

A Double-Difference Earthquake Location Algorithm: Method and Application to the Northern Hayward Fault, California

by Felix Waldhauser and William L. Ellsworth

Abstract We have developed an efficient method to determine high-resolution hypocenter locations over large distances. The location method incorporates ordinary absolute travel-time measurements and/or cross-correlation *P*- and *S*-wave differential travel-time measurements. Residuals between observed and theoretical travel-time differences (or double-differences) are minimized for pairs of earthquakes at each station while linking together all observed event-station pairs. A least-squares solution is found by iteratively adjusting the vector difference between hypocentral pairs. The double-difference algorithm minimizes errors due to unmodeled velocity structure without the use of station corrections. Because catalog and cross-correlation data are combined into one system of equations, interevent distances within multiplets are determined to the accuracy of the cross-correlation data, while the relative locations between multiplets and uncorrelated events are simultaneously determined to the accuracy of the absolute travel-time data. Statistical resampling methods are used to estimate data accuracy and location errors. Uncertainties in double-difference locations are improved by more than an order of magnitude compared to catalog locations. The algorithm is tested, and its performance is demonstrated on two clusters of earthquakes located on the northern Hayward fault, California. There it collapses the diffuse catalog locations into sharp images of seismicity and reveals horizontal lineations of hypocenters that define the narrow regions on the fault where stress is released by brittle failure.

Introduction

Seismicity analysis for the study of tectonic processes, earthquake recurrence, and earthquake interaction requires knowledge of the precise spatial offset between the earthquake hypocenters. This is particularly the case for crustal faults that are most readily investigated using microseismic activity. The location uncertainty of routinely determined hypocenters is typically many times larger than the source dimension of the events itself, thus putting limits on the study of the fine structure of seismicity.

The accuracy of absolute hypocenter locations is controlled by several factors, including the network geometry, available phases, arrival-time reading accuracy, and knowledge of the crustal structure (Pavlis, 1986; Gomberg *et al.*, 1990). The use of a one-dimensional reference velocity model to locate the earthquakes limits the location accuracy, since three-dimensional velocity variations can introduce systematic biases into the estimated travel times. One can partially account for the velocity variations by including station and/or source terms in the location procedure (e.g., Douglas, 1967; Pujol, 1988; Hurukawa and Imoto, 1992; Shearer, 1997) and/or by jointly inverting the travel-time data for hypocenters and velocity structure (e.g., Crosson,

1976; Ellsworth, 1977; Roecker, 1981; Thurber, 1983; Michael, 1988; Kissling *et al.*, 1994).

The effects of errors in structure can also be effectively minimized by using relative earthquake location methods (Poupinet *et al.*, 1984; Fréchet, 1985; Frémont and Malone, 1987; Got *et al.*, 1994) (for a discussion on relative location errors see Pavlis [1992]). If the hypocentral separation between two earthquakes is small compared to the event-station distance and the scale length of the velocity heterogeneity, then the ray paths between the source region and a common station are similar along almost the entire ray path. In this case, the difference in travel times for two events observed at one station can be attributed to the spatial offset between the events with high accuracy. This is because the absolute errors are of common origin except in the small region where the raypaths differ at the sources.

We can further improve location precision by improving the accuracy of the relative arrival-time readings using waveform cross-correlation methods. Two earthquakes produce similar waveforms at a common station if their source mechanisms are virtually identical and their sources are collocated so that signal scattering due to velocity heterogeneity

ties along the ray paths is small. A useful rule of thumb for maximum event separation, after which local earthquake signals become incoherent, is about 1/4 wavelength of the highest frequency of importance in the seismogram, which is related to the first Fresnel zone (Geller and Mueller, 1980). The similarity in the waveforms can be exploited either in the frequency domain (Poupinet *et al.*, 1984) or in the time domain (Deichmann and Garcia-Fernandez, 1992) to obtain highly accurate time shifts. With a relative timing precision of about 1 msec (compared to about 10–30 msec for routinely picked phase onset), such differential travel-time data allow one to calculate the relative location between earthquakes with errors of only a few meters to a few tens of meters.

Two techniques are commonly used to obtain relative locations from differential travel-time measurements. In the master event approach, each event is relocated relative to only one event, the master event (Ito, 1985; Scherbaum and Wendler, 1986; Frémont and Malone, 1987; Van Decar and Crosson, 1990; Deichmann and Garcia-Fernandez, 1992; Lees, 1998). As the linking occurs only through one event, errors due to the correlation of noise in the master event may propagate through the entire cluster and effect the location of all other events. This approach also puts limits on the maximum spatial extension of the cluster that can be relocated, as all events must correlate with the master event.

Got *et al.* (1994) overcame these restrictions by determining cross-correlation time delays for all possible event pairs and combining them in a system of linear equations that is solved by least-squares methods to determine hypocenter separations (see also Fréchet, 1985). For simplicity, a constant slowness vector was used for each station from all sources. Because only cross-correlation data is considered, this approach cannot relocate uncorrelated clusters relative to each other.

Dodge *et al.* (1995) converted cross-correlation delay times derived from a master event approach (Van Decar and Crosson, 1990) to absolute times by assigning them to ordinary time picks. Using a joint hypocenter determination (JHD) method, the travel times are then inverted to simultaneously estimate hypocentral parameters, velocity model corrections, and station corrections. This approach allows groups of correlatable events with accurate interevent distances to move relative to one another by weaker constraints. In order to retain the accuracy of the cross-correlation data, it is assumed that the events are clustered in a small volume so that the unmodeled velocity structure can be completely absorbed in the station corrections.

In this article, we present an efficient earthquake relocation technique that allows the simultaneous relocation of large numbers of earthquakes over large distances. We combine *P*- and *S*-wave differential travel times derived from cross-spectral methods with travel-time differences formed from catalog data and minimize residual differences (or double differences) for pairs of earthquakes by adjusting the vector difference between their hypocenters. Thus we are

able to determine interevent distances between correlated events that form a single multiplet to the accuracy of the cross-correlation data while simultaneously determining the relative locations of other multiplets and uncorrelated events to the accuracy of the absolute travel-time data, without the use of station corrections. This article describes the double-difference algorithm and applies it to earthquakes on the northern Hayward fault to illustrate the performance of the method and to develop the error model. The relocation results were recently discussed in Waldhauser *et al.* (1999).

Double-Difference Algorithm for Earthquake Location

The arrival time, T , for an earthquake, i , to a seismic station, k , is expressed using ray theory as a path integral along the ray,

$$T_k^i = \tau^i + \int_i^k u ds, \quad (1)$$

where τ is the origin time of event i , u is the slowness field, and ds is an element of path length. Due to the nonlinear relationship between travel time and event location, a truncated Taylor series expansion (Geiger, 1910) is generally used to linearize equation (1). The resulting problem then is one in which the travel-time residuals, r , for an event i are linearly related to perturbations, $\Delta\mathbf{m}$, to the four current hypocentral parameters for each observation k :

$$\frac{\partial t_k^i}{\partial \mathbf{m}} \Delta\mathbf{m}^i = r_k^i, \quad (2)$$

where $r_k^i = (t_k^{\text{obs}} - t_k^{\text{cal}})^i$, t_k^{obs} and t_k^{cal} are the observed and theoretical travel time, respectively, and $\Delta\mathbf{m}^i = (\Delta x^i, \Delta y^i, \Delta z^i, \Delta \tau^i)$.

Equation (2) is appropriate for use with measured arrival times. However, cross-correlation methods measure travel-time differences between events, $(t_k^i - t_k^j)^{\text{obs}}$, and as a consequence, equation (2) can not be used directly. Fréchet (1985) obtained an equation for the relative hypocentral parameters between two events i and j by taking the difference between equation (2) for a pair of events,

$$\frac{\partial t_k^{ij}}{\partial \mathbf{m}} \Delta\mathbf{m}^{ij} = dr_k^{ij}, \quad (3)$$

where $\Delta\mathbf{m}^{ij} = (\Delta dx^{ij}, \Delta dy^{ij}, \Delta dz^{ij}, \Delta d\tau^{ij})$ is the change in the relative hypocentral parameters between the two events, and the partial derivatives of t with respect to \mathbf{m} are the components of the slowness vector of the ray connecting the source and receiver measured at the source (e.g., Aki and Richards, 1980). Note that in equation (3) the source is actually the centroid of the two hypocenters, assuming a constant slowness vector for the two events. dr_k^{ij} in equation (3)

is the residual between observed and calculated differential travel time between the two events defined as

$$dr_k^{ij} = (t_k^i - t_k^j)^{\text{obs}} - (t_k^i - t_k^j)^{\text{cal}}. \quad (4)$$

We define equation (4) as the double-difference. Note that equation (4) may use either phases with measured arrival times where the observables are absolute travel times, t , or cross-correlation relative travel-time differences.

The assumption of a constant slowness vector is valid for events that are sufficiently close together, but breaks down in the case where the events are farther apart. A generally valid equation for the change in hypocentral distance between two events i and j is obtained by taking the difference between equation (2) and using the appropriate slowness vector and origin time term for each event (see Figure 1):

$$\frac{\partial t_k^i}{\partial \mathbf{m}} \Delta \mathbf{m}^i - \frac{\partial t_k^j}{\partial \mathbf{m}} \Delta \mathbf{m}^j = dr_k^{ij}, \quad (5)$$

or written out in full

$$\begin{aligned} \frac{\partial t_k^i}{\partial x} \Delta x^i + \frac{\partial t_k^i}{\partial y} \Delta y^i + \frac{\partial t_k^i}{\partial z} \Delta z^i + \Delta \tau^i - \frac{\partial t_k^j}{\partial x} \Delta x^j \\ - \frac{\partial t_k^j}{\partial y} \Delta y^j - \frac{\partial t_k^j}{\partial z} \Delta z^j - \Delta \tau^j = dr_k^{ij}. \end{aligned} \quad (6)$$

The partial derivatives of the travel times, t , for events i and j , with respect to their locations (x , y , z) and origin times (τ), respectively, are calculated for the current hypocenters and the location of the station where the k th phase was recorded. Δx , Δy , Δz , and $\Delta \tau$ are the changes required in the hypocentral parameters to make the model better fit the data.

