TESTING EVENT LOCATION CAPABILITY WITH GROUND TRUTH EVENTS IN KAZAKSTAN

Clifford Thurber, Florian Haslinger, Chad Trabant,
Department of Geology and Geophysics, University of Wisconsin-Madison

Sponsored by U.S. Department of Defense
Defense Threat Reduction Agency
Contract No. DSWA01-98-1-0008

ABSTRACT

We are investigating seismic event location capability in Kazakstan using first-P arrivals from nuclear explosions with exact ground truth information recorded on a sparse network of digital seismic stations. Our first step involves the application of a waveform cross-correlation method to extract higher precision first-P arrival times. Comparison of our revised picks with picks from the International Seismological Centre (ISC) catalog reveals significant discrepancies. Occasional large discrepancies (several seconds) may be due to clock errors (either not incorporated in the digital data or erroneously applied to the ISC data), but most are due to 0.5-second and 1-second round-off at various ISC stations (that is, stations reporting times rounded off to integer seconds or half-seconds), leading to errors up to 0.25 and 0.5 seconds, respectively. Though the round-off problem is not a direct concern for future CTBT monitoring, it does point to one source of error in global relocation and tomography studies.

We are using the improved arrival times to examine several issues regarding location accuracy and precision. Because the true explosion locations are known precisely, we can evaluate both relative and absolute location capabilities. Our initial focus is on relative locations. Aspects we are investigating include relative location accuracy using sparse regional-distance observations, the spatial variability of derived source-specific station corrections (also known as path corrections) to regional versus teleseismic stations, and the influence of source-region heterogeneity.

Relative relocations for most events are extremely accurate using all available picks from digital data. The vast majority of events are relocated within 1 to 2 km of ground truth, consistent with estimated location uncertainties of 1 to 2 km. Restricting the data to stations within 40° cuts the number of stations by a factor of about 3 and the number of observations by a factor of about 4, but both mislocations and estimated uncertainties increase only by modest factors.

For location using path corrections to be accurate and effective in practice, the rate at which the corrections change with changing source location must be relatively modest. We determined sets of path corrections for three subsets of explosions in adjacent regions each about 10 to 15 km in diameter. The sets of corrections are highly correlated, although with systematic differences that we attribute to uncertain origin times. Accounting for the systematic differences, correction differences are generally +/- 0.2 s or less.

The alternative is to use a model-based approach. We will examine the contributions of 3-D structure to mislocation. We will present the results of teleseismic tomography beneath the Kazakstan test site using the explosions as a source array. We will investigate the effect of source-region heterogeneity on the travel times. We will also assess the degree to which a regional or global 3-D velocity model can account for some of the travel time differences that are indicated by the path corrections and their spatial differences.

Key Words: location, ground truth, corrections, calibration, waveform cross-correlation, 3-dimensional structure

OBJECTIVE

We are carrying out investigations of seismic event location capability, using events that have ground truth information to allow a direct assessment of actual location errors and computed precision estimates. The objectives are to determine the most effective approaches for precise and accurate event location for CTBT monitoring purposes. Waveform cross-correlation methods have been applied to determine more accurate relative arrival times for available digital seismic data from nuclear explosions at Shagan River (Balapan), Kazakstan. These arrival times, along with the original bulletin times, are being used for a comparison of two strategies for improving absolute location accuracy and reducing estimated location uncertainty, one "model-based" and the other "calibration-based." The calibration-based approach involves joint hypocenter determination (JHD), including the estimation of path corrections. The model-based approach involves the use of regional and/or global three-dimensional (3D) velocity models to provide improved travel time estimates.

RESEARCH ACCOMPLISHED

Our current region of investigation is the former nuclear testing site at Shagan River, Kazakstan. Ground-truth information for Shagan River has been compiled from publicly available sources: absolute locations for 100 nuclear explosions (NNCRK, 1999) and origin times for 7 of them (Bocharov et al., 1989). This information was used to derive origin time estimates for the other 93 nuclear explosions using International Seismological Centre (ISC) data and JHD, assuming depths of 0 km for all the explosions. The event locations are shown in map form in Figure 1 and the locations and origin time results (including estimated origin time uncertainties) are presented in Table 1.

We analyzed the residuals from the JHD results, characterizing the residual distribution and evaluating the quality of the stations in terms of self-consistency of the estimated travel times (reported arrival time minus estimated origin time). The ISC residuals show a broad distribution with numerous +/- 5 and 10 second residuals. The distribution is well fit by the sum of a Gaussian term with a standard deviation of 0.3 seconds and a smaller non-Gaussian term. It is apparent that using raw ISC data, in this case mainly from reasonably large events, will be problematic for location calibration work as well as for tomography studies (see below), unless outliers are dealt with carefully.

Waveform cross-correlation can significantly improve pick quality (VanDecar and Crosson, 1990). We have applied the Xadjust package (Dodge et al., 1995) to the available digital data for Shagan River explosions (mainly from 1980 on). The cross-correlation technique is particularly valuable for extracting more accurate relative arrival times when some waveforms are noisy - see Figure 2 for an example. Comparing picks from the digital data to the ISC picks (see example in Figure 3) shows the effect of round-off in the reported ISC times (that is, stations reporting times rounded off to integer seconds or half-seconds). In Figure 4 we compare the set of cross-correlation picks to the corresponding ISC times, showing the frequency of time differences. Most of the time differences are less than 0.5 seconds. We attribute much of the scatter to round-off in the times reported to ISC.

