage statistic and length of the semi-major axis of the 95 per cent confidence ellipse (smajax). The use of model errors has almost no effect on mislocation, but most realizations fail in coverage (triangles). With model errors, most realizations satisfy the coverage criterion (inverted triangles). Even with large numbers of data, RCA analysis mislocates the event centroid by about 1 km, presumably due to unmodelled heterogeneity in the source region.

centroid) and mislocation (km from the true centroid location) and with and without the distance-dependent model error (Fig. 2).

Fig. 9 shows the results for mislocation as a function of the number of station–event pairs, with and without model errors. The introduction of model errors has a very small effect on the mislocations and does not change the fact that nearly all realizations are mislocated by less than 5 km; the exceptions are only for cases with a small number of station–event paths. We also note that the mislocation stabilizes at about 1 km for large data sets. This represents the location bias for unmodelled lateral heterogeneity in the Pahute Mesa source region.

In Fig. 10, the grey region indicates the cases in which coverage is satisfied (coverage < 1.0), with and without model errors. Without model errors, most realizations of the Monte Carlo test fail in coverage. The exception is for small numbers (less than about 40) of station–event pairs. With smaller number of data, the confidence ellipses are larger, which improves coverage even in the presence of systematic location bias. Fig. 10 shows that, with our simple distance-dependent form of model errors, confidence ellipses are inflated enough to provide good coverage statistics, regardless of the number of station–event pairs.

A more quantitative representation of the Monte Carlo tests is shown in Fig. 11, in which histograms and cumulative percentages are shown for coverage statistics and mislocation for the case of no model errors. All mislocations are less than 5 km, but in only about 10 per cent of the cases does the 95 per cent confidence ellipse cover the true centroid. The average mislocation is about 2.5 km.

Fig. 12 shows the same representation as Fig. 11 for the case where we have included model error. Coverage is now greatly improved: the histogram shows that most realizations satisfy the coverage criterion and the cumulative percentage shows that the coverage is about 95 per cent, in agreement with our statistical model (i.e. our confidence ellipses are scaled to 95 per cent confidence). As expected, the histogram of mislocations is virtually unchanged by the introduction of model errors. In this paper, we use the generic distance-dependent model error function to account for uncertainties in traveltime predictions.

Correlated error structure

Another assumption upon which the RCA algorithm (and most other location algorithms) depends is that the observations are

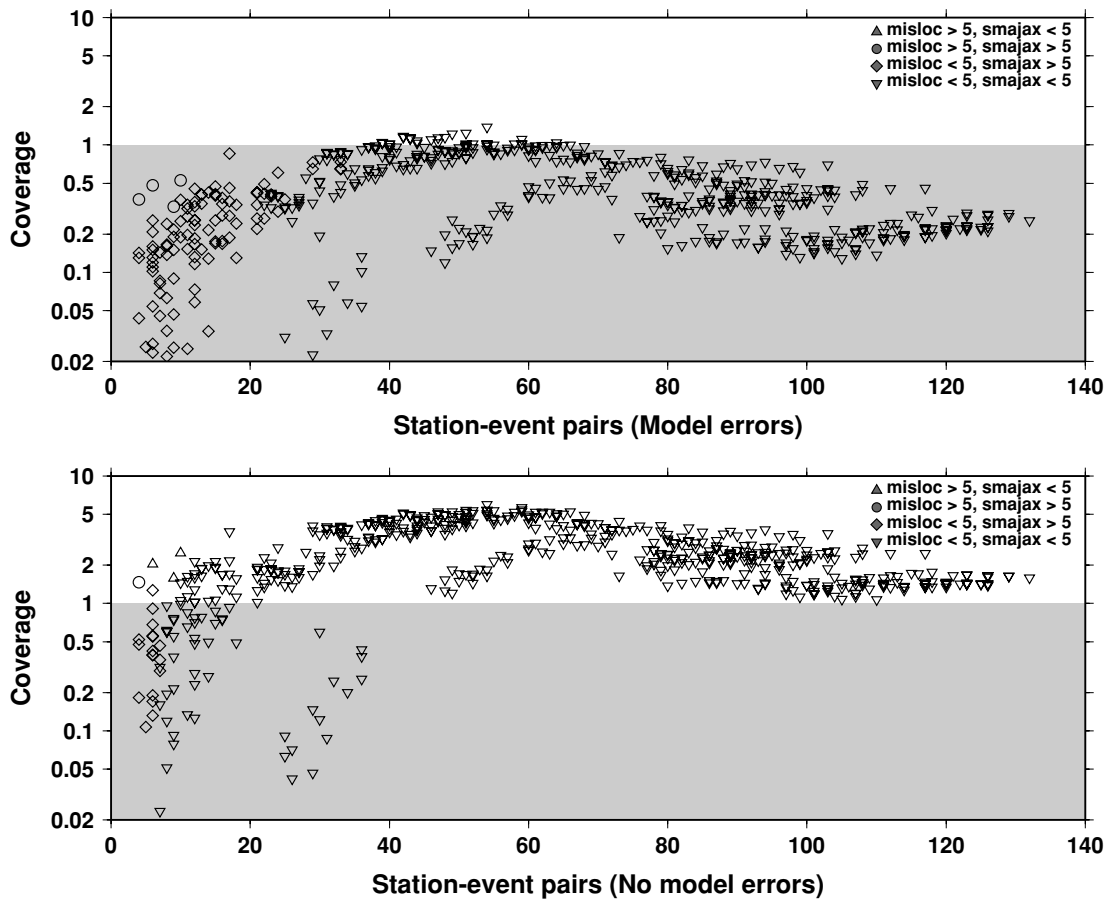


Figure 10. Same data set as in Fig. 9, except the coverage statistic is plotted instead of mislocation. Without model error, most realizations have small mislocation and confidence ellipses, but fail to cover the true centroid. Model error inflates the confidence ellipses enough to provide coverage in most cases.

