JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 112, B05303, doi:10.1029/2006JB004463, 2007

Full waveform earthquake location: Application to seismic streaks on the Calaveras Fault, California

Justin L. Rubinstein[1,2] and Gregory C. Beroza[1]

Received 23 April 2006; revised 2 January 2007; accepted 19 January 2007; published 9 May 2007.

[1] We use a novel technique based upon source array analysis to locate three moderate earthquakes that occur at the edge of previously identified streaks of seismicity on the Calaveras Fault, California. Our method determines centroid locations for earthquakes, in addition to the hypocenters previously determined using first-break picks. Application of the method to smaller earthquakes indicates that the errors associated with the locations are on the order of 100 m, much less than the rupture dimensions of the M > 4.5 events that we have analyzed. We treat high-precision locations of microearthquakes near the earthquakes that we want to locate as source arrays and compute the slowness of waves leaving these source arrays. We then use the slowness parameters to locate the earthquakes of interest. We find that the medium-magnitude events nucleate on the streaks and rupture into a zone devoid of seismicity. On this basis, we argue that streaks represent the boundary between creeping and locked sections of a fault. Our location technique has the potential for wide application, including circumstances where it may be necessary or desirable to locate earthquakes without using direct arrivals.

Citation: Rubinstein, J. L., and G. C. Beroza (2007), Full waveform earthquake location: Application to seismic streaks on the Calaveras Fault, California, *J. Geophys. Res.*, *112*, B05303, doi:10.1029/2006JB004463.

1. Introduction

[2] Recently, the seismology community has made a considerable effort to improve the precision and accuracy of earthquake locations. To do this, seismologists have addressed the two largest sources of uncertainty in earthquake locations: unknown earth structure and inaccurate arrival time measurements.

[3] To remove the influence of unknown earth structure, *Waldhauser and Ellsworth* [2000] developed the double-difference method, a method that relies on differential arrival time measurements instead of absolute measurements. *Zhang and Thurber* [2003] took the method one step further and jointly solved for both earthquake locations and velocity structure. Double-difference-based methods work well for earthquakes of all sizes.

[4] Standard methods of earthquake location rely on arrival times that are picked manually or automatically generated, making them subject to either operator error and/or error due to emergent onset. Measurements based on these techniques are also limited in precision by the sampling rate of the seismometers, as picks (manual or automatic) cannot be more precise than the sampling rate. The timing precision of first arrivals is usually on the order of 10–30 ms for Northern California Seismic Network (NCSN) data with 100 samples per second digitization. The measurement of relative arrival times of earthquakes using waveform cross correlation [*Poupinet et al.*, 1984], on the other hand, can reduce the error of relative arrival time measurements to subsample precision ($\sim$1 ms) for similar or nearby earthquakes. This improvement allows relative earthquake location with errors on the order of meters to tens of meters. The locations produced by cross-correlation-based methods are treated as centroids, as cross correlation focuses on amplitudes which are controlled by the location of the highest moment rate (slip rate). For microearthquakes, the dimensions of which are small, this location is also treated as the hypocenter, as the distance between the hypocenter and centroid will be small, likely within the uncertainty of the measurement. Obviously, this does not apply for larger earthquakes, which have longer durations and physical extents, such that there can be a considerable distance between the hypocenter and centroid. The technique of applying cross correlation for improved relative arrival times has been thoroughly explored by *Schaff et al.* [2004].

[5] Although cross-correlation-based methods have proven very useful in improving the precision and accuracy of earthquake locations in many areas around the world, the data available cannot always support using these methods. Like standard earthquake location techniques, cross-correlation-based methods can have large uncertainties in their locations should there be a limited range of source-receiver azimuths available. Clipped data also pose a problem for cross-correlation methods and are typically avoided. Clipping, in particular, affects medium-magnitude and large earthquakes as their strong shaking can saturate

[1]Department of Geophysics, Stanford University, Stanford, California, USA.
[2]Now at Department of Earth and Space Science, University of Washington, Seattle, Washington, USA.

0148-0227/07/2006JB004463$09.00

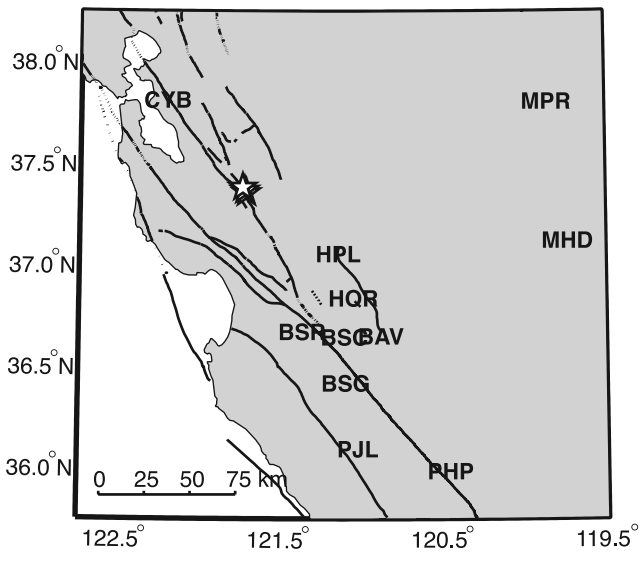

Figure 1. Map of the study region and the seismometers used. Stars indicate locations of the medium-magnitude earthquakes we are locating. Lines indicate faults.

both telemetry systems and the short-period networks typically used to locate earthquakes. Any combination of clipping, poor station geometry, or a general paucity of receivers (and therefore recordings of earthquakes) can seriously limit the utility of or even completely prevent the usage of the cross-correlation-based methods.

[6] In the interest of improving the locations of earthquakes where cross-correlation-based methods are not useful, we developed a new technique based upon seismic array analysis to improve earthquake locations for sparsely recorded earthquakes. A number of studies have recently used array analysis, based upon receiver arrays, to locate seismic sources [*Kao and Shan*, 2004; *Kao et al.*, 2005, 2006] and image larger events [*Ishii et al.*, 2005; *Krüger and Ohrnberger*, 2005a; 2005b; *Walker et al.*, 2005; *Fletcher et al.*, 2006]. In our method, we treat the precise microearthquake locations made possible by cross-correlation and double-difference relocation methods as an array of seismic sources [*Niazi*, 1969]. Using standard array analysis methods, we are able to determine the velocity and propagation direction of energy that is recorded at a single station for many different time windows. With this description of the slowness parameters at the source array, we are able to relocate nearby events by finding the location where its waveform best matches that predicted by the source array.

2. Geologic Setting and Motivation

[7] We focus our attention on the interplay between recently discovered microearthquake streaks and medium-magnitude earthquakes that occur within them. Streaks are lineations of seismicity within an individual fault that were first discovered on the south flank of Kilauea volcano, Hawaii [*Gillard et al.*, 1996], and have since been identified on the central San Andreas Fault [*Rubin et al.*, 1999; *Waldhauser et al.*, 2004], the Hayward Fault [*Waldhauser et al.*, 1999; *Waldhauser and Ellsworth*, 2002], and the Calaveras Fault [*Schaff et al.*, 2002]. In all of these locations the orientation of the streaks is approximately parallel to the direction of slip on the fault, which suggests that streaks are a slip-controlled process. Streaks are also consistently found in regions where a significant portion of the slip budget appears to be accommodated by creep. This suggests that the interplay between creep-slip and stick-slip behaviors may play a key role in the generation of this highly organized seismicity pattern. Thus, understanding streaks is important as we expect they will provide clues to the physics underpinning fault mechanics and how slip is partitioned on faults between creep and earthquakes.

[8] For this study, we are interested in determining precisely the locations of three medium-magnitude earthquakes. These earthquakes occurred on 26 May 1996, 10 November 1988, and 13 June 1988 and are the largest events that occurred in our study region over a period of 15 years. It follows that these events have had a strong influence on other earthquakes and creep within the area. The preliminary locations of these earthquakes (M_d4.6, M_L4.8, and M_L5.3) indicate that they are located at the ends of three streaks on the Calaveras Fault (Figures 1 and 2) [*Schaff et al.*, 2002]. These locations were determined using the double-difference method HypoDD [*Waldhauser and Ellsworth*, 2000], but cross correlations were not computed for the waveforms of these earthquakes as the high-gain channels on the NCSN were strongly clipped. As a result, only first breaks were used in the relocations of these events, so we expect the locations of these earthquakes will represent their hypocenters, but with larger location errors than for surrounding events as they lack the benefit of precise relative arrival time measurements. This differs from the locations for the cross correlation measured microearthquakes, which represent centroid locations. Fortunately, within the NSCN, some stations have both low-gain channels and high-gain channels, such that the medium-magnitude earthquakes we are interested in were recorded unclipped at a handful of stations.

