Epicentral location based on Rayleigh wave Empirical Green’s Functions from ambient seismic noise

M. P. Barmin, A. L. Levshin, Y. Yang and M. H. Ritzwoller

Center for Imaging the Earth’s Interior, Department of Physics, University of Colorado at Boulder, Boulder, CO 80309-03901, USA.
E-mail: levchin@colorado.edu

Accepted 2010 October 30. Received 2010 October 20; in original form 2010 July 1

SUMMARY
A new method to locate the epicentre of regional seismic events is developed with strengths and limitations complementary to existing location methods. This new technique is based on applying empirical Green’s functions (EGFs) for Rayleigh waves between 7 and 15 s period that are determined by cross-correlation of ambient noise time-series recorded at pairs of seismic receivers. The important advantage of this method, in comparison with standard procedures based on use of body wave traveltimes, is that it does not rely on an earth model. Rather it is based on interpolating the EGFs to arbitrary hypothetical event locations. The method is tested by locating well known ‘Ground Truth’ crustal events in the western United States (magnitude 4 earthquakes in California, magnitude 3 earthquakes in Utah, and a mine collapse) as well as locating seismic stations as virtual events. In these applications, location errors average less than 1 km, but are expected to vary with event mechanism and depth. Numerical simulations show that the method optimizes when source depth is less than 1 km or more than 5 km, and when the source mechanism is nearly purely strike-slip, thrust or normal.

Key words: Earthquake source observations; Seismic monitoring and test-ban treaty verification; Surface waves and free oscillations.

1 INTRODUCTION
In principle, seismic surface waves possess valuable information about seismic source epicentral location and depth. Surface waves travel slower than body waves, they are excited by even relatively small crustal events, and the frequency dependence of their amplitudes depends strongly on event depth. Nevertheless, with a few exceptions (e.g. Ammon et al. 2001; Ekström 2006), surface waves are not commonly employed to constrain epicentral location for small or moderate sized earthquakes, although long period surface waves are used on a regular basis to determine the centroid of large earthquakes (Dziewonski et al. 1981). This is not because standard procedures for determining the location of individual seismic events are perfect. Standard methods rely predominantly on body wave traveltimes and most depend on the ability to predict traveltimes through a reference earth model. Such models are often 1-D and in some cases may not be tuned regionally. 1-D models such as IASPEI91 (Kennett & Engdahl 1991), for example, are commonly used even for locations based exclusively on regional phases. 1-D models, however, do not capture the effects of 3-D structural variations on traveltimes, which for regional phases such as Pg and Pn can be very strong. The inability to model 3-D traveltime effects produces location bias which, in the context of body wave location methods, can only be mitigated by using an accurate underlying 3-D model (e.g. Shapiro & Ritzwoller 2002), by applying empirical path corrections (e.g. Ritzwoller et al. 2003; Yang et al. 2004), by utilizing multiple event methods (e.g. Richards et al. 2006), or in the presence of exceptional azimuthal coverage (which commonly only occurs for large events). Irrespective of the shortcomings of traditional event location methods, the effect of 3-D structural variations on surface waves is even more profound than that on body waves. This has led seismologists largely to ignore the potential for surface waves to improve the epicentral and hypocentral location of small to intermediate magnitude seismic events.

In this paper, we develop a new method and present pilot tests to locate regional seismic events based on surface wave information. This technique is based on applying empirical Green’s functions (EGFs) for Rayleigh waves between 7 and 15 s period, which are produced from ambient noise cross-correlation analysis (e.g. Shapiro & Campillo 2004; Snieder 2004; Shapiro et al. 2005; Bensen et al. 2007). In this period band, ambient noise is strong and shallow crustal earthquakes are energetic. Elastic EGFs are determined by cross-correlating ambient noise time-series recorded at pairs of stations. The important advantage of the application of ambient noise EGFs is that they do not depend on an underlying velocity reference model for detection and location. Rather, the EGFs contain full structural information without reference to a 3-D model. Location errors are, therefore, expected to be largely unbiased by structural effects. In this paper, by ‘location’ we refer to epicentral location. Hypocentral location is the planned subject of a future contribution.

The idea of the new location method is straightforward, but it rests on the existence of at least a temporary dense local array, termed
the base stations, that are deployed in the ‘region of interest’ where interesting seismic events may occur. Second, it is also assumed that there is a more distant long duration (but potentially sparse) regional network of stations termed the remote stations. This layout is depicted schematically in Fig. 1. Using what are now well established methods (e.g. Bensen et al. 2007), the EGFs from every base station to every remote station are computed. For this paper, we use only the EGFs from the vertical component of each station-pair, which are dominated by Rayleigh waves. The set of EGFs observed at the base stations in the neighbourhood of a seismic event to a given remote station are used to produce an interpolated ‘Composite’ EGF (CEGF) to the remote station for any hypothetical event location enclosed by the base stations. These CEGFs then can be used to locate events within the region of interest using waveform correlation with event (e.g. earthquake) records observed at the remote stations. In this paper, we use a simple grid search method for location, but the CEGFs could be used to compute partial derivatives with respect to the event coordinates, if desired. The key to the location method is the transformation of the observed EGFs to a hypothetical event location, which is described in detail in Section 3. The transformation improves the signal-to-noise ratio (SNR) and minimizes the effect of waveform differences between event records and CEGFs.

There are two principal positive aspects of the method worth discussing at the outset. First, as mentioned above, the EGFs are empirical and, therefore, contain the effect of 3-D structure. Second, the base stations do not have to be installed when the event of interest occurs. The base and remote stations must be emplaced simultaneously for some period of time in order to produce the set of EGFs between the base and remote stations, but the base stations need not be in place during the event. A temporary deployment of base stations will suffice to produce the EGFs for the grid of points within the footprint of the base stations to each of the remote stations. Thus, when an interesting event occurs, observations of the event are only needed at the remote stations, which are assumed to be ‘permanent’. The longer the base stations can be operated the better, because longer time-series will produce higher SNR EGFs, although quite short deployments can produce surprisingly good EGFs at periods below about 15 s (Bensen et al. 2007). The base stations can be installed even after an event of interest has occurred.

