Bulletin of the Seismological Society of America, Vol. 99, No. 5, pp. 2736–2748, October 2009, doi: 10.1785/0120080294

Near-Real-Time Double-Difference Event Location Using Long-Term Seismic Archives, with Application to Northern California

by Felix Waldhauser

Abstract We present a real-time procedure that uses cross-correlation and double-difference methods to rapidly relocate new seismic events with high precision relative to past events with accurately known locations. Waveforms of new events are automatically cross correlated with those archived for nearby past events to measure accurate differential phase arrival times. These data, together with delay times computed from arrival time picks, are subsequently inverted for the vector connecting the new event to its neighboring events using the double-difference algorithm. The new seismic monitoring technique is applied to earthquakes recorded in northern California, using near-real-time data feeds from the Northern California Seismic Network (NCSN) and the Northern California Earthquake Data Center, and a locally stored copy of the NCSN seismic archive. New events are automatically relocated in near-real time (tens of seconds) relative to a high-resolution double-difference earthquake catalog for northern California. Back testing using past events across northern California indicates that the real-time solutions are on average within 0.08 km laterally and 0.24 km vertically of the double-difference catalog locations. We show that the precision with which new events are located using this technique will improve with time, helped by the continued increase in density of recorded earthquakes and growth of the digital seismic archives. Real-time double-difference location allows for monitoring spatiotemporal changes in seismogenic properties of active faults with unprecedented resolution and therefore has considerable social and economic impact in the immediate evaluation and mitigation of seismic hazards.

Introduction

The rapid location of seismic events is a fundamental task in real-time monitoring of earthquake activity and compliance of the Comprehensive Nuclear-Test-Ban Treaty. Estimates of location (i.e., latitude, longitude, and depth) and origin time of a new event are crucial for efficient emergency and disaster reduction response as they provide information on the area of potential damage due to strong ground shaking. Furthermore, a location is needed to estimate the size of an earthquake. In general, hypocenter locations are computed on a routine basis, by the operator of seismic networks, within minutes after an earthquake occurred and subsequently disseminated to local authorities and the public (see the Data and Resources section for more information). The hypocentral information and associated metadata such as waveforms and phase arrival time readings produced during routine operations are typically archived in digital databases, which, in the case of a few networks, can be accessed in near-real time. These archives are the primary resources for the scientific and engineering community to study the structure of the Earth, the physics of earthquakes, or the hazard they impose.

Most seismic networks routinely process seismograms from new events in near-real time using some standard procedure to detect seismic signals, measure phase arrival times (either automatically or manually, or a combination of both) and associate them to individual earthquakes, and locate the hypocenter of the events. Events are commonly located one at a time by a linearized inversion of the phase arrival times for changes to some trial source computed from an *a priori* known velocity model (e.g., Geiger, 1910; Douglas, 1967; Klein, 2002). Such single-event location algorithms are fast and relatively robust, but the uncertainties in the hypocenter locations are often significant because of uncertainties in the phase onset picks and phase association and deviations of the true Earth structure from the model used to predict the data (Pavlis, 1986; Gomberg *et al.*, 1990). Even in well-monitored regions it is often difficult to associate events with known faults, for example.

There have been numerous studies that improve on the low-spatial resolution of cataloged routine single-event locations by using multiple-event location methods that compute phase delay times between nearby events to determine their relative hypocenter position (e.g., Poupinet *et al.*, 1984; Ito, 1985; Frémont and Malone, 1987; Deichmann and Garcia-Fernandez, 1992; Got *et al.*, 1994; Waldhauser and Ellsworth, 2000; Shearer, 2002). These approaches are able to minimize the bias in routine locations due to errors in the model by relocating multiple events simultaneously. Furthermore, delay times can be precisely measured via waveform cross correlation, reducing relative uncertainties in the phase arrival time picks (Poupinet *et al.*, 1984; Schaff *et al.*, 2004, among others). With their ability to improve the resolution in routine catalog locations by orders of magnitudes, multiple-event location methods have been used extensively to study the detailed structure of active faults and to characterize the spatiotemporal behavior of seismicity in areas of special interest.

Despite their ability to produce significant improvement in relative-event locations, such advanced location methods have not yet been implemented into the real-time production of earthquake catalogs. One of the reasons for this may be that the use of multiple-event location methods in real time requires fast access to a network's archive of past seismic data, especially waveforms and phase arrival time information. Such archives, however, often reside on storage media that are detached from the monitoring system. Only in recent years has it been possible to store and access them locally or over the internet due to the growth in storage, computing, and bandwidth capacity. Another reason may be that the accuracy of routinely computed hypocenter locations at local and regional networks has been generally sufficient in the past for immediate postseismic response efforts and analysis. Recent advancement in rupture modeling and strong ground motion prediction, however, could benefit from more detailed source information such as, for example, knowledge of the active fault as revealed by the precise location of a new event relative to the fault's past seismicity. Furthermore, the automated computation of precision locations would enable the monitoring and analysis of spatiotemporal evolution and changes in seismogenic properties of active faults in real time.

Using data from the Hawaiian Volcano Observatory (HVO) network in Hawaii, Got *et al.* (2002) presented a near-real-time procedure that relocates new events relative to previously determined clusters of correlated events (or multiplets) if their waveforms are sufficiently similar. In this article we report on a more general solution to the real-time use of high-precision event location methods in a monitoring context. We present an automated double-difference (DD) procedure that determines, in near-real time, the precise location of new events relative to an *a priori* known high-resolution DD catalog of past seismicity (Fig. 1). The procedure uses both waveforms and phase picks of new events immediately after they are available and of past events by accessing the seismic archives. This approach allows for rapid and comprehensive evaluation of the spatiotemporal behavior of new earthquakes relative to the historic seismicity at high resolution (tens to hundreds of meters) and can produce a continuously updated DD catalog. The performance of the new seismic monitoring technique is demonstrated in northern California with events recorded by stations of the Northern California Seismic System (NCSS), using near-real-time feeds of parametric and waveform data from the Northern California Seismic Network (NCSN) and the Northern California Earthquake Data Center (NCEDC), respectively.

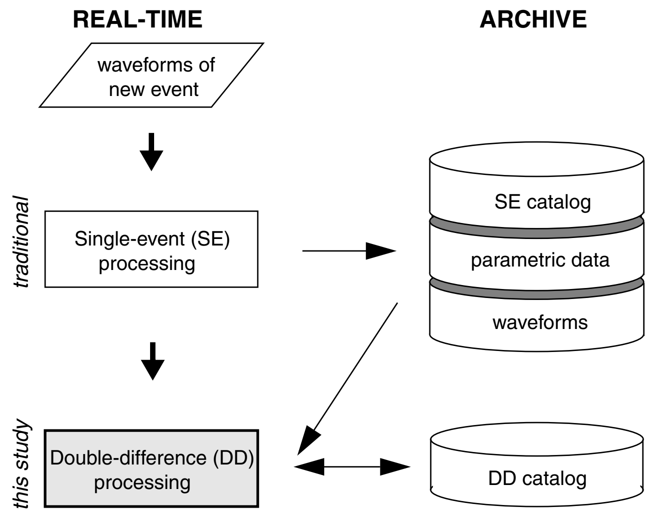

Figure 1. General overview of data flow and processes involved in the automated DD-RT relocation procedure.

Method

Data and Process Flow

The near-real-time double-difference (DD–RT) process requires real-time access to routinely produced location and phase information of a new event and the associated waveforms if cross-correlation timing is desired (Fig. 2). It uses the location information to search in an existing DD base catalog for the best-suited reference events in the neighborhood of the new event and computes differential arrival times between the new event and these reference events from phase picks and via waveform cross correlation. Subsequently, the delay times are inverted for the relative location vectors that connect the new event to its reference events. We note that this procedure significantly improves the precision of a new event's location relative to the background seismicity. The accuracy of the absolute location depends on the location accuracy of the reference events in the DD base catalog. We refer to Waldhauser and Schaff (2008) for a detailed description and location error analysis of the DD base catalog for northern California that is used in this study.

