The effect of distribution of stations upon location error: Statistical tests based on the double-difference earthquake location algorithm and the bootstrap method

Ling Bai[1], Zhongliang Wu[2], Tianzhong Zhang[2], and Ichiro Kawasaki[1]

[1]*Disaster Prevention Research Institute, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan*
[2]*Institute of Geophysics, China Earthquake Administration, Beijing 100081, China*

In this letter, we investigate the effect of station distribution (including the number and azimuthal gap of stations) upon location error based on the field data observed at Northern California Seismic Network (NCSN) using a double-difference earthquake location algorithm and a bootstrap method. The earthquakes relocated by all 117 stations are set as reference and the error of location is defined as the RMS of the difference to the reference. We find that the location error has a nonlinear relationship with the distribution of stations. The results may be used as guidelines for building seismic network.
Key words: Number of stations, azimuthal gap of stations, location error, double-difference earthquake location algorithm, bootstrap method.

1. Introduction

The effect of the distribution of seismic stations upon the location error and optimizing seismic network has long been an important issue in seismology. To study the focal mechanism, strong ground motion and seismic disaster assessment, and to make accurate predictions on the seismic tendency after strong earthquakes, it is required to set up a dense seismic network to observe aftershocks. The most accurate hypocenter can be obtained for events inside dense, local networks. The portable digital seismographs are generally combined to improve detective capability of the current permanent network. However, it is unrealistic to cost-effectively set up many portable seismographs close to the hypocenter in a short period of time. Therefore, as an optimization problem in engineering, the location error needs to be estimated with a given number of stations and their azimuthal gap (the largest open azimuth between recording stations) beforehand.

In recent years, there has been considerable interest among seismologists in assessing the potential resolution capability of seismic networks. Kijko (1977) suggested applying ideas from optimal experimental design to the problem of seismographic configuration. Rabinowitz and Steinberg (1990) studied optimal configuration of networks using a statistic approach. Hardt and Scherbaum (1994) designed optimum networks for observing aftershocks using simulated annealing. Bartal *et al.* (2000) simulated error ellipse by using genetic algorithm and differential evolution technique. Bondár *et al.* (2004) established location accuracy criteria based on bulletin data and Monte Carlo

simulations of network geometry. This letter studies the effect of distribution of stations upon the location error using the commonly used double-difference earthquake location algorithm (Waldhauser and Ellsworth, 2000) and the bootstrap method (Efron, 1979, 1982; Marcello *et al.*, 2001). Contrast to the above studies that are purely based on synthetic data, our study uses real seismic data observed at Northern California Seismic Network (NCSN).

2. Method, Data and Statistical Tests
2.1 Method

The double-difference algorithm is an efficient method to precisely determine relative hypocenter locations. It minimizes residuals between observed and theoretical travel-time differences for pairs of earthquakes observed at a common station. When the data recorded by nearby stations are available, the algorithm gives accurate relative locations of earthquakes, which is quite different from other relative location methods (Zhang *et al.*, 2005). A remarkable feature of the double-difference algorithm is the distance-weighting factor, which can exclude data from event pairs far apart. However, this weighting factor fails to do so in some instances and the data far apart can still be correlated in the inversion via a series of intermediate pairs (Zhang and Thurber, 2003).

The bootstrap method is used to estimate generalization error based on resampling. Our attention is focused on questions concerning the probability distribution of stations. The distribution of stations is stochastic for a certain earthquake. To obtain reliable standard errors, it is necessary to resample from the original data to create replicate datasets. The bootstrap method can be employed to implement resampling in nonparametric or in parametric mode.
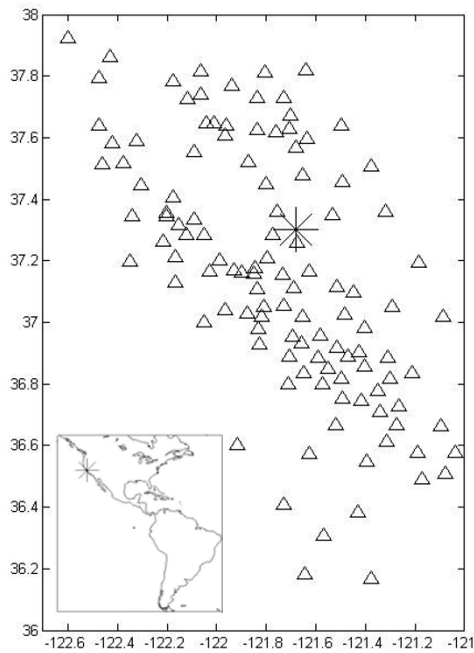
Fig. 1. Stations of NCSN (open triangles) used as population from which the statistical resamplings are taken. The asterisk indicates the location of earthquake sequence, which is also shown in the map inset in the lower left.