3. Method

[9] Our technique is based upon standard array processing techniques. A thorough review of array techniques is given by *Rost and Thomas* [2002]. In standard array analysis, seismologists typically assume plane wave propagation through an array of receivers with known locations to predict the relative arrival time of phases within the receiver array. We use the reciprocal geometry, where we have an array of seismic sources and one receiver to determine the propagation patterns within our source array. A schematic description of this technique can be found in Figure 3, which is Figure 2 from *Dodge and Beroza* [1997]. This method was first proposed by *Niazi* [1969] and developed by *Spudich and Bostwick* [1987]. Others have formulated techniques that utilize both source and receiver arrays (double beam analysis [*Krüger et al.*, 1993] and double beam imaging [*Scherbaum et al.*, 1997]) to produce refined images of the mantle and the core mantle boundary region. Here we use source-array beam-forming to determine slowness parameters for multiple time windows within source arrays centered on the earthquakes that we want to locate.

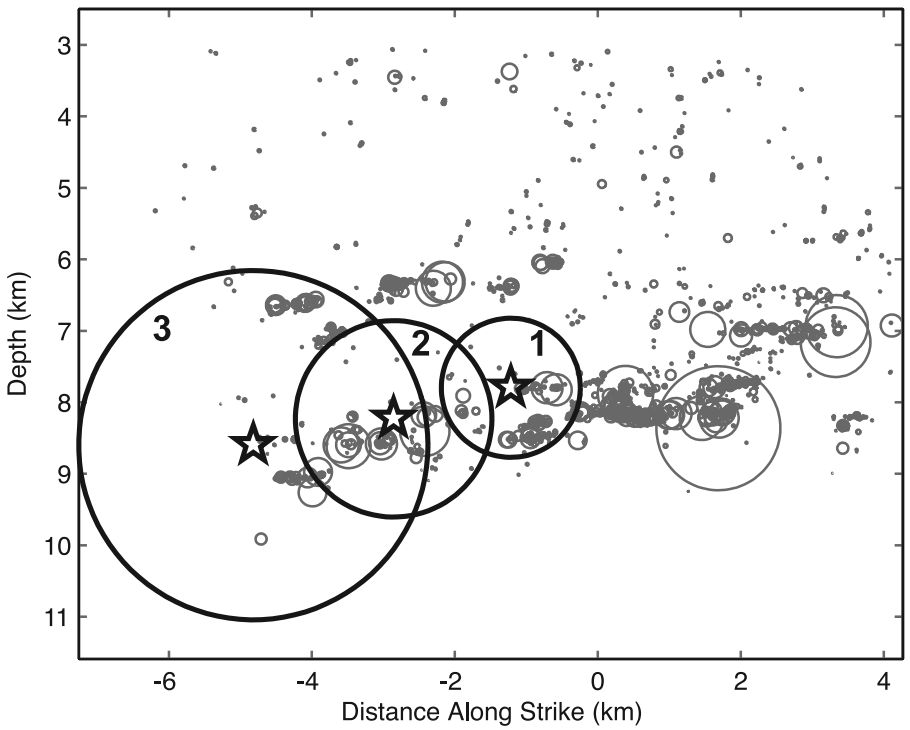

Figure 2. Cross section showing locations of seismicity on the Calaveras Fault determined by *Schaff et al.* [2002]. Circle size represents approximation of the source given the assumptions of circular rupture with a 30 bar stress drop. The dark, thick circles represent the approximate rupture extent of the three medium-magnitude earthquakes we are relocating. The stars represent hypocenters of these earthquakes as determined by double-difference relocation by *Schaff et al.* [2002]. The numbers indicate the earthquake number as referred to in the paper.

Once we have the slowness parameters, we search for the optimal location of the earthquake based on stacks of the source array and the waveforms of the earthquake we are locating. We discuss the method more fully below.

3.1. Precise Determination of Slowness Parameters

[10] We use the catalog of earthquakes produced by *Schaff et al.* [2002] as the basis of locations for this study. For each medium-magnitude earthquake that we are trying to locate, we choose a source array of earthquakes to determine wave propagation parameters. Our source arrays are defined as any earthquake in our catalog that falls within 3 km of the medium-magnitude earthquake that we are trying to locate. Each station has a slightly different source array, depending on which earthquakes it recorded. We require a minimum of 35 events contributing to the source array to ensure a sound basis for our slowness determination. The majority of the source arrays include many more than 35 events; in fact, many arrays include more than 100 events. We only assemble source arrays for stations where a reliable, low-gain recording of the medium-magnitude earthquake in question was available (the stations that we used to locate each earthquake are indicated in Table 1, see also Figure 1).

[11] Prior to beam-forming, we band-pass filter the seismograms from 1 to 6 Hz and normalize the power of the traces to 1 for a 24 s window starting 2 s before the *P* arrival. Traces where the signal-to-noise ratio does not exceed 3 and seismograms that are clipped are removed.

[12] Given the source arrays, we then use delay and sum beam-forming to determine wave propagation parameters for 20 time windows. The time windows are 2 s long, with the first window centered on the *P* arrival. The windows are stepped forward at increments of 1 s, such that they overlap with adjacent windows by 1 s. We determine the optimal slowness parameters for each of these 20 independent windows. For any trial slowness parameters $\tilde{p}_{acd}$, the source array stack $\tilde{S}_{acd}$ is determined to be

$$\tilde{S}_{acd}(\tilde{p}_{acd}, t_c) = \sum_{b=1}^{N} w_{ab}(t_c - \tilde{p}_{acd} \cdot x_b) + n_{abc} \quad (1)$$

with

- a station number;
- b event number in the source array;
- c time window;
- d bin number for trial slowness;
- t time;
- w the waveform [assumed constant over the whole array];
- $\tilde{p}$ trial wavefield slowness;
- x the relative position of the source;
- N the number of events recorded by station a in this source array;
- n noise.

[13] We use tildes to indicate that a parameter/variable has a trial value, we search over trial parameters to optimize a value.

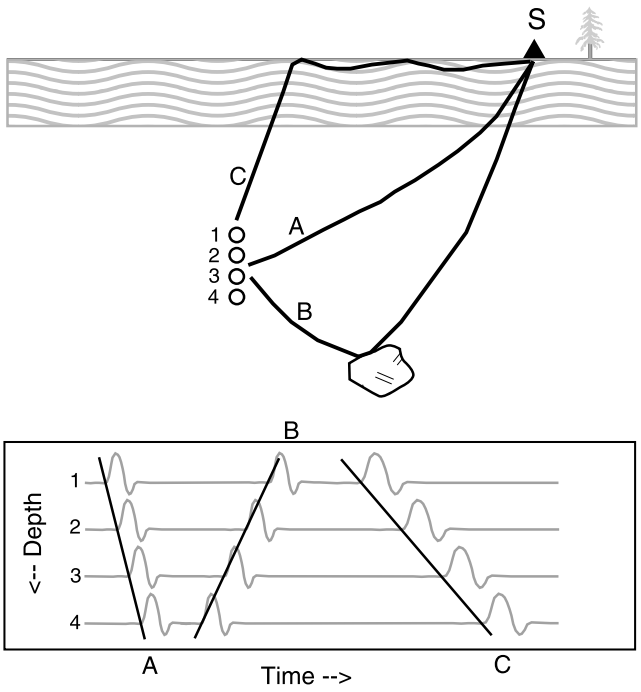

Figure 3. (top) Schematic representation of source array analysis from *Dodge and Beroza* [1997], which has a vertical column of earthquake hypocenters numbered 1 to 4 and the raypaths for three arrivals A, B, and C. (bottom) Simplified waveforms for these three arrivals and their differing moveouts that depend on the path that they took leaving the source array.

[14] Referring back to Figure 3, we assume we know our source locations very well (i.e., 1–4), but the propagation velocity and paths leaving the source array (i.e., A–C), and hence the slowness of arrivals, we treat as unknowns. We use a grid search method to determine the optimum slowness p_{ac}, by searching over velocity, azimuth, and angle of incidence and compute the trial slowness $\tilde{p}_{acd}$. For azimuth and angle of incidence, we use a grid spacing of 5°. We search over two velocities (3175 m/s and 5500 m/s), which are the *S* and *P* velocities that *Schaff et al.* [2002] used at the depths of the medium-magnitude earthquakes we are interested in. The velocities at these depths will not perfectly match these numbers, but they are consistent with the relative locations of the events within our source arrays as originally determined by *Schaff et al.* [2002]. The optimum slowness parameters p_{ac} for any station *a* and time window *c* are selected to be those where the power of our source array stack $\tilde{S}_{acd}$ is maximized.

[15] For constructive interference to occur and our stacking procedure to work, the relative errors in our earthquake locations must be less than one fourth of the shortest wavelength that we examine. Thus, given that the *S* velocities within our source region are believed to be 3175 m/s, the error in the locations cannot exceed 130 m for us to be certain of constructive interference. *Schaff et al.* [2002] cite the relative location error within small regions to be on the order of meters to tens of meters. On the basis of this criterion, our methodology should work.