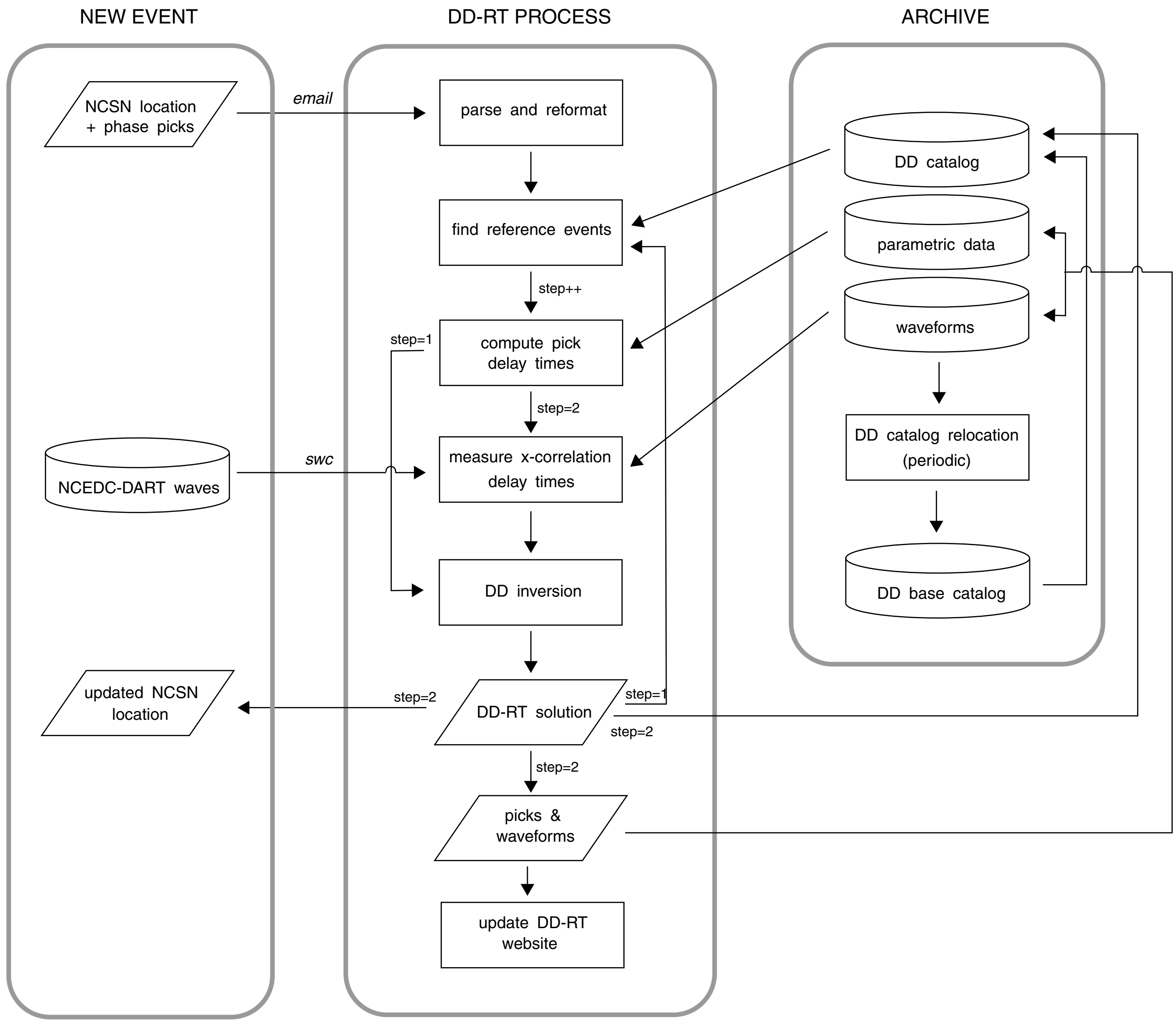

Figure 2. Detailed flow chart of the DD-RT procedure as implemented at the LDEO for automated near-real-time double-difference relocation of earthquakes in northern California with real-time data feeds from the U.S. Geological Survey (USGS) in Menlo Park (NCSN) and the University of California, Berkeley (NCEDC). swc: Simple Waveform Client (Doug Neuhauser, personal comm., 2008).

The following main steps are carried out once a network's routine information and metadata is available (Figs. 2 and 3):

1. Find reference events in the DD base catalog that are near the new event's routine location (1 in Fig. 3);
2. Compute a preliminary DD location for the new event (2 in Fig. 3) by using pick delay times between new event and reference events;
3. Find reference events in the DD base catalog that are near the new event's preliminary (pick-based) DD location;
4. Compute delay times between new event and reference events from phase picks and by cross-correlating pairs of seismograms recorded at common stations;
5. Invert pick and cross-correlation delay times for final DD solution (3 in Fig. 3).

Because network locations are often mislocated by distances greater than distances over which events typically correlate, a first relocation step is carried out in which a preliminary DD location is produced based on phase picks alone (location 2, Fig. 3). This step aims at moving a new event to within a few hundred meters of its neighboring events in a DD base catalog, assuring that the neighboring reference events are within distances over which correlated events produce similar seismograms. During the second relocation step, the first step is repeated but now using the improved location to find new reference events in the DD base catalog for which cross-correlation delay times are measured and used in addition to the pick delay times in the inversion for relative locations. In addition, the correlation output is screened to identify highly correlated neighbors to help

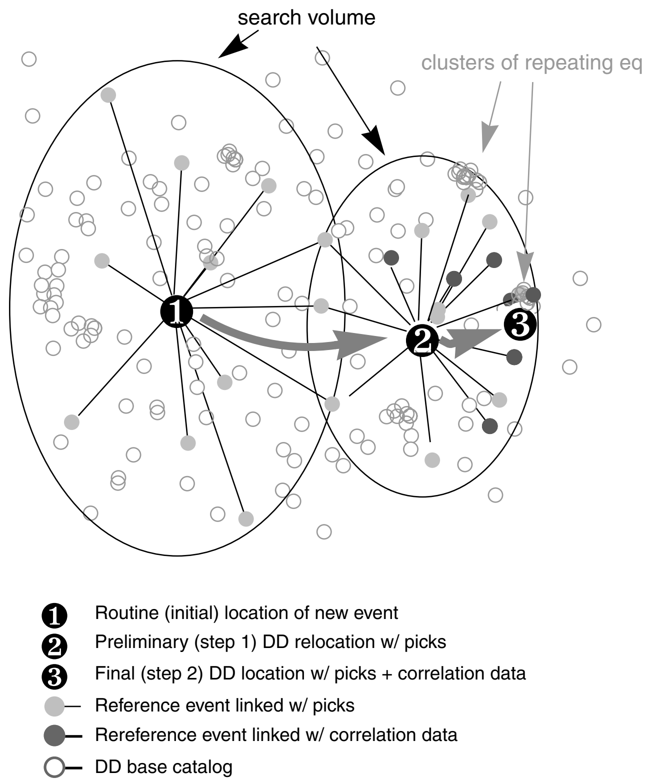

Figure 3. Schematics illustrating the two-step relocation approach. Location 1 is the initial location of a new event determined from routine network operation. Location 2 is the DD location based on phase picks from reference events near location 1. Location 3 is the final DD location based on picks and cross-correlation data from reference events near location 2. Large ellipses represent the search volume for reference events.

narrow the search for best reference events. If none of the reference events correlate with the new event, or if no waveforms are available, then the second relocation step is performed on picks alone. There is the possibility to iterate on step two, at the cost of processing speed, to refine final locations. However, the following applications were carried out using only the two relocation steps as additional iterations have shown to yield only minor improvements.

Selection of Neighboring Reference Events

The goal when searching for neighboring reference events is to find events in the DD base catalog that correlate with the new event and/or are nearest to the new event's final DD location. In areas of dense seismicity the required search volume can include thousands of potential reference events, and a subset needs to be found in order to limit the computational load in the subsequent DD analysis. Finding that subset rapidly can be difficult, especially in areas with large clusters of correlated events where a pure nearest neighbor search, for example, may collect all reference events within a single large multiplet of which the new event is not a member. To assure a spatially homogeneous subsampling, reference events are selected within each of five concentric, vertically elongated ellipsoidal layers of increasing thickness (Fig. 4). The dimension of the ellipsoid (D in Fig. 4) is based on the assumed error in the routine locations and the maximum interevent distance over which model errors can still be effectively minimized (see Waldhauser, 2001). The layers are twice as thick at the poles than they are at the equator to accommodate the generally larger error in the depth of the routine location of a new event. Each layer is split up into its eight quadrants (or cells), and only a limited number of events, in our application in northern California we choose the most recent five, are selected from each of the 40 cells. Pick information for the 200 selected events are then collected from the locally stored bulletin, and the 100 events