2.2 Data

There were 308 earthquakes recorded from 117 stations of the NCSN on the Northern Hayward fault between 1984 and 1997 (Fig. 1). The data consisted of both cross-correlation and catalog *P*- and *S*-wave. Stations were distributed within about 100 km from epicenters. The resolution of relocations using all 117 stations is about 0.21 km in the north-south direction, 0.23 km in the east-west direction, and 1.20 km in depth.

2.3 Experiments

Two statistical tests were implemented in this study: the first one focuses on the effect of the number of stations and

the second one deals with the effect of azimuthal gap of stations.

For the first test, 50–250 iterations may be needed to obtain a stable result (Efron, 1982), if the population is rather large. In our implementation, we use 100 sampling iterations for each test, which shows a stable RMS (root mean square) residual. We gradually decrease the number of stations from a total number of 117 down to 4 and observe the change of the location error. Figure 2(a) shows a subset of 40 stations.

In the second test, there are 12 subsets in the angle of 360° if we sample with an interval of 30°. According to the ergodic hypothesis in statistics, we use 12 sampling iterations for each test since the azimuthal gaps we discussed in this study are larger than or equal to 30°. Figure 2(b) shows a subset with an azimuthal gap of 100°.

3. Results

3.1 The criteria to assess the location error

We use two kinds of criteria to assess the location error: one is the number of earthquakes that can be relocated with a small RMS residual. Another criterion is the errors in longitude, latitude and depth of relocated earthquakes. The earthquakes relocated by all 117 stations are set as reference and the error of location is defined as the RMS of the difference to the reference.

3.2 The effect of the number of stations

This study shows that the location error decreases with increasing number of stations. The randomly sampling is repeated 100 times, and the cumulative results for 308 events (100∗308 samples) are presented in Fig. 3. Ellipses are derived from their distribution and contain 95% of the points. When the number of stations increases to 15, the number of relocated earthquakes is about 290 in average, which approaches 95% of the total number of 308 events to be relocated (Fig. 4(a)), and the resolution were improved to less than 0.5 km for epicenters and 1.0 km for depths (Fig. 4(b), (c) and (d)). The log-log and semi-log plots of the results indicate that the location error can be expressed
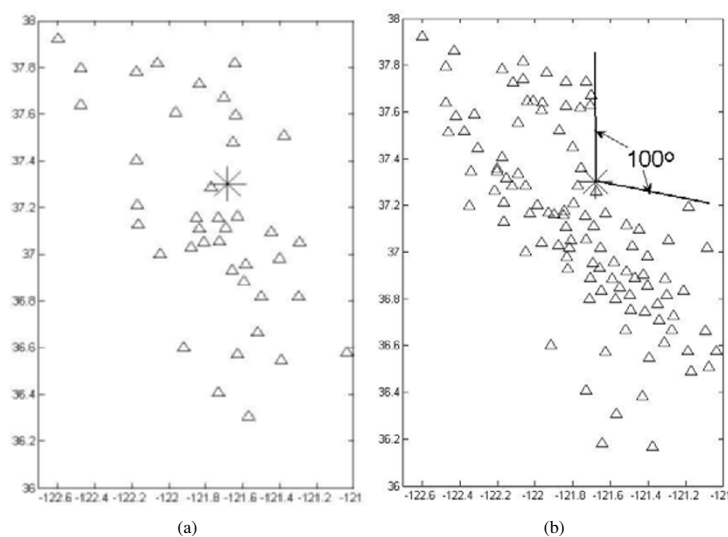

