EVALUATION OF CROSS-CORRELATION METHODS ON A MASSIVE SCALE FOR ACCURATE RELOCATION OF SEISMIC EVENTS

Won-Young Kim, Paul G. Richards, David P. Schaff, Felix Waldhauser, and Jian Zhang

Lamont-Doherty Earth Observatory of Columbia University

Sponsored by Air Force Research Laboratory

Contract No. F19628-03-C-0129

ABSTRACT

We are evaluating a method of locating seismic sources (earthquakes, explosions) based on the use of waveform cross-correlation (WCC) measurements instead of using the conventional measurements of seismic wave arrival time (phase picks). WCC measurements have been demonstrated to be 10 to 100 times more accurate, where they can be used. The principal issue we are exploring is the extent to which a significant fraction of seismicity can be located using WCC measurements. In this second year of the project we have focused on studies of intraplate seismicity in New Madrid (Central United States) and Charlevoix (Eastern Canada). We have also compared our results for these regions, with the results for a separately-funded project to study the seismicity of Northern California.

Datasets have been assembled from scratch working in conjunction with regional network operators. For New Madrid, three datasets have been acquired. The first is from stations of the network originally installed by Otto Nuttli and his colleagues at St. Louis University. This is an archive of 51,541 phase picks, associated with events from 1974 to 1998. The second is a waveform archive, associated with 42 PANDA stations, 918 events, and 17,598 phase picks for events from October 1989 to August 1992. The third is also a waveform archive, associated with 85 stations, 614 events, and 16,461 events from January 2000 to October 2003, obtained from the Center for Earthquake Research and Information (CERI) network, operated by the University of Memphis. We are currently processing the data from this station network for an additional time period, from 1995 to 1999, specifically in order to associate the waveforms with phase pick information and bulletin locations. It may be noted that several of the station locations operated by St. Louis University are the same as those used by CERI, and thus we expect to be able to present a unified picture of New Madrid seismicity for a period of about thirty years from 1974. Preliminary WCC results of the PANDA network indicate that about 76% of 809 relocated events have five or more P-wave cross-correlations ($CC \geq 0.7$). Similar statistics have been acquired for CERI data at New Madrid. Both PANDA data and CERI allow the identification of clearly defined lineations (faulting), previously identified but not seen so well in the traditional bulletin locations.

For Charlevoix, using data acquired from the Geological Survey of Canada, we have now relocated 2272 events using phase-pick data and waveform data from a 46-station network that had just eight stations close to the seismicity. Here we find that only 5% of the events had five or more P-wave cross-correlations ($CC \geq 0.7$). In the north-east part of this region, fault structures can clearly be seen in cross sections of the relocated seismicity.

We also comment on the results of a seismicity study of Northern California, for which 95% of 225,000 events cross-correlate at 4 or more stations. In general, we conclude that the benefits of using cross-correlation methods and multi-event relocation analysis are so substantial, that they should be considered for application to the work of routine bulletin publication.

The basic reason for a much lower percentage of Charlevoix events cross-correlating, is the paucity of stations close in to the seismicity.

OBJECTIVE

We are evaluating a method of locating seismic sources (earthquakes, explosions) that is based on the use of waveform cross-correlation (WCC) measurements instead of using the conventional measurements of seismic wave arrival time. WCC measurements are ten or a hundred times more accurate, where they can be used. The principal issue we shall explore is the extent to which a significant fraction of seismicity can be located using WCC measurements.

RESEARCH ACCOMPLISHED

We have completed and published waveform-based studies of the seismicity of China (Schaff and Richards, 2004ab), and are currently conducting such studies for New Madrid, Central United States; Charlevoix, Eastern Canada; and, in a separate study not funded by the Air Force Research Laboratory (AFRL) of Northern California. About 10% of the earthquakes in China, as monitored at far regional distances, cross-correlated. In the next subsection, we briefly report preliminary results for New Madrid and Charlevoix, and summarize key differences between these two regions, and a recent study of Northern California. Because there are significant differences in the fraction of seismic events that can be located precisely by modern methods (i.e., using relative arrival times measured accurately by waveform cross-correlation, followed by application of a multi-event relocation algorithm such as "double difference") in each region, we then discuss in a separate subsection the likely causes of these differences. Further discussion of these results has recently been submitted for publication (Richards et al., 2005).