We combine equation (6) from all hypocentral pairs for a station, and for all stations to form a system of linear equations of the form

$$\mathbf{W}\mathbf{G}\mathbf{m} = \mathbf{W}\mathbf{d}, \quad (7)$$

where \mathbf{G} defines a matrix of size $M \times 4N$ (M , number of double-difference observations; N , number of events) containing the partial derivatives, \mathbf{d} is the data vector containing the double-differences (4), \mathbf{m} is a vector of length $4N$, $[\Delta x, \Delta y, \Delta z, \Delta T]^T$, containing the changes in hypocentral parameters we wish to determine, and \mathbf{W} is a diagonal matrix to weight each equation. We may constrain the mean shift of all earthquakes during relocation to zero by extending (7) by four equations so that

$$\sum_{i=1}^N \Delta \mathbf{m}_i = 0 \quad (8)$$

for each coordinate direction and origin time, respectively. Note that this is a crude way to apply a constraint, but ap-

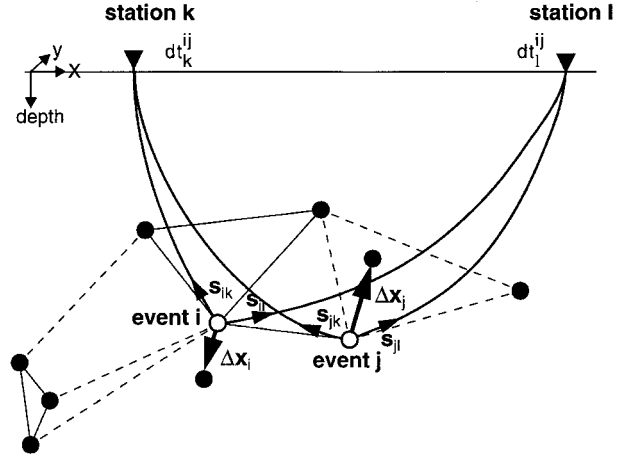


Figure 1. Illustration of the double-difference earthquake relocation algorithm. Solid and open circles represent trial hypocenters that are linked to neighboring events by cross-correlation (solid lines) or catalog (dashed lines) data. For two events, i and j , the initial locations (open circles) and corresponding slowness vectors, \mathbf{s} , with respect to two stations, k and l , are shown. Ray paths from the sources to the stations are indicated. Thick arrows ($\Delta \mathbf{x}$) indicate the relocation vector for events i and j obtained from the full set of equations (5), and dt is the travel-time difference between the events i and j observed at station k and l , respectively.

propriate for a solution constructed by conjugate gradients (see Lawson and Hanson [1974] for more exact solutions of constrained least squares). As shown later, the double-difference algorithm is also sensitive to errors in the absolute location of a cluster. Thus, equation (8) is usually down-weighted during inversion to allow the cluster centroid to move slightly and correct for possible errors in the initial absolute locations.

Matrix Structure and Regularization

The matrix \mathbf{G} is highly sparse as each equation links together only two events, i.e., of the $4N$ columns in each of the M rows of \mathbf{G} , only 8 have nonzero elements. To enhance the numerical stability of the solution, we scale \mathbf{G} by normalizing the L2-norm of each column of \mathbf{G} , i.e., $|\mathbf{G} \cdot \mathbf{e}_i| = 1$, for $i = 1, \dots, 4N$. If one event is poorly linked to all other events, then \mathbf{G} is ill conditioned, and the solution to equation (7) may become numerically unstable depending upon the solution method. One way to regularize such ill-conditioned systems is by prefiltering the data by only including events that are well linked to other events. In general this is achieved by only allowing event pairs which have more than a minimal number of observations. This number, however, also depends on the geometrical distribution of the stations observing the two events. A threshold value is usually found by trial and error. Another way of regularizing

moderately or severely ill-conditioned systems is damping of the solution. The problem then becomes

$$\mathbf{W} \begin{bmatrix} \mathbf{G} \\ \lambda \mathbf{I} \end{bmatrix} \mathbf{m} = \mathbf{W} \begin{bmatrix} \mathbf{d} \\ \mathbf{0} \end{bmatrix}, \quad (9)$$

with λ being the damping factor.

Solution

A standard approach to solving equation (7) in a weighted least-squares sense (i.e., minimizing the L2-norm of the residual vector) is the use of normal equations

$$\hat{\mathbf{m}} = (\mathbf{G}^T \mathbf{W}^{-1} \mathbf{G})^{-1} \mathbf{G}^T \mathbf{W}^{-1} \mathbf{d}, \quad (10)$$

with \mathbf{W} containing *a priori* quality weights. The *a priori* weights express the normalized quality of the data; that is, the precision of the cross-correlation measurements (e.g., proportional to the squared coherency; Fréchet, 1985; Fréchet and Malone, 1987), the quality and consistency with which first-motion arrival times are determined on a routine basis, and the relative accuracy between the two data sets.

For small clusters, and for well-conditioned systems, we can solve equation (7) by the method of singular value decomposition (SVD):

$$\hat{\mathbf{m}} = \mathbf{V} \mathbf{\Lambda}^{-1} \mathbf{U}^T \mathbf{d} \quad (11)$$

where \mathbf{U} and \mathbf{V} are two matrices of orthonormal singular vectors of the weighted matrix \mathbf{G} , and $\mathbf{\Lambda}$ is a diagonal matrix of the singular values of \mathbf{G} . We use SVD to investigate the behavior of small systems, as the matrices \mathbf{U} , $\mathbf{\Lambda}$, and \mathbf{V} in equation (11) store information on the resolvability of the unknown parameters \mathbf{m} and the amount of information (or lack thereof) supplied by the data \mathbf{d} . We estimate least squares errors, e_i , for each model parameter i by

$$e_i^2 = C_{ii} \cdot \text{var}, \quad (12)$$

where C^{ii} are the diagonal elements of the covariance matrix $\mathbf{C} = \mathbf{V} \mathbf{\Lambda}^{-2} \mathbf{V}^T$, and var is the variance of the weighted residuals calculated by

$$\text{var} = \frac{\sum_{i=1}^M (d_i - \bar{d})^2 - \frac{\left(\sum_{i=1}^M d_i - \bar{d}\right)^2}{M}}{M - (4N)}, \quad (13)$$

\bar{d} being the mean of the residual vector and d_i the residual of the i th observation.

As the system to be solved becomes larger, SVD is inefficient. In these cases we find the solution $\hat{\mathbf{m}}$ by using the conjugate gradient algorithm LSQR of Paige and Saunders

(1982) that takes advantage of the sparseness of the design matrix. LSQR solves the damped least-squares problem

$$\left\| \mathbf{W} \begin{bmatrix} \mathbf{G} \\ \lambda \mathbf{I} \end{bmatrix} \mathbf{m} - \mathbf{W} \begin{bmatrix} \mathbf{d} \\ \mathbf{0} \end{bmatrix} \right\|_2 = 0 \quad (14)$$

to find $\hat{\mathbf{m}}$ (see equation 9).

LSQR is very efficient and with a minimum storage requirement of $\geq 25M$ it is applicable to very large problems. A system of 10,000 earthquakes and two million equations is solved in less than 5 minutes on a Sun UltraSparc-II clocked at 300 Mhz. Typically 5 to 10 iterations are necessary, depending on the accuracy of the initial locations and the condition of the system.

Iteration

The initial solution to equation (7) is obtained from the starting locations and the *a priori* quality weights. The process is then iterated by updating the locations, the residuals, and the partial derivatives in \mathbf{G} . We iterate using the *a priori* weights until a stable solution is obtained (usually 2 to 3 iterations using SVD, 5 using LSQR) and then iterate further with reweighting of the data. The data is reweighted by multiplying the *a priori* quality weights with weights that depend on the misfit of the data from the previous iteration and on the offset between events.

Solving equation (7) in a least-squares sense assumes a Gaussian error distribution if one wants to make statements about the estimator. Errors in seismic travel-time data, however, have longer tails due to outliers in the data (Douglas *et al.*, 1997). The standard approach is to detect and remove the outliers by cutting off the non-Gaussian tail (Jeffreys, 1973) and to use M -estimators that iteratively weight each equation of condition by a function of its residual size (e.g., Anderson, 1982; Chave *et al.*, 1987). We use a biweight function (Mosteller and Tukey, 1977) to reweight the data in order to reject/downweight observations with large residuals:

$$W_i = \max^2 \left(0, 1 - \left(\frac{dr_i}{\alpha \cdot \frac{\mathbf{dr}_{\text{MAD}}}{\sigma_{\text{MAD}}}} \right)^2 \right). \quad (15)$$

dr_i represents the i th residual difference, \mathbf{dr} is the vector of residuals dr_i , $\mathbf{dr}_{\text{MAD}} = \text{med}(|dr_i - \text{med}(\mathbf{dr})|)$ is the median absolute deviation from the median (MAD), $\sigma_{\text{MAD}} = 0.67449$ is the MAD for Gaussian noise, and α is a factor that defines the rejection level at α standard deviations. Choices for α are typically between 3 and 6. Misfit weights are separately obtained from the residual distribution of the cross-correlation and the catalog data.

To downweight data for event pairs with large inter-event distances we use

$$W_i = \max^b \left(0, 1 - \left(\frac{s_i}{c} \right)^a \right), \quad (16)$$

where s is the interevent distance, c is a cutoff value to dismiss observations of event pairs with interevent distances larger than c , and a and b are exponents that define the shape of the weighting curve. The choice of a biexponential function is physically based and the function's shape needs to be adapted for cross-correlation and catalog data separately. In the former case, the parameters a , b , and c depend on the data quality, and in the latter they depend on the heterogeneity of the local velocity structure. Typically, unit weights are desired for events that are closeby, with a rapid/slow decrease with separation distance for cross-correlation/catalog data. Note that, although the correlation coefficient between two signals is already a measure of the distance between two events (e.g., Menke, 1999), we apply the distance weight to cross-correlation data too. The additional weighting is necessary to suppress outliers that arise from cross correlating imperfectly aligned phases and to permit absolute travel-time data, which are more reliable over large distances, to control the linkage between distant events.

Iteration of the process is terminated when the rms residual reaches a threshold defined by the noise level of the data, the change in the solution vector is below a chosen threshold, or the maximum number of iterations is reached.