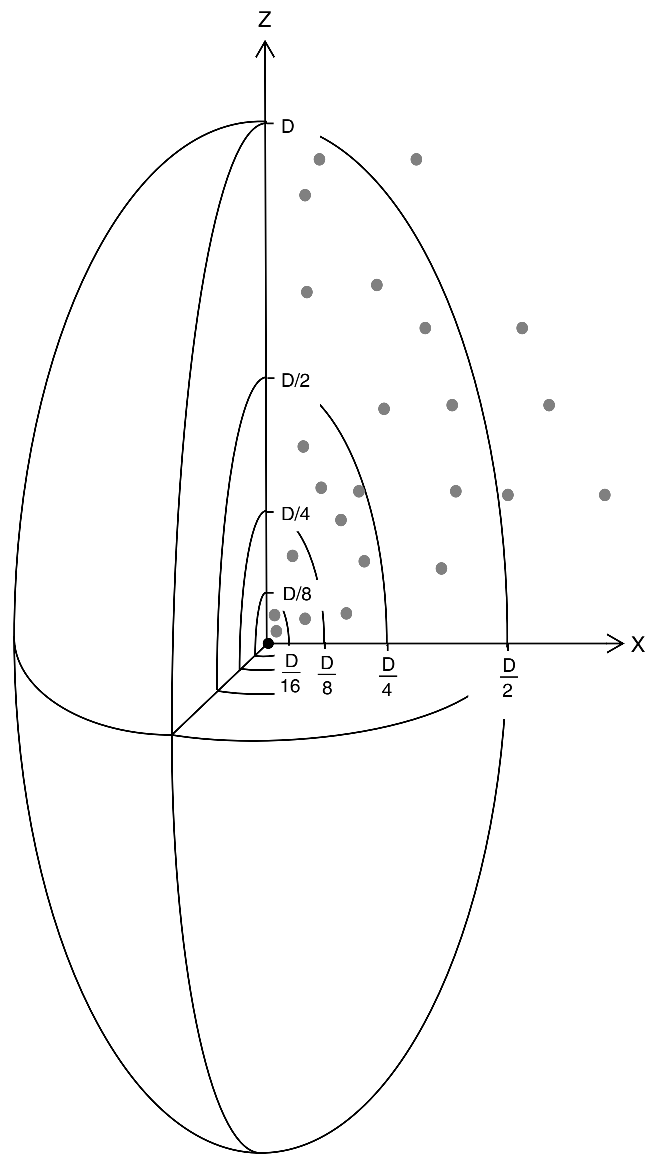

Figure 4. Schematic illustrating the search scheme for nearest reference neighbors. A limited number of events (dots) are collected within each quadrant of increasingly thicker ellipsoidal layers. D is typically between 5 and 15 km. The number of layers can vary. See text for discussion.

with the strongest links to the new event (as determined by a combination of number of stations, pick quality, and distance) are chosen as reference events for further processing. The choice of parameters for reference event selection depends on the characteristics of the archived seismicity data and its distribution and the configuration of the network.

Double-Difference Inversion

Differential arrival times between a new event and its reference events, measured at common stations from routine phase picks and/or from waveform cross correlation, are simultaneously inverted in an iterative weighted least-squares procedure for adjustments in the location of the new event relative to its reference events in the DD catalog. We use a modified version of the DD algorithm hypoDD (Waldhauser and Ellsworth, 2000; Waldhauser, 2001) that uses only the equations that constrain a new event, i, relative to its $nref$ reference events, j, in the DD catalog:

$$\frac{\partial t_k^{i=1}}{\partial \mathbf{m}} \Delta \mathbf{m}^i - \frac{\partial t_k^{j=1,nref}}{\partial \mathbf{m}} \Delta \mathbf{m}^j = dr_k^{ij}, \qquad (1)$$

where dr_k^{ij} is the residual between observed (dt^{obs}) and predicted (dt^{cal}) phase travel-time difference between a new event and its reference events observed at a common station, k (see also equation 18 in Waldhauser and Ellsworth, 2000). $\Delta \mathbf{m}$ are changes in the vector connecting their hypocenters through the partial derivatives of the travel times, t, for each event with respect to the vector, $\mathbf{m}$, of the four unknowns. The data is weighted according to an *a priori* quality weight and a dynamically adjusted weight that accounts for residual performance and interevent distance during individual iterations (see Waldhauser and Ellsworth, 2000).

Changes in the hypocentral parameters of the reference events, $\Delta \mathbf{m}^j$, are damped by adding to the system of DD equations four additional equations per event. The system then becomes

$$\mathbf{W}\begin{bmatrix} \mathbf{G} \\ \lambda \mathbf{I} \end{bmatrix}\begin{bmatrix} \mathbf{m}^{i=1} \\ \mathbf{m}^{j=1,nref} \end{bmatrix} = \mathbf{W}\begin{bmatrix} \mathbf{dr} \\ \mathbf{0} \end{bmatrix}, \qquad (2)$$

where $\mathbf{G}$ is the design matrix holding the partial derivatives, $\partial t/\partial \mathbf{m}$, $\mathbf{W}$ is a diagonal matrix with weights for each equation, and $\mathbf{m}$ and $\mathbf{dr}$ are the vectors that contain the model parameters and data residuals, respectively. λ is a vector of length $4(1 + nref)$ containing the damping values. We set $\lambda(1, 4) \ll 1$ to allow for changes in the parameters of the new event, while $\lambda[5, 4(1 + nref)]$ is set to ~ 1 to constrain changes in the parameters of the reference events that are assumed to be well located. In general, this formulation results in very small changes in the reference events while the new event moves towards the location that minimizes the delay-time residuals. In cases, however, where a new event is recorded with higher quality data than its reference events, it is possible that the new data can improve on the relative locations between the reference events and thus can produce significant changes in the location of the reference events. This is especially true in areas with sparse historic seismicity or during an aftershock sequence. We note that the changes in reference-event locations do not get propagated to the DD base catalog during real-time DD processing. However, changes in reference-event locations are monitored during real-time processing as they may indicate the need for ad hoc updating of the DD base catalog in a particular area to account for new higher quality data.

The number of rows in $\mathbf{G}$ is relatively small during real-time DD processing because only delay-time links between a new event and its reference events are used (compared to links between all possible pairs of events in regular DD runs). Therefore, the system of DD equations can be efficiently solved by singular value decomposition, and the full covariance matrix is analyzed to derive formal errors in relative location and origin time. The real-time DD algorithm has some similarity to the master-event location method (e.g., Evernden, 1969), except that in the master-event approach many slave events are relocated relative to one master event, while here we relocate one slave event (new event) relative to a set of master events (reference events).

Test-Bed Implementation and Results

A prototype system that implements the DD-RT process is currently running on a test-bed server at Lamont-Doherty Earth Observatory (LDEO) with near-real-time data feeds from earthquakes recorded at stations of the Northern California Seismic System (NCSS) (Figs. 2 and 5a). The NCSS, which assimilates data from 13 seismic networks (see the Data and Resources section), records an average of ~50 earthquakes on 1200 channels each day. The majority of earthquakes occur in diverse and complex tectonic settings such as the San Andreas fault system, the volcanic region of Long Valley Caldera, and the Mendocino Triple Junction. In addition, large numbers of anthropogenic earthquakes are induced at the Geysers geothermal field by geothermal production activities.

Parametric data of new events, including phase picks and hypocentral information, are automatically sent from the NCSN in Menlo Park, California, to the LDEO server, in the form of Hypoinverse archive files (Klein, 2002). Seismograms of new events, 2 min long and starting 10 sec before the origin time, are accessed from the real-time DART server at NCEDC in Berkeley via the Simple Waveform Client (swc) protocol (Romanowicz *et al.*, 1994; D. Neuhauser, personal comm., 2008) for those stations that also have archived seismograms for the reference events (Fig. 2). The parametric and waveform data for the reference events are retrieved from a locally stored copy of the entire NCSN archive from 1984 to 2003 (the period for which we currently have a DD base catalog), totaling about 700 Gb in size. The archive consists of ~6 million NCSN phase arrival time picks (mostly P