Preliminary relocations for New Madrid and Charlevoix, and their differences from Northern California

We are studying the seismicity of the New Madrid region, Central United States, for three different eras of station deployment. Data consists of phase picks as well as waveforms. The first era, is from 1974 to about 1998 when personnel at St. Louis University operated analog stations and used traditional phase pick methods (over 52,000 phase picks). The second era is associated with the operation of a PANDA network by the Center for Earthquake Research Information (CERI) of the University of Memphis, from October 1989 to August 1992 (42 stations, 918 events, 17,598 phase picks). The third era is also associated with CERI, but utilizing a different network (up to 85 stations, about 1000 events) from 1994 to October 2003. In this short paper we show only one example of relocations for New Madrid. Thus, Figure 1 shows relocation of 809 events monitored by the New Madrid PANDA network: 616 (76%) of these events cross-correlated (with CC $\geq$ 0.7) at 5 or more stations; 695 (86%) at 4 or more stations; and 735 (91%) at 3 or more stations. The two cross-sections in Figure 1 show a fault plane dipping to the west, with dip that steepens to the south.

We are studying the seismicity of the Charlevoix region, Eastern Canada, from January 1988 to December 2003 involving 27,976 events with 33,423 phase picks recorded at 46 stations operated by the Geological Survey of Canada. Although we have waveforms from many of these stations, as we shall see in a later section we find that only 8 of these stations, close to the source region, are the basis for almost all the successfully cross-correlated waveforms. Many of these stations are 100s to 1000s of km away and record only the larger events. Figure 2 shows the relocation of 2272 events: only 242 of them (5%) cross-correlated (with CC $\geq$ 0.7) at 5 or more stations; 622 (25%) at 4 or more stations; and 1439 (57%) at 3 or more stations. These preliminary relocations, as well as the original bulletin locations for this region, indicate that active faulting is simpler and more clearly defined in the northeastern part of the zone, compared to more complex features in the southwestern part. Cross-section 2–2′ shows a fault dipping about 50º to the southeast, whereas the other cross-section shows a fault dipping more steeply, at about 75º.

It may be noted that the percentage of events whose waveforms cross-correlate at enough stations to apply modern methods of event location differs significantly between New Madrid (higher %), and Charlevoix (lower %). Schaff and Waldhauser (2005) have described results from an application of cross-correlation methods to process the complete digital seismogram data base for Northern California to measure accurate differential times for correlated earthquakes observed at common stations. Their results are even better than those for New Madrid. Their waveform database includes about 15 million seismograms from 225,000 local earthquakes between 1984 and 2003. A total of 26 billion cross correlation measurements were performed on a 32-node (64 processor) Linux cluster. All event pairs with separation distances of 5 km or less were processed at all stations that recorded the pair. A total of about 1.7 billion P-wave differential times had cross correlation coefficients (CC) of 0.6 or larger. The P-wave differential times are often on the order of a factor of ten to a hundred times more accurate than those obtained from routinely picked phase onsets. 1.2 billion *S*-wave differential times were measured with CC > 0.6, a phase not routinely picked at the Northern California Seismic Network because the onset of *S*-phases is often obscured by *P*-wave coda. These results show a surprisingly high degree of waveform similarity for most of the Northern California catalog, which is very encouraging for improving earthquake locations. For about 95% of the events, waveforms have CC values that are greater than 0.7 for at least four stations with one or more other events. 90% of the events meet this criterion at eight or more stations, and 82% of the events in the catalog cross-correlate at twelve or more stations. Even tectonically complicated zones exhibit these favorable statistics, such as Long Valley Caldera and Geysers Geothermal Field, where mechanisms are quite variable. Apparently, as long as the earthquake density is high enough there is a high probability that at least one other event occurs nearby with a similar focal mechanism, enabling very precise relative locations.

For the four regions we have examined, China was studied with far-regional signals and the other three regions were studied with local stations. Yet we found significant differences, in that about 85% of New Madrid events could be relocated with modern methods (using waveform cross-correlation), and 95% of Northern California events; but only about 25% of Charlevoix events. In the next sub-section we comment further on these differences and discuss underlying causes.

Comparisons, to explore the applicability of wave-based methods in different seismic regions

Figure 3 shows a map for each of three regions: part of Northern California (around the Calaveras fault); New Madrid; and Charlevoix. The Figure shows relocated events and the stations at which cross-correlated signals are found. Note that primarily seven of the Charlevoix stations cross-correlate. The other stations of the network in Eastern Canada are at greater distance, and typically can contribute only phase picks plus just a few correlations to the location of Charlevoix events. These maps all share the same distance scale, and it is interesting that New Madrid and Northern California have approximately the same station density (about one station per 100 sq. km) and distances between stations (about 12 km) whereas Charlevoix is a factor of two smaller in density and a factor of two greater in station distance. From the station plot for Charlevoix it is easy to see why a criterion of “4 or more” or “5 or more” stations is hard to achieve for the events: first, because greater epicentral distances are involved to reach four stations, and second because at least half the stations had to record the event with good signal-to-noise ratio (SNR) and high similarity. This may not happen, due to radiation pattern, etc. Four out of 38 stations at New Madrid is much easier to achieve. It appears that increasing the number of stations near Charlevoix would increase the percentage of events that cross-correlate in that region.