independent. With dense local networks, such as the one in Fig. 5, there is an increased chance that multiple observations will be from similar ray paths that produce similar traveltime prediction errors due to unmodelled local heterogeneities. By ignoring the correlated error structure, the number of independent observations may be overestimated and therefore the location uncertainties would be underestimated. The Monte Carlo experiment described above shows that the generic model error curve is fairly robust against a reasonably wide range of network geometries. Nevertheless, when we use the entire local station network of 38 stations in the Pahute Mesa cluster, the stations south of the cluster conspire to pull the RCA centroid slightly to the southeast, and the 95 per cent absolute error ellipses (combined HDC and RCA location uncertainties) do not cover 95 per cent of the true locations (Fig. 13a). However, when we use subnetworks with acceptable combined secondary azimuthal gaps, the 95 per cent error ellipses do cover GT0 locations (Fig. 13b) because the station networks now better satisfy the assumption of independent errors and the confidence ellipses, albeit larger, are more accurate.

Fig. 13 illustrates another important point: since not every station recorded every event, had we only used the sparse seven-station subnetwork, we could have only located four events with any conventional single event location algorithm. RCA, on the other hand, used 7 stations and 13 events to determine the HDC location bias, which located all events within 5 km of GT0 and identified 50 (out of 52) locations as GT5. Thus, in favourable conditions, RCA may produce GT5 locations with sparse local networks.

Cross validation with GT based on single event location

In this section, we carry out HDC-RCA analysis on the 1980 Campania-Lucania, Italy earthquake sequence for which there are enough local stations to determine GT5 locations in a single event location study, using the criteria developed by Bondár *et al.* (2004a). Owing to the dense regional network in Europe, the single event bulletin locations (and EHB locations) are quite good. HDC did not move the EHB locations very much, but it was still able to tighten the cluster (Fig. 14a), which consisted of 60 events recorded by 633, mostly regional stations. A total of 59 events, together with 34 local stations, passed the RCA connectivity criteria, forming a cluster with a combined secondary azimuthal gap of 23° (Fig. 14b). In the HDC processing, the depths of the events were fixed at 5 km, based on an analysis of reported depth phases. We used a local velocity model published by Amoroso *et al.* (2005) to obtain traveltime predictions for local phases. RCA shifted the cluster by 7 km to the SE (Fig. 14c) and moved the centroid depth up to 2.5 km. We identified 55 events as GT5. Note that 13 of these events also pass the Bondár *et al.* (2004a) GT5 selection criteria. Comparison of the bulletin-based GT5 locations (diamonds) and the HDC-RCA GT5 locations (circles) shows a good agreement, as the 5 km uncertainty radii around the single-event GT5 events overlap or fully contain the absolute error ellipses obtained from the HDC-RCA method (Fig. 14d).

Events for which GT5-level location accuracy is possible in a single event location analysis, are typically recorded by dense

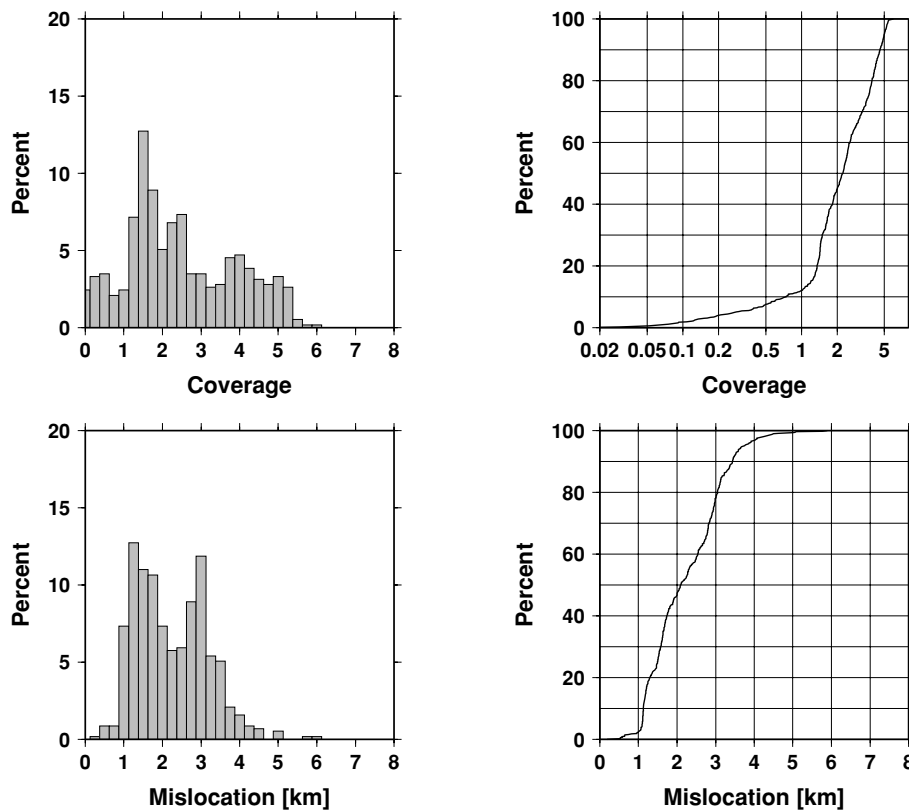


Figure 11. Ability of RCA to recover the true event centroid, based on the Monte Carlo experiment displayed in Figs 9 and 10. Histograms of coverage and mislocation, with corresponding cumulative percentages, for the case in which no model (traveltime) error is used. Almost all realizations have mislocation less than 5 km, but only about 12 per cent of them satisfy the coverage criteria.