[16] Using the above described methodology, we compute the ideal slowness parameters p_{ac} describing the departure velocity and angles of 20 windows of energy leaving the source region for every station/source array pair. Some example results are shown in Figures 4–6. Figure 4 shows that the beam-forming is working; when the waveforms are aligned by origin time, there does not appear to be any correlation between the waveforms, but taking location and preferred propagation direction into account shows a coherent waveform, the stack of which has much more power than those aligned on origin time. In Figure 5 we see that the stack is most powerful in a narrow range of azimuths and departure angles. This is typical, the majority of the energy we examine for both *S* and *P* is coming from a very limited range of azimuths and incidence angles that are close the azimuth from the station to the receiver (Figure 6). Other studies examining the early coda also find that the majority of energy appears to be scattered locally near the receiver [*Scherbaum et al.*, 1991; *Dodge and Beroza*, 1997].

3.2. Precise Determination of Earthquake Locations

[17] Once we have computed the slowness estimates, p_{ac}, for each window for each source array/station pair, we follow a very similar method to determine the location for the medium-magnitude earthquake. We take the preferred slowness parameters p_{ac} determined by the above method for an individual window of one source array at one station, and compute a stack specific to that window-array-station combination. This stack represents our best approximation of what the waveform should look like for each window at a specific station for the earthquake in question. We refer to this as the microearthquake stack, M_{ac}:

$$M_{ac}(p_{ac}, t_c) = \sum_{b=1}^{N} w_{ab}(t_c - p_{ac} \cdot x_b + n_{abc}) \qquad (2)$$

Note the microearthquake stack M_{ac} is simply $\tilde{S}_{acd}$, given the preferred slowness parameters d. We then search over earthquake location parameters (x, z, and t), shift the trace of the large event accordingly, and sum this with the microearthquake stack. If we refer to Figure 3, schematically we could say that we now know our propagation velocities and directions (i.e., A–C), but the location of our large earthquake is unknown so we search over locations (i.e., 1–4) to identify the ideal earthquake location. This

Table 1. List of Earthquakes Relocated by Each Station

Station[a]	Medium-magnitude Earthquakes Relocated
BAV	1,2,3
BSC	2,3
BSG	1,3
BSR	1
CYB	1
HPL	1,3
HQR	2
MHD	1
MPR	1
PHP	1
PJL	1

[a]Station locations are indicated in Figure 1.

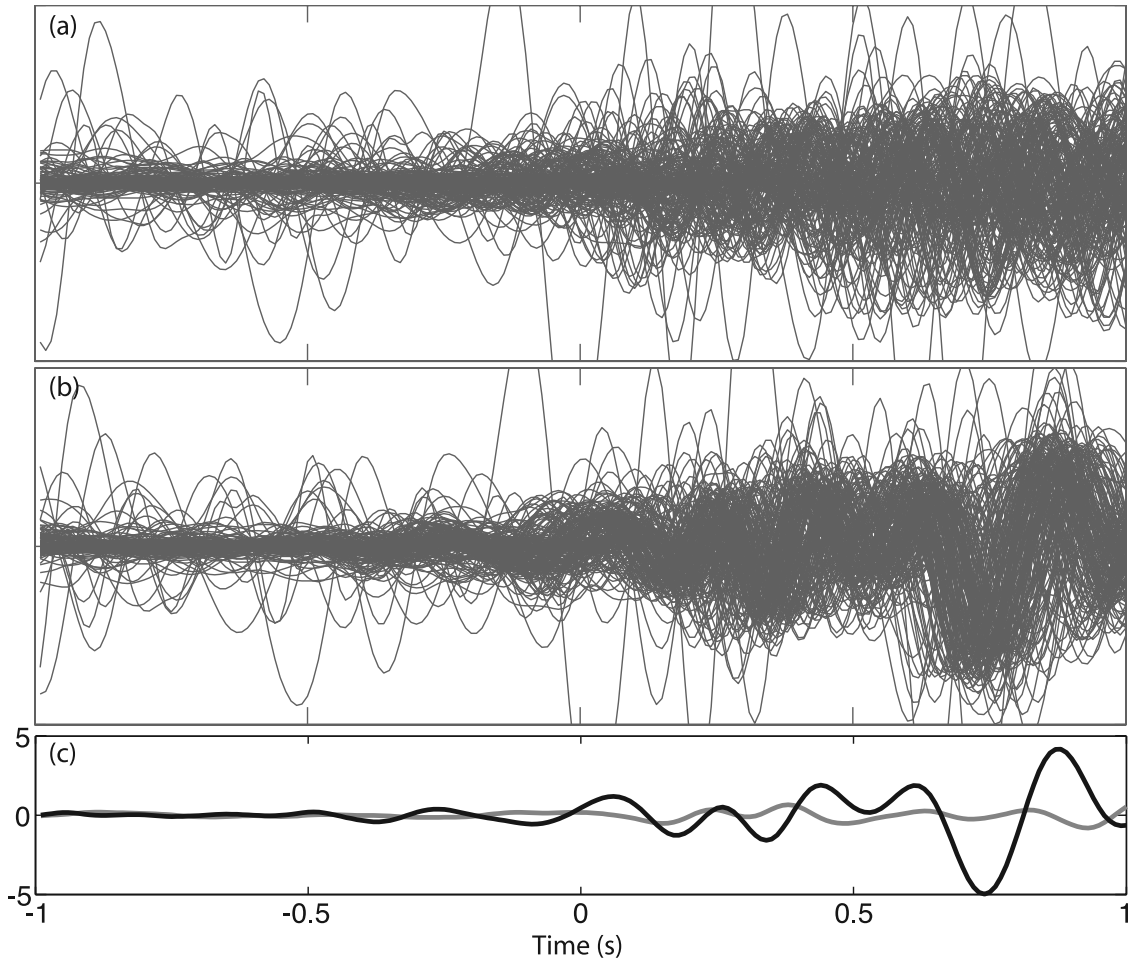

Figure 4. Waveforms from the first window for station BAV for medium-magnitude earthquake 1. (a) Waveforms aligned based on origin time only. (b) Waveforms aligned using ideal slowness parameters. (c) Stack of all the waveforms aligned based on origin time only (gray) and aligned using ideal slowness parameters (black).

sum of the microearthquake stack and the waveform of the medium-magnitude earthquake, we refer to as a location stack $\tilde{L}_{acef}$, where

$$\tilde{L}_{acef} = G_{ac}M_{ac} + \tilde{H}_{acef}W_{ac}(t_c - p_{ac}\cdot(X_0 + \tilde{X}_e) + \tilde{t}_{0f}) + n^0_{ac} \quad (3)$$

with

W the seismogram of the earthquake that we are relocating;
X_0 the initial, relative position of the earthquake that we are relocating;
G weighting factor of the microearthquake stack;
$\tilde{H}$ weighting factor for the earthquake that we are relocating;
$\tilde{X}$ trial movement of the earthquake that we are locating;
$\tilde{t}_0$ trial time offset for the centroid time of the earthquake that we are locating;
e bin number for the movement of the earthquake along fault strike and vertically;
f bin number for the centroid time offset;
n^0 noise for the earthquake being located.

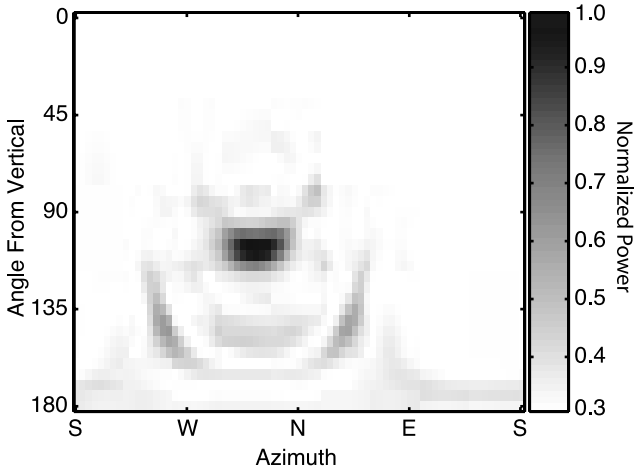

Figure 5. Power of the stack of the second window for medium-magnitude earthquake 1 at CYB. Angle from the vertical represents the difference from an upgoing wave; that is, 0° represents an upgoing wave, 90° represents a horizontal wave, and 180° represents a downgoing wave. For this plot the power is normalized to the maximum power for this specific window.