Figure 4 shows distributions of event depth and magnitude for the three regions. Here we see that while the magnitude distributions are more or less similar, the Charlevoix events tend to be deeper than both California and New Madrid by 5 km or more. Since depths as well as epicentral distances (see Fig. 3) are greater for many of the Charlevoix events, it is likely that fewer stations would have good SNR, and hence fewer stations would cross-correlate. Concerning earthquake densities, it is of interest that New Madrid and Charlevoix have about the same value (around 1 earthquake per sq. km), whereas Northern California (or at least the Calaveras fault) has about 160 events per sq. km (for the time periods studied). The inter-event separations are likewise similar for New Madrid and Charlevoix (averaging about 2 km), as shown by the distributions depicted in Figure 5, whereas for the Calaveras fault it is only about 100 m.

Figure 6 shows the distribution of cross-correlation values with distance. These curves tend to flatten out beyond 2 km, and similar results are obtained for Northern California. It is for this reason that the scientifically most interesting results of streaks and repeating events in Northern California can be documented — because most of the good correlations are for inter-event separations of less than 2 km. For New Madrid and Charlevoix, where many of the nearest neighbors are 2 km away or more, we expect lower CC values and less collapsing of structures into clusters, except for a few isolated clusters with higher CCs.

Differences in the number of cross-correlation measurements per event pair are mainly controlled by the number of available stations. In both the Charlevoix and the New Madrid region 90% or more of the events correlate with at least one other events at one station. That number decreases to 10% for the Charlevoix and 76% for the New Madrid region when 5 or more correlation measurements per event pair are considered. This difference reflects the availability of stations with digital waveforms. The Charlevoix network consists effectively of only 8 stations which provided digital waveforms for cross correlation, while the Panda network includes 38 such stations.

CONCLUSIONS AND RECOMMENDATION

Cross-correlation measurements for earthquakes in the New Madrid seismic zone (NMSZ) and the Charlevoix seismic zones (CSZ) indicate that a much higher number of events (about 85%) correlate in the NMSZ, compared to only 5 to 25% of the events that correlate in the CSZ. The reason for this discrepancy may be due to: differences in the network geometry and number of stations; differences in the type and diversity of faulting associated with the events; the variation of geophysical properties in general; and the degree of structural heterogeneity in particular, within the areas of investigation; or a combination of all.

We recommend that for regions of high seismicity within which a high percentage of events cross-correlate at enough stations to achieve precise relocations, consideration be given to a wholly different paradigm for event location — namely, a framework in which events are located using cross-correlation measurements obtained from relevant portions of the waveform, rather than using phase picks.

ACKNOWLEDGEMENTS

We greatly appreciate the assistance of the Geological Survey of Canada, and personnel at St. Louis University and the University of Memphis. In particular we thank Drs. David McCormack and Jim Lyons at the Geological Survey of Canada, Ottawa, who helped us assemble phase and waveform data from earthquakes in the Charlevoix region. We thank Drs. Jer-Ming Chiu of CERI, University of Memphis and Jiakang Xie at Lamont-Doherty who helped us to acquire phase and waveform data from the PANDA New Madrid deployment. We also thank Drs. Mitch Withers at CERI, University of Memphis for New Madrid seismic network data and Robert Herrmann at St. Louis University for the Central Mississippi Valley seismic bulletin.

REFERENCES

Chiu, J. M., A. J. Johnston, and Y. T. Yang (1992). Imaging the active faults of the central New Madrid seismic zone using PANDA array data, *Seism. Res. Let.* 63: 375–393.

Richards, P. G., F. Waldhauser, D.P. Schaff, and W.-Y. Kim (2005). The applicability of modern methods of earthquake location, submitted to *Pure and Applied Geophysics*.

Schaff, D. P., and P. G. Richards (2004a). Lg-wave cross correlation and double-difference location: application to the 1999 Xiuyan, China, sequence, *Bull. Seism. Soc. Amer.* 94: 867–879.

Schaff, D. P., and P. G. Richards (2004b). Repeating seismic events in China, *SCIENCE* 303: 1176–1178.