local networks. However, only a subset of events in such regions is recorded by a sufficient number of local stations with favourable azimuthal coverage, which limits the number of events that can be promoted to GT5 status. HDC-RCA, being a multiple event location technique, is less constrained by these limitations, provided that there is strong connectivity between stations and events. Therefore, the HDC-RCA methodology is usually able to produce more GT events from a cluster than the bulletin-based single event GT selection criteria developed by Bondár *et al.* (2004a).

EXAMPLES OF HDC-RCA CLUSTERS

In this section, we give an example of an earthquake cluster for which HDC-RCA is able to produce GT5 locations, and where no other current method would be successful. Finally, we give an example in which HDC-RCA method fails to produce GT5 events at a high confidence level.

Rogun cluster, Tajikistan

The Rogun, Tajikistan cluster is located between the South Tien Shan and the Northern Pamir mountains. This region is characterized by shallow seismicity along the Vaksh and Darvaz faults (Pegler & Das 1998). The HDC cluster (Fig. 15a) consisted of 24 events with 376 regional/teleseismic stations. Altogether, 13 events recorded by only three stations passed the RCA connectivity tests (Fig. 15b). Even though we have only three stations, the RCA geometry is so favourable that the combined secondary azimuthal gap is only 110° . We used a local velocity model by Hamburger *et al.*

(1993) to predict traveltimes for the 44 *Pg* and *Sg* readings. Again, we had no close-in stations; so, we ran RCA with depth fixed at 12 km. The RCA analysis shifted the entire cluster to the Vaksh river valley (shown as grey line in Fig. 15c), which is the surface expression of the Vaksh fault. The HDC-RCA analysis produced 17 GT5 events.

The Rogun cluster illustrates a strength of the HDC-RCA methodology that is manifested when only a few local stations are available. Richards *et al.* (2006) point out that to achieve improved locations for more than 25 per cent of the events in a cluster, using the double-difference algorithm with differential times from waveform cross-correlation, a high local station density (one station per 100 km^2 and about 12 km distance between stations) is required. Because of the natural separation of tasks in the HDC-RCA methodology (HDC resolves the event pattern using regional and teleseismic stations, RCA reduces the bias in the hypocentroid using local data), the applicability of RCA is not restricted by such strong conditions on station density. As long as the RCA geometry is favourable, HDC-RCA is capable of producing GT events even with very sparse local station coverage, in principle, even a single station.

Terceira cluster, Azores Islands

Finally we show an example where HDC-RCA fails to produce GT events at a high confidence level. Fig. 16(a) shows the RCA geometry for a cluster near Terceira Island in the Azores. We used the Ak135 oceanic velocity model (Kennett *et al.* 1995) to predict traveltimes to the local stations, used in the RCA. Note that the combined secondary azimuthal gap is 148° and the closest station

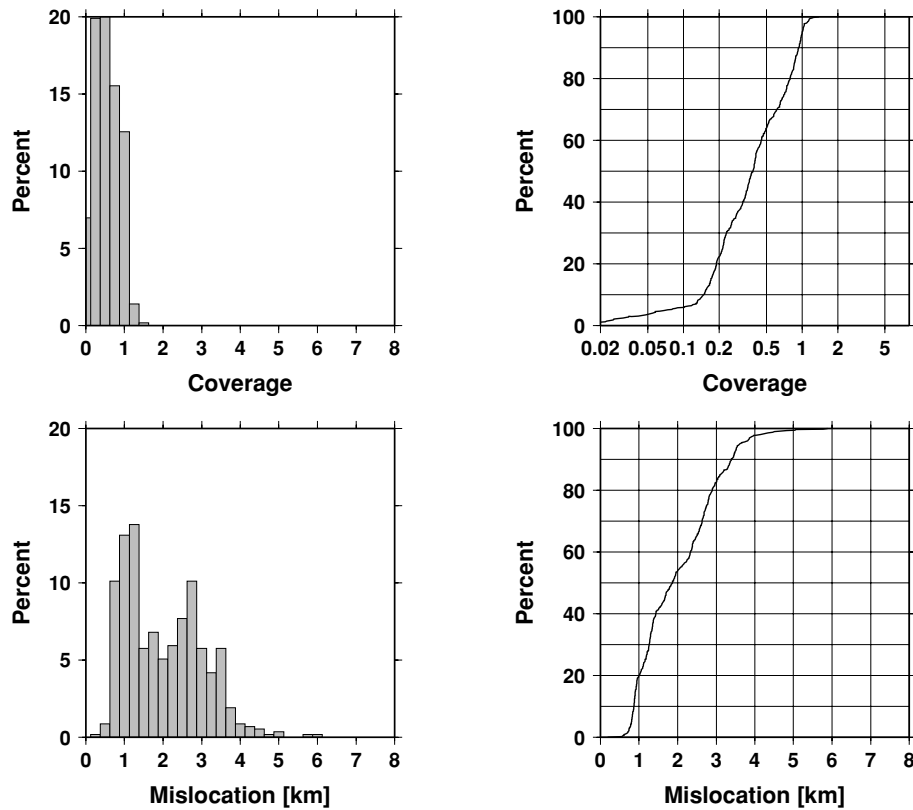


Figure 12. Ability of RCA to recover the true event centroid, based on the Monte Carlo experiment displayed in Figs 9 and 10. Histograms of coverage and mislocation, with corresponding cumulative percentages, for the case in which the model (traveltime) error in Fig. 2 is used. See Fig. 11 for comparison. The distribution of mislocations is unchanged, but approximately 95 per cent of the realizations now satisfy the coverage criteria, which is based on a 95 per cent confidence ellipse.