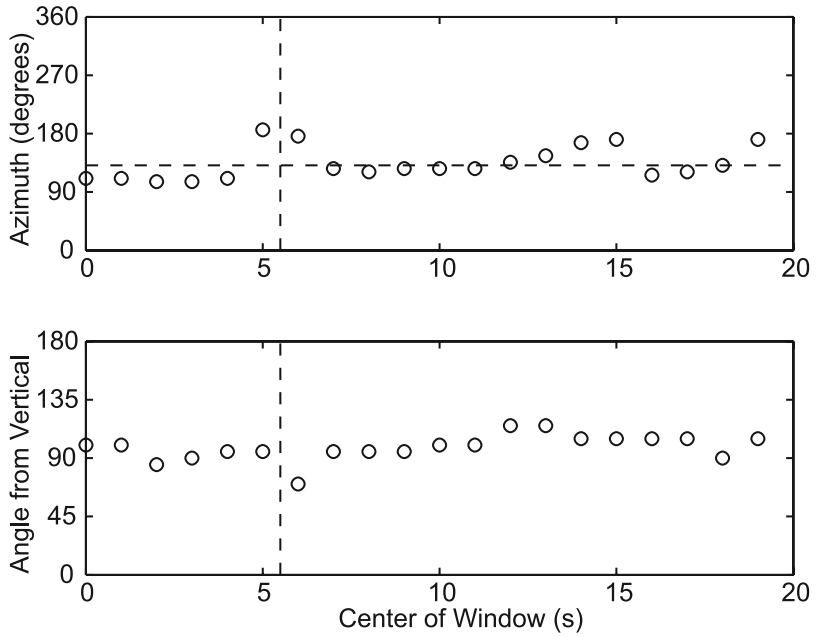

Figure 6. (a) Optimal propagation azimuth and (b) angle from the vertical for all 20 time windows at station HPL for medium-magnitude earthquake 1. The vertical dashed lines approximate the *S* arrival. The horizontal dashed line in Figure 6a indicates the approximate azimuth of the source arrays to HPL.

[18] The nodes of our grid are separated by 20 m for x and z and 0.025 s for origin time. We search over a range of ±1 km in x and z and ±1 s in t. The variable x represents distance along the strike of the fault and z represents depth; we do not search for locations of the events that are off the fault; that is, the y coordinate is fixed a priori. We do this for two reasons. First, the Calaveras has been shown to be particularly thin in this region, typically less than 75 m at depth [*Schaff et al.*, 2002], and it is unlikely that the largest ruptures in the vicinity occur off the main fault plane. Second, our resolution perpendicular to the plane of the fault is weak, as the majority of our stations are located at azimuths similar to that of the fault plane.

[19] For the location stacking that we do, we normalize the power of the window of both the microearthquake stack and the large earthquake seismogram to 1, using weighting factors G_{ac} and $\tilde{H}_{acef}$, respectively. This way, the power Q_{acef} of any location stack $\tilde{L}_{acef}$ represents the similarity in phase between the microearthquake stack and the waveform of the large earthquake. It gives secondary importance to amplitudes, but each window has the potential to sum to the same power. We show an example of the location stacking in Figure 7. Examining Figure 7, one finds a high-power stripe for each window, the orientation of this stripe is perpendicular to the propagation direction of the waves leaving the source array. We find that the orientation of these stripes is generally very similar. This means that for this event/station pair there is not much variation in the angle that the waves are departing the source region as we see in Figure 6. For the window that shows energy departing at significantly different angles than the others (18), we can see that there is an overlap of its high-power region and the high-power region of the other widows. This shows that the locations determined are consistent over all of our windows. Further examining Figure 7, one can see that the maximum power for some windows is larger than others. This is not a result of differing amplitudes (as each window has the potential to sum to the same power), but instead the variation in power is indicative of waveform similarity, and some windows have lower waveform similarity. This variation likely arises from low signal-to-noise ratios for these windows.

[20] To determine the optimum location at one station, we sum the power of the location stacks Q_{acef} for all 20 time windows. For any individual station *a*, this produces a matrix describing the power of the location stacks for all x, z, t combinations Z_{aef}, where

$$Z_{aef} = \sum_{c=1}^{20} Q_{acef} \tag{4}$$

The maximum value of Z_{aef} indicates the preferred centroid (x, z, t) of the large earthquake as determined by an individual station. We can further refine this estimate by summing these matrices over all of our stations. The result is a matrix R_{ef} that describes the total power for any location (x, z, t) over all the stations for all 20 time windows

$$R_{ef} = \sum_{a=1}^{S} Z_{aef} \tag{5}$$

where *S* represents the number of stations used to locate the earthquake in question. The values of Z_{aef} should be comparable from station to station as the normalization emphasizes similarity in phase and not relative amplitude. We treat the location determined from this stack as the centroid location. We call it the centroid because the stack will prefer to put the earthquake in a location where the apparent moment rate (slip rate) was largest. Some

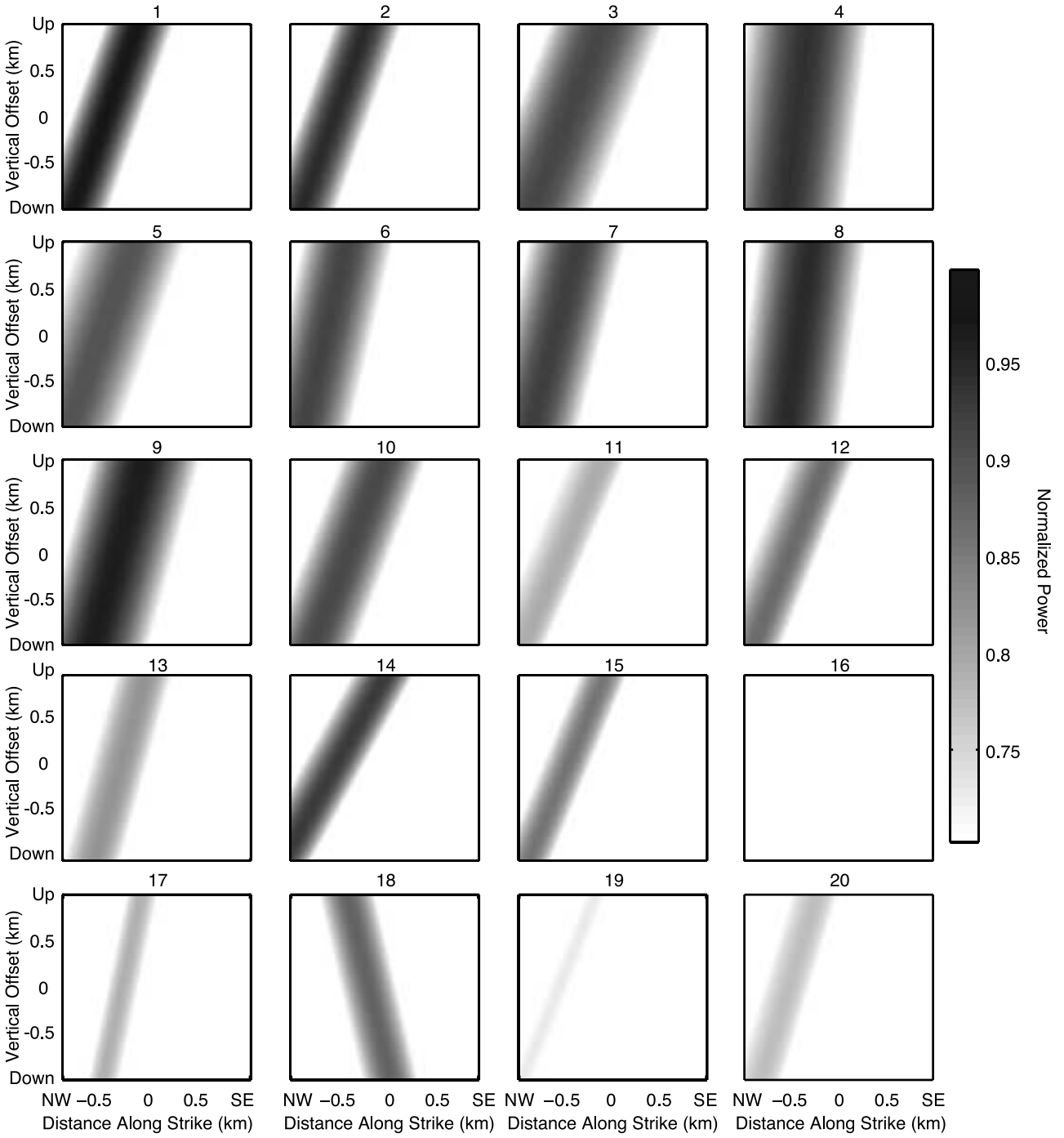

Figure 7. Power Q_{acef} of the location stacks $\tilde{L}_{acef}$ for medium-magnitude event 1 as recorded by station BAV. Each pane represents a window of the stack, the number above them specifying which window. Window 1 is the first window, which is centered on the *P* arrival, having 1 s before and after the *P* arrival. Shading indicates the normalized power of the stack, darker shades indicating larger powers. A power of 1 would indicate that the waveforms match perfectly. In the stacking procedure, each window of both the source-array stack and the waveform of the large earthquake is normalized such that the power of them is 1. The result of this is that each window can sum to the same power. We can see that the power of this normalized stack varies from window, which suggests that waveform similarity varies from window to window, which may be a result of low signal to noise or differing waveform characteristics.

examples of the results of the stacking are shown in Figure 8. From Figure 8 we see that there is a location where the energy appears to be coming from and individual stations are consistent with the location predicted by a stack of all the stations. We can also compare Figure 8 (top right) (which examines medium-magnitude earthquake 1 as observed by BAV) to Figure 7 (which examines each window for that station-event pair) and see that region of