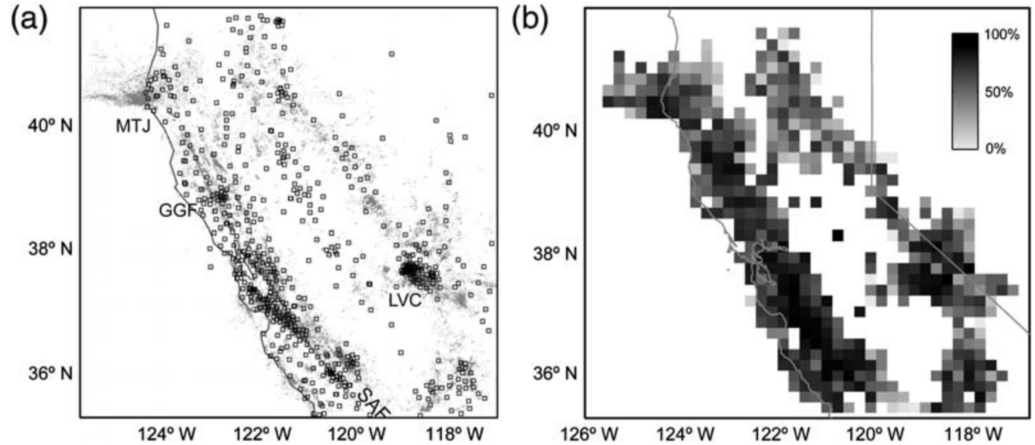

Figure 5. (a) Seismicity (gray dots) and stations (squares) of the Northern California Seismic System (NCSS). Event locations are from the DD base catalog described in Waldhauser and Schaff (2008). The catalog includes 311,273 earthquakes that occurred between January 1984 and May 2003. SAF, San Andreas fault; LVC, Long Valley Caldera; GGF, Geysers geothermal field; MTJ, Mendocino Triple Junction. (b) Percentage of correlated earthquakes across northern California, displayed within cells of 20 × 20 km. Only cells with 10 or more events are shown. Gray lines denote coast and state line.

phases) and some 15 million waveforms in seismic analysis code format. The waveforms, each including the *P*- and *S*-wave train of a particular event recorded at a particular station, are stored as individual files in a simple file system structure (Schaff and Waldhauser, 2005), with each file name including event ID, station name, and station component. This storage approach allows fast access to individual waveforms on an event/station pair basis.

A recently computed high-resolution DD earthquake hypocenter catalog for northern California (Fig. 5a; Waldhauser and Schaff, 2008) is used as the DD base catalog relative to which new events are relocated. The catalog includes ~300,000 events that occurred between 1984 and 2003. About 90% of the events are correlated (Fig. 5b), indicating that most of the new events are expected to share similar seismograms with at least one of their neighboring events. A maximum of 100 reference events are collected from the DD base catalog near the location of a new event following the search procedure outlined previously. The dimension of the search ellipsoid (*D* in Fig. 4) is chosen to be 10 km during the first relocation step with pick data and 8 km during the second relocation step with cross-correlation data.

Seismograms of new events are cross correlated with those of its reference events at common stations to compute differential arrival times using the time-domain correlation function described in Schaff *et al.* (2004) (see also Van Decar and Crosson, 1990; Dodge *et al.*, 1995). We choose correlation window lengths of 1.25 sec for both *P* and *S* phases and use delay times computed from seismograms with correlation coefficients $\mathrm{Cf} \geq 0.7$ for subsequent relocation (see Waldhauser *et al.*, 2004). The correlation windows are aligned on the routine picks and on predicted arrival times (using a simple 1D model) when no picks are available. The cross-correlation times are combined with the delay times from phase picks in the DD inversion. Local 1D layered velocity models are used to rapidly predict the partial derivatives, $\partial t/\partial \mathbf{m}$, and the differential times, dt^{cal} (see equation 1). The models are the same as used by Waldhauser and Schaff (2008) to relocate the NCSN catalog and are based on those used by the NCSN to locate events in northern California on a routine basis (Oppenheimer *et al.*, 1993).

Pick-based DD solutions generated during the first relocation step are computed within a few tens of seconds after receiving the initial location and phase information, using a single Sun UltraSparc IIIi processor. Final DD-RT solutions computed during the second relocation step using cross-correlation data (if available) are typically available in less than a minute. New events that occur in areas of dense seismicity with numerous strongly correlated reference events may take longer because of the additional time spent on accessing and cross correlating the seismograms and on inverting the larger system of DD equations. Real-time DD locations are obtained for all new events that have at least one reference event with at least four *P*-phase picks at common stations. New events that cannot be relocated, such as events that occur in areas with no historic activity or are recorded by new stations only, are flagged as such and added to the continuous DD catalog. Subsequent events that occur nearby will be automatically relocated relative to the initial (i.e., single-event) location of the first event. Periodic relocation of the continuous DD catalog will eventually produce a new DD base catalog for this previously aseismic area.

The DD-RT process automatically generates figures and tables that summarize key parameters describing the solution quality and characterizing the hypocenter's location relative to the past seismicity. Figure 6 shows an example summary figure for an M_{L} 1.5 event that occurred on the San Andreas fault near Parkfield on 8 May 2008. It displays the final DD-RT location (red dot in Fig. 6) relative to its routinely determined solution (green star) and relative to the historic

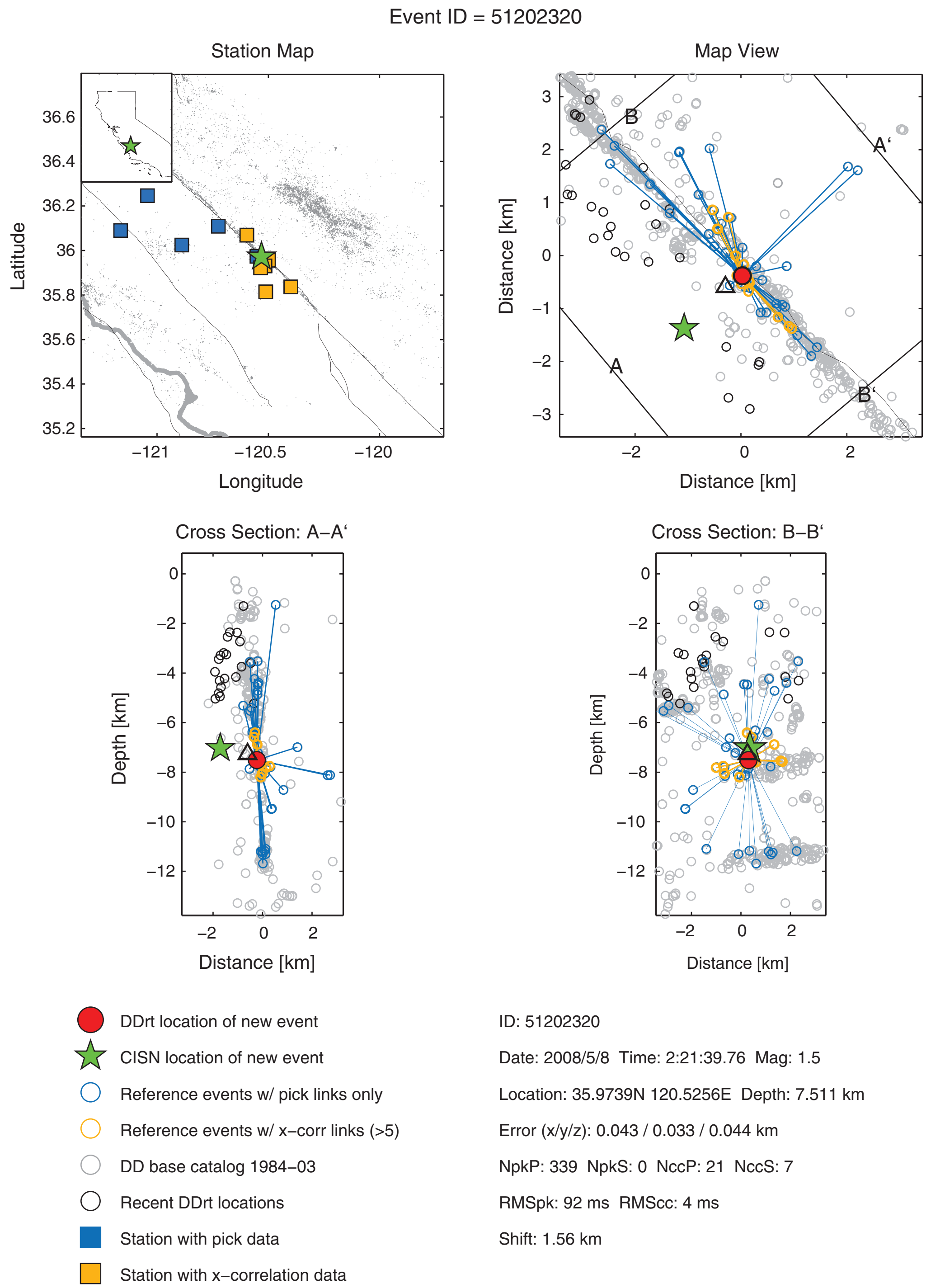

Figure 6. Example of automatically generated figure displaying and summarizing the DD-RT relocation performance and results for each event relocated. The event shown occurred on 8 May 2008 and had a magnitude of M_L 1.5.

seismicity as present in the DD base catalog (gray circles). The example illustrates how the hypocenter of the event, initially located about 1.5 km to the southwest of the San Andreas fault, falls onto the main fault after relocation. The preliminary pick-based DD solution for this event (shown as a triangle in Fig. 6) is less than 300 m from the final cross-correlation based DD location. The focal mechanism of this event, which would be difficult to estimate from first motion data alone because of its size, can be inferred from the subvertical dip and northwest strike of the main San Andreas fault near Parkfield. In the future, focal mechanisms for such small events may be computed on a routine basis by combining first motion data from highly correlated past events to form composite solutions (Reasenberg and Oppenheimer, 1985).