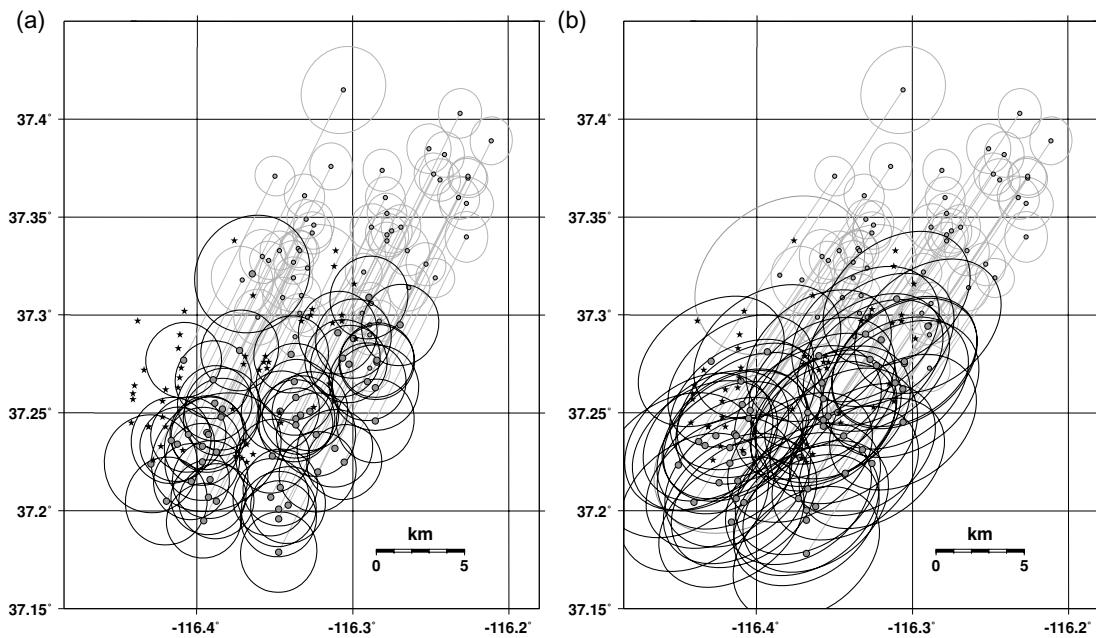


Figure 13. (a) RCA result when using all stations in the local network shown in Fig. 5. Small grey circles, HDC; large circles, RCA and stars, GT0 locations. While the RCA event mislocations average about 3 km, the 95 per cent error ellipses do not cover the GT0 locations. (b) RCA results using a 7-station subnetwork. The error ellipses now cover the true locations, and 50 out of 52 events are identified as GT5.