Relocation analysis was carried out using JHD for the entire group of events and for 3 subsets (Northeast, Central, West groups). Relative relocations for most events are extremely accurate using all available picks from digital data (fixing one event location and origin time in each case). Many events are located within 1 km of ground truth, and almost all locate within 2 km. Estimated location uncertainties are of the same order. The number of truly regional-distance observations (within 20°) with reliable timing for these events is extremely small. Thus to simulate a regional monitoring situation in a reasonable manner, we relocated the events using stations within 40°; an example is shown in Figure 5.

For location using path corrections to be accurate and effective in practice, the rate at which the corrections change with changing source location must be relatively modest. We determined sets of path corrections for three subsets of explosions in adjacent regions each about 10 to 15 km in diameter. The sets of corrections are highly correlated, although with systematic differences that we attribute to uncertain origin times. Accounting for the systematic differences, correction differences are generally +/- 0.2 s or less.

In order to image the 3D crustal structure beneath the Shagan River test site using teleseismic traveltime residual analysis, we follow the Aki-Christofferson-Husebye (ACH) approach (Aki et al., 1977). The unique characteristic in our case is that the nuclear explosions (seismic events) are treated as a station array and the stations from which arrival time data are available are treated as seismic events. For a general overview and mathematical details of ACH type teleseismic tomography see for example Evans and Achauer (1993); here we will just briefly discuss the peculiarities of the inversion when using events as stations and vice versa.

The travel time residual r_{ij} for a ray traveling from source j to receiver i can be expressed as

$$r_{ij} = dO_j + \sum_{k=1,K} \left(- \int_{base}^{receiver} \frac{1}{V_0^2} \frac{\delta V}{\delta \boldsymbol{m}_k} ds \right) \delta \boldsymbol{m}_k + e_{ij}$$

The integral describes the change in travel time of the ray from the model base to the receiver with respect to change of the seismic velocity V in the K model parameters $\boldsymbol{m}_k$. In our case, as generally in ACH tomography, the model is constructed of blocks of constant velocity. e_{ij} is an error term, and dO_j is a source term which is introduced to account for absolute velocity differences between ray paths outside the model volume and other unresolvable terms (e.g. source location and origin time uncertainties) which may be present in the data. dO_j can be expressed as a weighted average residual travel time between source j and all receivers. Doing so allows to replace the absolute travel time residual r_{ij} by the relative travel time residual rr_{ij}. These relative residuals are the data used in the inversion, and ideally they solely contain the difference between true velocities and initial model velocities within the space of model parameters $\boldsymbol{m}_k$.

Reversing sources and receivers for the inversion does not change the influence of absolute velocity differences outside the model volume, but places all source uncertainties inside the model volume. To account for that, an additional source term would have to be introduced and inverted for. In our case, however, ground truth location information for all sources is available, and origin time uncertainties are in the order of 30 to 50 ms (see above). Source terms are therefore expected to have magnitudes comparable to the station corrections used in 'normal' ACH tomography, and there is no need to introduce extra parameters in the inversion. On the other hand, our 'sources' now are stations at the surface of the earth. Even though their locations are (at least in theory) precisely known, near surface velocity heterogeneities and topography effects will affect the absolute travel time residuals at least as much as source location and origin time uncertainties in normal ACH tomography. Thus in the case of using ground truth event information, the ACH tomographic scheme can be fully retained when reversing stations and receivers.

We use two different data sets to invert for the local 3D P-velocity structure of the crust beneath the Shagan River test site. One is the high quality arrival time data set described above (further referred to as C-C data), the other is the P-arrival data for the explosions reported in the ISC catalogs. In an initial stage the data sets are re-sorted such that every recording station is taken as a seismic event (pseudo-event), and all explosions for which arrival times are reported at this station are taken as pseudo-stations with arrival time information for this pseudo-event. Residuals for all travel time observations are then calculated with reference to the Herrin (1968) travel time tables. As relative residuals will be used in the inversion, and because our study area is very small (approx. 30 by 30 km), the choice of the reference earth model has no effect on the inversion results. Figure 6 shows the distribution of relative residuals sorted by nuclear explosions. Relative residuals are obtained by subtracting the mean residual for each pseudo-event from the single residuals for this pseudo-event. The C-C data set contains 783 relative residuals for P-wave arrival times from 78 nuclear explosions determined at 33 global network stations. The ISC data set contains 13147 relative P-arrival time residuals from 99 nuclear explosions observed at 351 stations, which were selected for azimuthal coverage, quality and number of observations. The difference in data quality is apparent from the comparison of Figure s 6a and b. It is to be expected that the use of the ISC data in an inversion for local structure will be problematic, as the magnitude of the travel time perturbation expected from local velocity heterogeneities is around the same or even smaller than the scattering for relative residuals at each pseudo-station.

Figure 7 shows map distributions of relative residuals for four stations at different azimuths for the C-C data set. On top of rather random fluctuations, distinct local patterns are visible in each map view. These patterns can be regarded as 'shadows' of the regional and local 3D heterogeneities when illuminated by wavefronts from different directions. On one hand they show that the residual distribution in the data set contains a considerable amount of structural signal, on the other hand they can be used to validate the 3D inversion results, which should mirror the observed pattern to some extent.

In the following, the pseudo-stations will be called 'stations', and the pseudo-events 'events'. The velocity model is parameterized with blocks of 3 by 3 km horizontal size in the upper 27 km, and 6 by 6 km horizontal size from 27 to 49 km. Vertical block size (layer thickness) increases from 4.6 km at 2 km depth to 5.8 km at the bottom. We use a special first layer with 2 km thickness to model pseudo-station corrections (Evans & Achauer, 1993). For that layer, every pseudo-station is assigned one inversion block, no matter how closely the pseudo-stations are spaced. The layering and the layer velocities approximately follow the models proposed by Walter & Ammon (1993, unpublished) for two locations in the wider area. The Moho in this region is at about 50 km depth (e.g. Antonenko, 1984), so due to the small aperture of our pseudo-station array we will not be able to resolve Moho topography or sub-Moho velocities.