Stepwise Relocation and Memory Handling

It is often desirable to relocate new events relative to already relocated events. In such an attempt the internal structure of the relocated events has to be retained while the new events are relocated relative to this structure. This is easily done by splitting the relocation vector for the relocated events in equation (5) into two components: one, $\bar{\mathbf{m}}$, which represents a shift common to all relocated events (i.e., shift of the entire cluster relative to the new events), while the other, \mathbf{m}' , represents individual shifts between the events. Thus, to link a new event, i , to a relocated event, j , equation (5) becomes

$$\frac{\partial t_k^i}{\partial \mathbf{m}} \Delta \mathbf{m}^i - \frac{\partial t_k^j}{\partial \mathbf{m}} \Delta (\mathbf{m}' + \bar{\mathbf{m}})^j = dr_k^{ij}. \quad (17)$$

To relocate a set of n new events relative to a cluster of already relocated events, equation (7) extends to

$$\begin{bmatrix} \mathbf{d}^1 & -\mathbf{d}^2 & \dots & 0 & 0 & \dots & 0 \\ 0 & \mathbf{d}^2 & \dots & -\mathbf{d}^{(n+1)} & 0 & \dots & -\mathbf{d}^{(n+1)} \\ \mathbf{d}^1 & 0 & \dots & -\mathbf{d}^{(n+1)} & 0 & \dots & -\mathbf{d}^{(n+1)} \\ \mathbf{d}^1 & 0 & \dots & 0 & -\mathbf{d}^{(n+2)} & \dots & -\mathbf{d}^{(n+2)} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & \dots & 0 \end{bmatrix}$$

$$\begin{bmatrix} \Delta \mathbf{m}^1 \\ \Delta \mathbf{m}^2 \\ \dots \\ \Delta \mathbf{m}'^{(n+1)} \\ \Delta \mathbf{m}'^{(n+2)} \\ \dots \\ \Delta \bar{\mathbf{m}} \end{bmatrix} = \begin{bmatrix} \mathbf{dr}_1 \\ \mathbf{dr}_2 \\ \mathbf{dr}_3 \\ \mathbf{dr}_4 \\ \dots \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \quad (18)$$

where $1 \dots n$ are the new events, each \mathbf{d} represents the four partial derivatives for one event, $\Delta \mathbf{m}$ represents the 4-vector changes for the new events, $\Delta \mathbf{m}'$ represents the 4-vector changes in the individual shifts of the relocated events, $\Delta \bar{\mathbf{m}}$ represents the 4-vector changes of the common shift of the relocated events, and \mathbf{dr} represents the residual vectors.

By appropriate weighting of the equations that constrain the individual shifts of the relocated events (last two equations in system 18), we may allow poorly constrained events to move relative to well-constrained events and the new events. Usually a weight inversely proportional to the squared rms error from the previous relocation is chosen. No constraint is put on the common shift parameters, $\bar{\mathbf{m}}$, allowing the relocated cluster to move as a whole relative to the new events.

Note that, although the design matrix \mathbf{G} , compared to equation (7), has four more columns that store the common shift parameters, the number of rows are dramatically decreased as only observations between new events and between new events and already located events are used, but not between already located events. Thus the aforementioned scheme can also be used to split off very large systems into smaller ones depending on the availability of computer memory.

Tests and Application

Data and Data Weighting

Catalog and waveform data from the Northern California Seismic Network (NCSN) for earthquakes recorded between 1984 and 1998 on the Northern Hayward fault near Berkeley and El Cerrito are used to test and demonstrate the performance of the relocation algorithm. The NCSN catalog lists 346 earthquakes in the magnitude range of M 0.7 to M 4.0 with routinely picked first arrival times. We obtain travel-time differences for each event pair with a separation distance less than 10 km at stations that locate within 200 km distance from the cluster centroid (Fig. 2). The catalog travel-time data are selected to optimize the quality and minimize the number of links between events.

In addition to the catalog-derived travel-time differences, we measure P - and S -wave differential travel times between all possible pairs of earthquakes with similar waveforms following the cross-spectral method of Poupinet *et al.* (1984) (Fig. 3). Waveforms recorded on vertical component seismometers located within 200 km from the cluster centroid are used. Most of the measurements, however, were

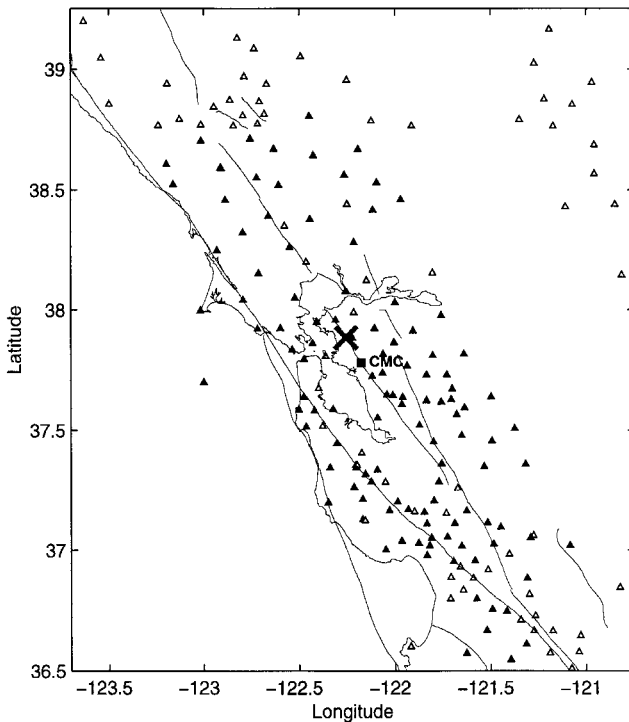


Figure 2. Map of northern California showing the location (x) of the event clusters near Berkeley and El Cerrito examined in this study and the distribution of stations (triangles and square symbols) used to relocate the events. Solid triangles represent stations that have at least five cross-correlation measurements in addition to the catalog data, and open triangles represent those that have catalog data only. The solid square symbol southeast of the earthquake cluster is station CMC. Thin lines indicate coast line and major faults.

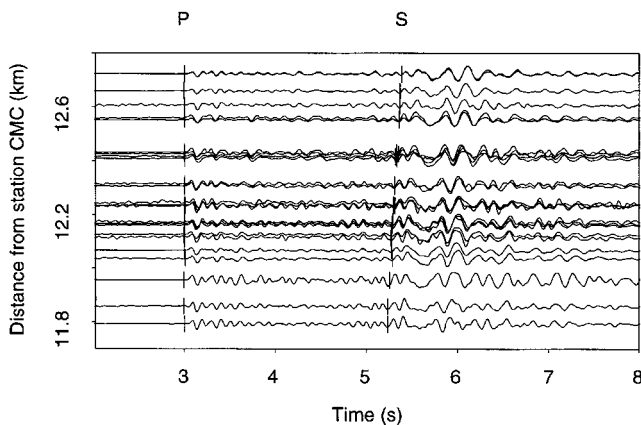


Figure 3. Seismograms (bandpass-filtered between 0.1 and 8 Hz) of events located in the Berkeley cluster and recorded at station CMC (see square in Fig. 2). Seismograms are aligned on the P arrival. Relative position of P - and S -wave picks derived by waveform cross correlation are indicated. The move-out of the S arrivals for events at greater distance from the station is clearly visible. From Waldhauser *et al.* (1999).

derived from stations closer than the crossover point (about 80 km) (Fig. 2). Two waveforms recorded at a specific station are considered similar when 50% of the squared coherency values, in the frequency range 2–10 Hz, of a tapered 2.56 sec (256 samples) window containing the P -wave (S -wave) train, exceed 0.9. For two similar waveforms, the time difference dt is proportional to the slope of the phase of the cross spectrum (Poupinet *et al.*, 1984). For sufficiently similar signals, a precision of 1 msec can be achieved for the measurement of time differences in the energy of P - and S -arrivals from data digitized at 10 ms intervals (Poupinet *et al.*, 1984; Fréchet, 1985). This is in particular important for S -waves for which absolute arrivals cannot be picked to even 200 msec. Contamination of the windowed P -wave signal with S waves at stations close to an event pair is avoided by moving the end taper of the P -wave window just before the S arrival, allowing an evaluation of the low-frequency content of such signals.

The cross-correlation data is *a priori* weighted by the squared coherency. Catalog data are assigned relative *a priori* weights of 1, 0.75, and 0.25 for the pick quality classes 0, 1, and 2 used by the NCSN. As cross-correlation delay time measurements are about an order of magnitude more precise than first-motion picks, we typically downweight the catalog travel-time data by a factor of 100 when mixing the two data types. Equal weights are kept for P -wave and S -wave cross-correlation data.

After inspection of the relocation results we define the reweighting functions as shown in Figure 4. Cutoff values, c (see equation 16), for interevent distances are 2 km for cross-correlation data and 10 km for catalog data. We chose a steep curve for the cross-correlation data ($a = b = 5$, see equation 16) in order to retain unit weights until the breakdown of cross-correlation, usually around 600 m. A few reliable cross-correlation measurements, however, have been obtained for event pairs with interevent distances up to 1700 m. A less restrictive curve is designed to weight catalog data based on event separation. We found a bicube function appropriate for the data on the northern Hayward fault (Fig. 4).

The weighting/reweighting scheme assures that the catalog data mainly constrain the relative position of events that do not correlate without sacrificing the highly accurate cross-correlation data that constrain the locations of closeby events that correlate. It also accounts for the different dependence of each of the data types on the interevent distance. Highest weights, therefore, are assigned to cross-correlation data of events that locate within their common first Fresnel zone. Beyond this distance, the weight on cross-correlation data decreases with increasing interevent distance, eventually allowing the catalog data to control the relative locations of events with larger separation distance (Fig. 4).

We solve the forward problem for this data set with a 1D-layered P -velocity model used for routine location by the NCSN. The S -velocity model is obtained by scaling the P -velocity model by a factor $1/\sqrt{3}$. Deviation of the model

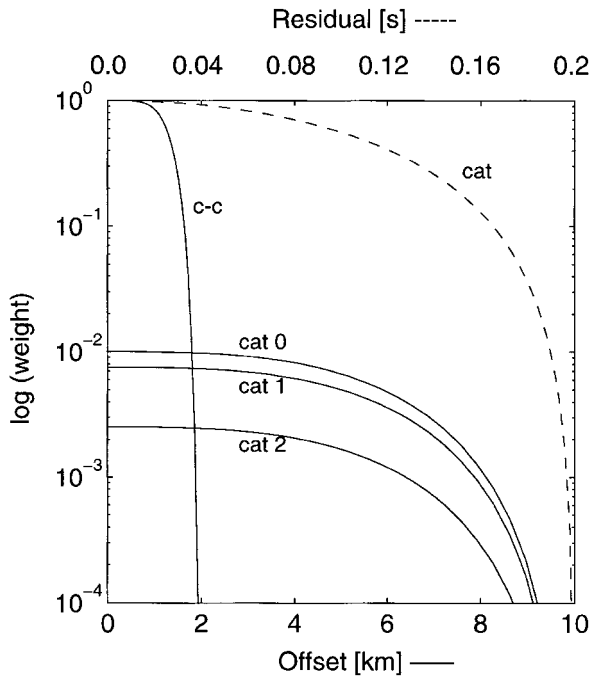


Figure 4. Reweighting functions used to down-weight/reject data with large distances between the events (shown for cross-correlation data and catalog data of different weight classes; solid lines) and/or large residuals (shown for catalog data only and a typical rejection threshold of 0.2 sec; dashed line). See text for explanation.

ratio from the real V_p/V_s ratio, which can be significant within fault zones, will predominantly effect the spatial scaling of a multiplet, rather than its internal structure.