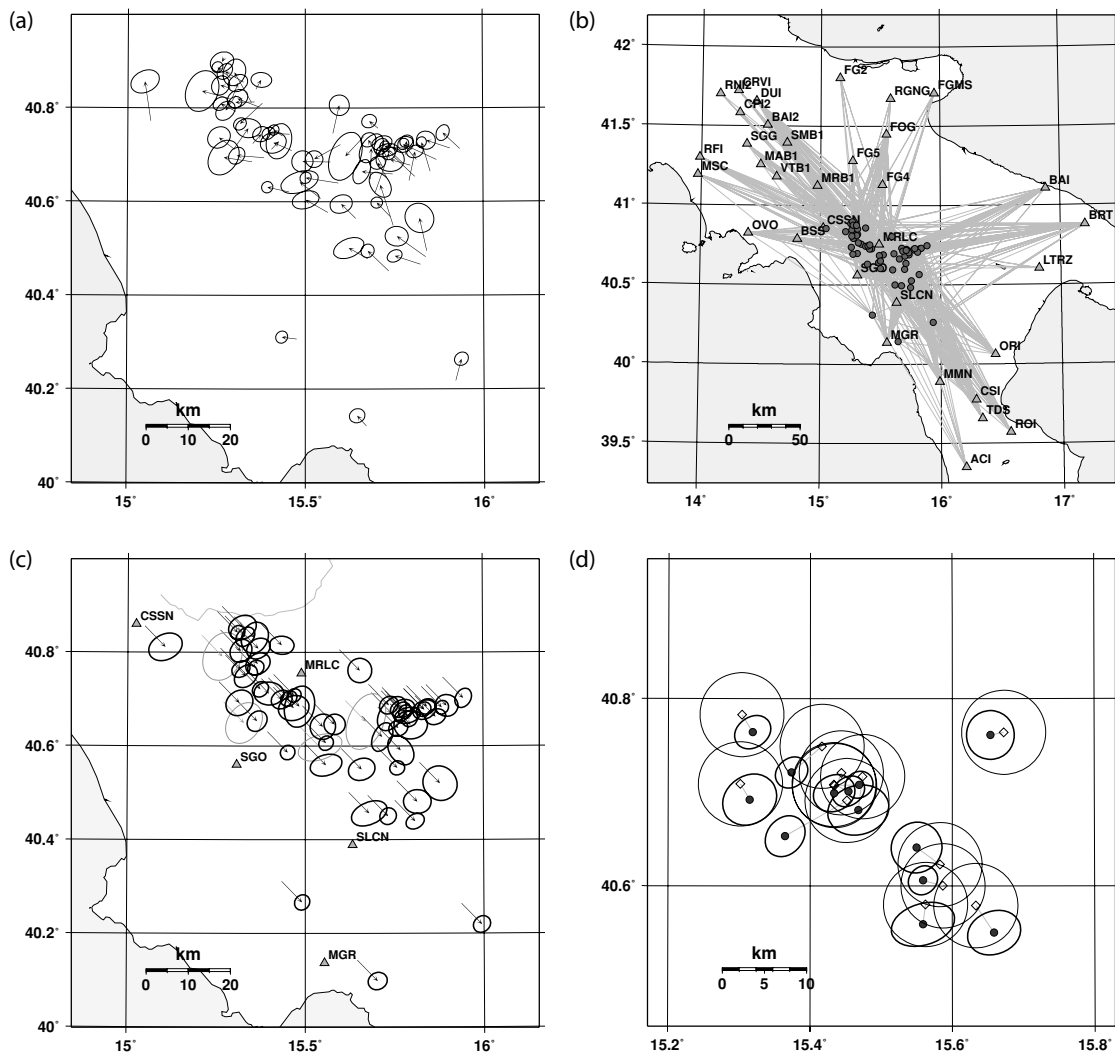


Figure 14. Campania-Lucania cluster, Southern Italy. (a) HDC relocations of 60 EHB events. Vectors point from the EHB locations to the HDC locations shown with their relative error ellipses. (b) RCA network geometry. (c) RCA shifts the entire HDC cluster by 7 km to the SE. The absolute 95 per cent error ellipses of the events identified as GT5 are plotted with thick black lines. (d) Comparison between the bulletin-based GT5 selection (open diamonds) and the HDC-RCA GT5 events (filled circles).

is at a distance of about 65 km, which prevents us from identifying any GT5 events at a high confidence level because the combined HDC and RCA uncertainties are too large. Nevertheless, as Fig. 16(b) illustrates, RCA moves the cluster toward the plate boundary between Africa and Europe. Thus, even if we could not identify GT events at a high confidence level, due to the improved relative locations and reduced bias, HDC-RCA locations are substantially improved over standard single event locations.

DISCUSSION

Depth resolution

Since regional and teleseismic first-arriving P phases lack the resolution to resolve the full depth pattern in a cluster, event depths are typically fixed in the HDC analysis to educated guesses, based on a preliminary analysis of individual events with reported depth phases and local P and S phases, as well as waveform analyses. RCA treats the HDC event cluster as if it were a rigid pattern and shifts the

cluster without changing the event pattern. Therefore, RCA may improve the hypocentroid (average) depth, but not the depth pattern. However, the local data may have enough resolution to resolve the depth, at least for a subset of events. In general, local data can resolve the depth if there are stations covering the distance range where the vertical partial derivative of the traveltime of the first arriving phase changes sign, or, in other words, if we have a mixture of downgoing and upgoing waves. This is expressed in the HDC-RCA applicability criteria by the somewhat rudimentary criterion that requires at least one close-in station to be able to resolve the hypocentroid depth at a high confidence level. One possible way to improve the depth pattern in the cluster is to identify those events for which the local stations provide depth resolution and for these events, perform a grid search in depth and origin time, by keeping the epicentre fixed. The depth of the rest of the events would be adjusted so that the hypocentroid depth, resolved by the HDC-RCA, remains the same. Clearly, focal depth (and origin time, which is closely coupled to depth) remains the most difficult of hypocentral parameter to resolve.