Schaff, D. P., and F. Waldhauser (2005). Waveform cross-correlation based differential travel time measurements at the Northern California Seismic Network, submitted to *Bull. Seism. Soc* .

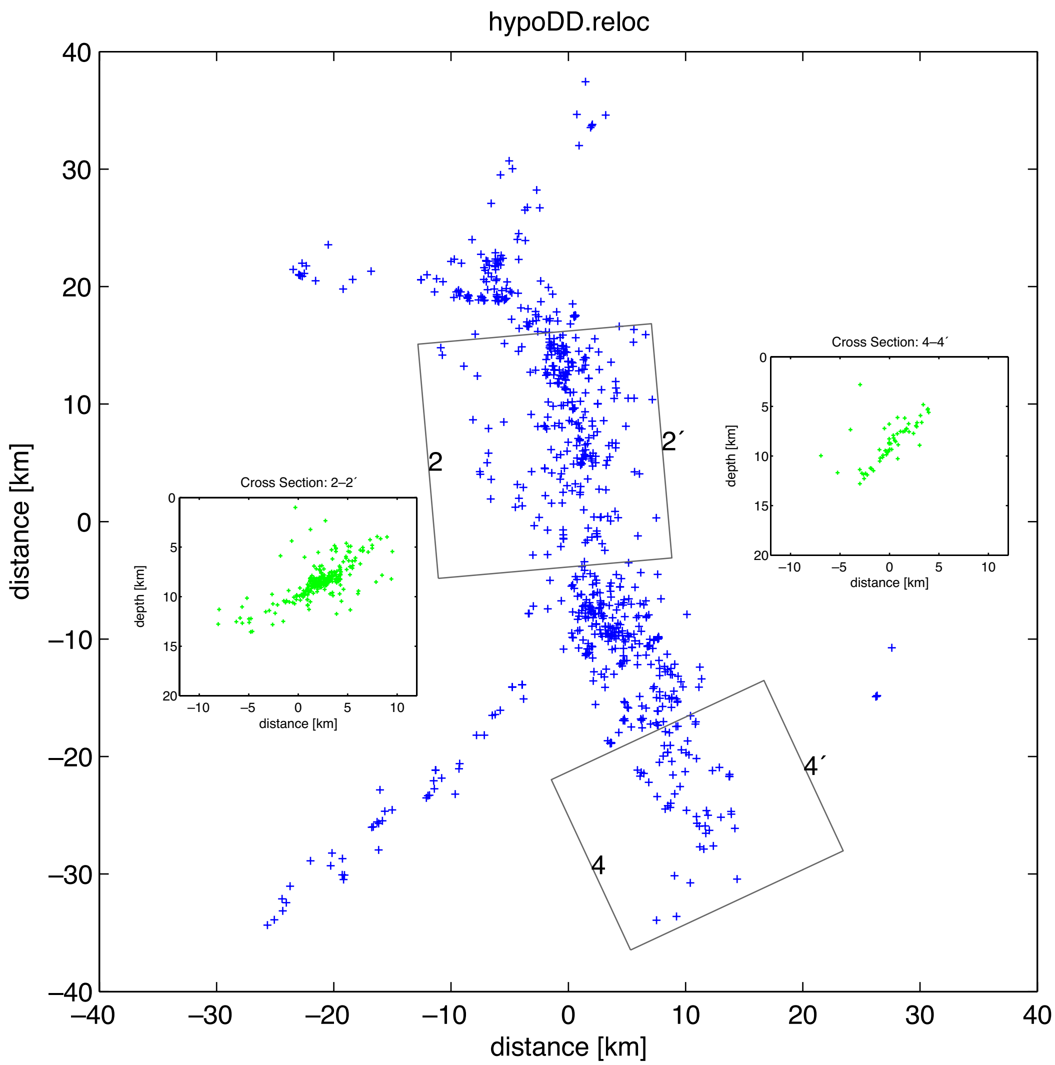

Figure 1. Relocation of 809 seismic events in the New Madrid region of te central United States from October 1989 to August 1992, using 42 stations of a PANDA network (see Chiu et al. 1992). This application of the double-difference algorithm was based upon phase pick pairs and cross-correlations, in each case using P-waves and S-waves. We have also obtained such relocations for the later period during which the University of Memphis operated a larger network of stations.

