# GLOBAL GROUND TRUTH DATA SET WITH WAVEFORM AND IMPROVED ARRIVAL DATA

István Bondár[1], Ben Kohl[1], Eric Bergman[2], Keith McLaughlin[1], Hans Israelsson[1], Yu-Long Kung[1], Jeff Given[1], and Eric R. Engdahl[3]

Science Applications International Corporation[1], Global Seismological Services[2], and University of Colorado at Boulder[3]

Sponsored by Air Force Research Laboratory

Contract No. FA8718-04-C-0020

## ABSTRACT

The main objective of this three-year research project is to produce a quality-controlled global GT0-5 event set, accompanied with waveform and groomed arrival time data sets. Our efforts are directed toward developing and refining methodologies for generating new ground-truth (GT) events through multiple-event location analysis.

To accomplish this goal, we have developed the hybrid HDC-RCA (hypocentroidal decomposition and reciprocal cluster analysis) methodology. The HDC analysis determines accurate event location patterns relative to a provisional cluster centroid using regional and teleseismic phases. The RCA analysis then determines the accurate location of the cluster centroid using local phases only. RCA accomplishes this by keeping the event and station patterns fixed and relocates the station centroid using the events as fictitious stations. Because both relative station and relative event patterns are fixed and multiple events are typically recorded at each station, solving for the cluster hypocentroid represents an overdetermined inversion problem that is robust with respect to strong local seismic-velocity biases.

We have extended our multiple-event location technique, RCA, to obtain unconstrained depth estimates for the hypocentroid of the event cluster. We have validated the methodology using synthetic and real event clusters. We performed a large-scale Monte Carlo experiment on a synthetic event cluster to develop applicability criteria for the HDC-RCA methodology. We sought local network geometries under which RCA produces GT5 events at a high confidence level. We found that if the combined secondary azimuthal gap (taking into account all event-station pairs entering the RCA inversion) is less than 180°, RCA recovers the true event centroid within 5 km at the 90% confidence level. Furthermore, if there are one or more stations within 30 km of the true cluster centroid, the hypocentroid (epicenter and depth of the cluster centroid) are both recovered within 5 km at the 95% confidence level. The hybrid HDC-RCA method may produce GT5 (epicenter and depth) event locations that are not produced by local or teleseismic network location methods alone.

We report new GT5 event locations (90% and 95% confidence) produced by the HDC-RCA methodology for event clusters from Romania, Honshu, South Africa, Hawaii, Mona Passage, Ethiopia, Southern Italy, Nepal, and Honduras.

## OBJECTIVE

The objective of the research project is to produce new high-confidence GT events of GT5 from an updated EHB (Engdahl et al., 1998) bulletin on a global scale. In order to achieve this goal we have developed a novel hybrid method, the HDC-RCA analysis, which allows us to identify new GT events without the reliance on dense local networks or prior GT information.

## RESEARCH ACCOMPLISHED

To generate new GT5 events we have developed the two-tier HDCA-RCA methodology. HDC (Jordan and Sverdrup, 1981; Engdahl et al., 2004) determines accurate event location patterns relative to a provisional hypocentroid using regional and teleseismic phases. RCA analysis (Bondár et al., 2005), using local phases only, determines the accurate location of the cluster centroid by keeping the event and station patterns fixed. Since regional and teleseismic data usually lack the resolution to resolve the full depth pattern in a cluster, event depths are typically fixed in the HDC analysis to a best educated guess, based on analysis of individual events with depth phases, waveform analyses, or prior local data. Fixing event depths to nominal values projects the depth errors into the origin times. We have extended the RCA algorithm so that it solves for the cluster hypocentroid (epicenter, depth, and origin time shift). We have also enabled RCA to use secondary phases (Sg, Sb) in the inversion process.

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. To ensure strong cluster connectivity, we employ a graph theory approach. An event cluster can be viewed as an undirected graph, where the vertices are the events and stations, and the edges are the ray paths. An undirected graph is biconnected if at least two different paths exist between any two vertices (Orwant et al., 1999). Biconnectivity ensures that there are neither isolated vertices in the graph nor bridges whose removal would cause the graph to fall into disconnected pieces. Graph density, the ratio between the actual number of edges in the graph and the number of edges in the fully connected graph, offers a metric to characterize the cluster connectivity. To build a cluster for HDC analysis, we extract the largest biconnected graph from the initial EHB cluster for stations in the 3°–90° epicentral distance range. Similarly, for the RCA cluster we extract the largest biconnected graph from the HDC output for stations in the 0°–1.5° distance range.