Data Consistency

Because seismograms of local seismic events are strongly influenced by the details of the local crustal velocity structure, we generally expect greater waveform similarity between events situated close together than between events situated farther apart. Fig. 3 shows seismograms with relative P - and S -wave picks of 28 highly correlated events that occurred in the cluster near Berkeley and were recorded at station CMC, located about 10 km to the SE of the cluster (see Figure 2). The seismograms are plotted at their actual (relocated) distance to the station. The relocated events form a horizontal line (streak) of about 1-km length in direction of station CMC at a depth of about 10 km. Note how neighboring seismograms are more similar than more distant seismograms. The decay of signal coherency with increasing distance between the events is due the presence of velocity heterogeneities that induce scattering of the signals for waves with different paths (e.g., Vernon *et al.*, 1998). Figure 3 also shows the move-out of the S -wave arrivals for events at greater distance to the station. This feature is observed for events that are only a few tens of meters apart from each

other, a scale that is far beyond the resolution capabilities of network picks.

We use the multiplet of 28 events located in the Berkeley cluster to evaluate the performance of the relocation algorithm. We first demonstrate the consistency of the data by independently relocating the multiplet using the P -wave catalog, the P -wave cross-correlation, and the S -wave cross-correlation data on their own (Fig. 5). No reweighting of the data is performed during relocation in order to allow comparison. NCSN locations (Fig. 5a) show a diffuse picture of the seismicity in both map view and cross section. Relocating the 28 events using only catalog travel-time differences (Fig. 5b) greatly improves the epicenter locations that align along a SE–NW trending zone. Depths are less well constrained and distributed over about 800 m. The weighted rms residual after four iterations is 36 ms, and the average 2σ relative location error is 70 m in horizontal direction and 250 m in depth.

Relocated events using only P -wave cross-correlation measurements collapses the hypocenters to a thin, SE–NW trending, horizontal line (Fig. 5c). The rms residual after four iterations is 9 msec. The average horizontal error is 13 m, and depth errors average 32 m. Remarkably, a very similar structure is obtained after relocation with S -wave data alone (Fig. 5d). The method's ability to relocate events using either P or S cross-correlation data demonstrates the high accuracy of the differential S travel-time data, and the robustness in cross correlating the proper phases at ranges beyond the crossover points. A small difference in the length of the streak between Figure 5c and Figure 5d is controlled by the V_p/V_s ratio in the model used to solve the forward problem.

Relocated events obtained using all three data types simultaneously with their appropriate *a priori* weights are shown in Figure 5e. The average horizontal and vertical error estimates decrease to 12 m and 25 m, respectively, and the weighted rms residual is 11 ms, indicating that we properly downweight the large amount of catalog data. Relocated events using P - and S -wave cross-correlation data only and reweighting during iteration reduces the rms residual to 3 msec and the average horizontal and vertical errors to 5 m and 10 m, respectively.

Relative Location Error Estimates

To assess the reliability of the least-squares-error estimates, we apply a statistical resampling approach (“bootstrap” method, Efron, 1982; Billings, 1994; Shearer, 1997) to the events along the lineation shown in Figure 5. For the final hypocenters in Figure 5b–d, we replace the final residuals in equation (7) by samples drawn with replacement from the observed residual distribution and relocate all events with these bootstrap sample data and unit weights to determine the shift in location with the resampled data vector. The process is repeated 200 times, and the cumulative results (200×28 samples) are presented in Figure 6a–c. Ellipses in Figure 6 contain 95% of the points and are derived from their distribution. Horizontal errors for the cata-

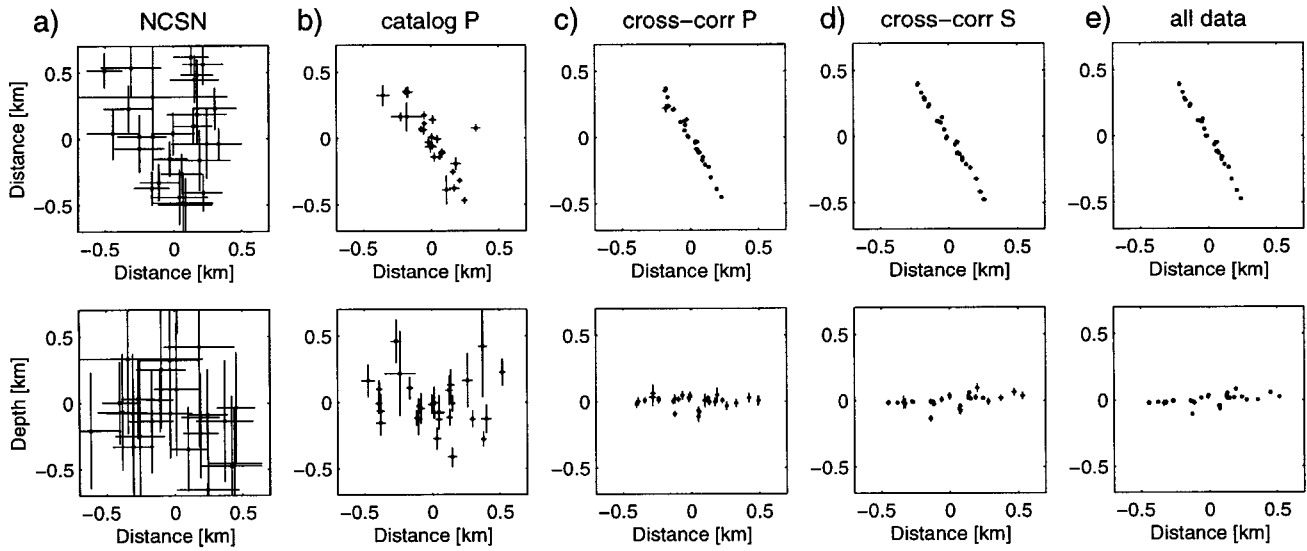


Figure 5. Relocation of 28 highly correlated events in the Berkeley cluster with three independent data sets. (a) NCSN locations, (b) Double-difference locations computed using the difference of NCSN catalog travel times for each event pair; (c) P -wave cross-correlation travel-time differences; (d) S -wave cross-correlation travel-time differences; and (e) all three data types. Top panels show map views, lower panels show NW-SE cross sections. Crosses indicate the 1σ (a), 2σ (b-e) error estimates given by the covariance matrix of the least-squares solution. Average number of phases per event used for relocation are 37, 47, and 41 for catalog P , cross-correlation P , and cross-correlation S data, respectively.

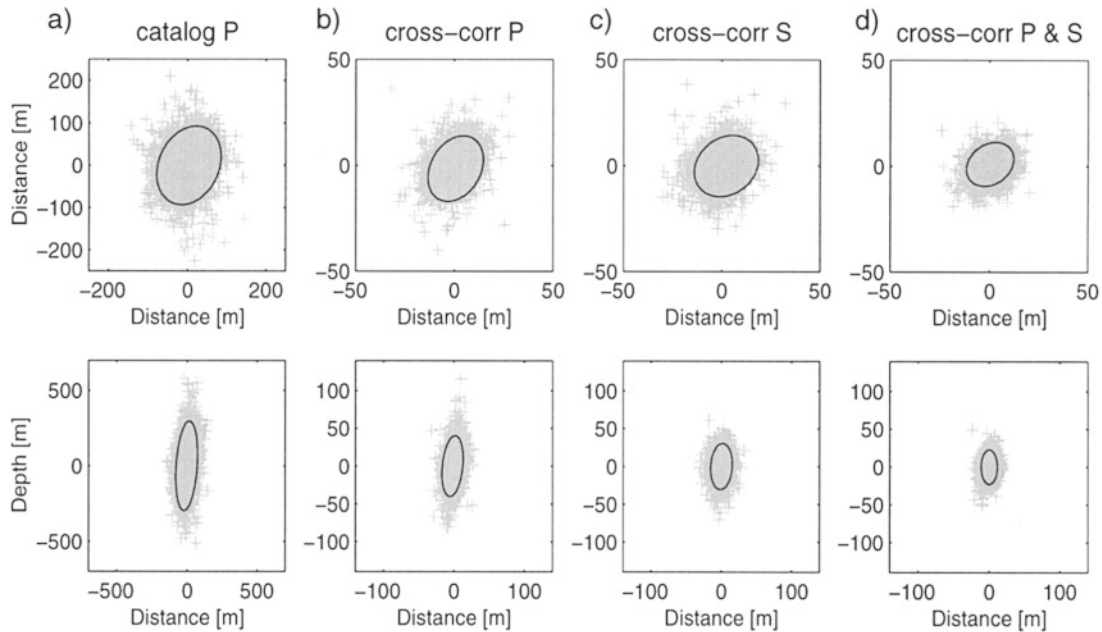


Figure 6. Bootstrap analysis of the relative location error obtained using residual travel times from the relocated events shown in Figure 5. Top panels show map view, lower panels show E-W cross section of the change in location from the least-squares location for all bootstrap samples (200×28 samples). Individual determinations shown by light crosses. (a) Catalog P -wave residuals, (b) cross-correlation P -wave residuals, (c) cross-correlation S -wave residuals, and (d) cross-correlation P - and S -wave residuals. Ellipses contain 95% of the points and are derived from observed distributions. Note the different spatial scale for (a).

log data are typically less than about 100 m (Fig. 6a), and horizontal errors for both the *P*- and *S*-wave cross-correlation data are typically less than 20 m (Fig. 6b,c). Depth errors are typically less than 250 m for the catalog data, and less than 25 m for the cross-correlation data. Figure 6d shows the results for the combined *P*- and *S*-wave cross-correlation data, with errors typically less than 15 and 20 m in horizontal and vertical direction, respectively. No elongated structure in direction of the observed streaks is observed, indicating that the observed streaks are not an artifact of the relocation procedure. In fact, the error ellipses are slightly oriented in a SW–NE (fault normal) direction, most likely due to the sparser station coverage in those areas.

