Relocating a Cluster of Earthquakes Using a Single Seismic Station

David J. Robinson, Malcolm Sambridge, Roel Snieder, Juerg Hauser

David Robinson: Earth Monitoring and Hazard Group
Geoscience Australia
GPO Box *383*
Canberra ACT *2601* Australia
E-mail: david.robinson@ga.gov.au

Abstract

~~Coda waves~~ Coda-waves arise from scattering to form the later arriving components of a seismogram. ~~Coda wave~~ Coda-wave interferometry is an emerging tool for constraining earthquake source properties from the intereference pattern of ~~coda waves~~ coda-waves between nearby events. A new earthquake location algorithm is derived which relies on ~~coda wave~~ coda-wave based probabilistic estimates of earthquake separation. The algorithm can be used with ~~coda waves~~ cod-waves alone or in tandem with ~~travel time~~ arrival-time data. Synthetic examples in 2D and 3D and real earthquakes on the Calaveras Fault, California are used to demonstrate the potential of ~~coda waves~~ coda-waves for locating poorly recorded earthquakes. It is demonstrated that ~~coda wave~~ coda-wave interferometry: (a) outperforms traditional earthquake location techniques when the number of stations is small; (b) is self-consistent across a broad range of station situations; and (c) can be used with a single station to locate earthquakes.

Introduction

Accurate earthquake location is important for many applications. Locations are required for: magnitude determination (*Richter*, 1935; *Gutenberg*, 1945); computing moment tensors (*Sipkin*, 2002); seismological studies of the Earth's interior (*Spencer and Gubbins*, 1980; *Kennett et al.*, 1995; *Curtis and Snieder*, 2002; *Kennett et al.*, 2004); understanding strong motion and seismic attenuation (*Toro et al.*, 1997; *Campbell*, 2003) and modeling earthquake hazard or risk (*Frankel et al.*, 2000; *Stirling et al.*, 2002; *Robinson et al.*, 2006). The accuracy required in earthquake location depends on the application. For example, imaging the structure of a fracture system from microseismicity requires greater ~~detail~~ location accuracy than determining whether a $M_w = 7.5$ earthquake occurs offshore for tsunami warning. This paper focuses on reducing location uncertainty for a cluster of events when they are recorded by a small number of stations.

Absolute location describes the location of an earthquake with respect to a global reference such as latitude, longitude (or easting/northing) and depth. Uncertainties associated with absolute locations are influenced by source to station distances, the number of stations and their geometry, signal-to-noise ratio, clarity of onsets and accuracy of the velocity model used in computing ~~travel times~~arrival-times. Uncertainties in absolute location are typically of the order of several kilometers, primarily because they are susceptible to uncertainty in the velocity structure along the entire path between the source and receiver. For example, *Shearer* (1999) states that location ~~uncertainty~~ uncertainties in the ISC (International Sesimological Centre) and PDE (National Earthquake Information Center) catalogues are generally around 25 km horizontally and at least 25 km in depth (Here the depth uncertainties of 25 km assume the use of depth dependent phases such as pP. Without such phases the uncertainty is higher). *Bondár et al.* (2004) demonstrate that ~~at the local scale,~~ absolute locations are accurate to within 5 km with a 95% confidence level when local networks meet ~~a number of station related criteria.~~ the following criteria:

1. there are 10 or more stations, all within 250 km,

2. an azimuthal gap of less than 110^o,
3. a secondary azimuthal gap of less than 160^o, and
4. at least one station within 30 km.

Such errors are too large for many applications, particularly those focussed on imaging rupture surfaces from aftershock sequences.

Relative earthquake location involves locating a group of earthquakes with respect to one another and was first introduced by *Douglas* (1967) who developed the technique commonly known as joint hypocenter determination (*Douglas* (1967) originally used the term joint epicentre determination. However, he was solving for hypocentre). In principle, relative locations can be computed by differencing absolute locations. However, *Pavlis* (1992) shows that inadequate knowledge of velocity structure leads to systematic biases when relative positions are computed in this way. To reduce errors from unknown velocity structure, relative location techniques typically compute locations directly from ~~travel time differences between two waveforms . By doing so , they remove~~ arrival-time differences computed by time-lag cross correlation of early-onset body waves (*Ito*, 1985; *Got et al.*, 1994; *Slunga et al.*, 1995; *Nadeau and McEvilly*, 1997; *Waldhauser et al.*, 1999). Doing so removes errors associated with velocity variations outside the local region, because such variations influence all waveforms in ~~the same~~ a similar manner (*Shearer*, 1999).

Reported location uncertainties from relative techniques are around 15 to 75 m in local settings with good station coverage (*Ito*, 1985; *Got et al.*, 1994; *Waldhauser et al.*, 1999; *Waldhauser and Schaff*, 2008). Here, ‘good coverage' implies multiple stations distributed across a broad range of azimuthal directions. Relative location techniques have been used to image active fault planes (*Deichmann and Garcia-Fernandez*, 1992; *Got et al.*, 1994; *Waldhauser et al.*, 1999; *Waldhauser and Ellsworth*, 2002; *Shearer et al.*, 2005); study rupture mechanics (*Rubin et al.*, 1999; *Rubin*, 2002); interpret magma movement in volcanoes (*Frèmont and Malone*, 1987); and monitor pumping-induced seismicity (*Lees*, 1998; *Ake et al.*, 2005).