### Preliminary RCA Applicability Criteria

In order to develop applicability criteria for the RCA similar to those of Bondár et al (2004), which would indicate whether the combined HDC-RCA analysis will reliably produce GT5 events from a cluster, we have performed a Monte Carlo experiment on a synthetic cluster. To generate the synthetic cluster, we distributed Yucca Flat GT0 nuclear explosions along a provisional fault plane (Figure 1a), hence creating a synthetic cluster of events of GT0 accuracy in both location and depth, with hypocentroid depth of 15.4 km. The arrival times are generated as iasp91 (Kennett and Engdahl, 1991) predictions at local (Pg), regional (Pn), and teleseismic (P) stations. We added distance and azimuth dependent delays (Figure 1b) to the arrival times to imitate separate local, regional, and teleseismic biases. We then relocated the events with EvLoc, using Pn and P phases in the 3°–90° distance range and keeping the depth fixed, introducing a 12 km location bias. This constitutes the input synthetic bulletin for the RCA Monte Carlo experiment. Figure 1c shows the RCA station network. Note that Pg arrival times are generated for every event at every station, thus producing a fully connected RCA cluster. Furthermore, the Pg arrival time bias is roughly equivalent to ±5% velocity perturbation with respect to iasp91, stations being increasingly slow to the SW and increasingly fast to the NE.

In the course of the Monte Carlo experiment, we model the HDC depth estimation procedures by fixing the entire cluster to a provisional depth of 5, 10, 15, 20, or 25 km. Since this introduces a depth error, we adjust the origin times in order to project the depth differences between the assumed and true depths into origin time errors. We randomly select stations (5, 6, 7, 8) and events (10, 12, 14, 16) and perturb the biased Pg arrival times with random Gaussian noise, assuming 0.5 s standard deviation normal picking errors. We generate 500 realizations for each combination for a total of 40,000 realizations. For each realization, we perform an RCA analysis and generate various metrics to measure the quality of the results.

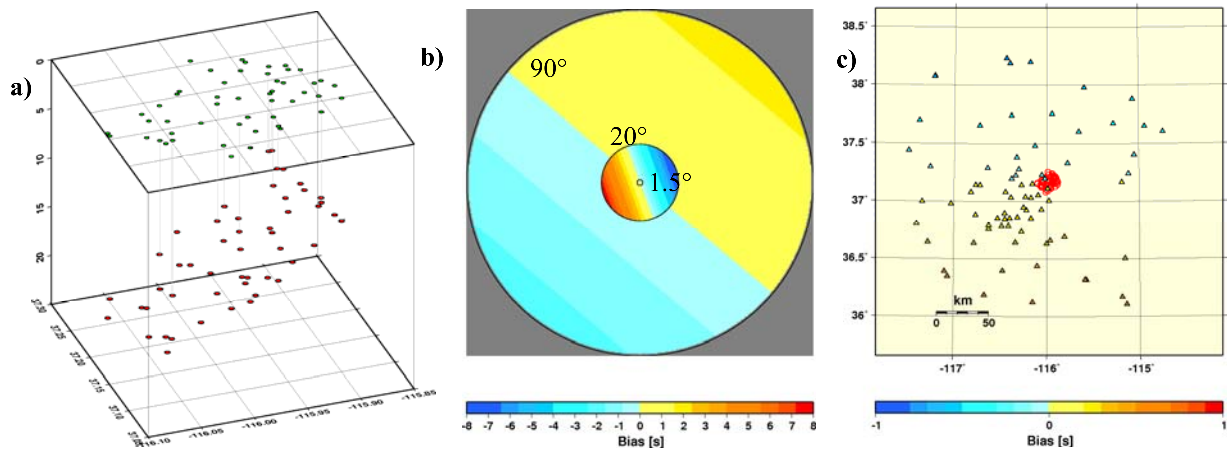


**Figure 1. (a) Synthetic event cluster to validate RCA depth inversion applicability criteria. (b) Arrival time delays (biases) applied to Pg, Pn, and P iasp91 predictions. (c) RCA local network geometry. The biased Pg arrival times represent ±5% crustal velocity perturbations in a worst case scenario.**