The damping value for the inversion is empirically selected by analyzing the trade-off between final data variance and model length (Figure 8). The very small decrease in data variance for the ISC data set is again an indication that this data set is probably not suitable for 3D inversion. For the C-C data set a damping value of 1 s^2 is chosen. This results in a total residual data variance improvement of 70%, from 0.11 s^2 to 0.03 s^2, after a one-iteration inversion.

Figure 9 shows the distribution of station corrections (as velocity perturbations in the uppermost layer) and the velocity distribution in the underlying layer (2 to 6.6 km) resulting from this first preliminary inversion. As a coarse estimate of reliability, the regions with resolution matrix diagonal elements larger than 0.5 are depicted with white contours in Figure 9b. No further refinement of the inversion or quality analysis has been undertaken so far, so these results should be viewed as preliminary. Nevertheless, the apparent systematic patterns in Figure 9b suggest that detailed local 3D velocity structure can be imaged with adequate quality using the C-C data set. The pattern of relatively low velocity in the central part of the model and higher velocity to the west is consistent with results from the "depth of burial" at the test site (Pearson et al., 1998). Heterogeneity is much less at greater model depths.

The station corrections (Figure 9a) do not allow us to distinguish a regional pattern. They can be interpreted as combined effects of near-station heterogeneities, differences in shaft depths (all set to 0 elevation initially), and remaining origin time errors. The large positive correction close to the model center is most likely due to a larger origin time error still present in the dataset. An experiment with a different set of origin time estimates (based just on the C-C data) result in very different station corrections, but very similar models in deeper layers.

The maximum variance reduction that can be achieved for the ISC data, on the other hand, is around 10%. This leads to a much larger model length than the C-C data result (see Figure 8), in other words, we are modeling noise. The conclusion is that, at least without substantial editing, ISC catalog data cannot be used to invert for small scale crustal structure.

CONCLUSIONS AND RECOMMENDATIONS

Ground truth information is a key to evaluating seismic event location capability. For the Shagan River explosions, where we have excellent ground truth, the high-quality arrival times from cross-correlation analysis yield excellent relative location results, even from extremely sparse stations at relatively close distances. Path corrections vary modestly for different sub-areas. Tomography using the high-quality picks and treating the explosions as a source array yields a model for the uppermost crust that is consistent with results from a recent on-site study. Significant velocity variations are found only for the very shallow crust, consistent with the lack of path correction variability.

REFERENCES

Aki, K., Christofferson, A., Husebye, E.S., 1977. Determination of the three-dimensional seismic structure of the lithosphere, J. Geophys. Res., 82, 277-296.

Antonenko, A. N., 1984. Structure of the Earth's crust in Kazakhstan, Academy of Sciences of the Kazakh Soviet Republic, Alma-Ata. 222p.

Bocharov, V. S., S. A. Zelentsov, and V. N. Mikhailov, 1989. Characteristics of 96 underground nuclear explosions at the Semipalatinsk test site, Atomnaya Energiya, 67.

Dodge, D. A., G. C. Beroza, and W. L. Ellsworth, 1995. Evolution of the 1992 Landers, California foreshock sequence and its implications for earthquake nucleation, J. Geophys. .Res., 100, 9865-9880.

Evans, J.R. & Achauer, U., 1993. Teleseismic Tomography using the ACH method, in Iyer, H.M and Hirahara, K. (eds): Seismic Tomography, Chapman & Hall, London, 319-360.

Herrin, E. (ed), 1968. 1968 seismological tables for P phases, Bull. Seismol. Soc. Am., 58, 1193-1241.

NNCRK, 1999. Technical documentation, contributed by the National Nuclear Centre of the Republic of Kazakhstan, Oslo Workshop on IMS Location Calibration, Jan 1999.

Pearson, D. C., J. L. Bonner, L. A. Glenn, M. D. Denny, and W. S. Phillips, 1998. The source depth of burial experiment at Shagan River, Kazakstan, 20th Annual Seismic Research Symposium Proceedings, 86-95.

VanDecar, J. C., and R. S. Crosson, 1990. Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least squares, Bull. Seismol. Soc. Am., 80, 150-169.

Table 1. Absolute locations for 100 Shagan River nuclear explosions (NNCRK, 1999) and our preliminary origin times and their estimated uncertainties (+/- value, 90% confidence). The first 7 events have fixed origin times obtained from Bocharov et al. (1989).