Section 2 describes the basic idea of the method and provides some explanatory illustrations for why the method succeeds. The formal description of the technique is given in Section 3. Examples of the application of the method are presented as tests in the following sections. These include locating stations as virtual events as well as locating the Crandall Canyon Mine collapse in Utah on August 6, 2007, fourteen (magnitude 4 and 5) GT events in California, and four smaller (magnitude 3) events in Utah. Results are summarized in Tables 1–3. In Section 5, we discuss several factors that affect location accuracy, most notably source mechanism and depth.

**2 EPICENTRAL LOCATION METHOD: THE IDEA**

Consider, for example, locating an earthquake, such as the mb = 4.5 event that occurred north of the San Francisco Bay area on 2006 May 12, Fig. 2(d). The idea of the method is to compare the event (earthquake, explosion, mine collapse, landslide, etc.) records observed at the remote stations to the ambient noise Rayleigh wave EGFs between the base stations and the remote stations. To do so, we systematically move the hypothesized location of the event until the match between the EGFs and the earthquake records is optimized. There are two principal wrinkles to the method as it currently exists. First, when we stack the individual EGFs over all of the base stations to produce a CEGF for each hypothesized event location, we currently ignore phase information by summarizing both the CEGFs and the event records with envelope functions. Second, the transformation or ‘interpolation’ of the individual EGFs to form a CEGF is performed in the frequency–time (FT) domain. These details are discussed in the following section. The basic idea of the use of Rayleigh wave CEGFs determined from ambient noise to locate regional events is illustrated in Figs 2–4.

An example vertical component earthquake (event) record observed at station V04C in southern California is shown in Fig. 2(a) together with the vertical–vertical EGF from a station near the epicentre (GASB) with station V04C in Fig. 2(b). Station GASB is about 100 km from the earthquake epicentre. Both time-series are bandpass filtered between 7 and 15 s period and are dominated by Rayleigh waves. The earthquake and the EGF Rayleigh waveforms are similar, but are off-set in time because the epicentral distance for V04C is smaller than the interstation distance V04C-GASB. The EGF can be shifted in time and deformed to match the epicentral distance, however, as seen in Fig. 2(c). The process of time shifting and deformation is described in detail in Section 3.

The phase content of the earthquake and EGF records differ appreciably, however, because earthquakes impart an initial phase to surface waves that depends on the hypocentral depth and the source mechanism. For this reason, at this stage we ignore phase information and summarize both the EGFs and the earthquake records with their envelope functions, as seen in Figs 2(a)–(c) as dashed lines. There are, however, several base stations from the EarthScope/USArray Transportable Array and the Berkeley Seismic Network near to this earthquake. Fig. 3 shows the envelope functions

---

**Table 1.** USGS, UUSS and CU event parameters for the Crandall Canyon mine collapse on 2007 August 6.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>USGS/USGS</th>
<th>UUSS</th>
<th>CU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magnitude</td>
<td>3.9</td>
<td>3.9</td>
<td>–</td>
</tr>
<tr>
<td>Origin time (UTC)</td>
<td>08:48:40</td>
<td>8:48:40</td>
<td>–</td>
</tr>
<tr>
<td>Latitude (N)</td>
<td>39.4650</td>
<td>39.4675</td>
<td>39.465</td>
</tr>
<tr>
<td>Longitude (W)</td>
<td>111.2370</td>
<td>111.2248</td>
<td>111.224</td>
</tr>
<tr>
<td>Depth (km)</td>
<td>1.6</td>
<td>0.6</td>
<td>–</td>
</tr>
</tbody>
</table>

Note: CU, University of Colorado, this study; UUSS, University of Utah Seismograph Stations.
The epicentral location of a seismic point source is specified by the time-space location vector \( \mathbf{z}_0 = (t_0, \mathbf{x}_0) \), whose components are source time, latitude, and longitude. In this paper we ignore source depth, and focus on determining epicentral location. Let \( \rho(\mathbf{x}, \mathbf{y}) \) be the great-circle distance between the surface points \( \mathbf{x} \) and \( \mathbf{y} \). Consider the set of observation points (seismic stations) deployed in the circular vicinity of radius \( \rho_1 \) centred on a hypothetical source region. This set of stations may be split into two subsets: base stations \( \mathbf{x}_i, i = 1, \ldots, n \), and remote stations \( \mathbf{x}_j, j = 1, \ldots, m \). The base stations are deployed in the vicinity of the source location (local area) \( \mathbf{x}_0 \), \( \rho(\mathbf{x}_0, \mathbf{x}_i) \leq \rho_2 \), where we define \( \rho_2 \) to be not greater than 100–150 km. The radius \( \rho_1 \) of the entire set of receiving stations is taken here to be between 400 and 500 km. The remote stations are deployed in the annulus formed by radii \( \rho_1 \) and \( \rho_2 \), \( \rho_2 < \rho(\mathbf{x}_0, \mathbf{x}_j) \leq \rho_1 \). This lay-out is depicted schematically in Fig. 1.

To locate a particular event we require two data sets: vertical component event records for all remote stations and vertical–vertical component EGFs determined from ambient noise. (In principle, horizontal components also could be used for both Rayleigh and Love waves.) The event records observed at the remote stations, \( w_j(t, \mathbf{x}_j) \), \( j = 1, \ldots, m \), or simply \( w_j(t) \), must include the Rayleigh waves and must be a continuous segment of data starting from some fixed time \( t_L \). The EGFs, \( r_{ij}(t, \mathbf{x}_i, \mathbf{x}_j) \), or simply \( r_{ij}(t) \), must be present for some subset of station indices \( i \) and \( j \), where \( i = 1, \ldots, n \) refers to the base stations and \( j = 1, \ldots, m \) to the remote stations. We use the symmetric representation of the EGFs, which is the half-sum of the positive and negative correlation lags. Thus, each EGF, \( r_{ij}(t) \), starts from zero time. For simplicity, we will assume that EGFs exist for all combinations of the indices \( i \) and \( j \), but this will generally not be the case in practice.