Fig. 2. (a) A subset of 40 stations. (b) A subset with an azimuthal gap of 100°. The asterisk indicates the location of earthquake sequence. Open triangles denote the stations of Northern California Seismic Network.
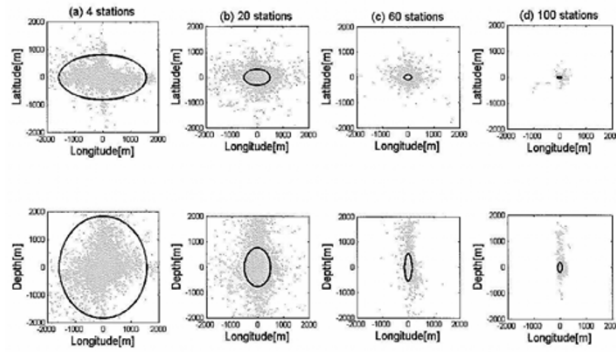
Fig. 3. Bootstrap analysis of the RMS of the difference to the reference with station's number of (a) 4, (b) 20, (c) 60 and (d) 100. Top panels show map view and lower panels show E-W cross section for all bootstrap samples (100∗308 samples). Ellipses contain 95% of the points.
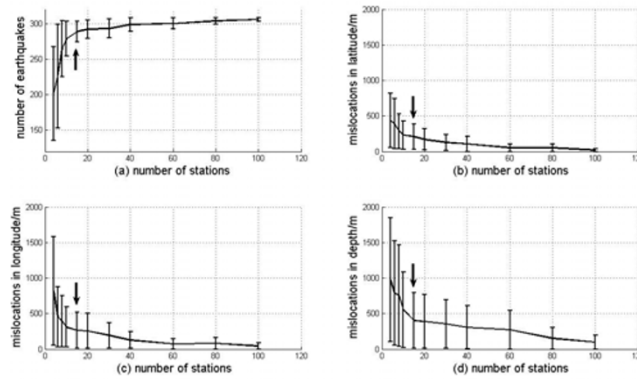

Fig. 4. Mislocation versus number of stations. In each case the median and spread with a percentage confidence of 95 are plotted. Curves linked median values. The arrows indicate the critical points. The horizontal axis is the number of stations. The vertical axis is (a) the number of earthquakes that can be relocated, and errors (b) in the N-S direction, (c) in the E-W direction and (d) in depth relative to reference in meter, respectively.
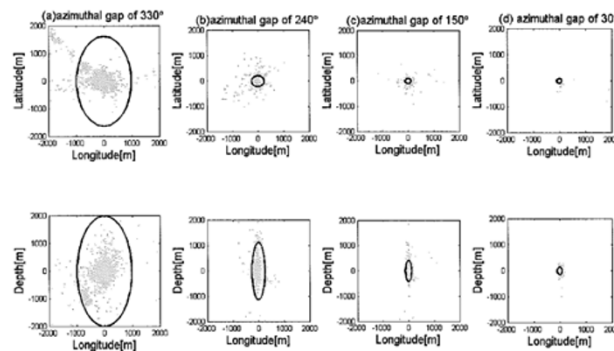

Fig. 5. Bootstrap analysis of the RMS of the difference to the reference with azimuthal gap of (a) 330°, (b) 240°, (c) 150° and (d) 30°. Top panels show map view and lower panels show E-W cross section for all bootstrap samples (12∗308 samples).

in a power law form when the number of stations is small, and the errors stay almost constant when the number of stations is larger than the critical point of 15.

3.3   The effect of azimuthal gap of station

Figures 5 and 6 show the relationship between the azimuthal gap and the location error. The location errors decrease with decreasing azimuthal gap. The randomly sampling is repeated 12 times, and the cumulative results (12∗308 samples) are presented in Fig. 5. Ellipses contain 95% of the points as in Fig. 3. When the azimuthal gap decreases to 210°, the number of relocated earthquakes is about 300 (Fig. 6(a)), and the resolution becomes stable in the horizontal direction and less than 1.0 km in depth

(Fig. 6(b), (c) and (d)). In log-log and semi-log plots, the location errors can be expressed as a power law form of the azimuthal gap when the azimuthal gap is large, and the errors tend to be stable when the azimuthal gap is smaller than critical point of 210°.

The stations used in this study are distributed in the NNW direction along the coastal line, which might cause the location error in longitude larger than that in latitude. The location error in depth is rather large due to the fact that its resolution depends not only on the distribution of stations, but also on the velocity model and the type of seismic phase to be used.
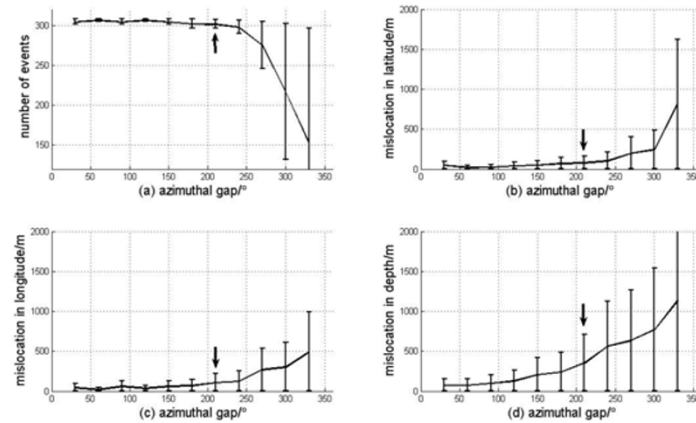
Fig. 6. Mislocation versus the azimuthal gap of stations. The horizontal axis is the azimuthal gap. Other symbols are the same as in Fig. 3.