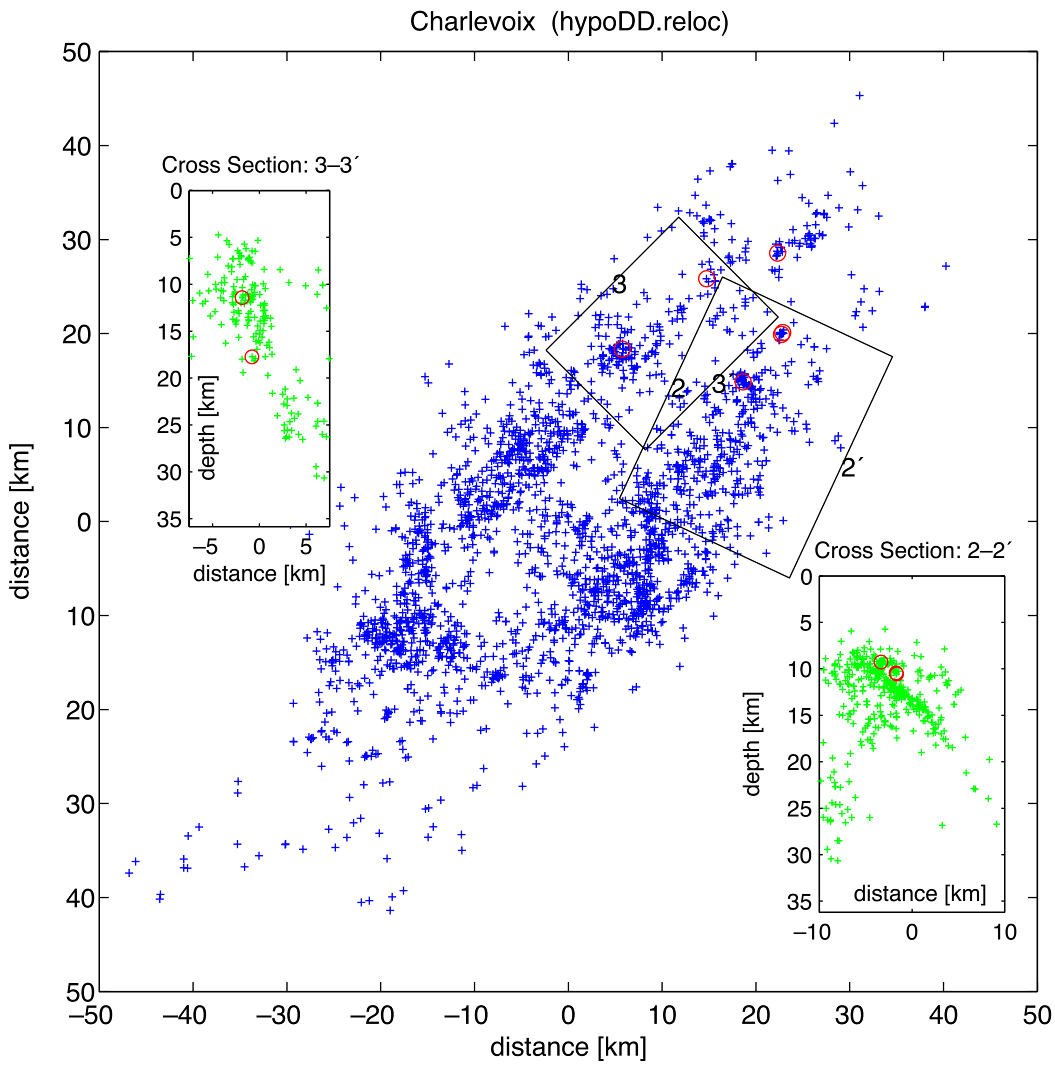

Figure 2. Relocation of 2242 events in the Charlevoix region of Eastern Canada from January 1988 to December 2003, using 46 stations operated by the Geological Survey of Canada. Red circles denote events of magnitude 4 or larger.