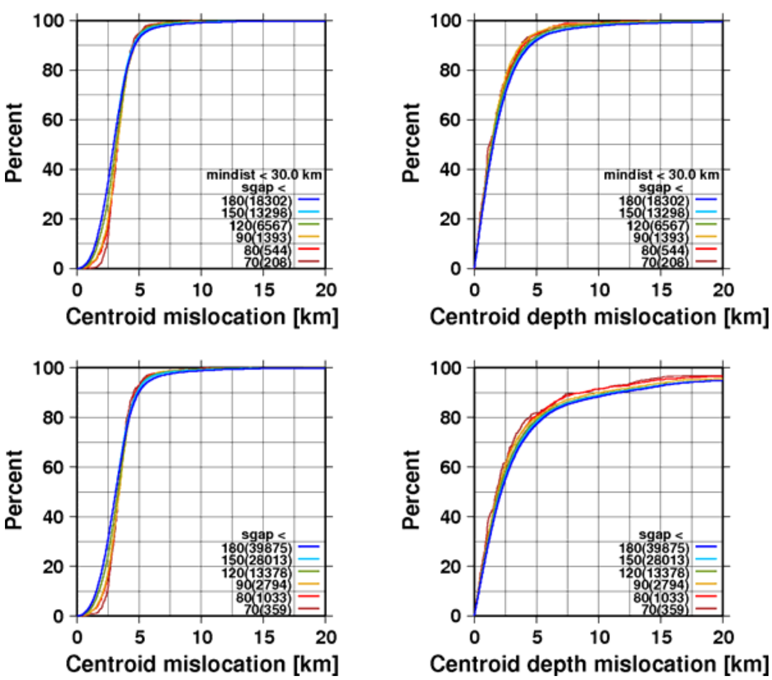


**Figure 2. Cumulative distributions of centroid mislocation and centroid depth mislocation for combined secondary gaps less than a specific threshold (bottom) and with station(s) within 30 km from the cluster centroid (top). When there is a nearby station, both the epicenter and the depth of the cluster centroid are recovered with high accuracy.**

We found that the combined secondary azimuthal gap, defined as the largest secondary azimuthal gap when considering the azimuths of all ray paths, provides a robust metric that predicts the quality of the RCA results. The lower panels in Figure 2 show the cumulative distributions of the centroid horizontal and depth mislocation for combined secondary gaps less than a specific threshold. The numbers in parentheses next to the combined secondary azimuthal gap threshold in the legend denote the corresponding number of realizations. The figure shows that we recover the true cluster centroid location within 5 km at the 90% confidence level when the combined secondary gap is less than 180°. However, this alone does not guarantee that the centroid depth is recovered with high accuracy and reliability. Not surprisingly, we need at least one station in the close vicinity of the cluster centroid to provide a constraint on the hypocentroid depth. This is illustrated in the upper panel of Figure 2. When there is at least one station within 30 km of the cluster centroid and the combined secondary gap is less than 180°, the epicenter of the cluster centroid is recovered within 5 km at the 95% confidence level, and the centroid depth is recovered within 5 km at the 90% confidence level. Hence, for a fully connected cluster, the above conditions provide GT5 applicability criteria for the cluster centroid, analogous to the GT5 criteria of Bondár et al. (2004) for single-event locations. The GT5 cluster centroid criteria are necessary (but not sufficient) conditions to generate GT5 events. Once the absolute location of the cluster centroid is pinned down with high accuracy, we promote events to the GT5 category if the semi-major axis of their combined absolute error ellipses (HDC+RCA), scaled to the 95% confidence level, is less than 5 km.