With high-precision DD locations available in near-real time, we can quickly put a new event into the spatiotemporal context of the nearby historic seismicity that may have been characterized in detail in previous separate studies. In our example case an onfault view of the Parkfield seismicity shows that the new earthquake (× in Fig. 7) ruptured within the upper streak of the two deep streaks just southeast of Middle Mountain (Waldhauser *et al.*, 2004, Bakun *et al.*, 2005; Thurber *et al.*, 2006) (Fig. 7). This upper streak is interpreted by Waldhauser *et al.* (2004) as the upper bound of the rupture area near the nucleation point of the 1934 and 1966 $M \sim 6$ earthquakes (Bakun and McEvilly, 1979) and appears to mark the upper bound of the rupture area at the tail end of the 2004 M 6 event that started about 30 km to the south (Bakun *et al.*, 2005). During the interseismic cycle, the area between the two deep streaks is believed to be locked (see Waldhauser *et al.*, 2004), and thus an increase in seismicity rate or earthquake size in either of the two streaks may indicate increased loading of the locked patch and the potential to bring the fault patch closer to failure in a large earthquake. Thus, the ability to rapidly place new microearthquakes within the spatiotemporal pattern of past seismicity on a given fault may prove valuable in assessing its short-term hazard.

Performance Evaluation

A back-testing scheme is employed to evaluate the overall performance of the DD-RT procedure in northern California. We treat past events as new events and relocate them relative to the DD base catalog while excluding data associated with the archived new event. The DD-RT locations are then evaluated by their deviation from the corresponding location in the DD base catalog. It is noted here that we assume that the events in the DD base catalog are located with high precision, and that their location accuracy is generally better compared to the routine locations because of the improvement in location precision (see Waldhauser and Schaff, 2008). We use two sets of back-testing events: 26 events in the San Andreas Fault Observatory at Depth (SAFOD) target clusters, and 2360 randomly selected events across northern California that occurred in the year 2002. These experiments were also used to fine-tune the various DD-RT parameters necessary to run the procedure automatically.

SAFOD Target Events

The San Andreas fault near Parkfield, and the site of the SAFOD in particular, is one of the most intensively monitored regions in the world. The SAFOD target area includes three

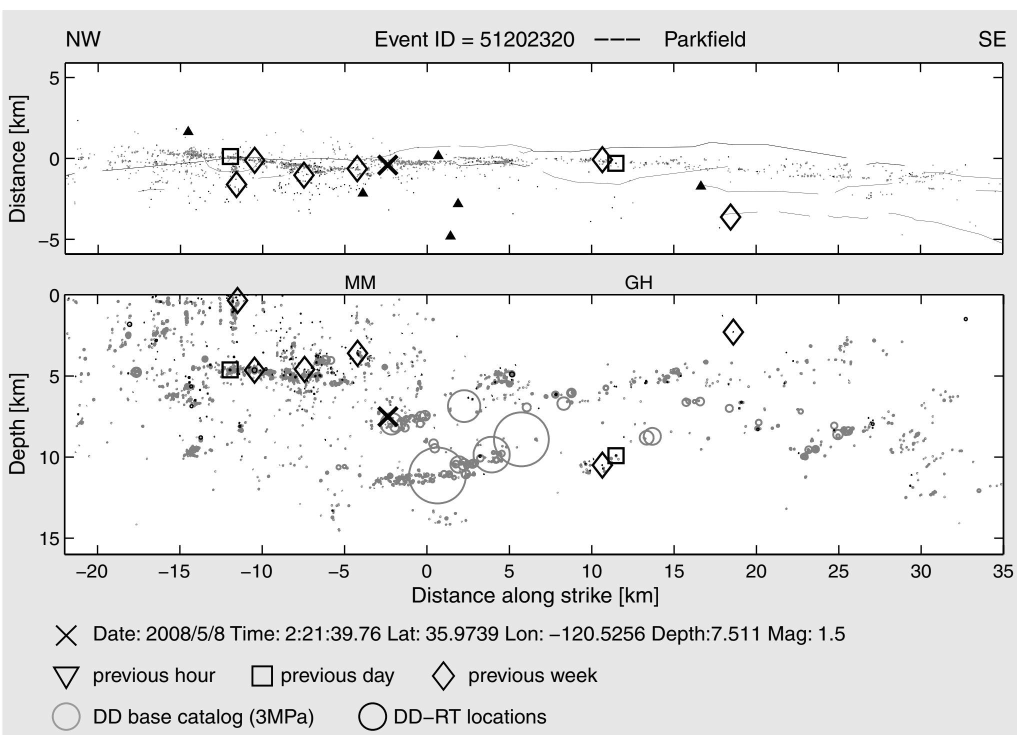

Figure 7. Automatically generated figure for example event showing map view (rotated into east–west direction) (top panel) and longitudinal cross section (bottom panel) of San Andreas fault seismicity near Parkfield. Gray circles, events in the DD base catalog (Waldhauser and Schaff, 2008) (size of circles represent 3 MPa circular rupture area); black circles, past DD-RT locations; diamonds, DD-RT locations from the last 7 days; squares, from the last 24 hr; inverted triangles, from the last hour. × denotes the most recent M 1.5 example event for which a DD-RT hypocenter location was computed shortly after it occurred on 8 May 2008 (see also Fig. 6). Lines in the top panel are mapped surface traces. Solid triangles are stations.

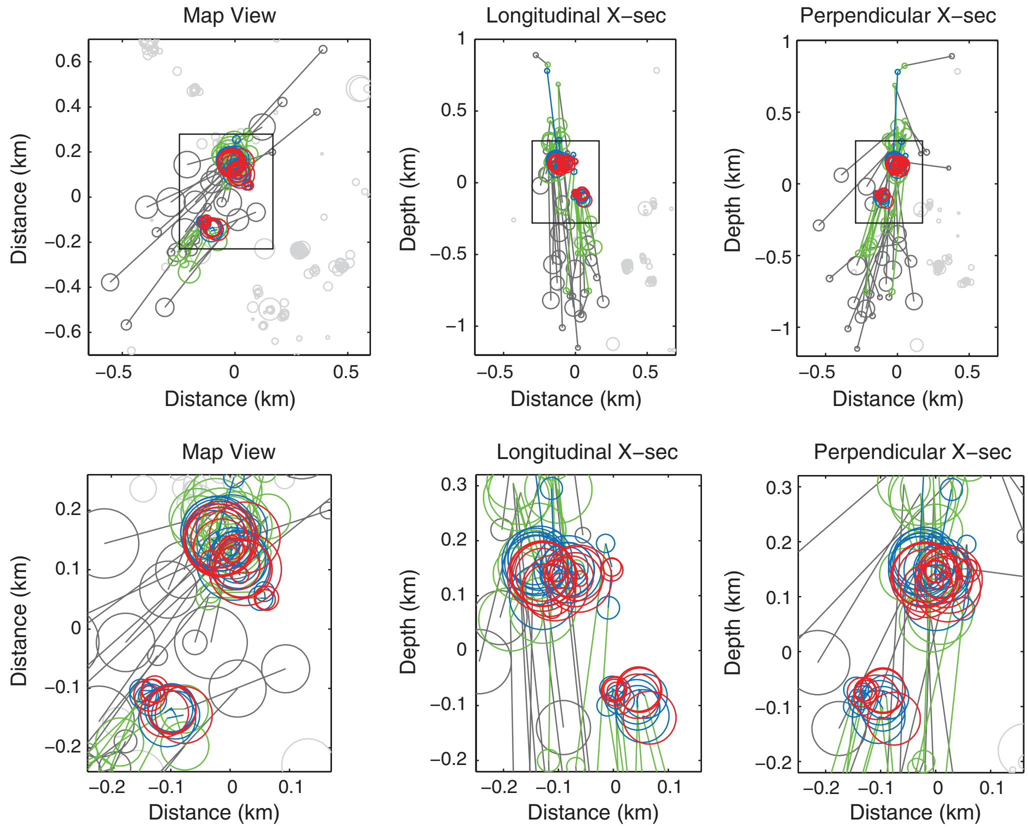

Figure 8. Back-testing results for 26 events in the SAFOD target clusters. Top panels show map view and the along fault and fault perpendicular cross sections; bottom panels show the same but zoom into the source region. Routine (NCSN) locations (black), pick DD-RT locations (green), and cross-correlation DD-RT locations (blue) are shown. DD base catalog locations are shown in gray, with the SAFOD events highlighted in red. Lines connect the hypocenters determined during different location steps.