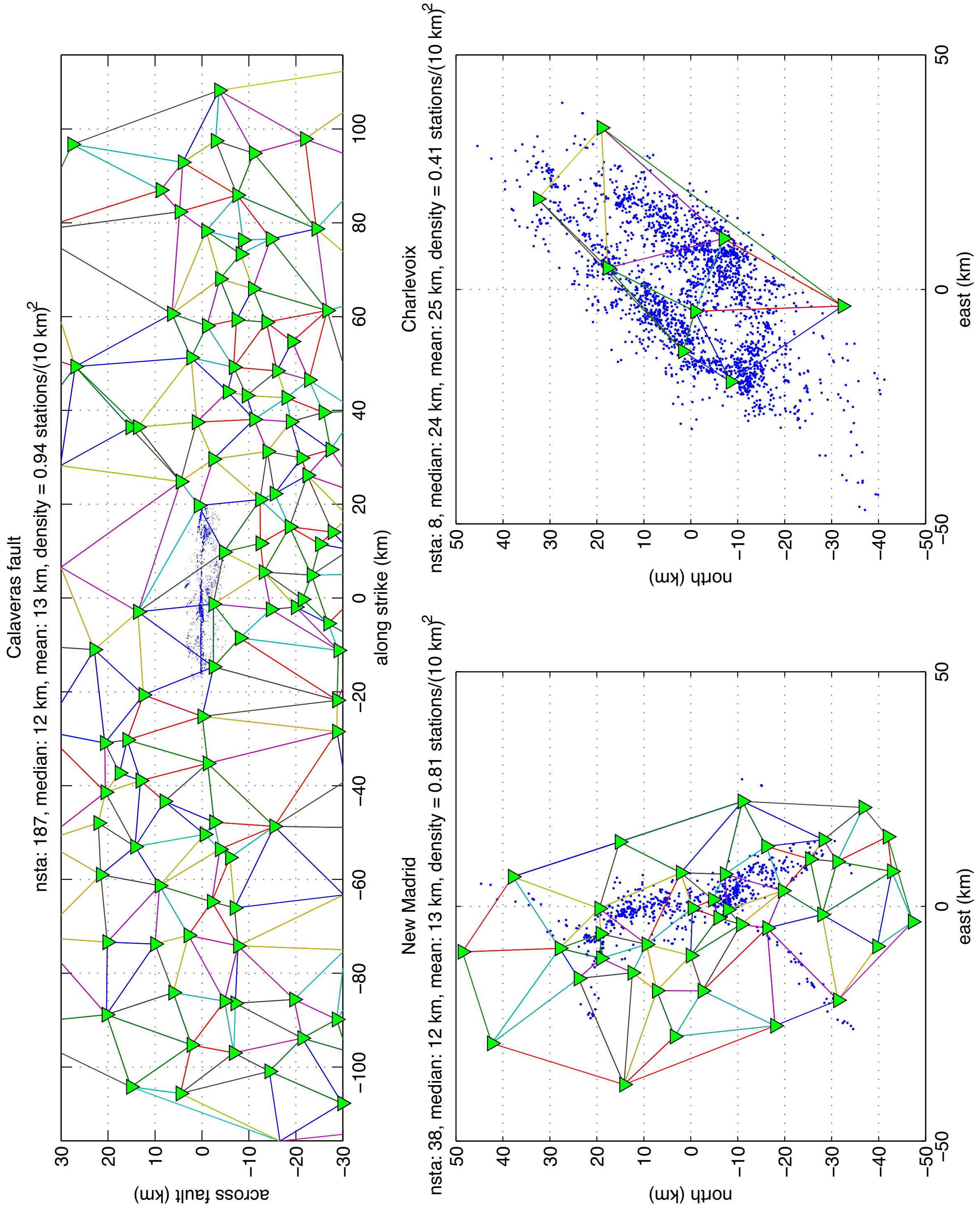

Figure 3. Station distance and density (Delaunay tessellation). Charlevoix has about half the station density of the Calaveras and New Madrid regions.