Variations in station distribution for each event pair can introduce errors in relative event locations that cannot be quantified directly. We apply the jackknife method (e.g., Efron, 1982) to estimate the variance of errors in each coordinate direction. The procedure involves repeated relocation, each time subsampling the data by deleting one station at a time. No outliers are removed during this process. Results summarizing the standard deviations for each event in the three spatial directions are presented in Table 1. Among the

Table 1
Standard Deviations (m) of Jackknife Errors

Direction	Median	Lower Quartile	Upper Quartile	Maximum
East	0.9	0.7	1.1	1.5
North	2.0	1.9	2.1	2.3
Vertical	4.1	3.9	4.4	5.4

28 relocated events, we observe median errors of 0.9, 2, and 4.1 m in the east, north, and vertical direction, respectively. Location errors range from 0 to about 15 m in horizontal direction and to 45 m in depth. This suggests that, in general, errors due to improper station geometry are smaller than errors that are introduced by noise in the data or model.

The jackknife method is also used to assess the influence of one event on the locations of all others. As each event is linked to others by direct measurements, a bad event may affect the relative locations between others. To investigate this possible source of error, we relocate the multiplet 28 times, each time leaving out a different event. The resulting

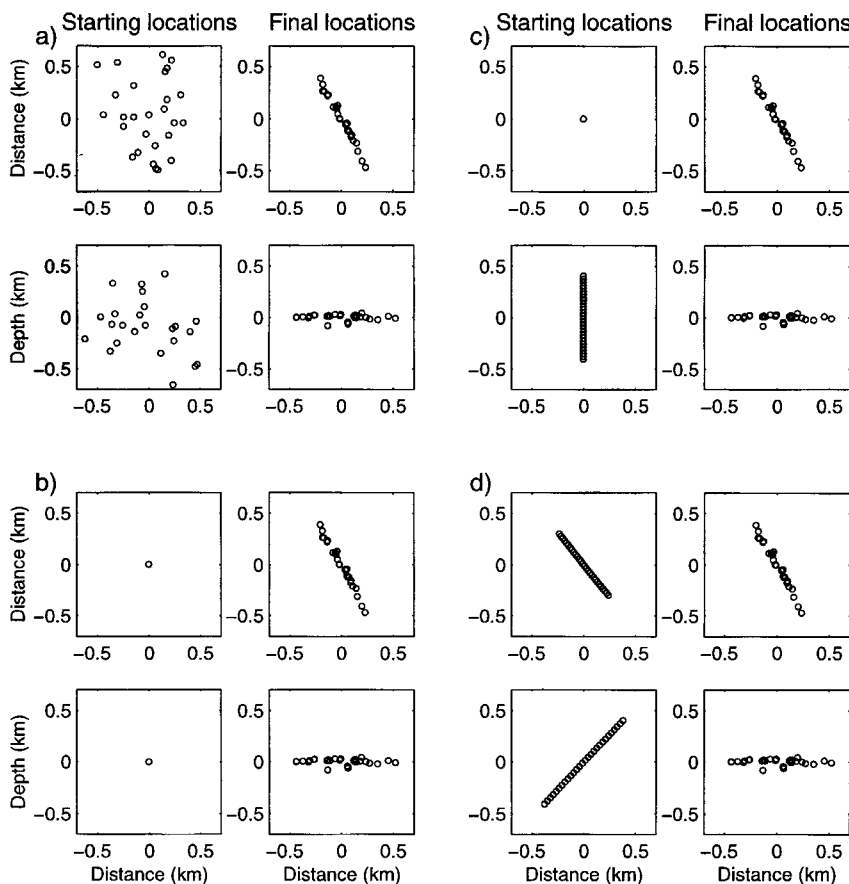


Figure 7. Relocation of the events in Figure 5 using different starting locations: (a) NCSN locations (on this scale randomly distributed), (b) common initial hypocenter, and initial hypocenters that are aligned (c) vertically and with a (d) 45° dip. In each of the four subfigures top panels are map view, lower panels are NW–SE cross sections. Cross-correlation *P*- and *S*-wave data are used in all cases.

28 sets of 27 locations show differences in relative locations between the sets that are typically less than a few meters. Such differences, however, may become significant in cases where data coverage and quality is not as good as in our test case.

Although the conjugate gradient method LSQR computes estimates of the standard error for each model parameter, the accuracy of these estimates is not guaranteed (Paige and Saunders, 1982). This is because the diagonal elements of the covariance matrix are only approximately computed and critically dependent on the proper convergence during the internal iterations. The reliability of the errors reported by LSQR, however, can be easily assessed using SVD and equation (12) or by the statistical resampling approaches described above.

Robustness Against Variations in Initial Locations

To demonstrate the robustness of the relocation algorithm to changes in the starting model, the 28 events were relocated using combined *P*- and *S*-wave cross-correlation data and four different sets of starting locations (Fig. 7). After relocation, the hypocenters form virtually the same structure, independent of their starting locations. Whereas with small multiplets for which the multiplet centroid might be an appropriate starting location, using catalog locations as trial positions give the most stable results for larger clusters (several kilometers dimension), as they are, on average, closer to the true locations.

Consistent errors in initial absolute locations may cause

changes in the orientation and scaling of the streak due to systematic differences in the takeoff angles and the partial derivatives. In Figure 8 and with perfect data, however, we show that the double-difference algorithm has some degree of sensitivity to errors in absolute locations. We compute synthetic travel times for five sources that are horizontally aligned over 2 km in a east–west direction at a depth of 10 km (Fig. 8a). We then shift the common source location by 2 km to the east and use this shifted location as initial trial source for the inversion of the error-free data. The relocated lineation (Fig. 8b) dips slightly to the east due to the systematic errors in the partial derivatives when the centroid is constrained to the wrong initial location (equation 8). The rms residual increases by a factor of 20. By relaxing the constraint on the mean shift, the relocated sources move to their correct absolute locations (Fig. 8c). Although the double-difference algorithm is able to properly correct for errors in absolute locations in the case of perfect data, this sensitivity is limited by the presence of errors in the case of real data.

Application to the Northern Hayward Fault

The 346 events on the northern Hayward fault near Berkeley and El Cerrito (i.e., 100% of the seismicity listed in the NCSN catalog) were relocated using the double-difference method. Figure 9a shows the NCSN locations of these events. The seismicity forms a northwest-striking zone associated with the Hayward fault, and a diffuse zone of earthquakes about 2 km northeast of the fault zone near Berkeley (Waldhauser *et al.*, 1999). In cross section, the cat-

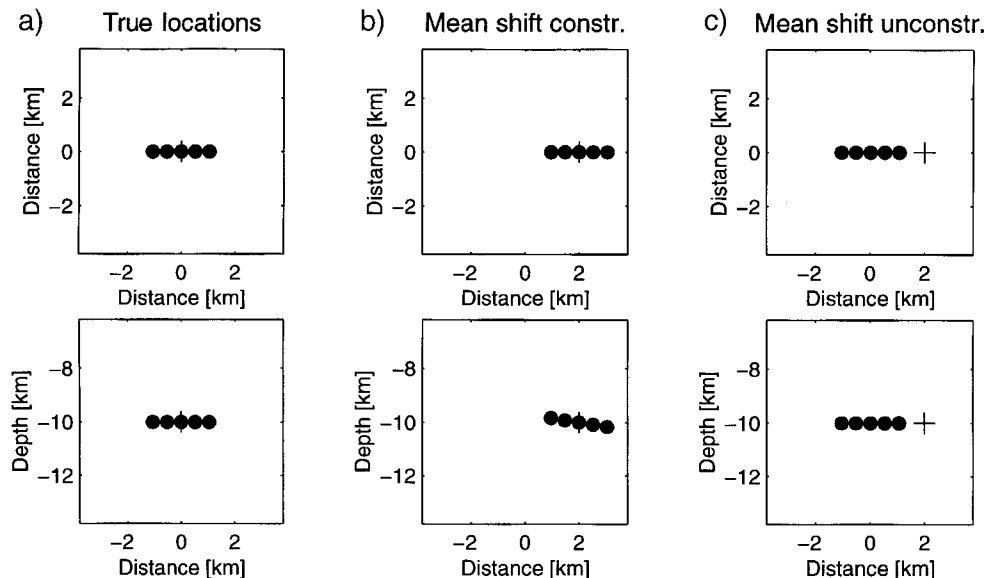


Figure 8. Effects of absolute source mislocations on double-difference locations. Relocation of five sources using error-free synthetic data. All events were given a common initial location +. Solid circles denote the relocated sources. Map view (top panels) and cross sections (lower panels) are shown. (a) Relocation with trial location at true cluster centroid. (b) Relocation with the trial location shifted to the right by 2 km with the mean relocation shift constrained to zero. (c) Relocation as in (b), but with the constraint on the mean relocation relaxed (by a factor of 1000).