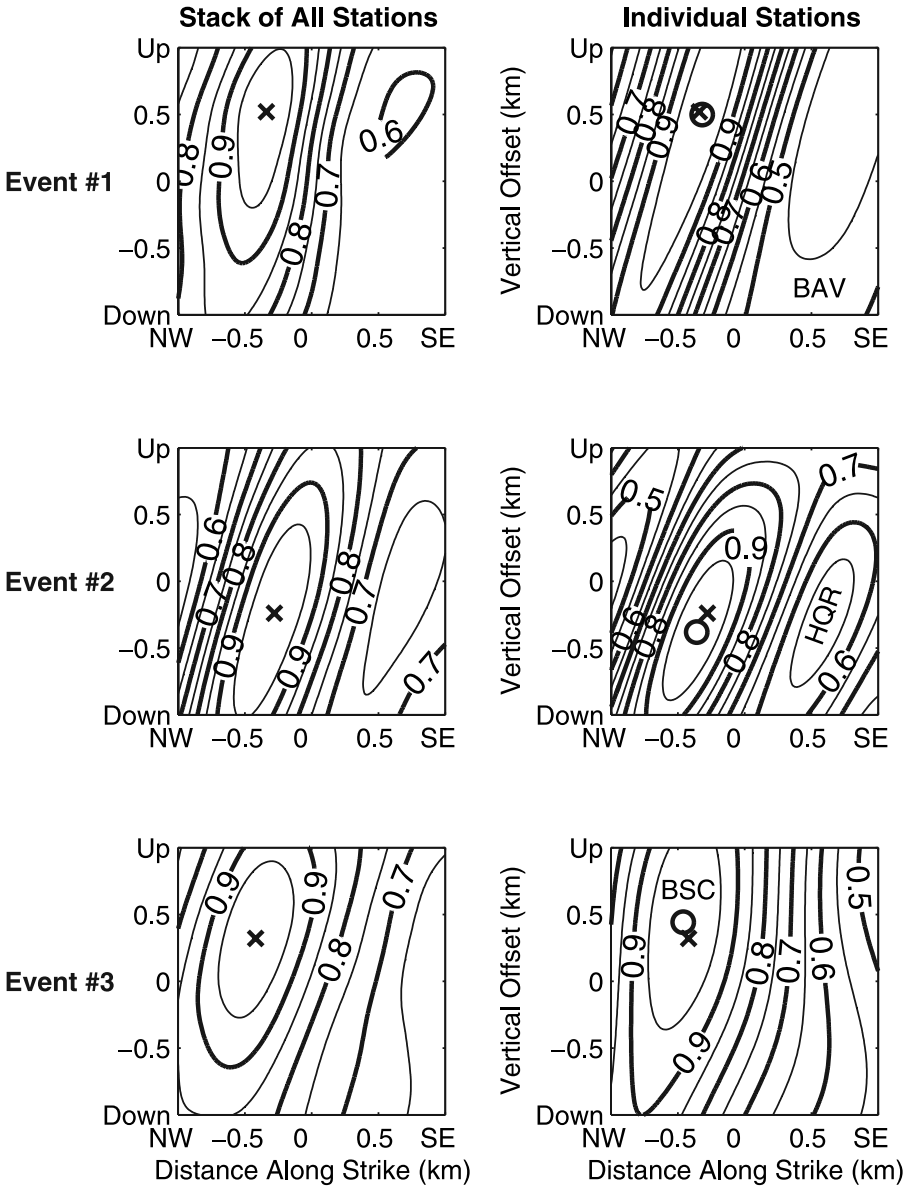

Figure 8. Contours of the normalized power for the location of each medium-magnitude earthquake as determined by (left) a stack of all our stations and (right) a stack of an individual station (1, BAV; 2, HQR; 3, BSC). Figure 8 (left) represents R_{ef}. Figure 8 (right) represents Z_{aef}. The preferred centroids determined by stacking all the stations are indicated for each event as a cross. The preferred centroids, as determined by the individual stations that we examine in Figure 8 (right), is indicated with a circle. We note that time is also a variable in our location stacking. Here we show cross sections using the preferred time offset, as identified by the maximum power of the stack, thus fixing f.

highest power (and thus predicted location) in Figure 7 is always a region of high power for each time window. The final centroid locations we determine using this technique are shown in Figure 9 and listed in Table 2.

3.3. Error Analysis

[21] We are interested in knowing the uncertainty in the locations of the three medium-magnitude events that we have relocated. To do this, we apply the same method we used to locate the medium-magnitude earthquakes to relocate 15 microearthquakes, the locations of which we know very well from the cross-correlation/double-difference locations computed by *Schaff et al.* [2002]. Any deviation of our relocations from our starting locations, which are believed to have uncertainty on the order of tens of meters [*Schaff et al.*, 2002], can be considered an estimate of the location uncertainty of the larger events.

[22] For each of our three medium-magnitude events, we relocate five nearby "trial events." For medium-magnitude earthquakes 2 and 3, we choose the five earthquakes closest to the catalog hypocenter of the large event that were recorded by all the stations that located the larger event. For medium-magnitude earthquake 1 it is much more difficult to find earthquakes that were recorded by all nine stations used to locate it. For this event we choose the five

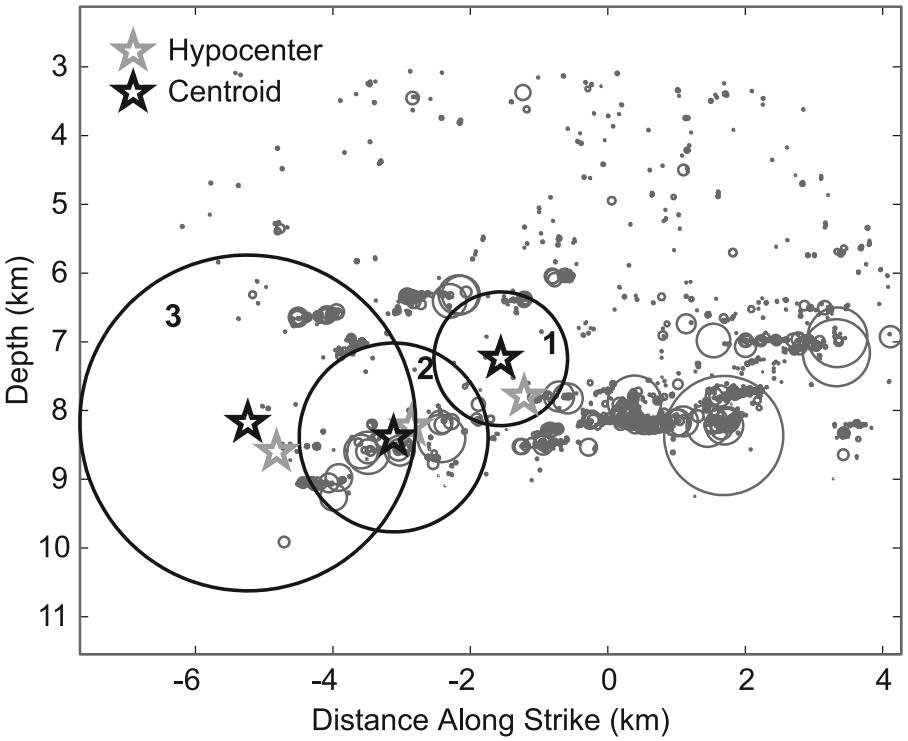

Figure 9. Relocations of the three medium-magnitude earthquakes. Gray stars indicate the hypocenters as determined by *Schaff et al.* [2002]. Black stars indicate centroid relocations. Circles represent an approximate rupture dimension assuming circular rupture and uniform 30 bar stress drop. The thick, black circles show approximate rupture extent given the new centroid locations. Numbers indicate the earthquake number referred to in this paper.

events closest to medium-magnitude event 1 that were recorded by at least five of the nine candidate stations. We also require that each station be used for at least one error estimate. We use the events believed to be closest to the center of our source arrays for error estimation because the waveforms for these events will be better approximated by our source array than those on the extremities. The average distance of our trial events to the catalog hypocenter for each source array is approximately 500 m, which is approximately the same distance as our relocations. This suggests that the error estimates from the trial events will be reasonable.

[23] We find the error estimates to be relatively consistent, both for all the test events for any medium-magnitude event and from medium-magnitude event to medium-magnitude event (Table 3). Because the estimates are relatively consistent for all the test events of any individual medium-magnitude earthquake, we take the mean of their values and treat it as the uncertainty in their centroid locations. These errors (approximately 100–150 m in x and z and 0.025 s in time) are substantially smaller than the movements of the medium-magnitude earthquakes (approximately 300 m in x and z and 0.4 s in time) found from our relocations (Table 2). Uncertainties in the centroid locations are not shown in our cross section showing the earthquake locations (Figure 9), as the size of the stars representing the centroids are approximately 500 m across, such that location uncertainty for all three earthquakes falls within the star representing their centroids.