### 3 Epicentral Location Method: The Technique

#### 3.1 Notation, observational setting and input data

The epicentral location of a seismic point source is specified by the time-space location vector \( \mathbf{z}_0 = (t_0, \mathbf{x}_0) \), whose components are source time, latitude, and longitude. In this paper we ignore source depth, and focus on determining epicentral location. Let \( \rho(\mathbf{x}, \mathbf{y}) \) be the great-circle distance between the surface points \( \mathbf{x} \) and \( \mathbf{y} \). Consider the set of observation points (seismic stations) deployed in the circular vicinity of radius \( \rho_1 \) centred on a hypothetical source region. This set of stations may be split into two subsets: base stations \( \mathbf{x}_i, i = 1, \ldots, n \), and remote stations \( \mathbf{x}_j, j = 1, \ldots, m \). The base stations are deployed in the vicinity of the source location (local area) \( \mathbf{x}_0 \), \( \rho(\mathbf{x}_0, \mathbf{x}_i) \leq \rho_2 \), where we define \( \rho_2 \) to be not greater than 100–150 km. The radius \( \rho_1 \) of the entire set of receiving stations is taken here to be between 400 and 500 km. The remote stations are deployed in the annulus formed by radii \( \rho_1 \) and \( \rho_2 \), \( \rho_2 < \rho(\mathbf{x}_0, \mathbf{x}_j) \leq \rho_1 \). This lay-out is depicted schematically in Fig. 1.

To locate a particular event we require two data sets: vertical component event records for all remote stations and vertical–vertical component EGFs determined from ambient noise. (In principle, horizontal components also could be used for both Rayleigh and Love waves.) The event records observed at the remote stations, \( w_j(t, \mathbf{x}_j) \), \( j = 1, \ldots, m \), or simply \( w_j(t) \), must include the Rayleigh waves and must be a continuous segment of data starting from some fixed time \( t_L \). The EGFs, \( r_{ij}(t, \mathbf{x}_i, \mathbf{x}_j) \), or simply \( r_{ij}(t) \), must be present for some subset of station indices \( i \) and \( j \), where \( i = 1, \ldots, n \) refers to the base stations and \( j = 1, \ldots, m \) to the remote stations. We use the symmetric representation of the EGFs, which is the half-sum of the positive and negative correlation lags. Thus, each EGF, \( r_{ij}(t) \), starts from zero time. For simplicity, we will assume that EGFs exist for all combinations of the indices \( i \) and \( j \), but this will generally not be the case in practice.

#### 3.2 Equalization of spectral amplitudes

As pointed out by Shapiro & Campillo (2004) and Bensen et al. (2007) as well as others, the procedure used to find the EGFs between a pair of stations does not preserve the amplitude spectrum...
Figure 2. Comparison of an earthquake record with an Empirical Green’s Function determined from ambient noise. (a) Earthquake record at station V04C in southern California band-pass filtered between 7 and 15 s period. (Earthquake 12 from Table 3: May 12, 2006; 38.816N, 122.816E; $m_b = 4.5$, located by the Northern California Earthquake Center). (b) EGF between stations GASB (north of the earthquake) and V04C, similarly band-passed. Red lines mark times corresponding to group velocities of 2.5 and 4 km s$^{-1}$. (c) Same as (b), but the EGF has been shifted and deformed to correspond to the earthquake location. Envelope functions are shown in (a), (b) and (c) with dashed lines. (d) Map showing earthquake and station locations.

Figure 3. Comparison of ambient noise Empirical Green’s Functions for the set of base stations. Similar to Fig. 2, but only the envelopes of the EGFs for the eight base stations are shown in (a). The envelope of the earthquake record is shown in red at top and the envelopes of the Composite EGF (sum of the individual EGFs transformed to the epicentral location) is on the second line.
of these functions. To overcome this problem, we equalize the amplitude spectra of the event record and the EGF without changing its phase spectrum. This could be done, of course, in a variety of ways; for example:

(1) by replacing the amplitude spectrum of the event with the amplitude spectrum of the EGF,

(2) by replacing the EGF amplitude spectrum with the event spectrum,

(3) by replacing both the EGF and event amplitude spectra with some average of the two, for example, the square root of the product of the amplitude spectra of the event and the EGF.

The first and the third equalization techniques provide quite similar results. The second technique may amplify noise in the case of weak events. For this reason, we have chosen procedure (1).

After amplitude equalization the event and CEGF signals typically become quite similar, as can be seen in Figs 5(c) and (d). To select the surface waves from both signals, cosine tapering in time is applied to the beginning and ending parts of both signals.

3.3 FT transformation

We transform the event record \( w_j(t) \) into the FT field \( W_j(t, \omega_l) \), and the EGF \( r_j(t) \) into the FT field \( R_j(t, \omega_l) \) by passing both equalized signals through a set of \( l = 1, 2, \ldots, L \) narrow-band frequency domain filters. For all event location examples in this paper \( L = 32 \), with filters within the bandpass period range from 7 to 15 s.