**HDC-RCA Clusters**

When we extract an event cluster from the EHB bulletin, we typically select a core set of clustered events that were recorded by both local (RCA) and regional/teleseismic (HDC) stations. We add regional/teleseismic-only events to the cluster to improve HDC performance. From this initial cluster we select a strongly connected graph of events and regional/teleseismic stations to facilitate robust HDC analysis. We then test for the RCA applicability criteria; if there is little hope of locating the cluster centroid with GT5 accuracy, we may skip the entire cluster and move on to the next cluster. Prior to HDC analysis, we establish the best depth estimates by waveform analysis or the analysis of individual events with depth phases and fix the event depths to the best depth estimates. The HDC analysis produces accurate relative locations and updates the phase identifications so that they are consistent with the fixed depth and ak135 (Kennett et al., 1995) predictions. HDC analysis may also remove observations (eventually entire events or stations) as outliers. From the HDC output cluster, we select a strongly connected graph of events and local stations for the RCA analysis and test again for the applicability criteria. If there are no stations within 30 km from the cluster centroid, we solve only for the horizontal shift of the cluster centroid by keeping the depths and origin times fixed to the HDC depth; otherwise, we solve for all model parameters (horizontal, vertical, and origin time shifts). We prefer to use local velocity models, especially for the depth inversion. We then shift the entire cluster to eliminate the HDC cluster centroid mislocation and identify GT5 events based on their combined absolute error ellipses (HDC+RCA), scaled to the 95% confidence level. Below we present some examples of HDC-RCA clusters.

Our first example is from the Welkom gold mines, South Africa. Figure 3a shows the RCA network geometry of 92 events and 8 stations. This is a very well connected cluster with a graph density of 0.53 and a combined secondary azimuthal gap of 66°. There are also two stations within 30 km of the cluster centroid. By using 553 Pb, Pn, Sb, and Sn phases, RCA shifted the HDC cluster west by 4.5 km (Figure 3b) and increased the hypocentroid depth from 7.5 km to 7.8 km. We promoted 84 events to GT5 level. In this cluster there are 3 EHB events that satisfy the Bondár et al. (2004) GT5 selection criteria for single-event locations, and their locations are consistent with the RCA results (Figure 3c). However, one might think that the 7.8 km hypocentroid depth is too deep for rockbursts. Hartnady (1990) and Richardson et al. (2005) note that natural seismicity does exist in the Kaapvaal craton, unrelated to mining activities. One of the events (1994/10/30, mb=5.7) in this cluster was studied in detail by Fan and Wallace (1995). Based on waveform inversion they concluded that the event was an earthquake with a source depth between 9 and 12 km. On the other hand, Bennett et al. (1996) identified this event as a rockburst, and Bowers (1997) reached the same conclusion, based also on waveform inversion, and put the event at 2.5 km depth. Given that the cluster most likely contains both earthquakes and rockbursts, the HDC-RCA cluster centroid depth is probably an acceptable compromise between the rockburst and earthquake populations. All of these published depth solutions fall within 5 km of the 7.8 km RCA hypocentroid depth.

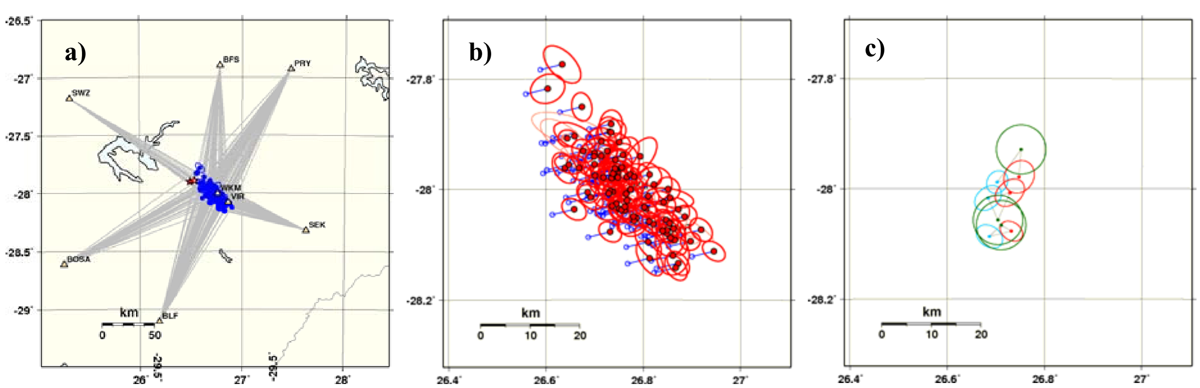


**Figure 3. (a) RCA geometry for the Welkom, South Africa, cluster. Triangles denote stations; blue circles represent HDC solutions. (b) RCA shifts the HDC (blue) locations to RCA (red) locations. Eighty-four RCA absolute error ellipses plotted with thick red were promoted to GT5 level. (c) HDC (blue) and RCA (red) locations are consistent with three existing GT5 events (green).**