[24] We note that there does appear to be a systematic bias in our locations, in that all the trial locations for each source array are located systematically in the same direction relative to their double-difference-determined locations (Table 3). For example, the locations of all the trial events for source array 1 are all to the southeast and above the locations determined by *Schaff et al.* [2002]. While the consistent mislocation of our trial events shows a bias, the errors introduced by this bias are small. As noted above, the approximate error in any parameter is much less than the movement of our medium-magnitude earthquakes and the bias is not strong enough to undermine our interpretation, particularly given the relative location of the hypocenters and centroids with respect to the streaks. There are a number of potential sources for this bias. It may result from the unfavorable station geometry; the majority of our stations are located at azimuths to the earthquake similar to

Table 2. Offset of Centroid From Hypocenter

Event	X Offset,[a] m	Z Offset,[b] m	t Offset,[c] s
1	−340	−520	0.55
2	−260	160	0.30
3	−420	−320	0.55

[a]Positive offset indicates movement to the SE.
[b]Positive offset indicates movement down.
[c]Positive offset indicates that centroid time is later than the hypocenter time.

Table 3. Error Estimates Determined by Relocating Test Events

Test Event	X Error,[a] m	Z Error,[b] m	T Error,[c] s
	Source Array 1		
1	60	−120	0.050
2	60	−240	0.050
3	180	−400	0.075
4	0	0	0.075
5	120	−100	0.000
Mean	84	−172	0.050
	Source Array 2		
1	−200	260	0.025
2	−220	200	0.000
3	−140	80	0.050
4	−100	60	0.050
5	−80	60	0.000
Mean	−148	132	0.025
	Source Array 3		
1	80	−40	0.000
2	60	0	0.000
3	120	−20	0.000
4	160	−20	0.000
5	180	−20	0.025
Mean	120	−20	0.005

[a]Positive offset indicates that the location that we determined is to the southeast of the HypoDD determined location.

[b]Positive offset indicates that the location that we determined is below the HypoDD determined location.

[c]Positive offset indicates that the centroid time that we determined is later than the centroid time determined by HypoDD.

that of the fault. This stems from the local network geometry, which has the majority of seismometers, in particular the low-gain stations, placed near the most active faults. The bias also may arise from the unfavorable geometry of our source arrays. The highly organized nature of streak seismicity may create an "array response," which could result in a large uncertainty in the determination of slowness parameters and potentially influences the locations determined by our method. Another alternative, though less likely, explanation is that the earthquake locations determined by *Schaff et al.* [2002] have an internal bias. We discount this explanation because the bias in our test locations is not consistent from source array to source array.

4. Comparison With Standard Methods of Earthquake Relocation

[25] To demonstrate the need for our alternative location methodology, we also use a more standard approach to locating the three medium-magnitude earthquakes, waveform cross-correlation and double-difference relocation (HypoDD). We cross correlate the available, unclipped low-gain waveforms of the three medium-magnitude earthquakes with high-gain recordings of all events determined to be within 3 km of them (i.e., the source arrays we used to relocate the earthquakes). These correlations were not included in previous relocations in the region [*Schaff et al.*, 2002] because they did not use data from low-gain channels. We place these correlations into HypoDD to relocate the catalog using our correlation-based difference times. HypoDD converges, and we come up with relocations for these three events.

[26] We find that the locations determined by HypoDD are highly unstable. We apply a jackknife test [*Efron*, 1979] to determine the variance (and thus standard deviation and 95% confidence bounds) in the locations for our earthquakes of interest. In using the jackknife, we remove the correlations determined for each medium-magnitude earthquake/station pair, one at a time, and then run HypoDD to determine new locations. In these runs of HypoDD, we hold the locations of the nearby microearthquakes constant, as we believe their locations to be well constrained previously by the cross correlations and locations computed by *Schaff et al.* [2002]. The uncertainties for each earthquake are only computed using those runs of HypoDD where a recording of that earthquake was removed; that is, we do not consider the influence of removing a recording of medium-magnitude earthquake 2 on the location of medium-magnitude event 1, as this effect should be negligible.

[27] From this analysis we find that the locations determined using cross correlation and HypoDD have 95% confidence bounds that regularly exceed 500 m in x, y, and z (Table 4). The 95% confidence bounds on t (the origin time) range from 0.04 to 0.36 s. We find these bounds unacceptably large. The fact that the locations are unstable and dependent upon what data are placed into HypoDD indicates that the cross-correlation-derived data do not add any significant constraint on the locations of these earthquakes. We do not consider this a condemnation of cross-correlation or double-difference-based relocations methods; this is merely demonstrating a known limitation of these methods. Like standard earthquake location methods, HypoDD requires a large range of azimuths to reliably locate earthquakes. Because HypoDD only uses direct arrivals, one gets a plane of uncertainty striking perpendicular to the azimuth of the stations to the events if there is a limited range of azimuths available (as in this case). HypoDD also works better with more data. As discussed by *Waldhauser* [2001], each cross-correlation measurement addresses eight unknowns (x, y, z, and t for both events), so if we have less than eight observations for each event/event pair our solution is potentially unstable. Because we fix the locations of the microearthquakes, the number of unknowns is reduced to four for any event/event pair, but we still regularly do not even have that many observations with acceptable correlation coefficients (CC > 0.7). From this analysis we can surmise that cross correlation applied to the data set that we are examining does not add useful information for the earthquake location problem. Were there more, better azimuthally distributed data available, we expect that these methods would work well. Because this kind of data is unavailable, we need to use our method.

[28] We feel that is important not only to show that the standard earthquake relocation methods are ineffective for

Table 4. Jackknife Test 95% Confidence Bounds for Earthquake Locations Determined Using Cross Correlation and HypoDD

	95% Confidence Bounds			
Event	x, m	y, m	z, m	t, s
1	±991	±667	±396	±0.18
2	±584	±207	±1298	±0.04
3	±2924	±2919	±1503	±0.36

Table 5. Jackknife Test 95% Confidence for Earthquake Locations Determined Using Beam-Forming Method[a]

	95% Confidence Bounds			
Event	x, m	y, m	z, m	t, s
1	±200	±300	±200	±0.05
2	±346	±346	±529	±0.09
3	±382	±200	±0	±0.06

[a]Confidence bounds were determined using a jackknife test. Note that these confidence bounds should only be used to compare to the confidence bound for HypoDD-based locations, as our other locations are forced to remain on the fault, these locations are not.

our given data but also to show that our method is superior to these methods. To demonstrate that the locations produced by our beam-forming method are superior, we also subject them to the same jackknife test. To make the comparison between our method and cross-correlation-based HypoDD completely fair, we search in three dimensions for the earthquake locations instead of the two dimensions that we normally search in; that is, instead of a fault-parallel component and z, we search over x, y, and z. To save computation time, we use a much coarser grid than we did previously; we now search over 100 m grid nodes, instead of 20 m previously. Applying the jackknife, we find that the 95% confidence intervals of the beam-forming-based locations (Table 5) are much smaller than those from HypoDD (Table 4), only once exceeding 500 m in x, y, or z, as opposed to only being less than 500 m twice. This is evidence that our methodology works better with this small and azimuthally limited data set than do more standard earthquake location methods. Our method works better with this limited data set because it does not rely solely upon direct arrivals. By sampling arrivals (i.e., coda) that depart the source region at varying angles, we are able to reduce the uncertainty in locations caused by only having a limited range of earthquake-station azimuths.

[29] We should note that we prefer the error estimates that we computed for our earthquake locations by using test earthquakes instead of the jackknife determined errors. We prefer these error estimates for three reasons. First, the errors we initially computed are specific to our two-dimensional (2-D) search that we use to locate the events, instead of the 3-D search that we use to compare with errors computed for HypoDD locations. Second, jackknife uncertainty estimates have been shown to be consistently biased upward [*Efron and Stein*, 1981], so we would prefer to use a more consistent method. Third, while we feel that the results of the jackknife test show that our method provides consistent results given varying quantities of data, we are more interested in showing that our method provides accurate results, considering that it is a new method. The location test that we devise shows that our methodology is accurate, as we compare the results of our methodology to known, high-quality locations. For these three reasons, in further discussions of the earthquakes and the errors associated with them, we use the errors computed from the microearthquake test events and not the jackknife errors.