Both FT fields are formed by a set of envelopes obtained using narrow-band output signals and their Hilbert transforms. Thus, for \( w_j(t) \), the field \( W_j(t, \omega_l) \) is given by

\[
W_j(t, \omega_l) = 2 \int_0^\infty K(\omega - \omega_l) \tilde{w}_j(\omega) e^{i\omega t} d\omega ,
\]

where \( \tilde{w}_j(\omega) \) is the Fourier transform of \( w_j(t) \), i.e.

\[
\tilde{w}_j(\omega) = \int_{-\infty}^{\infty} w_j(t) e^{-i\omega t} dt
\]

and \( K(\omega - \omega_l) \) is the narrow-band Gaussian filter centred around centre frequency \( \omega_l \)

\[
K(\omega - \omega_l) = e^{-\frac{(\omega - \omega_l)^2}{2\beta^2}} .
\]

The parameter \( \beta \) defines the width of the Gaussian filter and does not depend on \( \omega \). For more details see Levshin et al. (1989).

In the same way we introduce the FT field \( R_j(t, \omega_l) \) for the cross-correlation time-series \( r_j(t) \),

\[
R_j(t, \omega_l) = 2 \int_0^\infty K(\omega - \omega_l) \tilde{r}_j(\omega) e^{i\omega t} d\omega ,
\]
3.4 Interpolation of FT fields

For location purposes we need to interpolate the FT fields \( R_i(t, \omega) \) to an arbitrary geographical point \( x \) inside the area covered by the base station deployment. The simplest way to do this is to transform \( R_i(t, \omega) \) in the time domain as follows: \( \tilde{R}_i(t, \omega) = R_i(t/d_i, \omega) \),

\[
\tilde{R}_i(t, \omega) = R_i(t/d_i, \omega),
\]

(4)

where \( \tilde{R}_i(t, \omega) \) is the interpolated field at point \( x \), \( d \) is the distance between the base station \( i \) and the point \( x \) [or \( d = \rho(x_i, x_j) \)], and \( d_i \) is the distance between the base station \( i \) and the remote station \( j \) [or \( d_j = \rho(x_i, x_j) \)].

This transformation produces the correct position for the maxima of \( \tilde{R}_i(t, \omega) \) in time corresponding to the group traveltimes of the Rayleigh wave. More accurate interpolation can be performed by use of regional phase and group velocity maps obtained by tomographic inversion of EGF data (e.g. Shapiro et al. 2005; Lin et al. 2008). This more rigorous approach is beyond the scope of this paper.

By adding \( \tilde{R}_i(t, \omega) \) for different indices \( i \) and fixed \( j \) we obtain

\[
C_j(t, x, \omega) = \frac{1}{N} \sum_{i=1}^{N} a_i \tilde{R}_i(t, \omega),
\]

(5)

where \( C_j(t, x, \omega) \) is the FT-representation of the CEGF for a given point \( x \) within the base station deployment and remote station \( j \), the coefficients \( a_i \) are weights that depend on the distance between each base station \( i \) and location \( x \), and \( N \) is the sum of the weights.

3.5 Cross-correlation of FT-fields and evaluation of residuals

We assume that the two FT fields, the CEGF \([C_j(t, x, \omega)]\) and the event \([W_j(t, \omega)]\), are similar for each remote station \( j \) but shifted in time due to the difference in position between epicentre \( x_0 \) and a trial point \( x \), as well as the unknown difference between the origin time \( t_0 \) and the starting time of records \( t_a \). The search for the unknowns \( x_0 \) and \( t \) is performed by temporal cross-correlation of the two fields \( C_j(t, x, \omega) \) and \( W_j(t, \omega) \) for all values of the remote station index \( j \).

Let us introduce the cross-correlation function \( Q \) between fields \( C_j \) and \( W_j \) as

\[
Q_j(\tau, x) = \int_0^\infty \int_0^\infty C_j(t, \omega)W_j(t+\tau, \omega)\,dt\,d\omega.
\]

(6)

We define the value of \( \tau \) at the maximum of the resulting cross-correlogram \( Q_j(\tau, x) \) as \( \tau_{j,\text{max}}(x) \). The final expression for the residual between the observed and predicted traveltimes is as follows:

\[
s_j(t, x) = \frac{1}{\tau_{j,\text{max}}(x)} - (t_d - t).
\]

(7)

Here \( t_d - t \) is the time-shift due to the unknown origin time \( t_0 \). Repeating this procedure for all \( j \), \( j = 1, 2, \ldots, m \), we obtain the vector of traveltime residuals for the given point \( x \). This vector should be considered as equivalent to the vector of traveltime residuals in standard location procedures. The location problem is the minimization in space–time of the weighted squared norm of the residual vector given by

\[
\min_{t, x} \sum_{j} V_{ji}^{-1}s_j^2(t, x),
\]

(8)

where \( V \) represents the \((m \times m)\) diagonal covariance matrix (weights) of errors in data measurements. Usually, the residuals are evaluated across the area around an event of interest, for times \( t \) limited by the lengths of both records.

3.6 Grid search

In this paper, we apply a simple brute force method to find the minimum of the functional (8). To do this we build a 2-D regular rectangular grid \( x_k \), \( k = 1, 2, \ldots, K \), in the region surrounding a preliminary solution \( x_0 \) and compute the full residual
at each gridpoint. The minimum value of the function $f$ at some gridpoint $x_{k0}$ provides the solution. (Examples of full residual functions are plotted later in Figs 10b and 11b.) Due to the ability to find the residual for any geographical point in the area covered by the base station deployment, it is possible to use more sophisticated techniques to search for the minimum, such as the conjugate gradient method (e.g. Atkinson 1988).