Our next example (Figure 4) is from the south flank of Kilauea Volcano, Hawaii. The RCA cluster consists of 58 stations and 56 events, with a combined secondary gap of 12° and with 27 stations within 30 km of the cluster centroid. We used a local velocity model (Klein, 1981) for Pg and Sg travel times. In the HDC analysis, the event depths were fixed to prior local network solutions. Despite the very dense local network, only 22 events satisfy the original Bondár et al. (2004) criteria. RCA, however, promoted to GT5 status all 56 events, including the two offshore events near the underwater volcano, Loihi, off the coast of Hawaii and more than 20 km outside the local network.

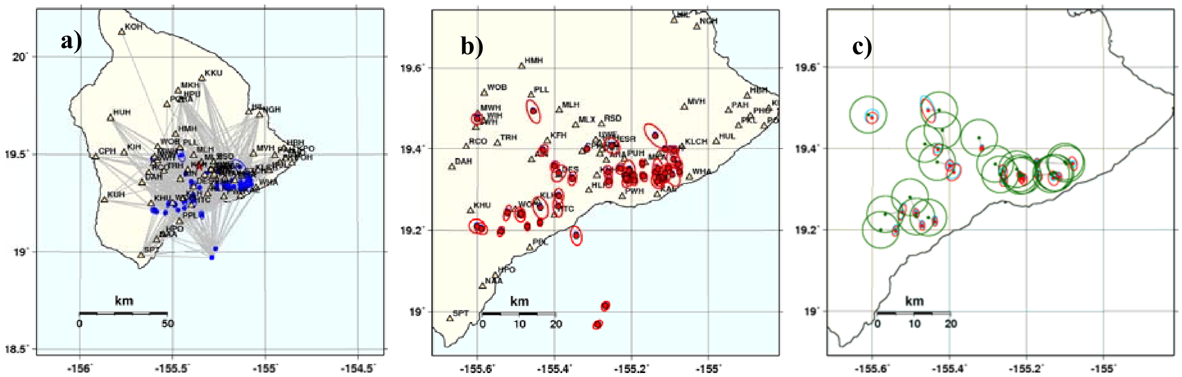


**Figure 4. (a) RCA geometry for the Kilauea Volcano south flank, Hawaii cluster. Triangles denote stations; blue circles represent the HDC solutions. (b) RCA promoted all 56 events to GT5 category. (c) HDC (blue) and RCA (red) locations are consistent with 22 existing GT5 events (green).**

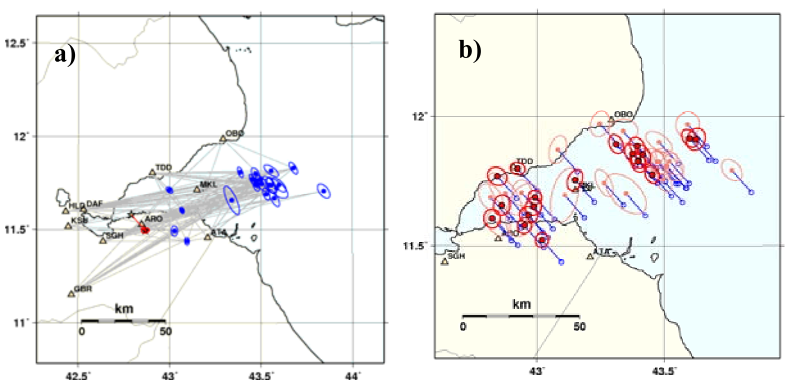


**Figure 5. (a) RCA geometry for the Gulf of Tadjoura, Djibouti cluster. Triangles denote stations; blue circles represent the HDC solutions. (b) RCA shift between the HDC (blue) and RCA (red) locations. RCA absolute error ellipses plotted with thick red represent 19 events promoted to GT5 level.**