5. Interpretation

[30] The earthquake relocations that we have computed for our three medium-magnitude earthquakes suggest that streaks represent a rheologic boundary between creeping and locked (or partially locked) portions of a fault. Others have also suggested that the boundary between creeping and locked sections of faults may be responsible for the generation of streaks [*Waldhauser et al.*, 1999, 2004]. The hypocenters of events 1 and 3, as determined by *Schaff et al.* [2002], are located on a streak and the centroids are located updip of the hypocenters, suggesting that the rupture propagated updip (Figure 9). *Boatwright and Seekins* [2004] have also found a predominance of updip rupture propagation for earthquakes in this same region. The centroid locations of the medium-magnitude earthquakes places them directly between two streaks in a region that is relatively devoid of seismicity. The lack of microseismicity and the presence of larger events that appear to rupture into and perhaps across the region between the two streaks, suggests that it is a locked zone, where slip is accommodated seismically. This behavior (moderate earthquakes occurring in regions devoid of microseismicity) has been observed for the Calaveras Fault before [*Oppenheimer et al.*, 1990; *Manaker et al.*, 2003]. We argue that the region above the upper streak and below the lower streak accommodate much of their slip through aseismic creep. This is supported by the fact that the lower streak marks the bottom of the seismogenic zone on the Calaveras Fault, suggesting that ductile processes accommodate slip on the fault below. Moreover, *Schaff et al.* [1998] found that the recurrence intervals of repeating earthquakes on the Calaveras Fault following the 1984 Morgan Hill earthquake was consistent with them being driven by postseismic creep under steady state velocity strengthening friction. There is also significant evidence for surface creep in this region [*Galehouse and Lienkaemper*, 2003], which is consistent with the notion that the region above the upper streak accommodates much of its slip in creep. Taken together, this evidence suggests that streaks mark the boundary between aseismically slipping regions and regions that accommodate their slip seismically. Similar observations have been made for streaks on the San Andreas Fault in the Parkfield region, where medium-magnitude earthquake nucleate on a deeper streak and rupture into a region devoid of microseismicity [*Waldhauser et al.*, 1999, 2004]. On the basis of the observations that there appear to be locked and creeping segments of a fault juxtaposed across a streak, we argue the sharp lineation of seismicity that we see (i.e., the streak) represents the freely slipping region slipping past the locked section of the fault.

[31] The location of medium-magnitude earthquake 2 does not fit the above model. Unlike events 1 and 3, the interpretation of rupture propagation for event 2 is less clear. One possibility is that the location of the centroid very close to the hypocenter indicates bilateral rupture. This event appears to nucleate and propagate in a section of the lower streak that is dominated by seismicity that is larger than the majority of events near the hypocenters of events 1 and 3. This difference in the local seismicity may explain the difference in the rupture propagation if, for example, a greater proportion of the fault is locked here, or if the boundary between creeping and locked portions of the fault here is less distinct.

[32] We also note that all three earthquakes had a centroid time that was significantly later than their origin times (Table 2). We consider this further evidence that our method is working; however, the combination of temporal centroid

Table 6. Earthquake Offsets and Errors Determined Using Data From the Coda Only[a]

Event	X, m	Z, m	T, m
1	−360 ± 100	−480 ± 140	0.55 ± 0.045
2	−260 ± 080	100 ± 092	0.30 ± 0.030
3	−320 ± 132	−440 ± 008	0.55 ± 0.005

[a]Errors computed as the mean of the errors determined using the test events.

shifts are too small to be consistent with unilateral rupture propagation at typical earthquake rupture velocities given the shifts in the spatial centroids. On the basis of the centroid time shift, event 2 appears to be less unilateral than events 1 and 3. This parallels our observations of the distance between the hypocenters and centroids of these events.

6. Locating Earthquakes Without Direct Arrivals

[33] In this section we demonstrate that it is possible to locate earthquakes using the coda alone, i.e., without direct

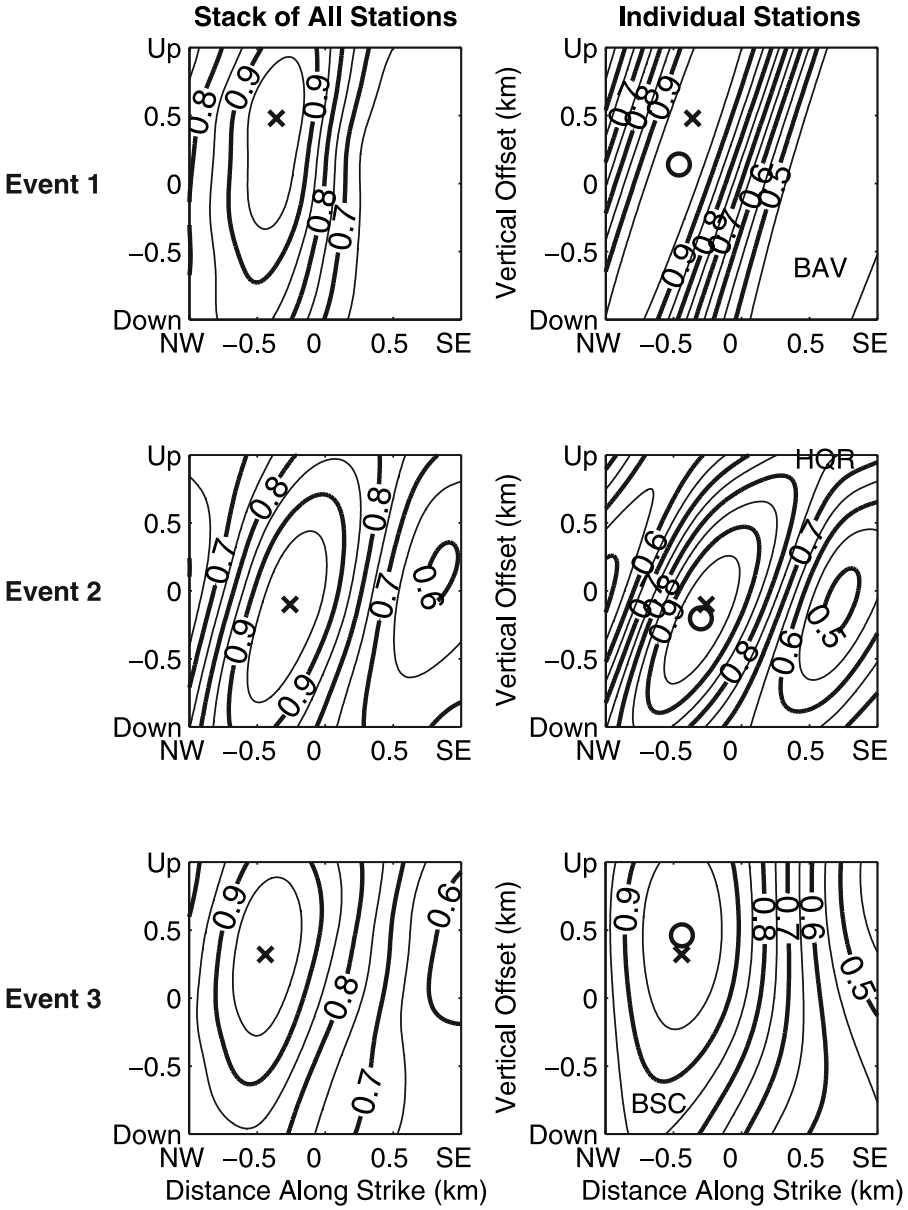

Figure 10. Contours of the normalized power for the location of each medium-magnitude earthquake using only the coda as determined by (left) a stack of all our stations and (right) a stack of an individual station (1, BAV; 2, HQR; 3, BSC). Figure 10 (left) represents R_{ef}. Figure 10 (right) represents Z_{aef}. The preferred hypocenters determined by stacking all the stations are indicated for each event as a cross. The preferred hypocenter, as determined by the individual stations we examine in Figure 10 (right), is indicated with a circle. We note that time is also a variable in our location stacking. Here we show cross sections using the preferred time offset, as identified by the maximum power of the stack, thus fixing f.

arrivals. To our knowledge, very little research has been conducted in locating earthquakes without using the *P* and/or *S* arrivals. *Schaff and Richards* [2004] and *Ekström* [2006] have shown that it is possible locate earthquakes using surface waves. *Snieder and Vrijlandt* [2005] demonstrated that the coda can be used to determine the relative positions of earthquakes. In that study, they show that using a moving window cross correlation one can approximate the distance between two earthquakes, but they do not actually locate the events. While these previous studies have shown that it is possible to locate earthquakes without using the *P* or *S* arrivals, we believe that this study is the first time that earthquakes have been located without using any known phases.

[34] To compute earthquake locations using only the coda, we remove those windows that include the direct arrivals. As a conservative estimate, we decide that the first 2 s of both the *P* and *S* arrivals represent directly propagating waves. We remove our first three windows (the first window is centered on the *P* arrival) and the first three windows that we determine to be dominated by *S* energy. To compute locations, we follow the same method as before and sum all of our windows at every station (now we only have 14 windows to work with). For each earthquake, we sum over all of our stations. The locations are remarkably similar to those determined using both the coda and the direct arrivals (Tables 2 and 6 and Figure 10). This is sensible in that the majority of the windows that we used in our full waveform locations are still included; however, it demonstrates that the coda alone can be used for earthquake location. We also compute error estimates for the coda-based locations. They behave similarly to those that used both the direct arrivals and the coda, so we only report the average error for each earthquake located in Table 6. We do see that the width of the high-power area of the location is slightly less when we include direct arrivals than when we use coda alone. This implies that the energy in the direct arrivals is helping to constrain the locations; however, the coda alone clearly provides enough information for earthquake location.