3.7 Confidence ellipsoid

Numerical trials demonstrate that the residuals $s_j(t, x)$ depend linearly on the spatial coordinates $x$. Time $t$ is also a linear term

\begin{equation}
    f(x_k) = \min_t \sum_j V_j^{-1} s_j^2(t, x_k)
\end{equation}

Figure 6. Illustration of location procedure based on cross-correlation of frequency-time (FT) representations of an event record and an EGF. (a) Event-station geometry. Event location (star) of the Crandall Canyon mine collapse and position of the base station P16A and the remote station T14A (triangles) are shown. The epicentre is surrounded by a 9-point grid (bold dots). (b) More detailed picture of the grid with gridpoints numbered from 1 to 9. The curved lines are circles centred on the remote station T14A. (c) FT-diagram of the event record for station T14A following the Crandall Canyon Mine mine collapse. (d) FT-diagram of the EGF for the station-pair T14A and P16A (transformed to the distance between gridpoint 9 and T14A). (e) Cross-correlogram (eqn. (6)) of the two FT-representations in (c) and (d) with the time delay of the maximum around 10 s. (f) Time delays of maxima of cross-correlograms as a function of distance from remote station T14A for all nine gridpoints. The minimum time delay corresponds to gridpoint 5, near the known epicentre of the mine collapse. Linear regression time-distance is shown by a solid line.
in the residual expression (7). Assuming that the linear part gives the principle contribution to the residuals and the rest is normally distributed noise, we follow the standard technique (Flinn 1965; Jordan & Swerdrup 1981) to construct a 2-D confidence ellipsoid at the point \((x_0, t_0)\) with a minimum of (8). (Examples of 90 per cent confidence ellipsoids are shown later in Figs 10b and 11b.)

### 3.8 Example of the Crandall Canyon Mine Collapse

The process described above is exemplified in Fig. 6 for a pair of USArray Transportable Array stations, station P16A which is near the Crandall Canyon mine collapse and a remote station T14A about 300 km to the southwest. In this example, we take only a single base station, and do not stack over several EGFs for simplicity of presentation. The amplitude equalized event record after the mine collapse observed at station T14A is shown in Fig. 5(c), and the associated EGF for the station pair T14A-P16A is in Fig. 5(d). The event-station geometry is shown in Fig. 6(a). For demonstration purposes, we impose a small regular nine-point grid (Figs 6a and b) with a grid step of 20 km where the Ground Truth position of the mine collapse is located at the centre of the grid, point 5. Gridpoints are numbered between 1 and 9, as shown in Fig. 6(b).

For each gridpoint we transform the FT representation of the EGF between stations T14A and P16A by means of eq. (4). The FT representations of the mine collapse and the EGF for gridpoint 9 are shown in Figs 6(c) and (d) and the cross-correlation between these representations (eq 6) for the band between 7 and 15 s period is shown in Fig. 6(e). Although the FT representations are similar, they are shifted in time by about 9 s. Moving across the grid from points 1 to 9 changes the distances to the remote station T14A from 285 to 340 km. The resulting time-shifts in the cross-correlograms for all nine gridpoints as a function of the distance from the remote station T14A are shown in Fig. 6(f). Positions of the maxima of the cross-correlograms between the FT fields correspond to time residuals. The time-shift for gridpoint 5 has the value closest to zero, indicating that this point of the grid is the nearest to the epicentre of the mine collapse.

Note that the times of the maxima of these cross-correlograms depend linearly on the distance between the gridpoint and the remote station. The linearity of the residuals versus small spatial perturbations around a gridpoint is evidence that the squared norm of residual vector is a quadratic form of the spatial coordinates. This justifies the statement above that the task of location is reduced to searching for the vector \((t_0, x_0)\) in eq. (8) that minimizes the squared sum of residuals for all remote stations.

---

**Figure 7.** Location errors of 445 ‘virtual sources’ (stations) in California and surrounding states. (a) Map of location errors for the stations. The distance between the known station location and the estimated location, in km, for each station position is shown by the colored circles. Grid spacing for location is 500 m. (b) Histogram of location errors (in km).
4 FURTHER EXAMPLES

4.1 Locating stations as virtual sources

Tests of the ability to locate seismic events are normally based on Ground Truth locations. However, such locations are themselves subject to uncertainties. To minimize this problem, we present a test to locate seismic stations as virtual sources. To do this, first we consider each USArray Transportable Array (TA) station in the western US to be a virtual earthquake and the estimated EGF to every other TA station to be the event records. We then attempt to locate the station based on those event records by applying EGFs between all of the other stations in the neighborhood of the virtual earthquake with the remote stations. Another advantage of this method is that it eliminates the effects of unknown source mechanism and depth on the location. We assess the effects of these variables in Section 5.

Through 2007 August, more than 100,000 EGFs for station pairs in California and surrounding states had become available (e.g. Moschetti et al. 2007; Lin et al. 2008). To obtain these data, 445 USArray Transportable Array (TA) and permanent stations were used as identified with colored circles in Fig. 7(a). The results of the station location test are presented in Figs 7(a) and (b). The following parameters were used in this test: the grid size is a relatively coarse 500 m, $\rho_1 = 400$ km, $\rho_2 = 100$ km. That is, all stations within 100 km of each virtual event (i.e. each station) are used as base stations and the remote stations are taken from 100 km to 400 km from each virtual event. 61 per cent of all events have location errors less than 0.5 km, and 28 per cent have errors ranging from 0.5 to 1 km. Seven stations (1.5 per cent of 445) are located with an error greater than 3 km, and all have exceptionally poor azimuthal coverage. In addition, they are near the coast and their EGFs tend to be biased by nearby oceanic noise.
To evaluate how the number of remote stations and their azimuthal distributions influence location accuracy we focus tests on the TA station S10A in Central Nevada as a virtual earthquake (Fig. 8a). Using a jack-knife procedure, we systematically reduced the number of remote stations participating in the location from 126 to 10, while preserving a more-or-less homogeneous azimuthal coverage. This procedure included 10 runs for each given number of remote stations, but with different ensembles of stations selected. In order to remove the effect of grid increment, we performed locations on a very fine grid (25 m). The average of the mislocations for the 12 runs is shown in Fig. 8(b) as a function of the number of stations, presented with standard deviations. Note that the expected location error is less than 600 m even when the number of remote stations reduces to 10. Mislocation asymptotically approaches about 250 m, which we take as the practical limit for the current location algorithm when source mechanism and depth are eliminated. As discussed in Section 5, this residual error appears to be determined largely by time discretization (1 sps per hour).