YRMODAHRMN	SEC	+/-	LATITUD	LONGITU	DEP
650115 6 0	0.80	0.00	49.9350	79.0094	0.0
680619 5 559	80	0.00	49.9802	78.9855	0.0
691130 33259	70	0.00	49.9243	78.9556	0.0
710630 35659	80	0.00	49.9460	78.9804	0.0
720210 5 3	0.00	0.00	50.0243	78.8780	0.0
721102 127	0.20	0.00	49.9270	78.8171	0.0
721210 42710	00	0.00	50.0270	78.9955	0.0
730723 123	0.16	0.03	49.9689	78.8174	0.0
731214 74659	67	0.03	50.0438	78.9856	0.0
740416 55259	84	0.09	50.0244	78.9264	0.0
740531 327	0.02	0.03	49.9606	78.8441	0.0
741016 633	0.14	0.04	49.9875	78.8941	0.0
741227 54659	49	0.04	49.9658	79.0032	0.0
750427 53659	78	0.03	49.9375	78.9035	0.0
750630 327	0.25	0.07	49.9856	78.8969	0.0
751029 44659	89	0.03	49.9539	78.8738	0.0
751225 51659	67	0.03	50.0439	78.8199	0.0
760421 5 259	75	0.04	49.9006	78.8307	0.0
760609 3 259	79	0.04	49.9936	79.0243	0.0
760704 25659	99	0.03	49.9042	78.8993	0.0
760828 257	0.04	0.03	49.9750	78.9262	0.0
761123 5 259	84	0.03	50.0131	78.9432	0.0
761207 45659	98	0.03	49.9439	78.8391	0.0
770529 257	0.13	0.03	49.9465	78.7715	0.0
770629 3 7	0.24	0.04	49.9995	78.8666	0.0
770905 3 259	96	0.03	50.0556	78.9141	0.0
771029 3 7	4.90	0.04	50.0522	78.9802	0.0
771130 4 659	94	0.03	49.9673	78.8743	0.0
780611 257	0.14	0.03	49.9133	78.8018	0.0
780705 247	0.05	0.03	49.9000	78.8665	0.0
780915 23659	97	0.03	50.0061	78.9670	0.0
780829 237	8.85	0.03	49.9283	78.8670	0.0
781104 5 559	86	0.03	50.0418	78.9471	0.0
781129 433	5.03	0.03	49.9533	78.7952	0.0
790201 413	0.17	0.03	50.0808	78.8532	0.0
790623 257	0.11	0.03	49.9148	78.8456	0.0
790707 34659	87	0.03	50.0332	78.9891	0.0
790804 35659	67	0.03	49.9031	78.8876	0.0
790818 25159	71	0.03	49.9481	78.9187	0.0
791028 31659	55	0.03	49.9967	78.9949	0.0
791202 437	0.06	0.03	49.9095	78.7843	0.0
791223 457	0.02	0.03	49.9322	78.7527	0.0
800425 357	0.06	0.03	49.9765	78.7593	0.0
800612 327	0.19	0.03	49.9887	78.9910	0.0
800629 233	0.24	0.03	49.9486	78.8180	0.0
800914 24241	71	0.03	49.9367	78.7974	0.0
801012 33416	65	0.03	49.9675	79.0224	0.0
801214 347	9.01	0.03	49.9089	78.9184	0.0
801227 4 910	63	0.03	50.0619	78.9752	0.0
810329 4 352	58	0.03	50.0182	78.9787	0.0
810422 11713	91	0.03	49.8989	78.8084	0.0
810527 35814	88	0.03	49.9869	78.9705	0.0
810913 21720	84	0.03	49.9133	78.8942	0.0
811018 357	5.22	0.03	49.9281	78.8445	0.0
811129 33511	20	0.03	49.9019	78.8488	0.0
811227 34316	72	0.03	49.9332	78.7781	0.0
820425 323	7.99	0.03	49.9170	78.8876	0.0
820704 11716	68	0.03	49.9587	78.8115	0.0
820831 131	3.19	0.03	49.9142	78.7613	0.0
821205 33715	12	0.03	49.9309	78.8095	0.0
821226 33516	67	0.03	50.0631	78.9938	0.0
830612 23646	12	0.03	49.9250	78.8979	0.0
831006 147	9.17	0.03	49.9247	78.7505	0.0
831026 155	7.42	0.03	49.9126	78.8215	0.0
831120 327	6.90	0.03	50.0508	78.9991	0.0
840219 357	5.92	0.03	49.8962	78.7429	0.0
840307 239	8.86	0.03	50.0501	78.9560	0.0
840329 51910	75	0.03	49.9112	78.9267	0.0
840425 1 9	6.09	0.03	49.9359	78.8503	0.0
840526 31314	92	0.03	49.9790	79.0054	0.0
840714 1 913	08	0.03	49.9095	78.8769	0.0
841027 15012	52	0.03	49.9347	78.9279	0.0
841202 319	8.94	0.03	50.0062	79.0087	0.0
841216 355	5.19	0.03	49.9458	78.8084	0.0
841228 35013	15	0.03	49.8803	78.7037	0.0
850210 32710	07	0.03	49.8993	78.7804	0.0
850425 057	9.10	0.03	49.9268	78.8805	0.0
850615 057	3.21	0.02	49.9086	78.8426	0.0
850630 239	5.13	0.03	49.8644	78.6685	0.0
850720 05316	98	0.03	49.9498	78.7837	0.0
870312 15719	63	0.03	49.9353	78.8287	0.0
870403 11710	36	0.02	49.9181	78.7801	0.0
870417 1 3	7.14	0.03	49.8779	78.6687	0.0
870620 053	7.16	0.02	49.9353	78.7440	0.0
870802 058	9.27	0.03	49.8806	78.8745	0.0
871115 331	9.17	0.03	49.8987	78.7579	0.0
871213 321	7.25	0.03	49.9632	78.7929	0.0
871227 3 5	7.22	0.03	49.8795	78.7248	0.0
880213 3 5	8.28	0.03	49.9367	78.8637	0.0
880403 133	8.21	0.02	49.9083	78.9081	0.0
880504 057	9.15	0.02	49.9495	78.7501	0.0
880614 227	9.00	0.03	50.0189	78.9605	0.0
880914 35959	77	0.02	49.8779	78.8229	0.0
881112 330	6.27	0.03	50.0431	78.9688	0.0
881217 418	9.24	0.03	49.8820	78.9245	0.0
890122 357	9.06	0.02	49.9395	78.8191	0.0
890212 415	9.26	0.03	49.9187	78.7109	0.0
890708 347	0.09	0.03	49.8679	78.7801	0.0
890902 41659	85	0.04	50.0058	78.9855	0.0
891019 94959	90	0.03	49.9223	78.9081	0.0