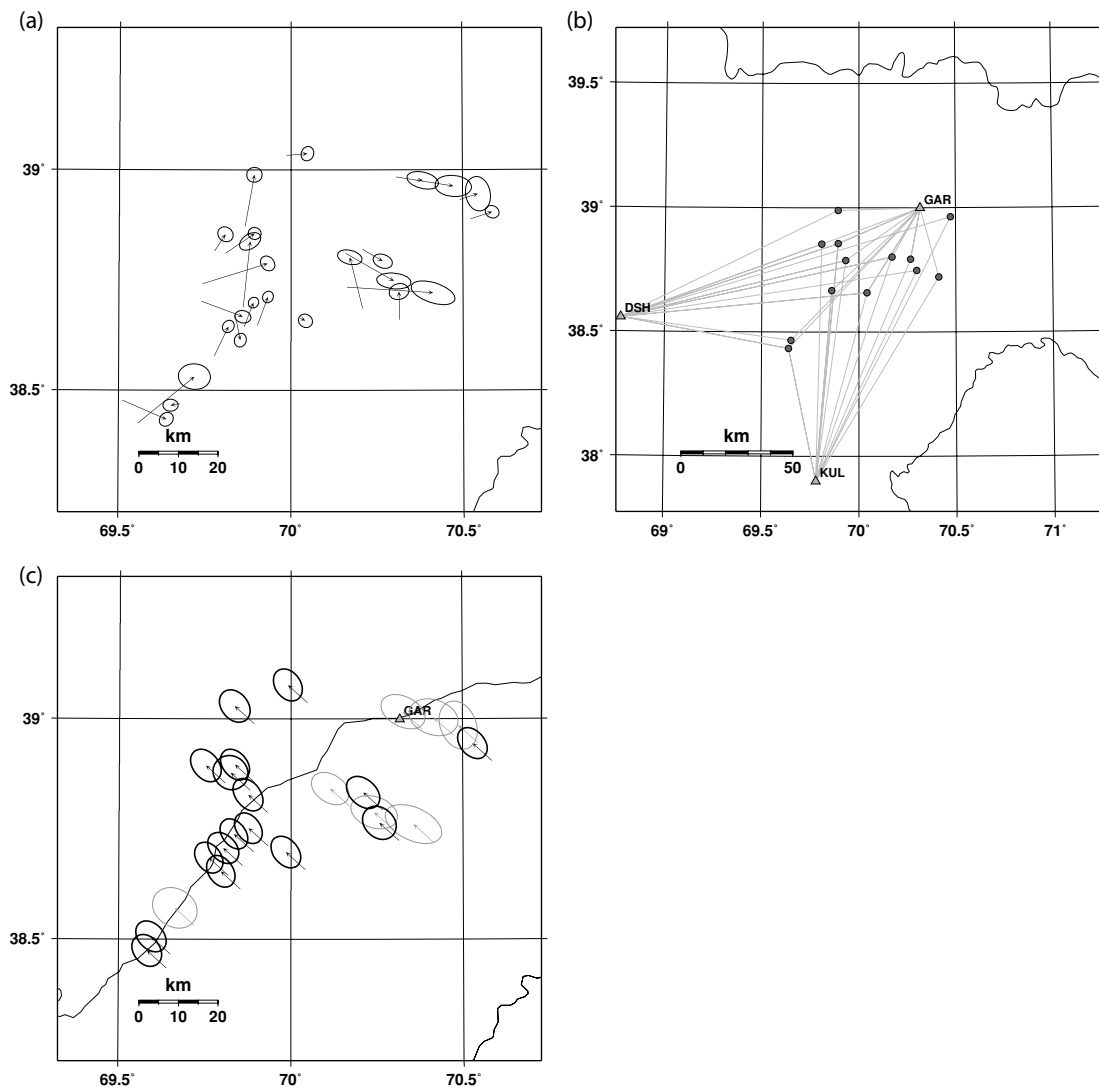


Figure 15. Rogun, Tajikistan cluster. a) HDC relocations of 24 EHB events. Vectors point from the EHB locations to the HDC locations shown with their relative error ellipses. (b) The RCA network geometry (3 stations, 13 events) defines a combined secondary azimuthal gap of 110° . (c) RCA shifts the cluster to the Vaksh river valley. The river (grey line) indicates the surface trace of the Darvaz-Vaksh fault. The absolute 95 per cent error ellipses of the events identified as GT5 are plotted with thick black lines.

Unmodelled heterogeneity in the source region

RCA corrects the regional/teleseismic bias in the HDC hypocentroid location, caused by unmodelled velocity heterogeneity outside the source region. However, the method is still susceptible to heterogeneity that is not accounted for by the 1-D local velocity models used in RCA. Our model error function (Fig. 2) is based on one particular cluster, Pahute Mesa, for which we have an unusually dense local seismograph network and also a full set of GT0 locations. For the other clusters on which we have carried out the HDC-RCA analysis, we have made the assumption that the location bias due to unmodelled heterogeneity and bulk velocity variations is comparable to that of Pahute Mesa. It should be remembered, however, that failure of this assumption does not change the actual mislocation of the RCA analysis. It only effects the scaling of the confidence ellipses, such that the coverage (if it could be known) may be less than or greater than the expected 95 per cent, depending on whether the unmodelled heterogeneity is greater than or less than that of Pahute Mesa, respectively.

The effects of local heterogeneities are especially troublesome if the RCA ray paths are not evenly distributed in azimuth. These may introduce correlated traveltime predictions errors due to 3-D velocity structure, unmodelled by the 1-D local velocity model. Correlated errors lead to underestimation of the uncertainty (size of the confidence ellipse) of the calibration shift estimated in RCA analysis, which, as most location algorithms do, assumes that the errors are independent. This in turn can lead to failure of the coverage statistics, even if the actual level of mislocation is not large.

The Pahute Mesa cluster is exceptional in having a very dense network of local stations, plus a large set of GT0 sources with which we can actually observe the effects of correlated errors on the location statistics. We tuned the model error function in such a way that it is fairly robust against conspiring stations in a dense local network (i.e. correlated traveltime prediction errors). In other words, the model errors are inflated so that they compensate for the underestimated uncertainty estimates due to the correlated error structure, for a wide range of network geometries. Hence, the generic model errors provide conservative estimates for a typical HDC-RCA