We also systematically investigated the effect of open azimuth by eliminating remote stations in azimuthal sectors that vary in size from 30° to 240°. The azimuthal position of each sector of a given size was changed 12 times by shifting the central azimuth of the sector by 30°. The average mislocations for 12 runs as a function of open azimuth plotted with the standard deviations are shown in Fig. 8(c). As with most location methods, open azimuth is more important than number of remote stations for location accuracy. In particular, as open azimuth grows above about 240°, location accuracy degrades rapidly. This explains the occurrence of large location errors near the Pacific coast in Fig. 7(a). As long as open azimuth is less than 240°, however, location errors are expected to be less than 500 m and approach 250 m as open azimuth shrinks.

4.2 Crandall Canyon mine collapse

The second test is to locate the magnitude \( m = 3.9 \) mine collapse that occurred on Monday 2007 August 6 at 08:48:40 UTC at the Crandall Canyon mine, Utah (Pechmann et al. 2008). The US Geological Survey (USGS) and the University of Utah Seismograph Stations (UUSS) determined that this event does not have the characteristics of a typical, naturally occurring earthquake, but was caused by the collapse of the coal mine, trapping several minders underground. The schema of the mine is shown in Fig. 9, which has been taken from Pechmann et al. (2008). USGS and UUSS summary information about this event is presented in Table 1.

When this event occurred, the USArray TA was situated in Utah. To locate the event we used seven stations as base stations and 66 remote stations (instead of the 33 used by the USGS) affiliated with different networks: USGS, University of Utah, and the TA (Fig. 10a). The other location parameters were: grid size 100 m, \( \rho_1 = 400 \) km, \( \rho_2 = 100 \) km. The location results are as follows: 39.465N (USGS 39.465N), 111.224W (USGS 111.237W). The origin time-shift is equal to \(-0.26 \) s. Confidence ellipsoid parameters (Fig. 10b) are: \( a = 0.52 \) km (major semiaxis), and \( b = 0.43 \) km.
(minor semiaxis). Our location is shifted relative to the preliminary USGS location to the east by 1.5 km, which is within the footprint of the mine and in better agreement with the location of the mine collapse and the latest UUSS location (Pechmann et al. 2008).

4.3 Ground Truth earthquakes in California

Table 2 presents Ground Truth (GT) locations and our locations for 14 California earthquakes with $m_b$ between 4.5 and 5.1 that occurred in 2005 and 2006. The location parameters used in this test were: grid size 500 m, $\rho_1 = 400$ km, $\rho_2 = 100$ km. An example for event #1 is shown in Fig. 11. The GT locations for events in southern California are determined by CalTech and for events in northern California by the Northern California Earthquake Center in Menlo Park. Both organizations use a combination of broad-band and short period instruments and 1-D models calibrated for their region. CalTech uses both $P$ and $S$ phases, whereas only $P$ phases are used in northern California. The accuracy of locations for these events is believed to be better than 1 km (Egill Hauksson, James Dewey, Bob Engdahl, private communications, 2008). If we accept GT locations as exact, differences with them are considered to be ‘errors’ in distance and origin time. Because the GT locations are not exact, however, the errors we report here are conservative.

The location and origin time errors are listed in the two last columns of Table 2. The average error for the 14 events is 1.2 km, and the average error in time is 0.04 s. At the 90 per cent confidence level, our locations are better than 2 km. These error estimates are larger than those reported for the station location (virtual earthquake) experiment and the location of the mine collapse because source mechanism and depth contribute group time-shifts to the earthquake records that can bias the location using our method. This effect is discussed further in Section 5.

4.4 Smaller Ground Truth earthquakes in Utah

To investigate whether our method is applicable to events smaller than about magnitude 4.0, we applied the same approach to the location of weaker earthquakes ($3 < m < 4.0$) that occurred in Utah in 2007 and 2008. The location parameters used in this test were: grid size 200 m, $\rho_1 = 400$ km, $\rho_2 = 100$ km. These events are well located and characterized by the University of Utah (UUSS) and by the Department of Earth and Atmospheric Sciences at Saint Louis University (SLU) (J. C. Pechmann, private communication, 2010). Information about these events is given in Table 3. The configuration of stations is very similar to that shown in Fig. 10(a). Average mislocation is less than 500 m, which indicates the method appears to work well with little modification for events with magnitudes between 3.0 and 4.0. The smaller mislocation for the Utah events compared with the California events may be due to source depth, as these smaller events may be either shallower or deeper than the earthquakes in California. The effect of event depth on mislocation is discussed further in Section 5. The method may be able to be improved further by tuning the band width of the method to the magnitude of the event.

5 DISCUSSION

We have described a method to locate seismic epicentres based on cross-correlation of FT representations of CEGFs with observed event seismograms in the period band between 7 and 15 s. The accuracy of the location depends on a number of factors that we discuss further here.

5.1 Technical factors affecting location accuracy

There are a number of technical variables that affect the accuracy of locations. We have used EGFs and event seismograms with a time sampling rate of 1 sps, which leads to errors up to ±0.5 s in timing of the resulting correlograms. Increasing sampling rates to 10 sps improves the location of well located events but has little effect for events mislocated by 1 km or worse. The size of the spatial grid spacing also affects accuracy, of course. The examples presented here have different spatial grids and the spacing of grid nodes is noted in each case. The current technique to interpolate the FT representations of the wavefields (eq. 4) assumes the constancy of surface wave group velocities in the frequency range considered. Significant frequency dependence of group velocities may distort

![Figure 11](image-url). California event #1 station configuration and misfit surface. (a) Symbols: base stations (green triangles), remote stations (blue triangles), and event location (red star). (b) Map of residuals: our event location (green star), USGS location (red star), 90 per cent confidence ellipsoid (yellow line). Gridpoints are marked with white crosses and are separated by about 500 m.
the FT representations. Corrections for the frequency dependence of dispersion may be introduced before the cross-correlation, however. This is particularly important in regions with strong lateral variability of crustal structure near the station, the effects of which can be decreased by using phase and group velocity maps, which now exist in the western US (e.g. Moschetti et al. 2007; Lin et al. 2008) and many other regions.