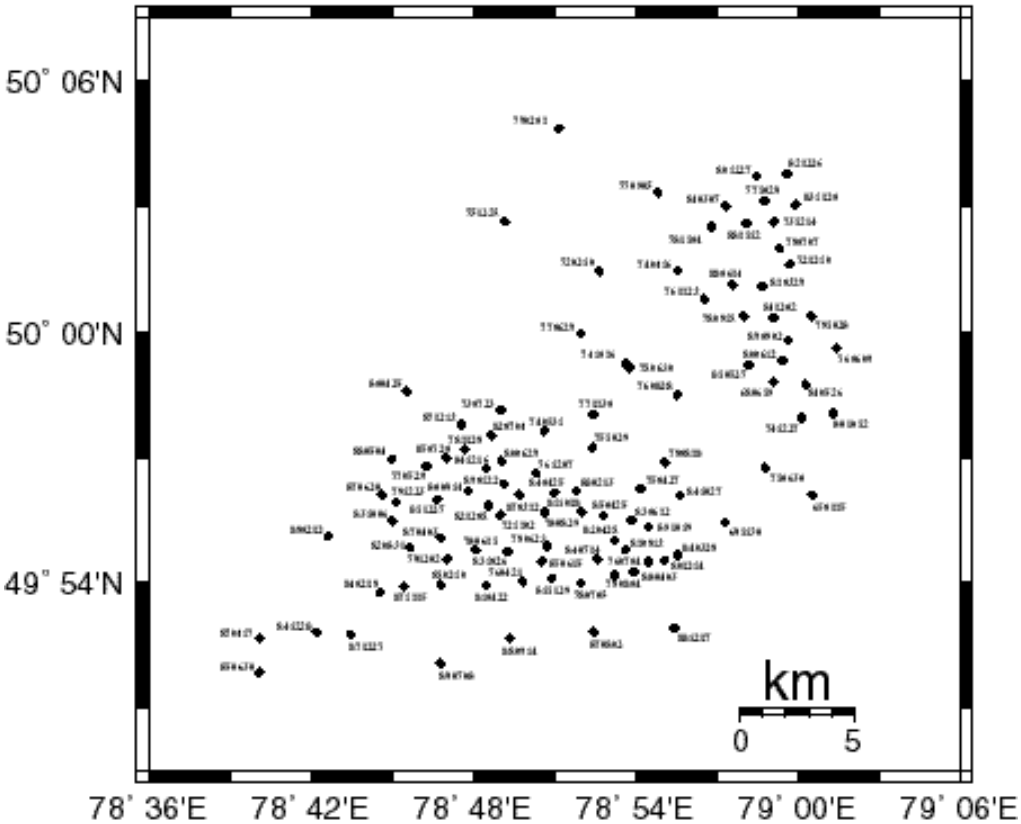

Figure 1. Map of absolute locations of 100 Shagan River nuclear explosions (NNCRK, 1999) used in our analysis.

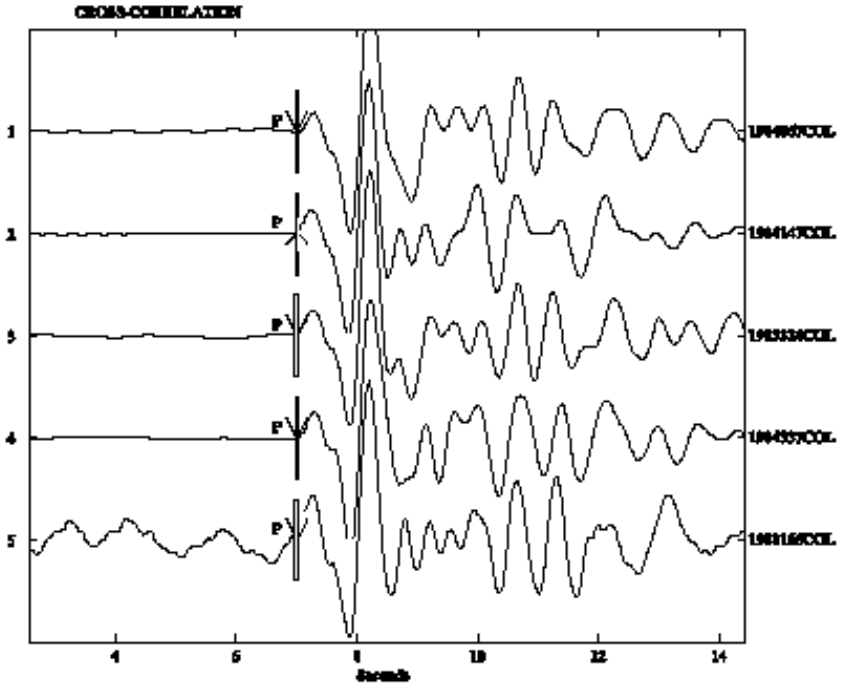

Figure 2. Illustration of cross-correlation arrival time estimation - note alignment of noisy trace at bottom.