clusters of repeating earthquakes at ~2 km depth with magnitudes M_L ~1.7–2.2. We select 15 tightly clustered and highly correlated events (Fig. 8). Locations for these events in the DD base catalog show that two clusters appear to rupture the same fault strand about 60 m apart, while a third cluster ruptures a different fault strand about 280 m to the southwest and 140 m below the former two clusters. We add 11 smaller magnitude events (M_L ~1) to the test dataset that locate mostly at the edge of the rupture areas of the larger repeating events as estimated from a 3 MPa constant stress drop model.

The DD-RT procedure uses an average of 82 reference events for each of the 26 SAFOD target events. The number of waveform pairs collected for each event ranges from 439 to 1463, from which 85–386 delay times were measured via cross correlation. In addition, between 440 and 900 delay times were computed for each event from differencing the travel-time picks. The median of the 26 differential time root mean square (rms) values drops from 35 msec before to 3 msec after relocation (Table 1).

Figure 8 shows the results for both the first relocation step using picks alone (green) and the second step using cross-correlation data in addition to the pick data (blue). The target events as located in the DD base catalog are shown in red. The pick-based relocations move the test events on average to within 0.05 km horizontally and 0.25 km

Table 1
Back-Testing Results I: Deviations in Horizontal and Vertical Directions of the NCSN and the Pick-Based and Cross-Correlation Based DD-RT Solutions from the Corresponding Locations in the DD Base Catalog for 26 SAFOD Target Events

	DX (km)			DY (km)			DZ (km)			rms (msec)		
	Median	Mean	Maximum	Median	Mean	Maximum	Median	Mean	Maximum	Median	Mean	Maximum
NCSN	0.19	0.20	0.57	0.12	0.19	0.54	0.70	0.61	1.30	35	35	43
DD-RT$_{pick}$	0.04	0.05	0.15	0.05	0.07	0.19	0.25	0.29	0.90	21	21	30
DD-RT$_{xcorr}$	0.01	0.01	0.04	0.01	0.02	0.14	0.02	0.07	0.70	3	5	22

DX is the west–east direction, DY is the south–north direction, and DZ is the vertical direction.

vertically of the corresponding locations in the DD base catalog (see Table 1). After adding the cross-correlation data during the second relocation step, the DD-RT solutions are within 0.01 km horizontally and 0.02 km vertically. In comparison, the deviations of the routine locations from their corresponding position in the DD base catalog have medians of 0.19 km horizontally and 0.70 km vertically. Pick-based solutions represent a factor of ~3 improvement over routine locations, and cross-correlation based locations improved by a factor of ~20 horizontally and ~35 vertically. The large improvement in the vertical direction in the later locations is mostly due to *S*-wave delay-time measurement obtained from cross correlating new with reference waveforms. Pick-based DD-RT locations are generally based on *P* phases only, as *S* waves have rarely been picked in the past at the NCSN.

Regional-Scale Back Testing

We extended the back-testing procedure to evaluate the DD-RT performance using events across all of northern California. We use a total of 2360 events in the year 2002, selecting them randomly within cells of 20 × 20 km (max 50 events in each cell) to achieve uniform aerial coverage. The median rms residual for all events after relocation is 4 msec for cross-correlation data and 30 msec for the pick data, down 94% and 52%, respectively, from the 64 msec seen in the routine locations (Table 2, Fig. 9).

A summary of the relocation results showing differences between the two DD-RT locations (picks alone and picks combined with cross-correlation data) and the corresponding locations in the DD base catalog is given in Table 2 and Figure 10. The median horizontal (vertical) deviation of the final DD-RT locations is 0.08 (0.24) km, compared to 0.34 (0.73) km in the routine locations (Table 2). The highest precisions are achieved along the San Andreas fault and in Long Valley where epicentral differences have medians generally less than 0.05 km (Fig. 10c, top panel). Differences in depths are generally less than 0.5 km and less than 0.1 km in well-monitored regions such as along the San Andreas fault and at the Geysers geothermal field (Fig. 10c, bottom panel). Pick-based DD-RT relocations are typically within 0.1 km horizontally and 0.5 km vertically along the San Andreas fault and Long Valley, and within 0.5 km horizontally and 1 km vertically elsewhere (Fig. 10b).

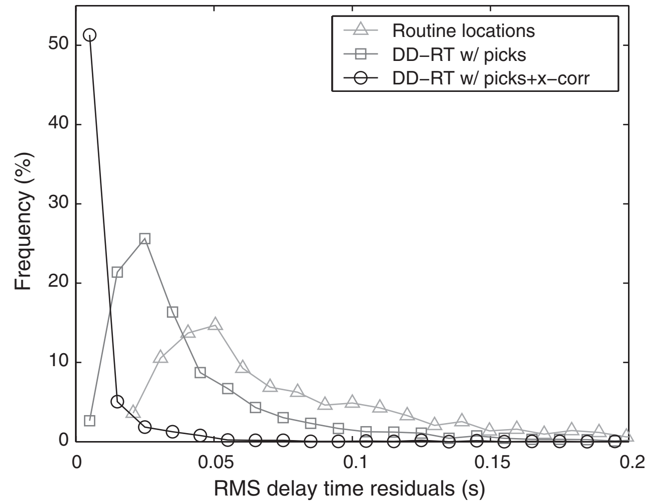

Figure 9. Distribution of rms residuals for the 2360 events used for back testing. Median values within bins of 0.01 sec are shown.

Discussion and Conclusions

The DD-RT locations of the 2360 back-testing events in 2002 represent a significant improvement over the corresponding routine locations across most of northern California (Fig. 11). We achieve horizontal (vertical) location improvement of a factor of 10 or better for 12% (3%) of the seismically active area (as calculated by the number of cells). For a factor of 5 or better these values are 41% (19%) and for a factor of 2 or better, 82% (57%). Epicenter locations improve in all areas, while depth locations appear to get worse in 3% of the area. The precision with which we can relocate new events relative to the background seismicity is increasing with time, as the density of earthquakes increases along active faults (Waldhauser and Schaff, 2008). This is especially true for events that occur in regions with low-seismicity rates. To illustrate this effect we back tested ~2000 events in each of the years 1998, 1994, and 1990, in each case only using the events in the DD base catalog that occurred before the respective year in which the test events were collected. (Note, however, that locations in this truncated catalog are still based on the full time period.) A comparison of the location improvement factors for the test events in each of the four years indicate

Table 2

Back-Testing Results II: Deviations in Horizontal and Vertical Directions of the NCSN and the Pick-Based and Cross-Correlation Based DD-RT Solutions from the Corresponding Locations in the DD Base Catalog for 2360 Events that Occurred Across Northern California in the Year 2002

	DX (km)			DY (km)			DZ (km)			rms (msec)		
	Median	Mean	Maximum	Median	Mean	Maximum	Median	Mean	Maximum	Median	Mean	Maximum
NCSN	0.34	0.65	8.9	0.33	0.61	7.36	0.75	1.34	15.56	64	88	1797
DD-RT$_{pick}$	0.14	0.40	5.77	0.14	0.37	6.58	0.44	1.02	15.55	30	44	824
DD-RT$_{xcorr}$	0.08	0.29	5.73	0.07	0.27	9.00	0.24	0.80	15.55	4	10	661

DX is the west–east direction, DY is the south–north direction, and DZ is the vertical direction.