5.2 Event source mechanism and depth
The most significant physical effects on location accuracy are source mechanism and depth. EGFs determined from ambient noise correspond to surface sources with a predominant source mechanism close to a vertical force. The event seismograms, however, may result from different types of sources at different depths: for example, earthquakes within Earth’s crust, volcanic explosions, meteoritic impacts, explosions, mine collapses, or other human-related excitation processes. The radiation patterns of earthquakes depend on the source mechanism and depth as well as local structure, and are quite variable. Because the envelopes of event records and EGFs are what we compare in the FT domain to determine the epicentre, it is group time-shift (rather than source phase) that is the relevant effect of event mechanism and depth. Levshin et al. (1999) discuss group time-shifts expected for different types of events in detail, and we only summarize briefly here.

![Diagram of event source mechanism and depth](image)

**Figure 12.** Simulation of the effect of source mechanism and depth on group time-shifts and mislocation. Combination of base-remote stations used in the numerical simulation is shown in Fig. 8(a). (a) Source mechanisms (upper) and corresponding polar diagrams of group time residuals (lower) for the indicated source depths. Mechanism: red, near normal; green, near thrust; navy, near vertical thrust; light blue, near strike-slip. (b) Event mislocation for the four different source mechanisms as a function of depth. Symbol color correspond to the source mechanisms shown in (a).
According to Aki & Richards (1980) the azimuthal dependence of the displacement spectrum for a Rayleigh wave at a given frequency excited by a point double couple source at the depth \( h \) in a laterally homogeneous Earth is given by the complex function \( E(\omega, \phi, h) \)

\[
E(\omega, \phi, h) = \frac{dr_2}{dz} \left| M_{zz} + k(\omega)r_1(h)[M_{xx} \cos^2 \phi + 2M_{xy} \sin \phi \cos \phi + M_{yy} \sin^2 \phi] + i \left[ \frac{dr_1}{dz} - k(\omega)r_2(h) \right] [M_{xz} \cos \phi + M_{yz} \sin \phi] \right|
\]

(10)

where \( k(\omega) \) is wave number which depends on frequency \( \omega \), \( M_{zz}, M_{xx}, M_{yy}, M_{xy}, M_{xz}, M_{yz} \) are moment tensor components, \( r_1 \) and \( r_2 \) are horizontal and vertical eigenfunctions, \( \phi \) is azimuth taken clockwise from north. The modulus \( |E| \) of the complex function \( E \) represents the source amplitude radiation pattern and the argument \( \varphi = \arg(E) \) represents the source phase delay. Both \( |E| \) and \( \varphi \) are real functions that depend on \( \omega, \phi \) and \( h \). The imaginary part of eq. (10) is proportional to the tangential component of stress and equals zero for the surface \( h = 0 \). As the phase of the radiation function (source phase delay) varies with azimuth, it produces different phase shifts in the spectra of the event seismograms. However, our location method measures source group (not phase) time delays, \( \delta t_U \), which are frequency derivatives of the source phase delays and will vary from station to station depending on azimuth.

The group time-shift may be expressed explicitly as

\[
\delta t_U = \frac{\partial \varphi}{\partial \omega} = \text{Im} \left\{ \frac{E^* E}{|E|^2} \right\},
\]

(11)

where the prime represents a frequency derivative. Substitution of \( E \) defined by eq. (10) into (11) shows that the source group time delay function can be rewritten as

\[
\delta t_U = \frac{A_1 \cos(\varphi + \alpha_1) + A_3 \cos(3\varphi + \alpha_3)}{1 + A_2 \cos(2\varphi + \alpha_2) + A_4 \cos(4\varphi + \alpha_4)},
\]

(12)

where \( A_i \) and \( \alpha_i (i = 1, \ldots, 4) \) are rather complicated functions of \( \omega \) and \( h \) whose details are not needed here.

Several conclusions can be drawn immediately from these equations.

(1) Levshin et al. (1999) showed that group time delays are generally much smaller than the source phase delays. This is because the phase delays tend to change relatively slowly with frequency and is one of the reasons we have ignored phase information in our location method. Nevertheless, group time delays are not negligible and can bias locations.

(2) It is evident from eq. (12) that group time-shifts are anti-symmetric in azimuth \( \phi \); that is, changing azimuth by 180° changes only the sign of the delay. Group time-shifts differing in azimuth

Figure 13. Theoretical prediction of group time delays at different periods. Azimuthal dependence of group time delays are presented for four periods and four different mechanisms (the same as in Fig. 12). The size of the time delays scale according to the bar at the lower right-hand side of each component of the figure, in seconds.
by 180°, therefore, will constructively interfere. Thus, averaging residuals over azimuth while searching for the source position only partially suppresses the effect of group time delays on location.

(3) For events at the free surface, group time delay are zero for any source mechanism due to zero amplitude of tangential stresses near the free surface. For very shallow depth events (less than 1 km) group time delays will be very small. Thus, shallow events will have very small location bias due to source mechanism. This is probably why the Crandall Canyon Mine collapse was located so well.

(4) For some mechanisms, such as a pure normal, pure thrust, or a pure strike-slip, group times delays are exactly zero. This can be seen from eq. (10). For pure normal and thrust faults (45° dip and ±90° slip) the real part of E is zero. For pure strike-slip faults (90° dip and 0° slip), the imaginary part of E (traction at the surface) is zero. In these cases, arg(E) and the frequency derivative of E will be zero, so the group time delay will be zero. As discussed by Levshin et al. (1999), the source mechanism for most earthquakes is relatively close to the pure mechanisms.