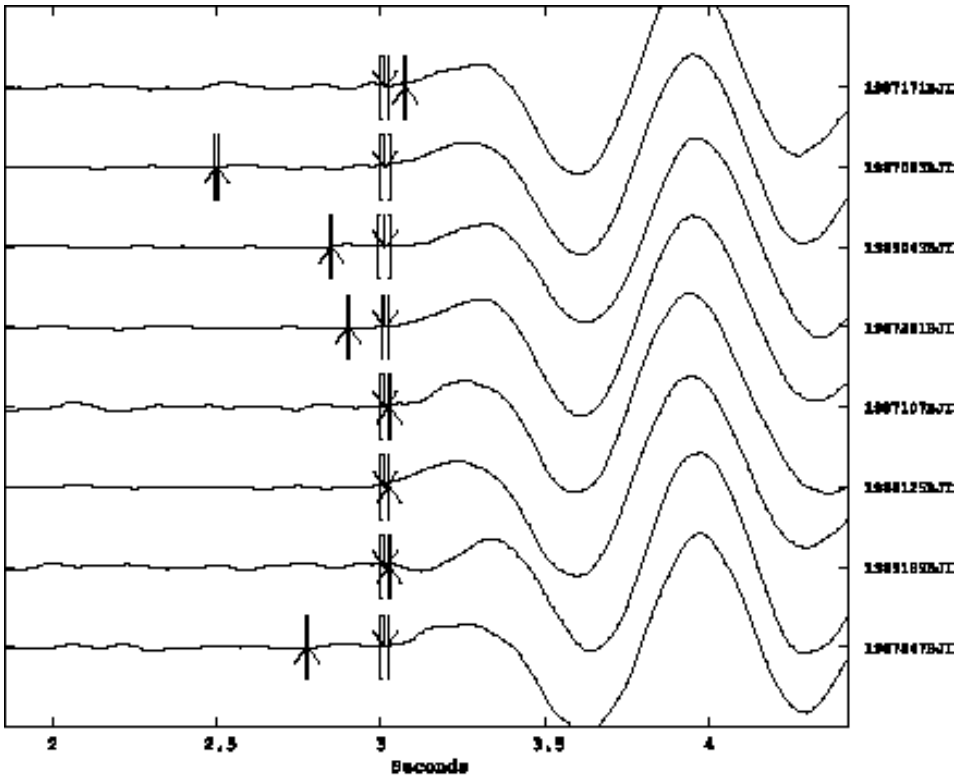

Figure 3. Comparison of arrival times reported by ISC (upward-pointing arrows) to picks determined using cross-correlation analysis (downward-pointing arrows). The differences are due to round-off of the ISC data in this case.

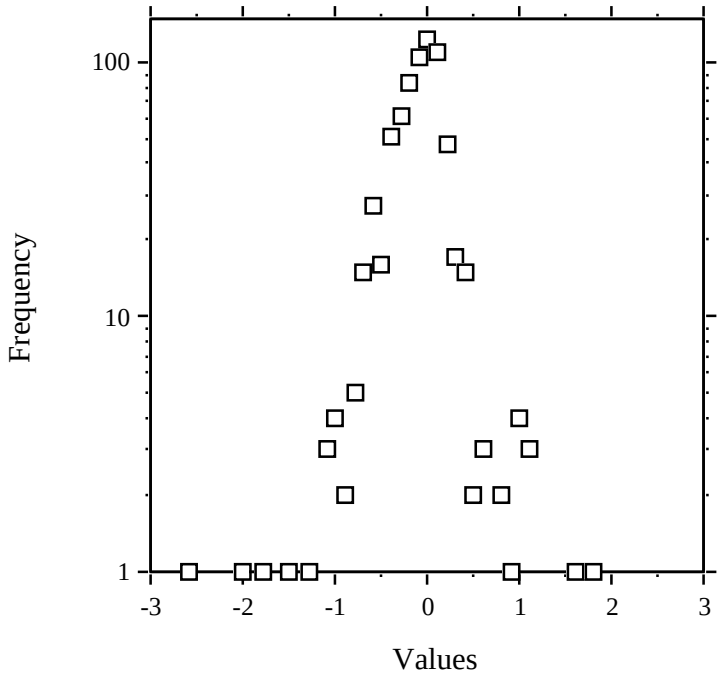

Figure 4. Frequency plot of the differences between ISC and cross-correlation arrival times (values in seconds, in 0.1 s bins). Note the logarithmic frequency scale. Most residuals are with +/- 0.5 s. Some outliers are not shown.

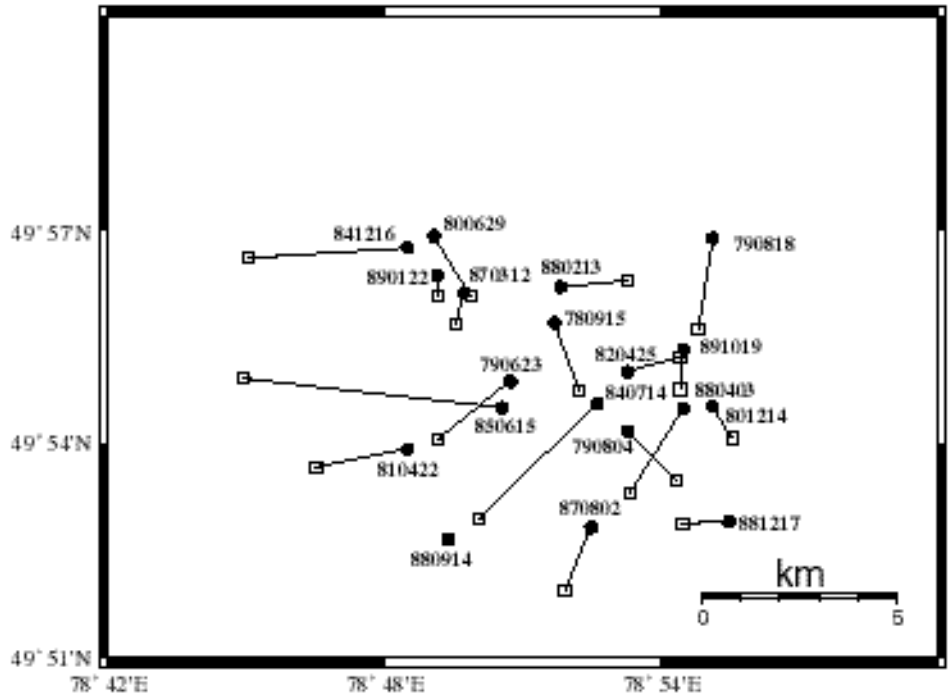

Figure 5. Mislocations (open squares) for explosions from the central Shagan River group using only stations within 40°. The mean mislocation distance is 2.2 km and the mean 90% confidence ellipse area is 8.6 km.