4. Concluding Remarks

In this letter, we discussed the effect of the distribution of stations upon location error using the data observed at Northern California Seismic Network (NCSN) and the methods of double-difference and bootstrap. The earthquakes relocated by all 117 stations are set as reference and the error of location is defined as the RMS of the difference to the reference. One of the interesting results is that the location error has nonlinear relationship with the distribution of stations. If we take the status parameters of station distribution (number and azimuthal gap of stations) as independent variables, and the number of earthquakes that can be relocated with a given small RMS residual, the relative mislocations in latitude, longitude and depth as dependent variables, the results can be expressed in a power law form between the independent variables and the dependent variables. Figures 4 and 6 show a critical point in the above relationship. To the one side of the critical point, i.e., the number of stations is large or the azimuthal gap of stations is small, the dependent variables tend to be stable, and to the other side of the critical point the curves show power law relationship.

Bondár *et al.* (2004) created a 'ground truth' $GTX_{C\ percent}$ classification that uses X to designate location accuracy in kilometer (the true epicenter lies within 'X' km of the estimated epicenter) and the subscript "C percent" to designate the percentage confidence. We modified the $GTX_{C\ percent}$ classification to $REX_{C\ percent}$, which represents the reference error (the epicenter lies with in 'X' km of the epicenter of reference). The confidence interval in Figs. 4 and 6 can be expressed as $REX_{95\ percent}$. We found that the relative hypocenter error, including focal depth can reach $RE1_{95\ percent}$ and the relative epicenter error can reach $RE0.5_{95\ percent}$ if the distribution of stations satisfies the following conditions:

(1) There are at least 15 ± 2 stations within $100\,km$ around the epicenter;

(2) The azimuthal gap is small than $210°\pm15°$.

The accuracy will not improve much with increasing number of stations or decreasing their azimuthal gap as long as the above conditions are satisfied. In fact, the problem of earthquake location can be explained with the percolation theory (Zallen, 1978) and the above results are comprehensible with this theory.

References

Bartal, Y., Z. Somer, G. Leonard, D. M. Steinberg, and Y. B. Horin, Optimal seismic network in Israel in the context of the Comprehensive Test Ban Treaty, *Bull. Seism. Soc. Am.*, **90**, 151–165, 2000.

Bondár, I., S. C. Myers, E. R. Engdual, and E. A. Bergman, Epicentre accuracy based on seismic network criteria, *Geophys. J. Int.*, **156**, 483–496, 2004.

Efron, B., Bootstrap methods: another look at the jackknife, *Annals of Statistics*, **7**, 1–26, 1979.

Efron, B., *The Jackknife, the Bootsreap, and Other Resampling Plans*, Philadelphia: SIAM, 92 pp, 1982.

Hardt, M. and F. Scherbaum, The design of optimum network for aftershock recordings, *Geophys. J. Int.*, **117**, 716–726, 1994.

Kijko, A., An algorithm for the optimum distribution of a regional seismic network, *Pure App. Geophys.*, **115**, 999–1009, 1977.

Marcello, V., A. Dario, and M. Enzo, Classificatin of seismic strain estimates in the Mediterranean region from a 'bootstrap' approach, *Geophys. J. Int.*, **146**, 399–415, 2001.

Rabinowitz, N. and D. M. Steinberg, Optimal configuration of a seismographic network: a statistical approach, *Bull. Seism. Soc. Am.*, **80**, 187–196, 1990.

Waldhauser, F. and W. L. Ellsworth, A double-difference earthquake location algorithm: method and application to the Northern Hayward fault, California, *Bull. Seism. Soc. Am.*, **90**, 1353–1368, 2000.

Zallen, R. Overview and application of percolation theory, *Ann. Israel Phys. Soc.*, **2**, 309–327, 1978.

Zhang, H. J. and C. H. Thurber, Double-difference tomography: the method and its application to the Hayward fault, California, *Bull. Seism. Soc. Am.*, **93**, 1875–1889, 2003.

Zhang, T. Z., Y. Huang, and B. Wu, The effect of data recorded at nearby stations on relative location algorithm, Abstract of AGU fall meeting, S51B-1003, 2005.

L. Bai (e-mail: bai@rcep.dpri.kyoto-u.ac.jp), Z. Wu, T. Zhang, and I. Kawasaki