These observations demonstrate that our location method will be biased minimally by events that are near the surface or that possess nearly strike-slip, thrust, or normal faulting mechanisms. It remains to be determined quantitatively how bias sets on as events deepen and mechanisms diverge from these pure types. Numerical simulations are needed.

Numerical simulations of group time delays for events between depths of 2 and 20 km show significant variability with depth, azimuth from the source, and period. Because source mechanisms at these depths typically are close to pure strike-slip, thrust, or normal fault mechanisms, they create very small group delays (from 0 to ±1 s) at periods between 6 and 15 s. More complicated mechanisms, however, can create delays up to ±(3–4) s, depending on azimuth, depth and period.

To quantify group time-shifts for a variety of source mechanisms and depths for crustal events we computed simulated Green’s functions for a laterally homogeneous model typical of central Nevada (Shapiro & Ritzwoller 2002) using a surface wave 1-D synthetic code (Herrmann 1979; Levshin et al. 1989). The base and remote stations used are those shown in Fig. 8(a). The simulated Green’s functions correspond to a vertical force acting at the Earth’s surface. We also calculated simulated event seismograms at remote stations for four events with a fixed geographical position within the base network. These events have the source mechanisms shown in the upper part of Fig. 12. These mechanisms are similar to pure normal (red) and thrust (green) faults, a vertical thrust fault (navy blue) and a strike-slip fault (light blue), but the corresponding angles characterizing the double couple mechanism (dip and rake) are 15°–20° different from the pure mechanisms, which is sufficient to produce a significant group time-shift. The polar diagrams in Fig. 12(a) show the azimuthal distribution of group time residuals for the corresponding mechanisms with source depths of 2, 5 and 10 km. Note

**Figure 14.** Theoretical prediction of group time delays at different periods. Same as Fig. 13, but for events at 5 km depth.
that our location method is based on cross-correlating FT diagrams between 7 and 15 s period. Thus, the results in Fig. 12(a) are frequency averaged.

Applying our location technique to the simulated data we find errors in source locations as a function of source depth for the four mechanisms, as seen in Fig. 12(b). Shallow focus events (depths less than about 1 km) produce very small group time-shifts, and therefore, minimal location bias. Similarly, events deeper than about 7 km have location errors smaller than 500 m, at least given the period band of this study (7–15 s). Generally, events with mechanisms similar to thrust or normal faults will have a larger mislocation than those similar to strike-strike. Therefore, non-strike-slip earthquakes with depths between 1 and 5 km provide the greatest challenge for our method.

Our current method is integrated over frequency and it is possible that in the location process the observation of the azimuthal distribution of the group time residuals as a function of period may be used to provide information about the accuracy of the epicentral location or provide information about source depth. Figs 13–15 present polar diagrams of theoretical group time-shifts as a function of period for three event depths for the same mechanisms as in Fig. 12. These results indicate that the azimuthal patterns of group time-shifts tend not to change appreciably with period, but the amplitudes can vary strongly. Thus, in the location procedure, the observation of frequency-independent group time residuals is evidence that the epicentre of the event has not been strongly biased by source mechanism. Conversely, the observation of strong frequency-dependence is evidence that the epicentre may be biased, but that the event probably occurred between depths of about 1 and 5 km.

6 CONCLUSIONS

The method to locate the epicentre of regional seismic events described here has several features that make it a useful addition to existing location methods.

(i) Its accuracy does not require knowledge of Earth structure.
(ii) It works for weak events where the detection of body wave phases may be problematic.
(iii) The EGFs computed during a temporary deployment of a base network (such as the USArray Transportable Array or PASSCAL deployments) may be applied to events that occurred earlier or later than the temporary network using permanent remote stations even when the temporary stations are not present.

In addition, the method has several evident limitations.

(i) It requires a relatively dense base network within an area of interest for at least a few months.
(ii) It provides the best accuracy when the frequency derivative of the source phase is relatively small; for example, when the source mechanism is a vertical force or a centre of compression, when the

Figure 15. Theoretical prediction of group time delays at different periods. Same as Fig. 13, but for events at 10 km depth.
source depth is less than 1 km or more than 5 km, and when the source mechanism is nearly purely strike-slip, thrust, or normal.

The technique described here is the first attempt at such a non-standard approach to locate regional events using ambient noise EGFs. Among possible directions for improvement are the incorporation of phase information, considering not only vertical but also transverse components of the EGFs which are dominated by Love waves, the transformation (interpolation) of EGFs into CEGFs based on tomographic group and phase velocity maps for the region of interest (e.g. Moschetti et al. 2007; Lin et al. 2008), and the application of information about the frequency-dependence of the residuals. Many parameters define the technical instantiation of the method, such as the frequency band, group velocity window, base/remote station geometry, the spatial grid, and the time-series sampling rate. These parameters can be varied systematically to tune the method to different observational settings and for events of different magnitude and character.

ACKNOWLEDGMENTS

The authors gratefully acknowledge helpful conversations with Stephen Myers and E. Robert Engdahl, constructive reviews from Toshiro Tanimoto and an anonymous reviewer, and thank James Dewey for providing the list of Ground Truth earthquakes in California and J.C. Pechmann for providing similar information about the smaller events in Utah. This research was supported by DoE/NNSA contract DE-AC52-09NA29326. The facilities of the IRIS Data Management System, and specifically the IRIS Data Management Center, were used to access the waveform and metadata required in this study. The IRIS DMS is funded by the National Science Foundation and specifically the GEO Directorate through the Instrumentation and Facilities Program of the National Science Foundation under Cooperative Agreement EAR-0552316.

REFERENCES