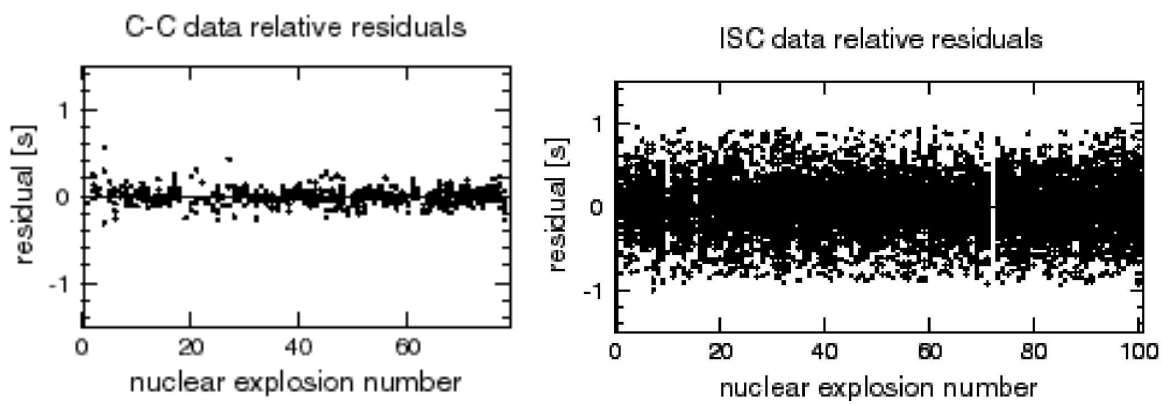

Figure 6. Relative residuals for each nuclear explosion (pseudo-station). Left: repicked (C-C) dataset. Right: ISC catalog dataset. Note the small scatter and apparent systematic variations for the C-C data versus the wide scatter with no apparent trend for the ISC data.

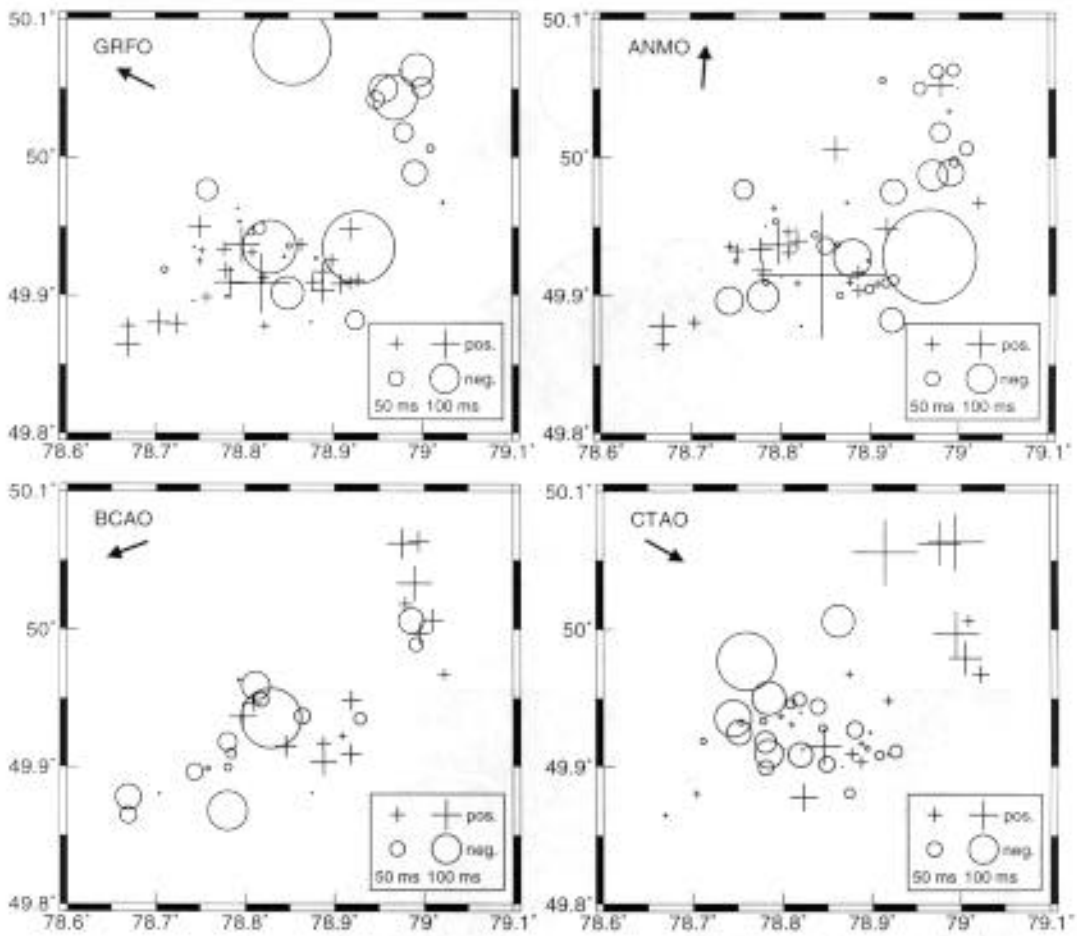

Figure 7. Map views of relative residuals for recording stations (pseudo-events) at different azimuths. Crosses denote negative residuals (i.e. travel times are shorter than average), open circles denote positive residuals (longer travel times). Station name and azimuth to station are given in the upper left corner of each map.