7. Conclusions

[35] In this paper we present a technique based upon source-array analysis to locate earthquakes reliably with data from only a few stations at a limited range of azimuths. Our results demonstrate rupture propagation of medium-magnitude earthquakes near streaks on the Calaveras Fault. We determine that these events nucleated on streaks and ruptured into a region relatively devoid of seismicity that we infer to be locked. Because we believe the regions above and below this region that lacks seismicity to be accommodating slip through creep, we argue that the streaks delineate the boundary between creeping and locked sections of a fault. We suggest that the stress concentration at the boundary between creeping and locked sections results in a lineation of microearthquakes.

[36] With our method we are able to locate earthquakes reliably when other methods cannot. We take advantage of the information contained within both the *P* and *S* codas, which is neglected by standard location methods, to compute precise earthquake locations. We demonstrate the power of the information contained within the coda, as we are able to locate earthquakes using only the coda (i.e., without direct arrivals). The uncertainties in the locations determined using the coda alone, are similar to those where we used both the coda and direct arrivals. Aside from the obvious applications of using the coda to locate earthquakes where low signal to noise or emergent onsets prevent high-quality picks, this technique also has the potential to locate other seismic sources that do not have a clear onset or strong, impulsive arrivals, e.g., nonvolcanic and volcanic tremor. In principle, the extra information in the coda might be further exploited to locate earthquakes using only one station, given a medium where strong scattering results in arrivals that leave the source at a wide range of angles. With our current data set, this is not possible as the large majority of energy is forward scattered, but if we extend the technique to longer times, it might work as later portions of the coda have been shown to be composed of more strongly scattered energy [*Scherbaum et al.*, 1991]. It should also be possible to determine relative locations of groups of earthquakes using coda beam-forming as the sole location criterion.

[37] **Acknowledgments.** We thank Jack Boatwright for many useful and spirited conversations. We acknowledge David Schaff for publicly sharing all his relocation results. The comments of Peter Shearer, Honn Kao, and an anonymous Associate Editor significantly improved this manuscript. All the data used in this study were made available through the NCEDC. The data were provided by the NCSN, via the USGS Menlo Park. J.R. was partially supported by the William K. Whiteford Fellowship.

References

Boatwright, J., and L. C. Seekins (2004), Inverting peak ground motions for rupture directivity in moderate and large earthquakes, *Eos Trans. AGU*, *85*(47), S24A-05 Fall Meet. Suppl., Abstract.

Dodge, D. A., and G. C. Beroza (1997), Source array analysis of coda waves near the 1989 Loma Prieta, California, mainshock: Implications for the mechanism of coseismic velocity changes, *J. Geophys. Res.*, *102*, 24,437–424,458.

Efron, B. (1979), Bootstrap methods: Another look at the jackknife, *Ann. Stat.*, *7*, 1–26.

Efron, B., and C. Stein (1981), The jackknife estimate of variance, *Ann. Stat.*, *9*, 586–596.

Ekström, G. (2006), Global detection and location of seismic sources by using surface waves, *Bull. Seismol. Soc. Am.*, *96*, 1201–1212.

Fletcher, J. B., P. Spudich, and L. M. Baker (2006), Rupture propagation of the 2004 Parkfield, California, earthquake from observations at the UPSAR, *Bull. Seismol. Soc. Am.*, *96*, S129–S142.

Galehouse, J. S., and J. J. Lienkaemper (2003), Inferences drawn from two decades of alinement array measurements of creep on faults in the San Francisco Bay Region, *Bull. Seismol. Soc. Am.*, *93*, 2415–2433.

Gillard, D., A. M. Rubin, and P. Okubo (1996), Highly concentrated seismicity caused by deformation of Kilauea's deep magma system, *Nature*, *384*, 343–346.

Ishii, M., P. M. Shearer, H. Houston, and J. E. Vidale (2005), Extent, duration and speed of the 2004 Sumatra-Andaman earthquake imaged by the Hi-Net array, *Nature*, *435*, 933–936, doi:10.1038/nature03675.

Kao, H., and S.-J. Shan (2004), The Source-Scanning Algorithm: Mapping the distribution of seismic sources in time and space, *Geophys. J. Int.*, *157*, 589–594, doi:10.1111/j.1365-246X.2004.02276.x.

Kao, H., S.-J. Shan, H. Dragert, G. Rogers, J. F. Cassidy, and K. Ramachandran (2005), A wide depth distribution of seismic tremors along the northern Cascadia margin, *Nature*, *436*, 841–844, doi:10.1038/nature03903.

Kao, H., S.-J. Shan, H. Dragert, G. Rogers, J. F. Cassidy, K. Wang, T. S. James, and K. Ramachandran (2006), Spatial-temporal patterns of seismic tremors in northern Cascadia, *J. Geophys. Res.*, *111*, B03309, doi:10.1029/2005JB003727.

Krüger, F., and M. Ohrnberger (2005a), Spatio-temporal source characteristics of the 26 December 2004 Sumatra earthquake as imaged by teleseismic broadband arrays, *Geophys. Res. Lett.*, *32*, L24312, doi:10.1029/2005GL023939.

Krüger, F., and M. Ohrnberger (2005b), Tracking the rupture of the Mw = 9.3 Sumatra earthquake over 1,150 km at teleseismic distance, *Nature*, *435*, 937–939, doi:10.1038/nature03696.

Krüger, F., M. Weber, F. Scherbaum, and J. Schlittenhardt (1993), Double beam analysis of anomalies the core-mantle boundary region, *Geophys. Res. Lett.*, *20*, 1475–1478.

Manaker, D. M., R. Bürgmann, W. H. Prescott, and J. Langbein (2003), Distribution of interseismic slip rates and the potential for significant earthquakes on the Calaveras fault, central California, *J. Geophys. Res.*, *108*(B6), 2287, doi:10.1029/2002JB001749.

Niazi, M. (1969), Use of source arrays in studies of regional structure, *Bull. Seismol. Soc. Am.*, *59*, 1631–1643.

Oppenheimer, D. H., W. H. Bakun, and L. A. G (1990), Slip partitioning of the Calaveras Fault, California, and prospects for future earthquakes, *J. Geophys. Res.*, *95*, 8483–8498.

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.

Rost, S., and C. Thomas (2002), Array seismology: Methods and applications, *Rev. Geophys.*, *40*(3), 1008, doi:10.1029/2000RG000100.

Rubin, A. M., D. Gillard, and J.-L. Got (1999), Streaks of microearthquakes along creeping faults, *Nature*, *400*, 635–641.

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

Schaff, D. P., G. C. Beroza, and B. E. Shaw (1998), Postseismic response of repeating aftershocks, *Geophys. Res. Lett.*, *25*, 4549–4552.

Schaff, D. P., G. H. R. Bokelmann, G. C. Beroza, F. Waldhauser, and W. L. Ellsworth (2002), High-resolution image of Calaveras Fault seismicity, *J. Geophys. Res.*, *107*(B9), 2186, doi:10.1029/2001JB000633.

Schaff, D. P., G. H. R. 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.

Scherbaum, F., D. Gillard, and N. Deichmann (1991), Slowness power spectrum analysis of the coda composition of two microearthquake clusters in northern Switzerland, *Phys. Earth Planet. Inter.*, *67*, 137–161j.

Scherbaum, F., F. Krüger, and M. Weber (1997), Double beam imaging: Mapping lower mantle heterogeneities using combinations of source and receiver arrays, *J. Geophys. Res.*, *102*, 507–522.

Snieder, R., and M. Vrijlandt (2005), Constraining the source separation with coda wave interferometry: Theory and application to earthquake doublets in the Hayward fault, California, *J. Geophys. Res.*, *110*, B04301, doi:10.1029/2004JB003317.

Spudich, P., and T. Bostwick (1987), Studies of the seismic coda using an earthquake cluster as a deeply buried seismograph array, *J. Geophys. Res.*, *92*, 10,526–10,546.

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

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

Waldhauser, F., and W. L. Ellsworth (2002), Fault structure and mechanics of the Hayward Fault, California, from double-difference earthquake locations, *J. Geophys. Res.*, *107*(B3), 2054, doi:10.1029/2000JB000084.

Waldhauser, F., W. L. Ellsworth, and A. Cole (1999), Slip-Parallel Seismic Lineations on the Northern Hayward Fault, California, *Geophys. Res. Lett.*, *26*, 3525–3528.

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.

Walker, K. T., M. Ishii, and P. M. Shearer (2005), Rupture details of the 28 March 2005 Sumatra Mw 8.6 earthquake imaged with teleseismic P waves, *Geophys. Res. Lett.*, *32*, L23303, doi:10.1029/2005GL024395.

Zhang, H., and C. Thurber (2003), double-difference tomography: The method and its application to the Hayward Fault, California, *Bull. Seismol. Soc. Am.*, *93*, 1875–1889.

G. C. Beroza, Department of Geophysics, Stanford University, 397 Panama Mall, Stanford, CA 94305-2215, USA.

J. L. Rubinstein, Department of Earth and Space Science, University of Washington, Box 351310, 4000 15th Avenue NE, Seattle, WA 98195, USA. (justin@ess.washington.edu)