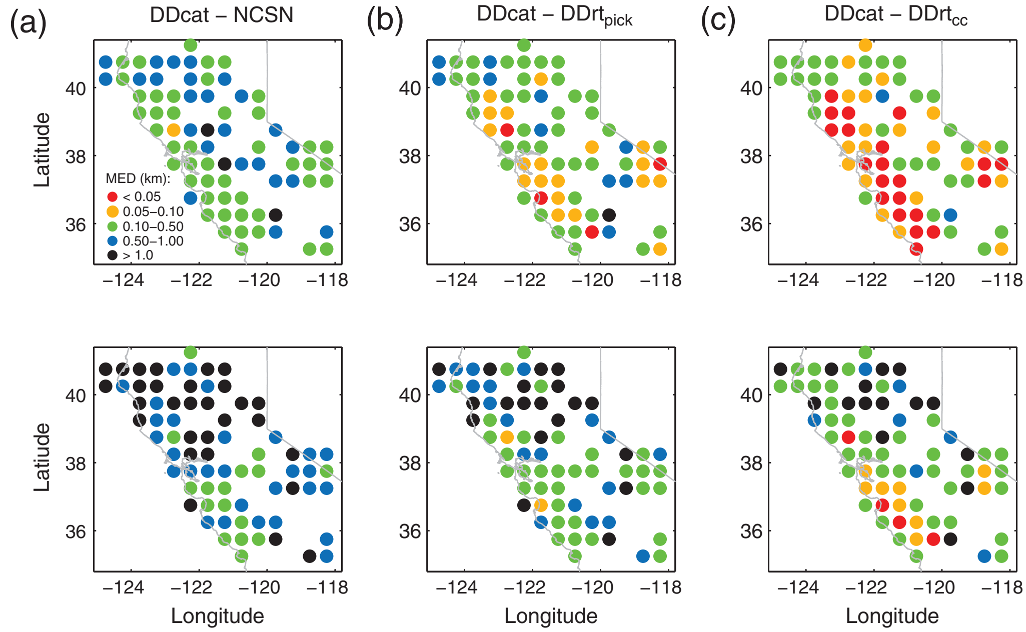

Figure 10. Summary of differences in epicenters (top panels) and depths (bottom panels) between 2360 locations in the DD base catalog and corresponding locations in the NCSN catalog (left-hand panels), the pick-based DD-RT locations (middle panels), and the cross-correlation based DD-RT locations (right-hand panels). Each dot indicates the median of all differences within 20 × 20 km cells. Only cells with at least five events are shown.

the continuously increasing precision with which new events are relocated relative to the background seismicity (Fig. 11).

The importance of having critical event density near a new event to achieve high-precision relocations is demonstrated in Figure 12 for the 2002 test events. It shows that the median mislocations of the DD-RT solutions from the corresponding locations in the DD base catalog are high for events with few reference events and flattens to less than 200 m horizontally when ~15 or more reference events are found and used for relocation. An average of 41 reference events, out of 100 searched for, were found for each event. While the percentage of test events with less than 15 reference events is 28% in the year 2002, that number was 31% in 1998, 37% in 1994, and 40% in 1990 (Fig. 13). The

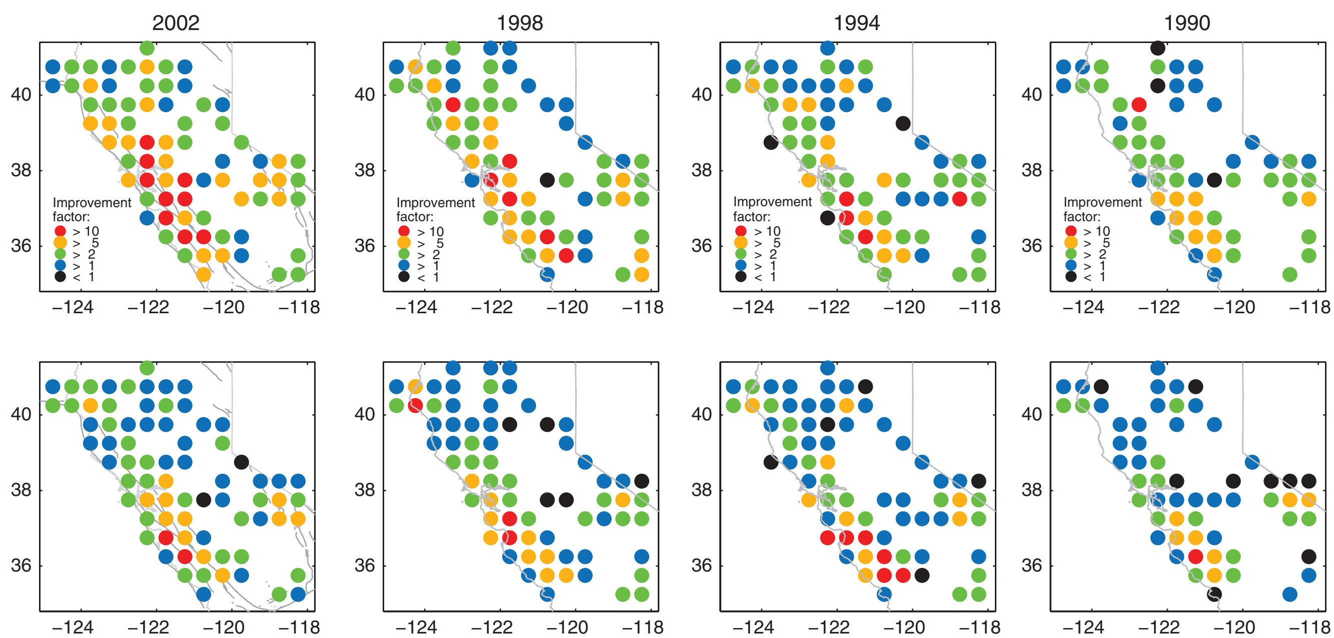

Figure 11. Improvement in horizontal (top panels) and vertical (bottom panels) locations of DD-RT solutions over the corresponding routine locations for the years 1990, 1994, 1998, and 2002, shown as median improvement factors within cells of 20 × 20 km. Only cells with at least five events are shown.

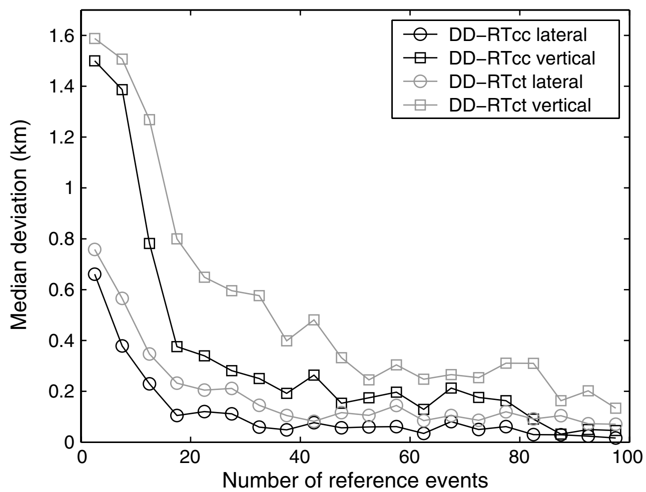

Figure 12. Median mislocations, computed within bins of 5 reference events, of back-tested DD-RT solutions for 2360 events that occurred in 2002, shown as a function of number of reference events available for relocation.

year-over-year increase in the number of reference events in any given area is a main contributor to the increase in location precision seen in Figure 11.

To assure continued improvement in real-time DD locations, periodic relocation of the continuous DD catalog to generate a new, updated, DD base catalog is required (Figs. 1 and 2). Periodic catalog relocation guarantees that the locations of events in the DD base catalog, which do not change once that catalog has been established, are adequately represented by the data of newer events. It is often the case that new events, with better quality data, can improve the relative location between nearby reference events in the base catalog. Therefore, in future work the cross-correlation database of

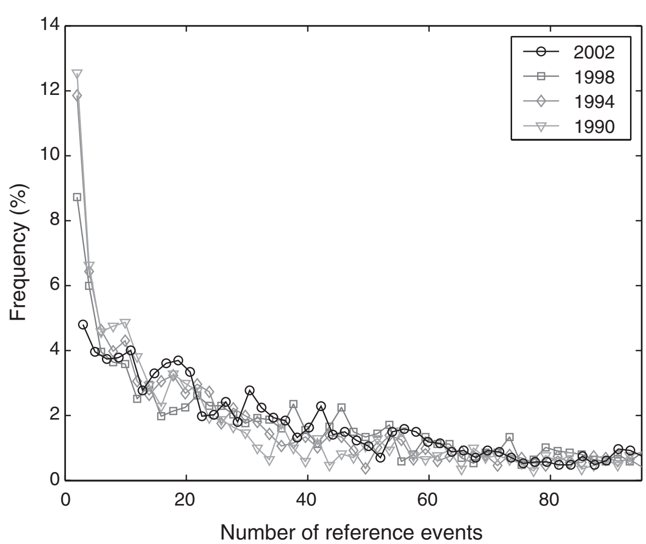

Figure 13. Distribution of the number of reference events per test event for four sets of back-testing events. Each set consists of ~2000 events that were collected in each of the years 1990, 1994, 1998, and 2002.