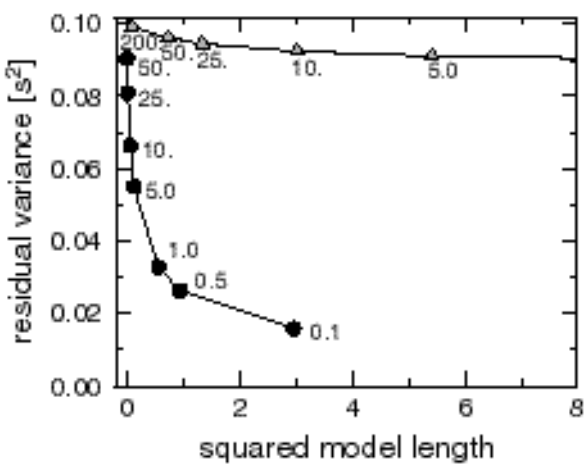

Figure 8. Damping test for the 3D velocity inversion; full circles denote the C-C data, grey triangles the ISC data. A damping of 1.0 was selected for the inversion of the C-C data from this test. The small decrease in residual variance together with the large increase in solution length for the ISC data indicates that this data set may not be adequate for the inversion.

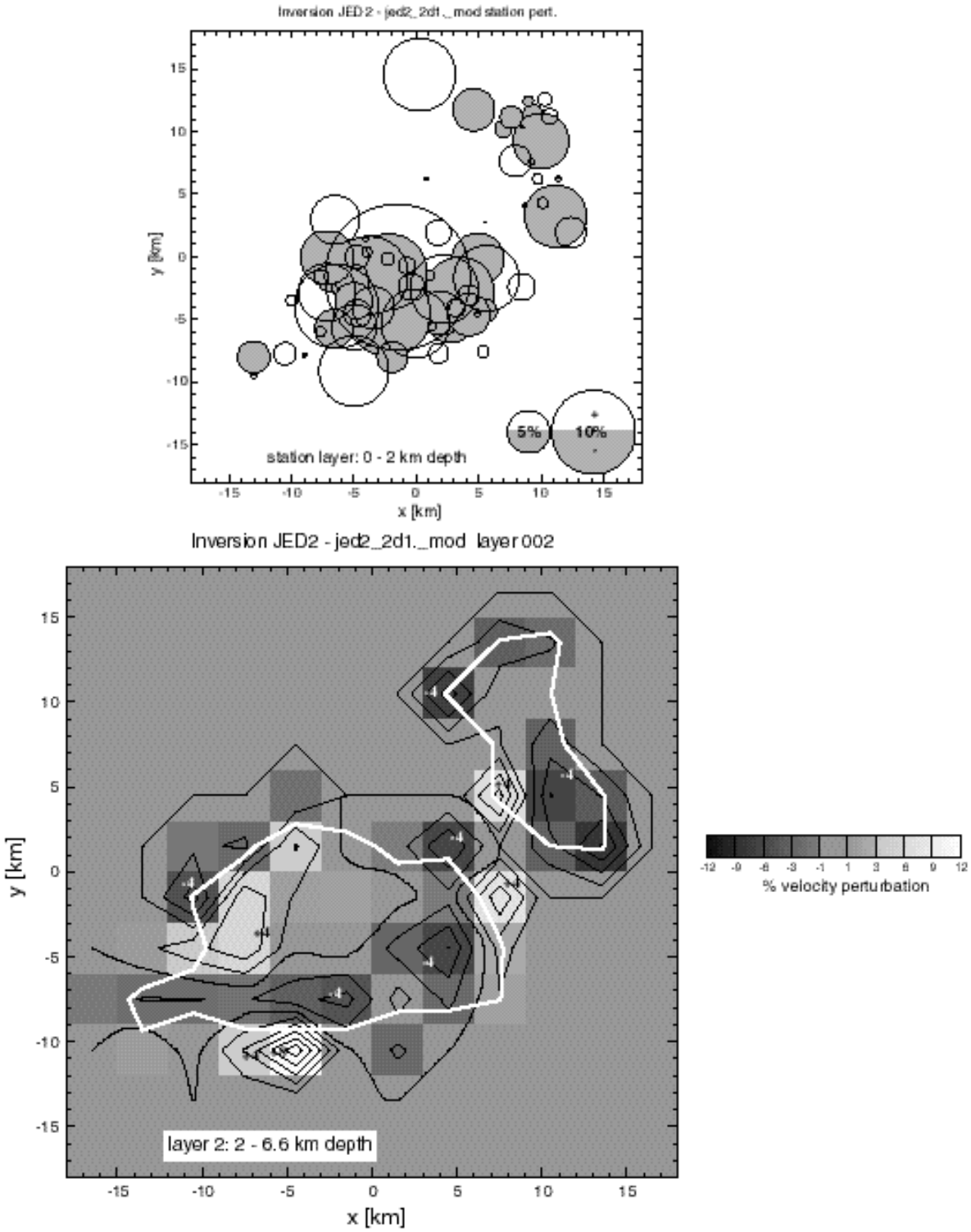

Figure 9. Preliminary 3D tomography results for the C-C dataset. Model center (0, 0) is at 49 57' N, 78 51' E. (a) Top panel: station perturbations (in %). Open circles are positive (faster), filled are negative (slower). (b) Bottom panel: layer 2 (2 - 6.6 km depth) velocity perturbations, shown as shaded blocks and smoothed contours (2% contours). White contouring (smoothed) indicates model areas with diagonal resolution element above 0.5.