Figure 5 shows an event cluster in the Gulf of Tadjoura, Djibouti. The cluster is located at the triple junction of the Gulf of Aden ridge and the Red Sea and East African rifts. Of the 40 HDC events, 21 events and 10 local stations were used to perform an RCA analysis. The combined secondary azimuthal gap is 126°, and there are two stations within 30 km of the cluster centroid. Again, we used local velocity model (Dugda and Nyblade, 2006) predictions for 115 Pg and Sg phases. RCA shifted the cluster 12.5 km NW and increased the cluster centroid depth from the HDC best estimate 10 km to 12.6 km. We promoted 19 events to the GT5 category.

As we already noted, we can resolve the hypocentroid depth with high confidence only if there are stations in the close vicinity of the cluster centroid. The choice of the local velocity model used in the inversion also plays some role. Figure 6 compares the local velocity models for several clusters to the iasp91 global 1D velocity model. For the clusters in Vrancea, Romania, and Taiwan, the iasp91 depth solutions were in quite good agreement with those obtained from local velocity models (Ma et al., 2001) for Taiwan and a 1D velocity model derived from the 3D tomographic model (Landes et al., 2004) by M. Popa. On the other hand, iasp91 produced non-sensical depth for the Gulf of Tadjoura cluster. Figure 6 indicates that the local velocity model derived from receiver function analysis for the Gulf of Tadjoura (Dugda and Nyblade, 2006) deviates the most from iasp91. Hence, depth is only as good as the local velocity model, and the proper choice of local velocity models is important for reliable depth determination.

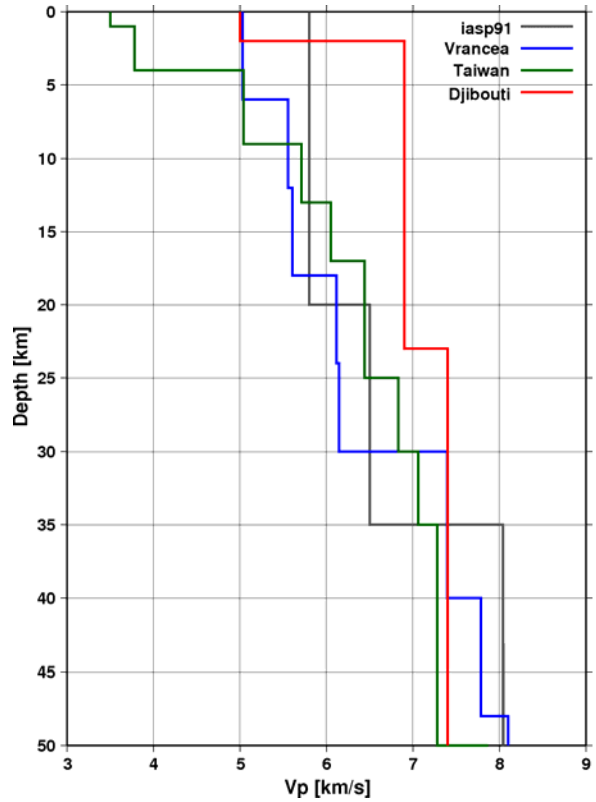


**Figure 6. Local velocity models compared to iasp91.**

It should be noted that failing the RCA applicability criteria does not mean that the locations are wrong; it only means that the cluster centroid cannot be recovered with GT5 accuracy at a high confidence level. This is illustrated in Figure 7. Figure 7 (a-b) shows the RCA geometry for an event cluster in the Mona Passage. All stations are located on Puerto Rico, and all the events occurred offshore, representing a combined secondary azimuthal gap of 275°. Therefore, even though some of the absolute error ellipses are small, we cannot promote any events to

the GT5 category because we cannot determine the cluster centroid location as GT5 with high confidence. Figure 7 (c-d) shows another example, this one for an event cluster in the West Azores. The stations are located on São Miguel Island, and all events are offshore. The combined secondary azimuthal gap in this case is 355°, representing the worst geometry we have encountered so far. Nevertheless, the locations are not inconsistent with the tectonic settings, indicated by the bathymetry and the plate boundaries.