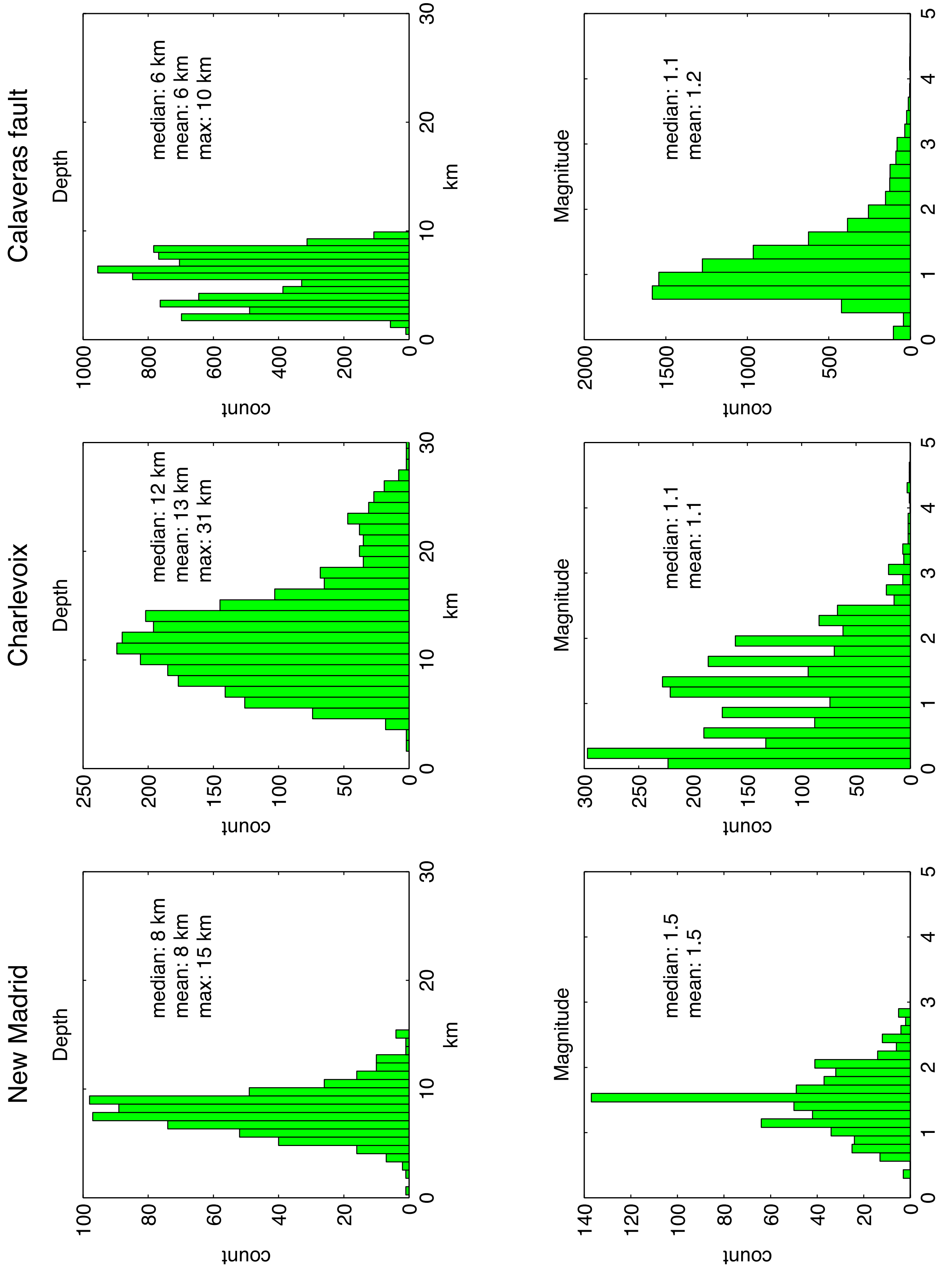

Figure 4. Depth and magnitude distributions, for three different regions of seismicity.

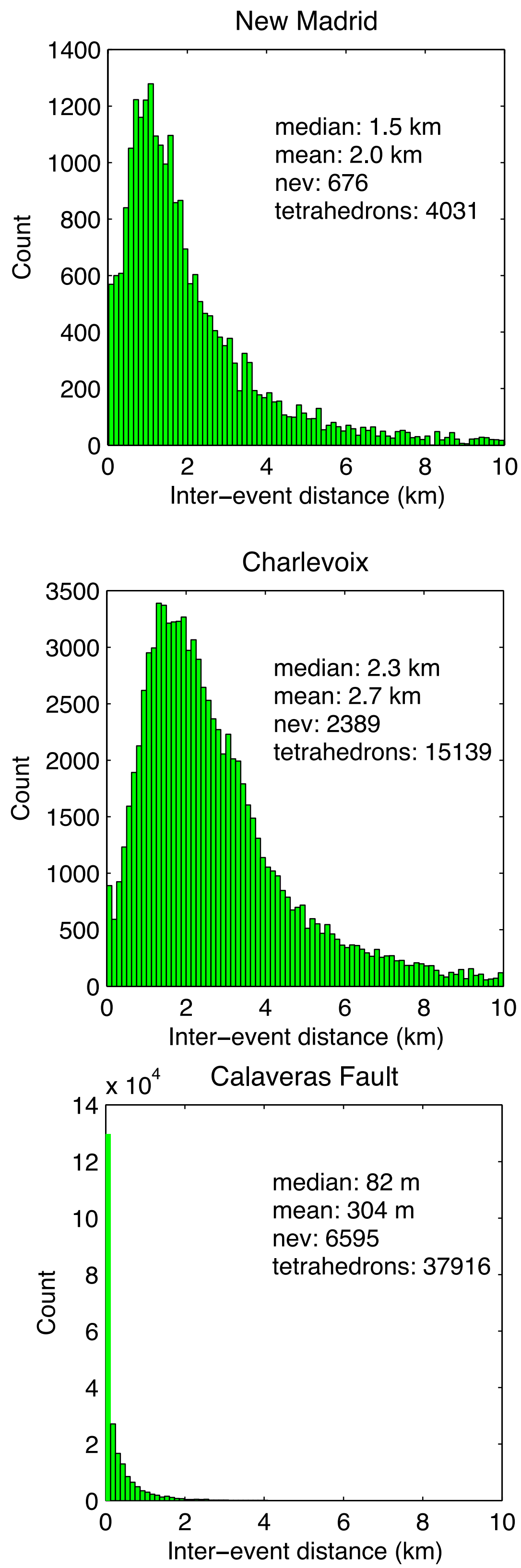

Figure 5. Distribution of nearest natural neighbor distances (based on 3-D Delaunay tessellation).