Schaff and Waldhauser (2005) needs to be updated with delay times measured between new events and all events in the DD base catalog that are within 5 km of the new event's location (and not just the reference events). In addition, cross-correlation delay times need to be measured between all new events added since the creation of the DD base catalog and separated by less than 5 km (see Schaff and Waldhauser, 2005, for details). Because the cross-correlation database is not needed during DD-RT processing, updating the cross-correlation database can be done periodically before generating a new DD base catalog or continuously while running as a background process. The updated seismic archive is eventually used for relocation of the entire catalog to create a new DD base catalog (see Waldhauser and Schaff, 2008, for details).

The DD earthquake monitoring procedure presented in this article is designed as an independent system running in post routine processing mode (see Fig. 2) and could be modified to work in other regions where a sufficient pick and/or waveform archive exists. Such an approach has the benefit of being able to be implemented, tested, and evaluated while assuring the uninterrupted and consistent production of catalogs and bulletins during existing routine network operations. Future work, however, needs to be dedicated to full integration of the DD-RT process into the real-time system at the California Integrated Seismic Network (CISN) to allow a more seamless interaction with the existing databases and archiving systems for more comprehensive data mining purposes. When vast seismogram archives can be rapidly accessed and mined via many CPUs in near-real time, then the task of picking phase onset times may become obsolete for correlated events as delay times could be computed directly by cross correlating new seismograms with hundreds of thousands of seismograms of potential reference events.

The availability of continuously updated DD earthquake catalogs in near-real time is expected to have considerable social and economic impact in the evaluation and mitigation of seismic hazards, and the catalogs are particularly valuable to the research community as they provide fundamental data in the geophysical sciences. For example, rapid knowledge of precise aftershock locations may be useful to delineate the rupture area of the mainshock, which may help scientists assess the potential for and size of future aftershocks as well as provide critical source information for strong ground motion prediction. Changes in the recurrence intervals of repeating events can be monitored to infer changes in the loading rate, or the short-term evolution of stress concentrations can be tracked in near-real time to better assess the potential occurrence of future events. Monitoring fine-scale changes in seismogenic behavior is particularly important along hazardous faults with well-characterized spatiotemporal behavior of past seismicity like the Hayward, Calaveras, and Central San Andreas faults.

The near-real-time DD hypocenter solutions for earthquakes in northern California produced by the DD-RT method described in this article are continuously being made available (see the Data and Resources section) on a beta-test basis.

Data and Resources

All parametric and waveform data used in this study are from the Northern California Earthquake Data Center (NCEDC) at University of California, Berkeley, and the Northern California Seismic Network (NCSN) at the USGS in Menlo Park, California, and can be accessed at www.ncedc.org (last accessed July 2009). The following institutions contributed with their seismic networks to the data used in this study: U.S. Geological Survey, Menlo Park; University of California, Berkeley; California Institute of Technology; University of Nevada, Reno; California Division of Water Resources; University of Utah; University of Southern California. Information on recent earthquakes is available at http://earthquake.usgs.gov/eqcenter/recenteqsus/ (last accessed July 2009). For near-real-time DD hypocenter solutions in northern California see www.ldeo.columbia.edu/~felixw/DDrtCISN (last accessed July 2009).

Acknowledgments

This work was supported by USGS–NEHRP Grant Number 06HQGR0054. We thank the NCSN and the NCEDC for making the real-time and archived seismic data available and easy accessible; David Oppenheimer, Doug Neuhauser, and Mahdad Parsi for help with data transfer and computing; David Oppenheimer, Jim Luetgert, Doug Neuhauser, Peggy Hellweg, David Schaff for discussions and suggestions; Cinna Lomnitz and an anonymous reviewer for comments that helped to improve the manuscript. This is LDEO Contribution Number 7289.

References

Bakun, W. H., and T. V. McEvilly (1979). Earthquakes near Parkfield, California: Comparing the 1934 and 1966 sequences, *Science* **205,** 1375–1377.

Bakun, W. H., B. Aagaard, B. Dost, W. L. Ellsworth, J. L. Hardebeck, R. A. Harris, C. Ji, M. J. S. Johnston, J. Langbein, J. J. Lienkaemper, A. J. Michael, J. R. Murray, R. M. Nadeau, M. S. Reichle, P. A. Reasenberg, E. A. Roeloffs, A. Shakal, R. W. Simpson, and F. Waldhauser (2005). The 2004 Parkfield (CA) Earthquake: Implications for prediction and hazard assessment, *Nature* **437,** 969–974.

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.

Evernden, J. F. (1969). Identification of earthquakes and explosions by use of teleseismic data, *J. Geophys. Res.* **74,** 3828–3856.

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.

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.

Got, J.-L., P. Okubo, R. Machenbaum, and W. Tanigawa (2002). A real-time procedure for progressive multiplet relative relocation at the Hawaiian Volcano Observatory, *Bull. Seismol. Soc. Am.* **92,** 2019–2026.

Ito, A. (1985). High resolution relative hypocenters of similar earthquakes by cross-spectral analysis method, *J. Phys. Earth* **33,** 279–294.

Klein, F. W. (2002), User's guide to HYPOINVERSE2000, a Fortran program to solve for earthquake locations and magnitudes, *U.S. Geol. Surv. Open-File Rept. 02-172*, 123 pp.

Oppenheimer, D. H., F. W. Klein, J. P. Eaton, and F. W. Lester (1993). The Northern California Seismic Network bulletin, January–December 1992, *U.S. Geol Surv. Open-File Rept. 93-578*, Menlo Park, California.

Pavlis, G. L. (1986). Appraising earthquake hypocenter location errors: A complete, practical approach for single-event locations, *Bull. Seismol. Soc. Am.* **76,** 1699–1717.

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.

Reasenberg, P. A., and D. Oppenheimer (1985). FPFIT, FPPLOT, and FPPAGE: FORTRAM computer programs for calculating and displaying earthquake fault-plane solutions, *U.S. Geol. Surv. Open-File Rept. 85-739*, Menlo Park, California, 109 pp.

Romanowicz, B., D. Neuhauser, B. Bogaert, and D. Oppenheimer (1994). Accessing northern California earthquake data via internet, *EOS Trans AGU* **75,** 258–261.

Schaff, D. P., and F. Waldhauser (2005). Waveform cross-correlation based differential travel-time measurements at the Northern California Seismic Network, *Bull. Seismol. Soc. Am.* **95,** 2446–2461.

Schaff, D. P., G. Bokelmann, W. L. Ellsworth, E. Zanzerkia, F. Waldhauser, and G. C. Beroza (2004). Optimizing correlation techniques for improved earthquake location, *Bull. Seismol. Soc. Am.* **94,** 705–721.

Shearer, P. M. (2002). Parallel fault strands at 9-km depth resolved on the Imperial fault, southern California, *Geophys. Res. Lett.* **29,** no. 14, doi 10.1029/2002GL015302.

Thurber, C., H. Zhang, F. Waldhauser, J. Hardebeck, A. Michael, and D. Eberhart-Phillips (2006). Three-dimensional compressional wavespeed model, earthquake relocations, and focal mechanisms for the Parkfield, California, region, *Bull. Seismol. Soc. Am.* **96,** 38–49.

Waldhauser, F. (2001). HypoDD: A computer program to compute double-difference earthquake locations, *U.S. Geol. Surv. Open-File Rept. 01-113*, Menlo Park, California.

Waldhauser, F., and W. L. Ellsworth (2000). A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, *Bull. Seismol. Soc. Am.* **90,** 1353–1368.

Waldhauser, F., and D. P. Schaff (2008). Large-scale relocation of two decades of northern California seismicity using cross-correlation and double-difference methods, *J. Geophys. Res.* **113,** B08311, doi 10.1029/2007JB005479.

Waldhauser, F., W. L. Ellsworth, D. P. Schaff, and A. Cole (2004). Streaks, multiplets, and holes: High-resolution spatio-temporal behavior of Parkfield seismicity, *Geophys. Res. Lett.* **31,** L18608, doi 10.1029/2004GL020649.

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. Seismol. Soc. Am.* **80,** 150–169.

Lamont-Doherty Earth Observatory
Columbia University
Palisades, New York 10964
felixw@ldeo.columbia.edu

Manuscript received 1 October 2008