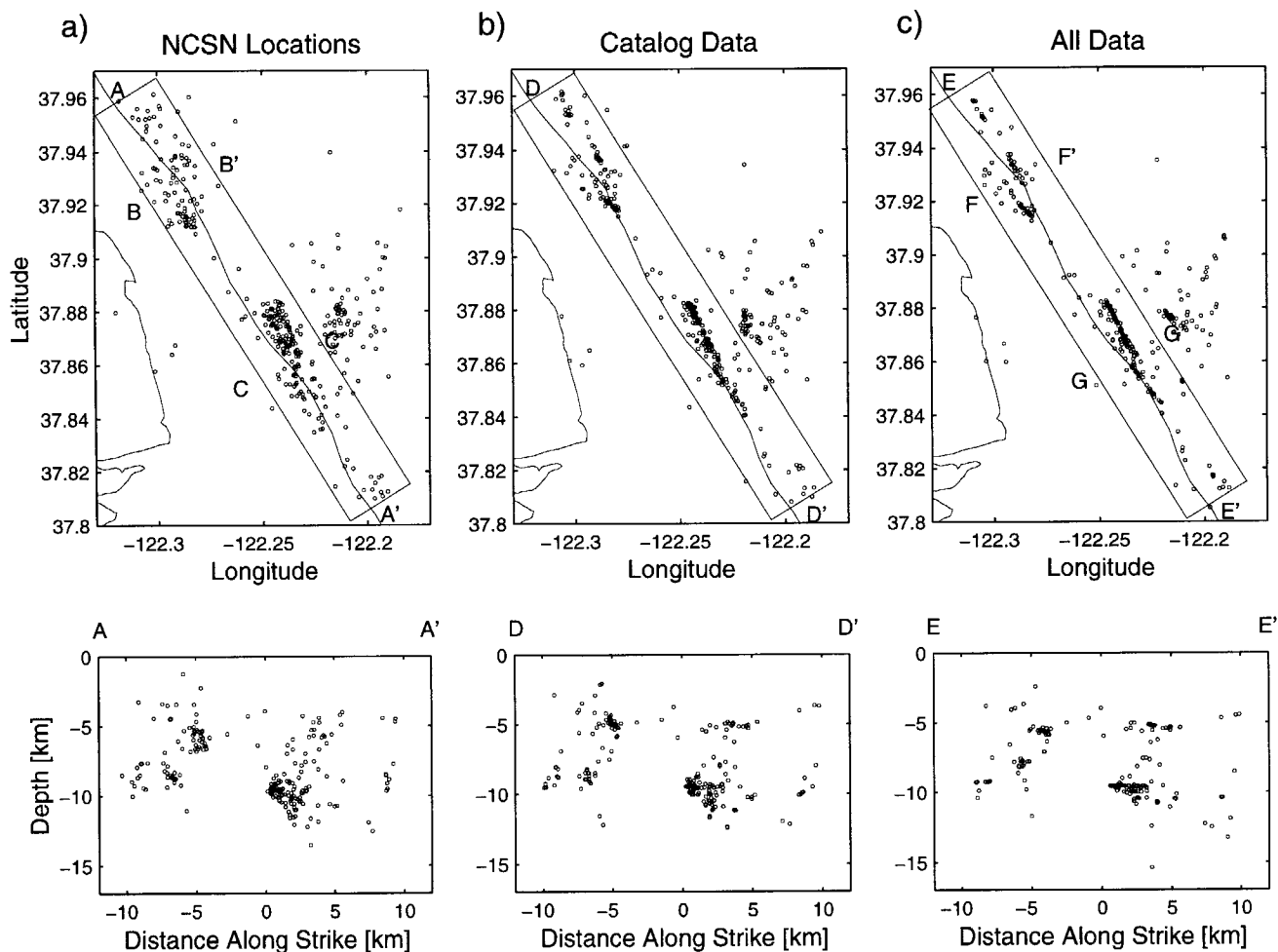


Figure 9. (a) NCSN locations and (b) double-difference locations obtained from catalog travel-time differences and (c) the combined set of catalog and cross-correlation data of events located in the El Cerrito and the Berkeley cluster. Top panels show map view, lower panels show cross sections along the fault in NW-SE direction. The surface trace of the Hayward fault and shore line is shown. Boxes indicate events included in the cross sections in this Figure and in Figure 11.

along locations of earthquakes along the fault zone are scattered but tend to concentrate at different depths along the fault zone.

Figure 9b shows the double-difference locations based solely on catalog data. The relocation method collapses the epicenters to a narrow zone of seismic activity, with the events clustered in depth. The rms residual decreases from 125 msec before to 41 msec after relocation. There is a significant reduction in the relative solution uncertainties. The 2σ standard errors in relative horizontal coordinates average about 50 m, and in depth about 130 m, compared to the 270 m and 590 m of absolute routine location uncertainties.

The relocated events derived with the combined set of catalog and cross-correlation data are shown in Fig. 9c. Re-weighting leads to the rejection of 6% of the data, most of which were catalog data with pick quality 2. The residuals for the final locations are individually examined for each

data type and weight class (P - and S -wave cross-correlation data, NCSN catalog weight classes 0, 1, and 2) (Fig. 10). The rms residual for cross-correlation derived locations decreases from about 96 ms (P -waves) and 147 ms (S -waves) before relocation to about 8.5 ms after relocation for both, P - and S -waves. The catalog rms residual decreases for weight class 0 from 83 ms to 60 ms, for weight class 1 from 100 ms to 77 ms, and for weight class 2 from 146 ms to 113 ms. The left panels in Fig. 10 shows the histograms of the residuals. The distributions look reasonably Gaussian-like, although they deviate from a Gaussian distribution in the tails where it is somewhat hard to see. The departures from Gaussian is made visible in quantile-quantile plots based on a Gaussian model scaled by the inner quartile of the sample distribution (right panels of Fig. 10). The straight line is the expected trend for the data if it were drawn from the Gaussian distribution. For all data types and weight classes, ex-

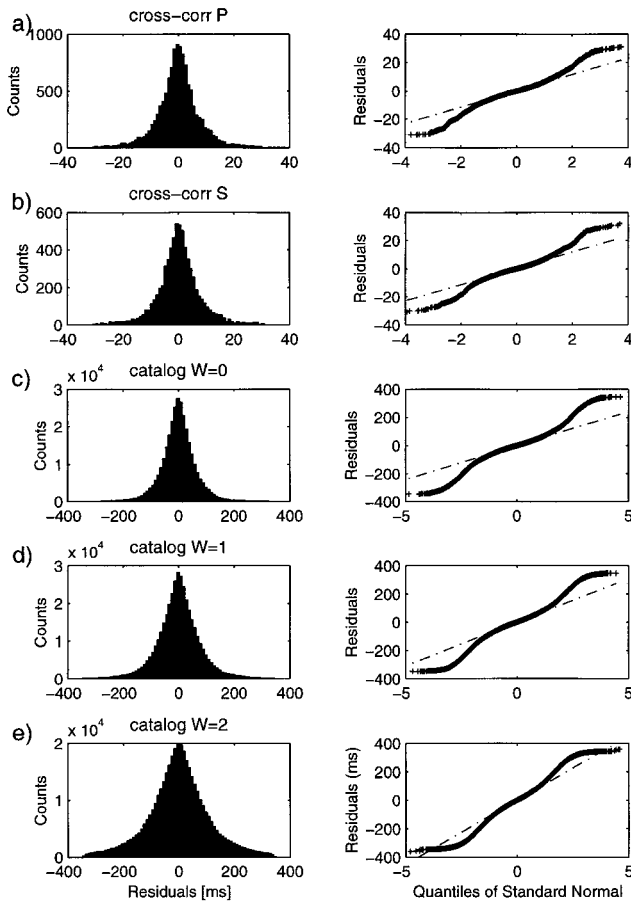


Figure 10. Distribution of delay time residuals after relocation. Shown are residuals of (a) cross-correlation P -wave and (b) cross-correlation S -wave data, and catalog data with (c) weights of class 0, (d) 1, and (e) 2. Left panels show histograms of the residuals. Right panels show quantile-quantile plots based on a Gaussian distribution. See text for explanation.

cept catalog weight class 2, the observed data bends away from the Gaussian distribution at about 2 standard deviations from the mean. For catalog data of weight class 2, the deviation shows up somewhat earlier. The errors, therefore, are nearly normal, but have much longer tails than the Gaussian distribution. Douglas *et al.* (1997) found a similar error distribution for ISC travel-time data. The observed long tails justify the down-weighting of large residuals during relocation to obtain a more Gaussian-like distribution.

Double-difference locations based on the combined data set (Fig. 9c) reveal a focused picture of seismicity, with most of the events aligning in depth along linear, horizontal streaks (Waldhauser *et al.*, 1999). The inclusion of highly consistent cross-correlation-derived differential times greatly reduced location uncertainty, in particular in depth due to the differential S -wave measurements for earthquakes that correlated. Figure 11 shows cross sections of the on-fault seismicity before and after relocation with 2σ uncer-

tainties indicated. Most of the on-fault activity consists of grouped similar events (solid circles in Fig. 11b) with strong cross-correlation coupling and location errors typically about a few meters to a few tens of meters. These multiplets contain more than half of the earthquakes recorded between 1984 and 1998. Weak cross-correlation coupling exists between earthquakes that belong to different multiplets or to the background seismicity. Some of the on-fault and most of the off-fault activity is constrained by catalog picks only and show typical errors of a few hundred meters.

Relocation of the individual multiplets with only the cross-correlation data reveal the same structure as seen in the relocated seismicity based on the combined data set, with rms residuals as low as the noise level of the data and location uncertainties of a few meters. This indicates that we obtain the desired high precision for interevent distances within multiplets, and the catalog precision for distance between multiplets and uncorrelated events.

Discussion

In ordinary JHD locations, one event influences the location of other events through a common station correction. In the double-difference approach, each event is coupled to its neighbors by direct measurements. Figure 12 displays the type and degree of coupling that exists between the relocated events on the Hayward fault from Fig. 11. A dense network of catalog-based observations (gray lines) constrains the relative locations of multiplets and uncorrelated events, whereas cross-correlation data (black lines) tie together events that occur within multiplets. A few event pairs show cross-correlation measurements over rather large distances (dashed black lines in Fig. 12). Such long-range correlations suggest that relatively little scattering occurs along the uncommon part of the raypath for these events. These measurements, although reliable, are rare (less than 5 per event pair), and the relative position between their hypocenters, therefore, is mainly constrained by catalog data.

The strong connectivity between closely correlated events is shown in a histogram of cross-correlation measurements in 200-m bins of interevent distances (Fig. 13a). Over 60% of the measurements relate to event pairs with a separation distance smaller than 200 m. The mean coherency for P and S phases calculated for the same bins show the expected decrease as a function of interevent distance (Fig. 13b). Note that the curve for the S -phase data is slightly steeper than the one for the P -phase data, most likely due to the contamination of the S -phases with earlier, out-of-phase P waves.

Figure 14 illustrates how the quality of the locations shown in Fig. 9c relate to data type and event separation. Event pairs constrained by cross-correlation data have binned rms travel-time errors that range from 6 to 13 msec (P waves) and 7 to 19 msec (S waves) between 0- and 2-km distance between the events (Fig. 14a). The rms residual errors for catalog data range from 43 to 89 msec for weight