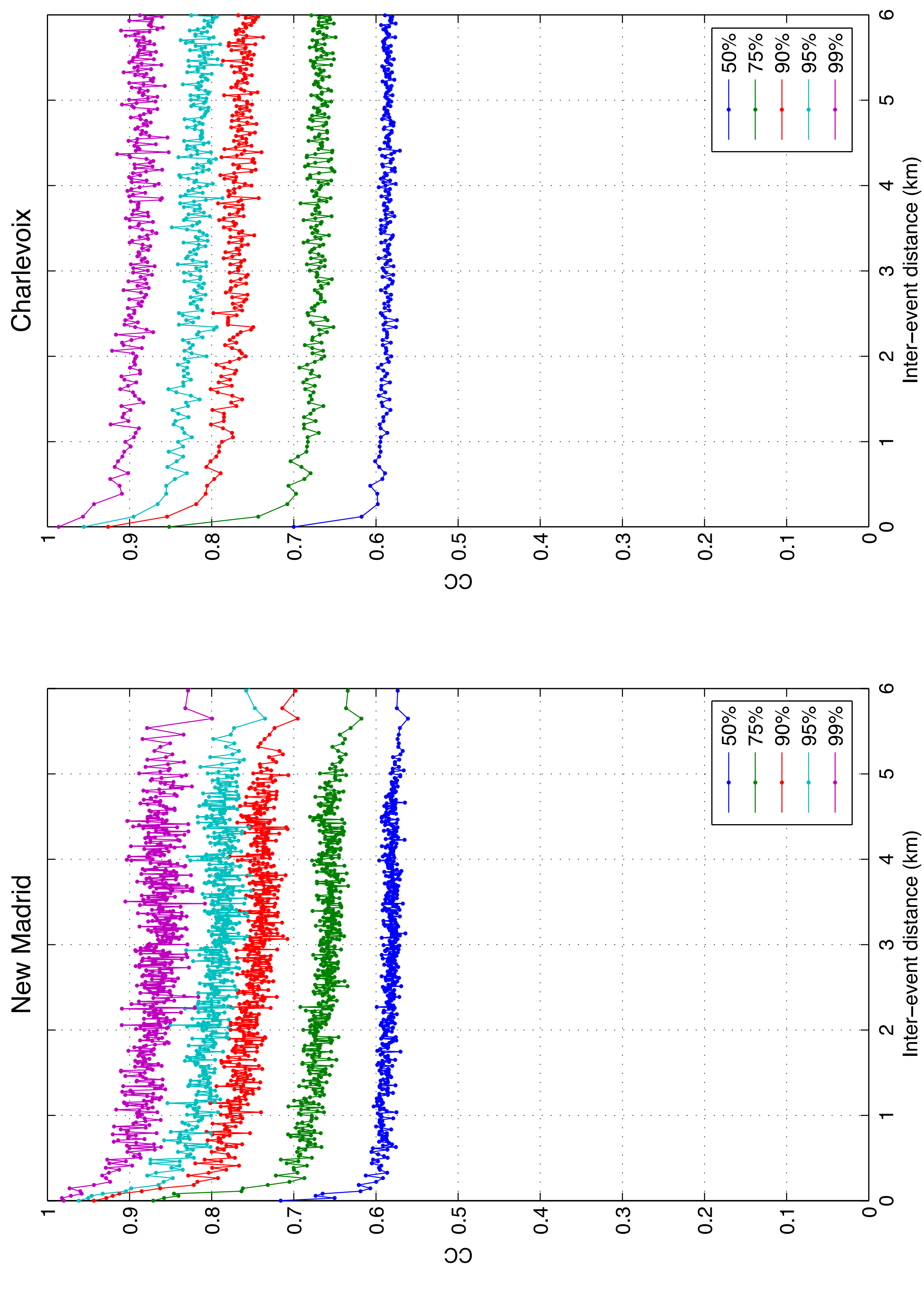

Figure 6. Distribution of cross-correlation values (CC) with distance