In traditional approaches to absolute and relative location only early onset body waves, typically P and/or S waves, are used. The data utilised may be the direct ~~arrival times; travel time~~ arrival-times; arrival-time difference computed between picked arrivals of two waveforms; or ~~time~~ arrival-time differences inferred from time-lagged cross correlation of relatively small windows around the body wave arrivals. In all three cases, the majority of the waveform is discarded. Furthermore, obtaining high accuracy with these techniques requires multiple stations with good azimuthal coverage. In this paper we demonstrate that it is possible to significantly reduce location uncertainty when few stations are available by using more of the waveform.

Coda refers to later arriving waves in the seismogram that arise from scattering (*Aki*, 1969; *Snieder*, 1999, 2006). Coda waves are ignored in most seismological applications due to the complexity involved in constraining complex hetergeneous velocity models in real settings. In this paper we develop an approach for locating earthquakes using ~~coda waves~~coda-waves. *Snieder and Vrijlandt* (2005) demonstrate that the coda of two earthquakes can be used to estimate the separation between them. Their technique, known as coda wave interferometry (CWI), is based on the interference pattern between the ~~coda waves. Unlike travel time~~ coda-waves. Unlike arrival-time based location techniques, CWI does not require multiple stations or good azimuthal coverage. In fact, it is possible to obtain estimates of separation using a single station (*Robinson et al.*, 2007a). This makes CWI particularly interesting for regions where station density is low such as intraplate settings. In this paper we demonstrate how CWI separation estimates can be used to constrain location with data from a single station. Our technique can be used on ~~coda waves~~ coda-waves alone or in combination with ~~travel times~~arrival-times. We begin by introducing the theory of CWI based earthquake location. This is followed by a demonstration of capability using synthetic examples and application to earthquakes on the Calaveras fault, California using CWI alone and CWI in combination with ~~travel time~~ arrival-time constraints.

Theory

Snieder and Vrijlandt (2005) introduce a CWI based estimator of source separation δ_{CWI} between two earthquakes

$$\delta_{CWI}^2 = g(\alpha, \beta)\sigma_\tau^2, \tag{1}$$

where σ_τ is the standard deviation of the ~~travel time~~ arrival-time perturbation between the ~~coda waves~~ coda-waves of two earthquakes, and α and β are the near-source P and S wave velocities, respectively. The function g depends on the type of excitation (explosion, point force, double couple) and on the direction of source displacement relative to the point force or double couple.

For example, for two double couple sources displaced in the fault plane,

$$g(\alpha, \beta) = 7\frac{\left(\frac{2}{\alpha^6} + \frac{3}{\beta^6}\right)}{\left(\frac{6}{\alpha^8} + \frac{7}{\beta^8}\right)}, \tag{2}$$

whereas, for two point sources in a 2D acoustic medium

$$g(\alpha, \beta) = 2\alpha^2 \tag{3}$$

(*Snieder and Vrijlandt*, 2005). *Snieder and Vrijlandt* (2005) also show that for two double couple sources that are not in the same fault plane

$$\sigma_\tau^2 = \frac{\left(\frac{6}{\alpha^8} + \frac{7}{\beta^8}\right)\delta^2_{\parallel fault} + 2\left(\frac{1}{\alpha^8} + \frac{2}{\beta^8}\right)\delta^2_{\perp fault}}{7\left(\frac{2}{\alpha^8} + \frac{3}{\beta^8}\right)}, \tag{4}$$

where $\delta^2_{\parallel fault}$ and $\delta^2_{\perp fault}$ are the separation parallel and perpendicular to the fault, respectively. In this paper we use equation 3 for the 2D examples. For the 3D examples we use equation 2 which assumes that the source mechanism of both events are identical, an assumption likely to be true for events in the same fault plane. *Robinson et al.* (2007b) explore the impact of a change in mechanism.

The σ_τ in equation (1) is related to the maximum of the cross correlation between the coda of the two waveforms, R_{max}, and hence can be computed directly from the recorded data. The original formulation by *Snieder and Vrijlandt* (2005) used a second-order Taylor series expansion of the waveform autocorrelation function to relate σ_τ and R_{max} by

$$R^{(t,t_w)}_{max} = 1 - \frac{1}{2}\overline{\omega^2}\sigma_\tau^2, \tag{5}$$

where $\overline{\omega^2}$ is the square of the dominant angular frequency

$$\overline{\omega^2} = \frac{\int_{t-t_w}^{t+t_w} \dot{u}_i^2(t')dt'}{\int_{t-t_w}^{t+t_w} u_i^2(t')dt'}, \tag{6}$$

and $\dot{u}_i$ represents the time derivative of u_i. In this paper we use the autocorrelation approach of *Robinson et al.* (2011) to relate the parameters directly and we apply a restricted time lag search when evaluating R_{max}. These extensions to the original technique of *Snieder and Vrijlandt* (2005) increase the range of applicability of CWI by 50% (i.e. from 300 to 450 m separation for 1 to 5 Hz filtered ~~coda waves~~coda-waves).