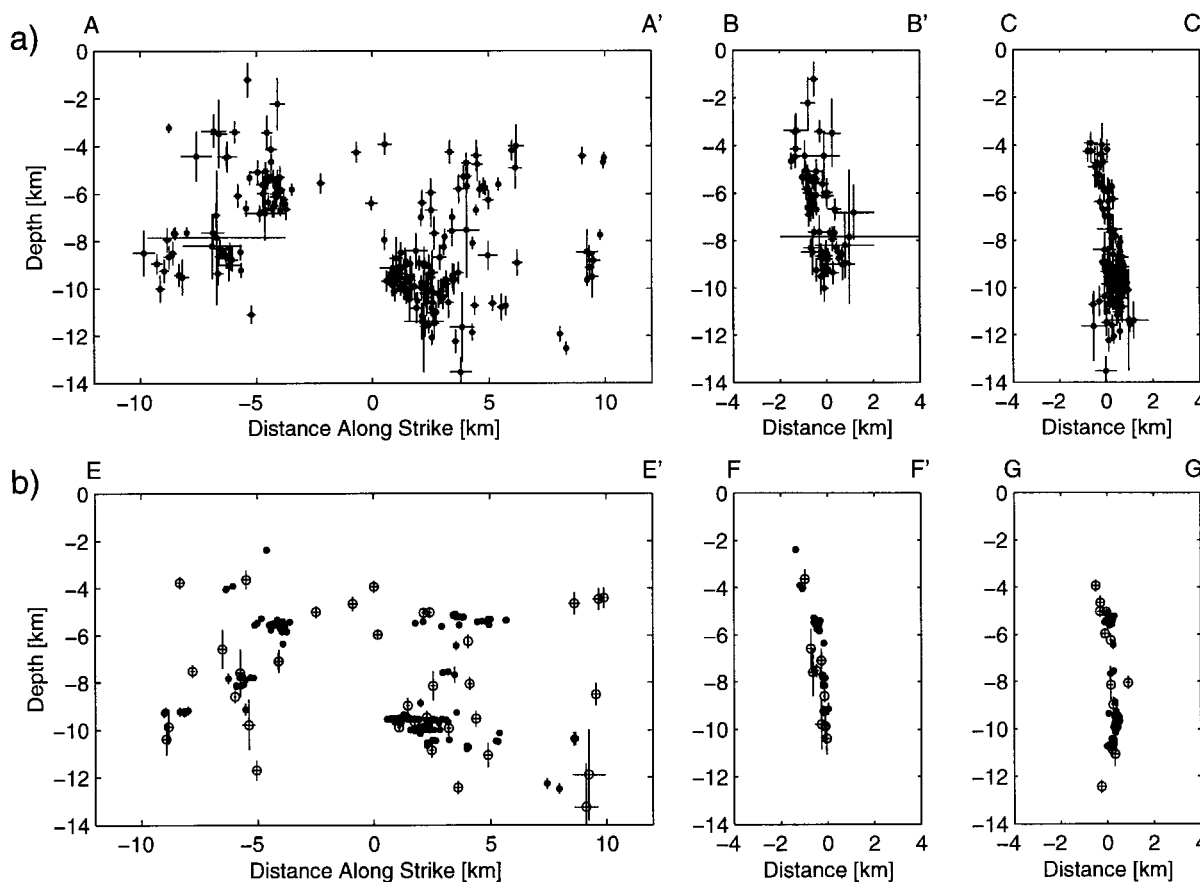


Figure 11. Cross sections of the on-fault seismicity along and across the Hayward fault. (a) NCSN locations with 1σ errors indicated. (b) Double-difference locations using all data and 2σ errors indicated. Events that are well constrained by cross-correlation measurements are represented by solid dots, and those with little or no cross-correlation data by open circles. For location of cross sections, see Figure 9.

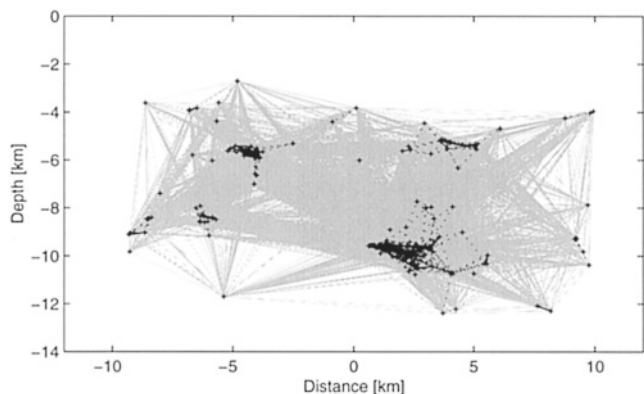


Figure 12. Cross section of the on-fault seismicity with the type of coupling between events indicated. Gray lines connect hypocenters constrained by catalog data, solid black lines connect hypocenters with more than five cross-correlation measurements, dashed black lines those with less than five cross-correlation measurements. + represents hypocenters.

class 0, from 61 to 100 ms for weight class 1, and from 97 to 120 ms for weight class 2 for separation distances between 0 and 10 km (Fig. 14b). In all cases, the rms residuals generally increase with increasing event offset. The observed increase for the cross-correlation data is most likely due to a decrease in data quality because scattering along the ray path and/or source orientation lower the coherency between the signals. In addition, noise may arise from the fact that the frequency bandwidth in the common Fresnel zone gets narrower as separation distance grows. In contrast, the increase in the rms residual for the catalog data is predominantly caused by heterogeneities in the velocity structure and thus depends on the area of investigation.

We can separate the effect of errors in the data from those caused by inadequacies of the velocity structure by looking at the residuals for very close (<50 m) events. The rms residual for these event pairs is about 2 msec for cross-correlation data, and 15 msec for weight class 0 catalog data. For a repeating earthquake source of 17 events on the Calaveras fault (Vidale *et al.*, 1994), we obtain double-difference locations with rms residuals for cross-correlation data of 1.5 msec and for the catalog data of 10 msec.

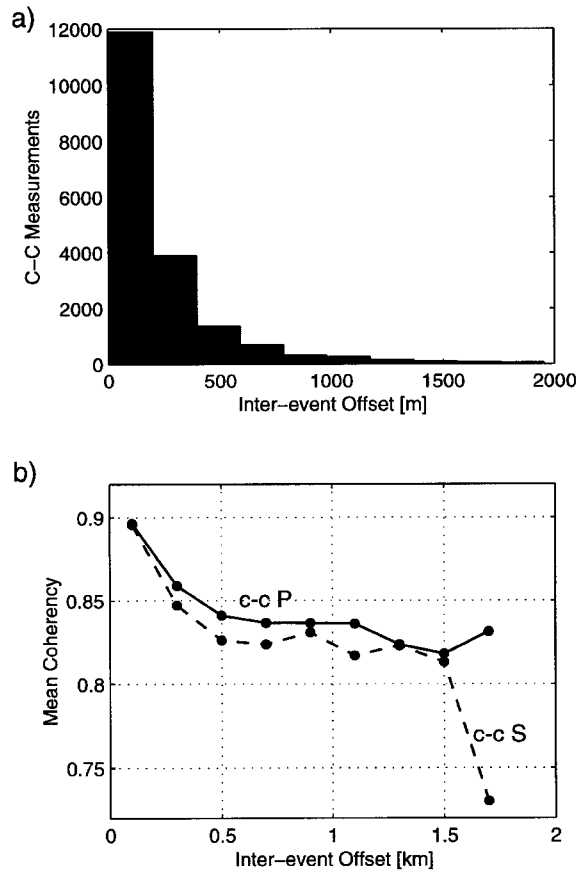


Figure 13. (a) Histogram showing the number of cross-correlation measurements in 200 m bins of inter-event distances. (b) Mean *P*- and *S*-wave coherency as a function of interevent distance.

When simultaneously using both first-motion picks and cross-correlation delay times, we have to keep in mind that the locations calculated using first breaks are hypocenter locations, but the locations calculated using cross correlation are relative centroid locations. In general these locations will not be the same. Without knowing the details of the rupture process for each earthquake, it is impossible to reconcile them. For sufficiently small earthquakes, the errors introduced by combining the two types of picks may be comparable to or even less than the uncertainty in the centroid locations. For larger earthquakes, however, the difference may become significant and may even exceed the uncertainty of the relative locations.

Figure 15 shows events that belong to a multiplet in the El Cerrito cluster (at 5 km depth and -4 km model distance in Fig. 11). The earthquakes are represented by their equivalent rupture area, modeled using a 30 bar constant stress drop, circular rupture model. They span a magnitude range from M 1.5 to M 4.0. In order to study the difference in source location between the two types of data, we relocate the hypocentroids (including the two largest events, M 3.9 and M 4.0) using cross-correlation data (solid circles) while

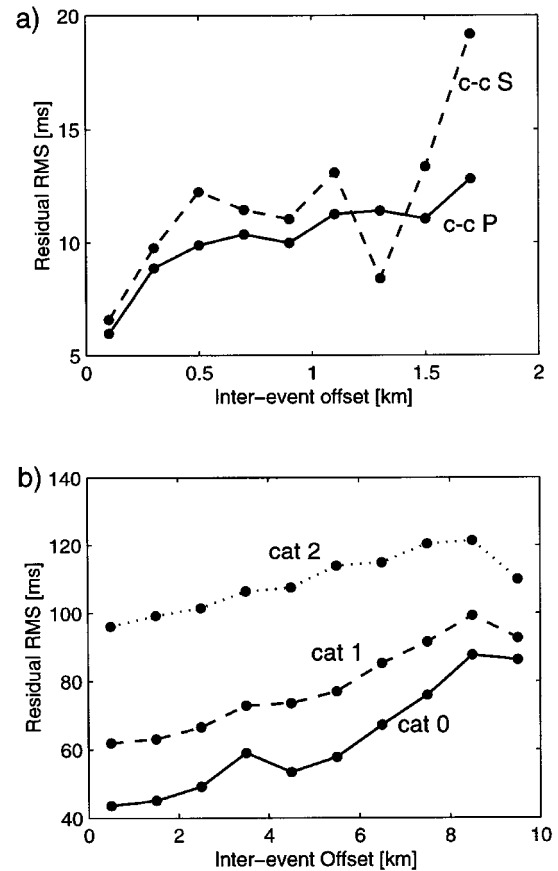


Figure 14. Rms travel-time residuals as a function of interevent distance for (a) cross-correlation data and (b) catalog data for different weights. Solid circles represent rms residuals derived from (a) 200 m and (b) 500 m bins of interevent distance.

simultaneously relocating the hypocenters of the M 3.9 and M 4.0 events with first-motion data only (dashed '+', Fig. 15). For the M 3.9 event, the hypocentroid location is insignificantly different from the hypocenter location, suggesting that it nucleated near the center of the final rupture. In contrast, the M 4.0 hypocenter locates deeper than the hypocentroid, suggesting that the rupture started near the base and spread upward.