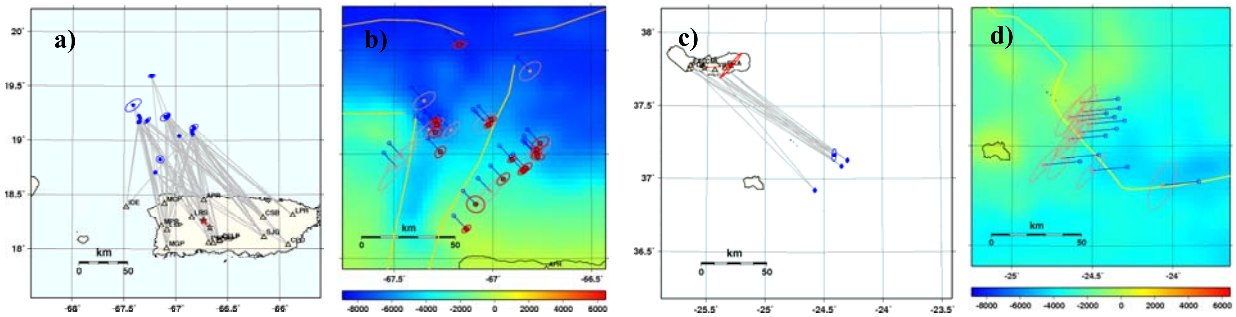


**Figure 7. (a) RCA geometry for the Mona Passage cluster (Puerto Rico). (b) HDC (blue) and RCA (red) locations plotted over the bathymetry map. Yellow lines show the Mona Passage transform faults and the Puerto Rico trough. (c) RCA network geometry for the West Azores cluster. (d) HDC (blue) and RCA (red) locations plotted over the bathymetry map. The yellow line shows the plate boundary between Europe and Africa. The HDC-RCA locations, while of unknown confidence, are consistent with the tectonic settings.**

The table below summarizes the clusters we have processed so far, ordered by their combined secondary azimuthal gaps. For those clusters that passed the RCA applicability criteria, we identify a significant number of GT5 events at a 90% or 95% confidence level.

| Cluster | Sgap | Centroid confidence | GT5 |
|---|---|---|---|
| Kileaua, Hawaii | 12 | 95%, with depth constraint | 56 |
| Picerno, S. Italy | 20 | 95%, with depth constraint | 58 |
| Vrancea, Romania | 29 | 95%, with depth constraint | 23 |
| Chi-Chi, Taiwan | 46 | 95%, with depth constraint | 25 |
| Welkom, S. Africa | 66 | 95%, with depth constraint | 84 |
| Owase, W. Honshu | 74 | 95%, with depth constraint | 14 |
| Gulf of Tadjoura, Djibouti | 126 | 95%, with depth constraint | 19 |
| Gulf of Fonseca, Honduras | 132 | 90%, no depth constraint | 9 |
| Galwa, W. Nepal | 203 | Stable solution, unknown confidence | None |
| Mona Passage, Puerto Rico | 275 | Stable solution, unknown confidence | None |
| São Miguel, Azores | 355 | Stable solution, unknown confidence | None |

## CONCLUSIONS AND RECOMMENDATIONS

We have extended the RCA algorithm to solve for all model parameters of the cluster hypocentroid (horizontal, vertical, and origin time shifts). Secondary phases (Sg, Sb) are also included in the RCA inversion. We use graph theory methods to ensure strong connectivity between stations and events in a cluster.

We have developed preliminary RCA applicability criteria for a fully connected cluster. The criteria are analogous to the single-event location GT5 selection criteria of Bondár et al (2004), but in this case they refer to the cluster centroid. According to the criteria,

- the cluster centroid epicenter is $GT5_{90\%}$ if the combined secondary azimuthal gap is less than 180°;
- the cluster centroid epicenter is $GT5_{95\%}$ and the centroid depth $GT5_{90\%}$ if the combined secondary azimuthal gap is less than 180° and there are stations within 30 km of the cluster centroid epicenter.