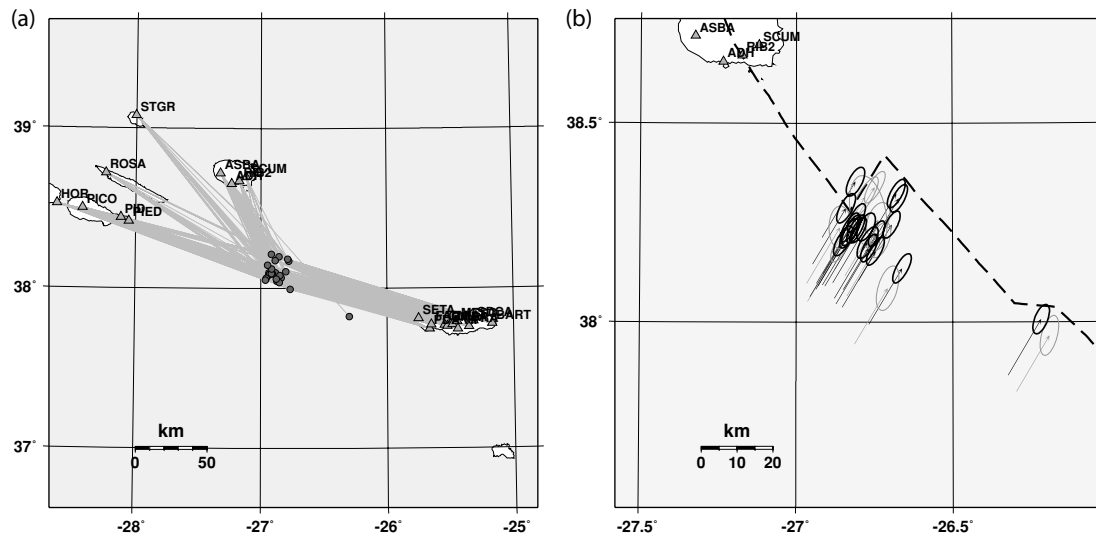


Figure 16. Terceira, Azores Islands. (a) The RCA network geometry defines a combined secondary gap of 148° ; therefore, no GT5 events can be identified at a high confidence level. (b) Nevertheless, RCA shifts the entire cluster toward the plate boundary (dashed line) between the Eurasian and African plates.

analysis, where there are fewer local stations so that the assumption of independence is not violated substantially. Recent studies by Rodi & Myers (2007) and Bondár *et al.* (2007) described methodologies to provide formal uncertainty estimates that account for correlated error structure in the data. Incorporating their results in the RCA uncertainty estimation procedures will lead to further improvements.

CONCLUSIONS

A hybrid seismic location method, HDC-RCA has been developed, which is able to produce minimally biased (relative to standard location procedures) locations for clusters of events. The method depends on the presence of small numbers (in the limit, one) of local seismic stations, and certain geometrical constraints must be satisfied. Seismic events must be large enough to be recorded at regional and teleseismic distances. The method is capable of producing large numbers of events with location errors of less than 5 km (GT5) in many regions of the world. Because it does not rely upon the existence of dense local networks, HDC-RCA can determine minimally biased locations where other methods would not. Even when a dense local seismograph network is present, HDC-RCA is usually capable of producing more high-quality locations than single event methods.

ACKNOWLEDGMENTS

This work was supported by the US Air Force Research Laboratory under contract FA8718-04-C-0020. We thank the International Seismological Centre (ISC) and the U.S. Geological Survey's National Earthquake Information Center (NEIC) for maintaining the catalogues of phase arrival times, used in this study. We are grateful to Jim Dewey and an anonymous reviewer for their efforts and commitment to help us improve the paper.

REFERENCES

- Amoruso, A., Crescentini, L. & Scarpa, R., 2005. Faulting geometry for the complex 1980 Campania-Lucania earthquake from leveling data, *Geophys. J. Int.*, **162**, 156–158.
- Biggs, J., Bergman, E., Emmerson, B., Funning, G.J., Jackson, J., Parsons, B. & Wright, T.J., 2006. Fault identification for buried strike-slip earthquakes using InSAR: the 1994 and 2004 Al Hoceima, Morocco earthquakes, *Geophys. J. Int.*, **166**, 1347–1362, doi:10.1111/j.1365-246X.2006.03071.x.
- Bondár, I., Myers, S.C., Engdahl, E.R. & Bergman, E.A., 2004a. Epicenter accuracy based on seismic network criteria, *Geophys. J. Int.*, **156**, 483–496, doi:10.1111/j.1365-246X.2004.02070.x.
- Bondár, I. *et al.*, 2004b. Collection of a reference event set for regional and teleseismic location calibration, *Bull. seism. Soc. Am.*, **94**, 1528–1545.
- Bondár, I., McLaughlin, K. & Israelsson, H., 2007. Improved event location uncertainties, in *Proceedings of the 29th Monitoring Research Review: Ground-Based Nuclear Explosion Monitoring Technologies*, Denver, Colorado, September 25–27.
- Dewey, J.W., 1972. Seismicity & tectonics of Western Venezuela, *Bull. seism. Soc. Am.*, **62**, 1711–1751.
- Douglas, A., 1967. Joint epicentre determination, *Nature*, **215**, 47–48.
- Engdahl, E.R., van der Hilst, R.D. & Buland, R.P., 1998. Global teleseismic earthquake relocation with improved travel times and procedures for depth determination, *Bull. seism. Soc. Am.*, **88**, 722–743.
- Fisk, M., 2002. Accurate locations of nuclear explosions at the Lop Nor test site using alignment of seismograms and IKONOS satellite imagery, *Bull. seism. Soc. Am.*, **92**, 2911–2925.
- Got, J.-L., Fréchet, J. & Klein, F.W., 1994. Deep fault plane geometry inferred from multiple relative relocation beneath the south flank of Kilauea, *J. geophys. Res.*, **99**, 15375–15386.
- Hamburger, M.W., Swanson, W.A. II & Popandopulo, G.A., 1993. Velocity structure and seismicity of the Garm region, Central Asia, *Geophys. J. Int.*, **115**, 497–511.
- IDC Documentation, 1999. IDC processing of seismic, hydroacoustic, and infrasonic data, [IDC5.2.1], Science Applications International Corporation, SAIC-99/3023.
- International Seismological Centre, 2001. On-line Bulletin, Internatl. Seism. Cent., Thatcham, United Kingdom, available at <http://www.isc.ac.uk/Bull>.
- Jordan, T.H. & Sverdrup, K.A., 1981. Teleseismic location techniques and their application to earthquake clusters in the South-central Pacific, *Bull. seism. Soc. Am.*, **71**, 1105–1130.
- Kennett, B. & Engdahl, E.R., 1991. Travel times for global earthquake location and phase identification, *Geophys. J. Int.*, **105**, 429–465.
- Kennett, B.L.N., Engdahl, E.R. & Buland, R.P., 1995. Constraints on seismic velocities in the Earth from travel times, *Geophys. J. Int.*, **122**, 108–124.
- Lin, G. & Shearer, P., 2005. Tests of relative earthquake location techniques using synthetic data, *J. geophys. Res.*, **110**, B04304, doi:10.1029/2004JB003380.