In summary, the double-difference earthquake locations of the northern Hayward fault seismicity reveal a sharp image of the seismicity. The on-fault seismicity defines a nearly planar and vertical fault zone beneath the surface trace of the Hayward fault. The longitudinal cross-section (Fig. 11b) shows that most of the seismicity is concentrated in elongated northwest oriented structures that are embedded within less clustered activity. These streaks are horizontal lineations of hypocenters that extend along the fault zone, outlining characteristic patterns of conditions on the fault where brittle failure conditions are met (for a detailed discussion on these lineations see Waldhauser *et al.*, 1999). As demonstrated in this article, the streaks are not an artifact of the relocation procedure. This is also supported by the existence of similar

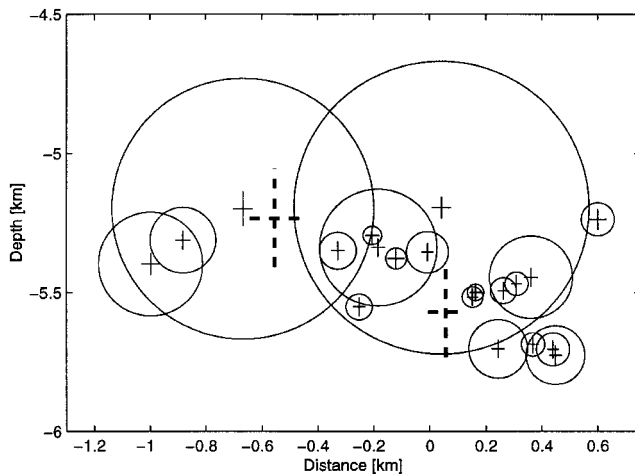


Figure 15. Relocation of 20 hypocentroids of the multiplet located in the El Cerrito cluster between 5 and 6 km depth. Hypocentroids denoted by + giving 2σ confidence intervals. Circles represent rupture areas modeled for a 30-bar constant stress drop source. The nucleation point (hypocenters) of the two largest events were simultaneously located using only catalog data linking them to each of the smaller events (dashed +). See text for discussion.

streaks that have recently been found on the San Andreas fault (Rubin *et al.*, 1999).

In transverse cross-sections, the seismicity clearly defines a narrow, steeply northeast dipping structure for the El Cerrito cluster (F-F', Fig. 11b) and a near-vertical structure for the Berkeley cluster (G-G', Fig. 11b). Both structures can be interpreted to represent the discrete active fault zone. The width of the seismicity band, and thus the discrete zone where the fault is accommodating slip by earthquakes, is very thin, about 200 m, rather than the about 2.5 km seen in the catalog locations. At some locations, an effective width of 25 m is measured.

An interesting feature to note is the acute angle between the direction of the surface trace of the fault near Berkeley and the direction of the lineation (Fig. 9c). Representative focal mechanisms computed for the events along the lineation show right lateral strike-slip motion consistent with the direction of the lineation, indicating that the imaged seismicity is real. Features like that and others, when imaged with very high resolution and along entire fault systems, permit the recognition of segment boundaries and fault bends that are believed to play important roles in the initiation and arrest of rupture.

Conclusion

We presented an efficient, double-difference technique to relocate large numbers of earthquakes with very high resolution. On the northern Hayward fault, this method collapses diffuse clouds of routinely located events into horizontal lineations of seismicity. We performed numerous

tests including statistical resampling approaches to assess the accuracy of the data and to estimate relative location errors. The results demonstrate that the horizontal lineations are not an artifact of the relocation procedure. Instead the double-difference locations image the very fine-scale structure of the seismicity along the fault zone.

The method presented is directly applicable to existing earthquake catalogs and/or digital waveform data as provided by almost any seismic network. The ability to consistently relocate the seismicity with high resolution along entire fault systems opens a wide range of research possibilities, from the scale of individual earthquakes to the scale at which tectonic processes take place.

The double-difference algorithm is implemented in the computer program hypoDD, which is available from the authors.

Acknowledgments

We greatly profited from discussions with Greg Beroza, Götz Bokelmann, Alex Cole, David Schaff, and Eva Zankerka. We thank Götz Bokelmann and Andy Michael for helpful comments on the manuscript, and Jose Pujol, Gary Pavlis, and Peter Shearer for their careful and constructive reviews. Gary Pavlis and Jim Larsen are thanked for useful comments on robust processing that lead to an improvement of the residual weighting function. Support of FW by the Swiss National Science Foundation is greatly appreciated.

References

- Aki, K., and P. G. Richards (1980). *Quantitative Seismology*, Vol. 2, W. H. Freeman, New York.
- Anderson, K. R. (1982). Robust earthquake location using M-estimates, *Phys. Earth Plan. Int.* **30**, 119–130.
- Billings, S. D. (1994). Simulated annealing for earthquake location, *Geophys. J. Int.* **118**, 680–692.
- Chave, A. D., D. J. Thomson, and M. E. Ander (1987). On the robust estimation of power spectra, coherences, and transfer functions, *J. Geophys. Res.* **92**, 633–648.
- Crosson, R. S. (1976). Crustal structure modeling of earthquake data. I. simultaneous least squares estimation of hypocenter and velocity parameters, *J. Geophys. Res.* **81**, 3036–3046.
- Deichmann, N., and M. Garcia-Fernandez (1992). Rupture geometry from high-precision relative hypocenter locations of microearthquake clusters, *Geophys. J. Int.* **110**, 501–517.
- Dodge, D., 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.
- Douglas, A. (1967). Joint epicenter determination, *Nature* **215**, 47–48.
- Douglas, A., D. Bowers, and J. B. Young (1997). On the onset of *P* seismograms, *Geophys. J. Int.* **129**, 681–690.
- Efron, B. (1982). The jackknife, the bootstrap, and other resampling plans, *SIAM*, Philadelphia, 92 pp.
- Ellsworth, W. L. (1977). Three-dimensional structure of the crust and mantle beneath the island of Hawaii, *Ph.D. Thesis*, Massachusetts Institute of Technology, Cambridge, 327 pp.
- Fréchet, J. (1985). Sismogenèse et doublets sismiques, *Thèse d'Etat*, Université Scientifique et Médicale de Grenoble, 206 pp.
- Frémont, M.-J., and S. D. Malone (1987). High precision Relative locations of earthquakes at Mount St. Helens, Washington, *J. Geophys. Res.* **92**, 10,223–10,236.

- Geiger, L. (1910). Herdbestimmung bei Erdbeben aus den Ankunftszeiten, *K. Ges. Wiss. Gött.* **4**, 331–349.
- Geller, R. J., and C. S. Mueller (1980). Four similar earthquakes in central California, *Geophys. Res. Lett.* **7**, 821–824.
- Gomberg, J. S., K. M. Shedlock, and S. W. Roecker (1990). The effect of S-wave arrival times on the accuracy of hypocenter estimation, *Bull. Seism. Soc. Am.* **80**, 1605–1628.
- Got, J.-L., J. Fréchet, and F. W. Klein (1994). Deep fault plane geometry inferred from multiplet relative relocation beneath the south flank of Kilauea, *J. Geophys. Res.* **99**, 15,375–15,386.
- Hurukawa, N., and M. Imoto (1992). Subducting oceanic crusts of the Philippine Sea and Pacific plates and weak-zone-normal compression in the Kanto district, Japan, *Geophys. J. Int.* **109**, 639–652.
- Ito, A. (1985). High resolution relative hypocenters of similar earthquakes by cross-spectral analysis method, *J. Phys. Earth* **33**, 279–294.
- Jeffreys, H. (1973). On travel times in seismology, in *Collected Papers of Sir Harold Jeffreys on Geophysics and Other Sciences*, H. Jeffrey (Editor), Vol. 2, Gordon and Breach Sci. Publ., London, 36–120.
- Kissling, E., W. L. Ellsworth, D. Eberhard-Phillips, and U. Kradolfer (1994). Initial reference models in local earthquake tomography, *J. Geophys. Res.* **99**, 19,635–19,646.
- Lawson, C. L., and R. J. Hanson (1974). *Solving Least Squares Problems*, Prentice-Hall, New Jersey, 340.
- Lees, J. M. (1998). Multiplet analysis at Coso geothermal, *Bull. Seism. Soc. Am.* **88**, 1127–1143.
- Menke, W. (1999). Using waveform similarity to constrain earthquake locations, *Bull. Seism. Soc. Am.* **89**, 1143–1146.
- Michael, A. J. (1988). Effects of three-dimensional velocity structure on the seismicity of the 1984 Morgan Hill, California, aftershock sequence, *Bull. Seism. Soc. Am.* **78**, 1199–1221.
- Mosteller, F., and J. W. Tukey (1977). *Data Analysis and Linear Regression*, Addison-Wesley, Reading, Massachusetts.
- Paige, C. C., and M. A. Saunders (1982). LSQR: Sparse linear equations and least squares problems, *ACM Transactions on Mathematical Software* **8/2**, 195–209.
- Pavlis, G. L. (1986). Appraising earthquake hypocenter location errors: a complete, practical approach for single-event locations, *Bull. Seism. Soc. Am.* **76**, 1699–1717.
- Pavlis, G. L. (1992). Appraising relative earthquake location errors, *Bull. Seism. Soc. Am.* **82**, 836–859.
- Poupinet, G., W. L. Ellsworth, and J. Fréchet (1984). Monitoring velocity variations in the crust using earthquake doublets: an application to the Calaveras fault, California, *J. Geophys. Res.* **89**, 5719–5731.
- Pujol, J. (1988). Comments on the joint determination of hypocenters and station corrections, *Bull. Seism. Soc. Am.* **78**, 1179–1189.
- Roecker, S. W. (1981). Seismicity and tectonics of the Pamir-Hindu Kush region of central Asia, *Ph.D. Thesis*, Massachusetts Institute of Technology, Cambridge, 294 pp.
- Rubin, A. M., D. Gillard, and J.-L. Got (1999). Streaks of microearthquakes along creeping faults, *Nature* **400**, 635–641.
- Scherbaum, F., and J. Wendler (1986). Cross spectral analysis of Swabian Jura (SW Germany) three-component microearthquake recordings, *J. Geophys.* **60**, 157–166.
- Shearer, P. (1997). Improving local earthquake locations using the L1 norm and waveform cross correlation: application to the Whittier Narrows, California, aftershock sequence, *J. Geophys. Res.* **102**, 8269–8283.
- Thurber, C. H. (1983). Earthquake locations and three-dimensional crustal structure in the Coyote Lake area, central California, *J. Geophys. Res.* **88**, 8226–8236.
- Van Decar, J. C., and R. S. Crosson (1990). Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least-squares, *Bull. Seism. Soc. Am.* **80**, 150–169.
- Vernon, F. L., G. L. Pavlis, T. J. Owens, D. E. McNamara, and P. N. Anderson (1998). Near-surface scattering effects observed with a high-frequency phased array at Pinyon Flats, California, *Bull. Seism. Soc. Am.* **88**, 1548–1560.
- Vidale, J. E., W. L. Ellsworth, A. Cole., and C. Marone (1994). Variations in rupture process with recurrence interval in a repeated small earthquake, *Nature* **368**, 624–626.
- Waldhauser, F., W. L. Ellsworth, and A. Cole (1999). Slip-parallel seismic lineations along the northern Hayward fault, California, *Geophys. Res. Lett.* **26**, 3525–3528.

U.S. Geological Survey
 345 Middlefield Rd., MS977
 Menlo Park, California, 94025
 felix@andreas.wr.usgs.gov
 ellsworth@usgs.gov

Manuscript received 21 August 2000.