We have pointed out the importance of local velocity models to obtain reliable depth estimates. We will further refine the above applicability criteria for more realistic, partially connected clusters (i.e., when each station records only a subset of events).

We have begun the systematic processing of HDC-RCA clusters extracted from an updated EHB (Engdahl et al., 1998) bulletin and demonstrated that once the RCA applicability criteria are met, we are able to produce larger numbers of high-confidence GT5 events that would not result from any previous analysis.

## ACKNOWLEDGEMENT

We thank Michaela Popa (National Institute for Earth Physics, Bucharest, Romania) for providing us with a 1D velocity model for the Vrancea region.

## REFERENCES

Bennett, T. J., M. E. Marshall, B. W. Barker, and J. R. Murphy (1996). Use of seismic signal characteristics to identify several recent rockbursts, *Seism. Res. Let*. 67: 32.

Bondár, I., S. C. Myers, E. R. Engdahl, and E. A. Bergman (2004). Epicentre accuracy based on seismic network criteria, *Geophys. J. Int.* 156: 1–14, doi: 10.1046/j.1365-246X.2004.02070.x.

Bondár, I., B. Kohl, E. Bergman, K. McLaughlin, H. Israelsson, Y-L. Kung, P. Piraino, and E. R. Engdahl (2005). Global ground truth data set with waveform and improved arrival data, in *Proceedings of the 27th Seismic Research Review: Trends in Nuclear Explosion Monitoring*, LA-UR-05-6407, Vol. 1, pp. 289–298.

Bowers, D. (1997). The October 30, 1994, seismic disturbance in South Africa: Earthquake or large rockburst?, *J. Geophys. Res*. 102: 9843–9857.

Dugda, M. T. and A. A. Nyblade (2006). New constraints on crustal structure in eastern Afar from the analysis of receiver functions and surface dispersion in Djibouti, in *The Afar Volcanic Province within the East African Rift System*, Yirgu, G., C. J. Ebinger and P. K. H. Maguire (Eds.), London: Geological Society, Special Publications, Vol. 259, pp. 239–251.

Engdahl, E. R., E. A. Bergman, S. C. Myers, and F. Ryall (2004). Improved ground truth in Southern Asia using in-country data, analyst waveform review and advanced algorithms, in *Proceedings of the 26th Seismic Research Review: Trends in Nuclear Explosion Monitoring*, LA-UR-04-5801, Vol. 1, pp. 257–266.

Engdahl, E. R., R. D. van der Hilst, and R. P. Buland (1998). Global teleseismic earthquake relocation with improved travel times and procedures for depth determination, *Bull. Seism. Soc. Am.* 88: 722–743.

Fan, G. and T. Wallace (1995). Focal mechanism of a recent event in South Africa: A study using a sparse very broadband network, *Seism. Res. Let*. 66: 5, 13–18.

Hartnady, C. J. H. (1990). Seismicity and plate boundary evolution in southeastern Africa, *S. Afr. J. Geol*. 93: 473–484.

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

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

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

Klein, F. W. (1981). A linear gradient crustal model for south Hawaii, *Bull. Seism. Soc. Am.* 71: 1503–1510.

Landes, M., W. Fielitz, F. Hauser, M. Popa, and CALIXTO Group (2004). 3-D upper crustal tomographic structure across the Vrancea seismic zone, Romania, *Tectonophysics*, 382: 85–102.

Ma, K-F., J. Mori, S-J. Lee, and S. B. Yu (2001). Spatial and temporal distribution of slip for the 1999 Chi-Chi, Taiwan, earthquake, *Bull. Seism. Soc. Am.* 91: 1069–1087.

Orwant, J., J. Hietaniemi, and J. Macdonald (1999). *Mastering Algorithms with Perl*. Sebastopol: O'Reilly & Associates, Inc.

Richardson, E. B. A. A. Nyblade, W. R. Walter, K. Mayeda, and L. Lowe (2005). Seismic source characterization and energy partitioning from in-mine and regional broadband data in South Africa, in *Proceedings of the 27th Seismic Research Review: Trends in Nuclear Explosion Monitoring*, LA-UR-05-6407, Vol. 1, pp. 622–631.