- Myers, S.C., Johannesson, G. & Hanley, W., 2005. Multiple-event location using the Markov-chain Monte Carlo technique, in *Proceedings of the 27th Seismic Research Review: Ground-Based Nuclear Explosion Monitoring Technologies*, Rancho Mirage, California, September 20–22.
- Myers, S.C., Johannesson, G. & Hanley, W., 2007. A Bayesian hierarchical method for multiple-event seismic location, *Geophys. J. Int.*, **171**, 1049–1063, doi:10.1111/j.1365-246X.2007.03555.x.
- Pan, J., Antolik, M. & Dziewonski, A.M., 2002. Locations of mid-oceanic earthquakes constrained by seafloor bathymetry, *J. geophys. Res.* **107**(B11), 2310, EPM8.1–EPM8.13, doi:10.1029/2001JB001588.
- Parsons, B. *et al.*, 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, *Geophys. J. Int.*, **164**, 202–217, doi:10.1111/j.1365-246X.2005.02655.x.
- Pavlis, G.L. & Booker, J.R., 1983. Progressive multiple event location (PMEL), *Bull. Seism. Soc. Am.*, **73**, 1753–1777.
- Pegler, G. & Das, S., 1998. An enhanced image of the Pamir-Hindu Kush seismic zone from relocated earthquake hypocentres, *Geophys. J. Int.*, **134**, 573–595.
- Richards, P.G., Waldhauser, F., Schaff, D. & Kim, W.-Y., 2006. The applicability of modern methods of earthquake location, *Pure appl. Geophys.*, **163**(2–3), 351–372, doi:10.1007/s00024-005-0019-5.
- Ritsema, J. & Lay, T., 1995. Long-period regional wave moment tensor inversion for earthquakes in the western United States, *J. geophys. Res.*, **100**, 9853–9864.
- Ritzwoller, M.H., Shapiro, N.M., Levshin, E.A., Bergman, E.A. & Engdahl, E.R., 2003. Ability of a global three-dimensional model to locate regional events, *J. geophys. Res.*, **108**(B7), 2353, doi:10.1029/2002JB002167.
- Rodi, W.L. & Myers, S.C., 2007. Modeling travel-time correlations based on sensitivity kernels and correlated velocity anomalies, in *Proceedings of the 29th Monitoring Research Review: Ground-Based Nuclear Explosion Monitoring Technologies*, Denver, Colorado, September 25–27.
- Rodi, W.L., Myers, S.C. & Schultz, C.A., 2003. Grid-search location methods for ground truth collection from local and regional seismic networks, in *Proceedings of the 25th Seismic Research Review: Nuclear Explosion Monitoring: Building the Knowledge Base*, Tucson, Arizona, September 23–25.
- Shearer, P.M., 2001. Improving global seismic event locations using source-receiver reciprocity, *Bull. seism. Soc. Am.*, **91**, 594–603.
- Springer, D.L., Pawloski, G.A., Ricca, J.L., Rohrer, R.F. & Smith, D.K., 2002. Seismic source summary for all US below-surface nuclear explosions, *Bull. seism. Soc. Am.*, **92**, 1806–1840.
- Waldhauser, F. & Ellsworth, W.L., 2000. A double-difference earthquake location algorithm: method and application to the Northern Hayward fault, California, *Bull. seism. Soc. Am.*, **90**, 1353–1368.
- Waldhauser, F. & Richards, P.G., 2004. Reference events for regional seismic phases at IMS stations in China, *Bull. seism. Soc. Am.*, **94**, 2265–2279.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. We processed 86 event clusters. The 66 clusters (green dots) yielded GT5 or better event locations. 20 clusters either failed (yellow dots) the RCA applicability criteria or their final 95 per cent error ellipses were too large to promote any events to GT5.

Table S1. List of the total number of events, the number of events promoted to GT5 and the reference to the local velocity model used in the RCA for GT-producing clusters.

Table S2. List of events promoted to GT5. Semi-major axes and strike describe the error ellipse scaled to the 95 per cent confidence level.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by authors. Any queries (other than missing material) should be directed to the corresponding author for the article.