Robinson et al. (2011) show that CWI leads to probabilistic constraints on source separation and introduce a Bayesian approach for describing the probability of true separation given the CWI data. Their approach is summarised by

$$P(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN}) \propto P(\widetilde{\delta}_{CWIN}|\widetilde{\delta}_t) \times P(\widetilde{\delta}_t) \tag{7}$$

where $P(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN})$ is the posterior function indicating the probability of true separation $\widetilde{\delta}_t$ given the noisy CWI separation estimates $\widetilde{\delta}_{CWIN}$; $P(\widetilde{\delta}_{CWIN}|\widetilde{\delta}_t)$ is the likelihood function (or forward model) giving the probability that the separation estimates $\widetilde{\delta}_{CWIN}$ would be observed if the true separation was $\widetilde{\delta}_t$; and $P(\widetilde{\delta}_t)$ is the prior ~~PDF~~ probability density function (PDF) accounting for all a-priori information. The use of N in δ_{CWIN} depicts CWI separations that include noise. The nomenclature is adopted here to remain consistent with *Robinson et al.* (2011) who study synthetically generated noise-free δ_{CWI} and relate them to noisy estimates δ_{CWIN}. The tilde above the separation parameters in equation (7) indicates the use of a wavelength normalised separation parameter

$$\widetilde{\delta} = \frac{\delta}{\lambda_d}, \tag{8}$$

which measures separation ($\delta = \delta_{CWIN}$ or δ_t) with respect to dominant wavelength λ_d. In this paper we consider a uniform prior over appropriate bounds to ensure that the posterior function is dominated by the recorded data. The procedure for computing the likelihood $P(\widetilde{\delta}_{CWIN}|\widetilde{\delta}_t)$ is derived by *Robinson et al.* (2011) and summarised in Appendix (The Likelihood) . With these two pieces in place we can compute the posterior $P(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN})$ (or PDF) for the separation between any pair of events directly from their coda waves.

We seek a ~~probability density function (PDF)~~PDF which links individual pairwise posteriors $P(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN})$ to describe the location of multiple events whose maximum corresponds to the most probable combination of locations. More importantly, however, the PDF shall quantify location uncertainty and provide information on the degree to which individual events are constrained by the data. For convenience, we begin with three earthquakes having locations $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$. Using a Bayesian formulation we write

$$P(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3|\mathbf{d}) \propto P(\mathbf{d}|\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) \times P(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3), \tag{9}$$

where $P(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3|\mathbf{d})$, $P(\mathbf{d}|\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ and $P(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ are the posterior, likelihood and prior functions, respectively. In equation (9) $\mathbf{d}$ represents observations that constrain the locations. They can be any combination of ~~travel times~~arrival-times, geodetic information or CWI separations. For example, if ~~coda waves~~ coda-waves are used we have $P(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3|\widetilde{\boldsymbol{\delta}}_{CWIN})$ and $P(\widetilde{\boldsymbol{\delta}}_{CWIN}|\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, where $\widetilde{\boldsymbol{\delta}}_{CWIN}$ are the wavelength normalised separation estimates. Alternatively, if we use CWI and ~~travel time~~ arrival-time data we may write $P(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 | \widetilde{\boldsymbol{\delta}}_{CWIN}, \boldsymbol{\Delta}_{TT})$ and $P(\widetilde{\boldsymbol{\delta}}_{CWIN}, \boldsymbol{\Delta}_{TT} | \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ where $\boldsymbol{\Delta}_{TT}$ ~~represent travel time~~ represents the arrival-time differences. In the following derivation and in the Synthetic Experiements and Relocating Earthquakes on the Calaveras Fault section we focus on the constraints imposed by ~~coda waves~~coda-waves, whereas in Combining ~~Travel Time~~ Arrival-Time and CWI Constraints we demonstrate how CWI and ~~travel time~~ arrival-time data can be combined.

For three earthquakes we have likelihoods; $P(\widetilde{\delta}_{CWIN,12} | \mathbf{e}_1, \mathbf{e}_2)$, $P(\widetilde{\delta}_{CWIN,13} | \mathbf{e}_1, \mathbf{e}_3)$ and $P(\widetilde{\delta}_{CWIN,23} | \mathbf{e}_2, \mathbf{e}_3)$. In writing these likelihoods we have replaced the conditional term on separation $\widetilde{\delta}_t$ with the locations (e.g. $\mathbf{e}_1$ and $\mathbf{e}_2$). This can be done because knowledge of location translates to separation. Note, however, that the reverse is not true. That is, knowledge of separation between a single event pair does not uniquely translate to location but rather places a non-unique constraint on location. ~~Furthermore, since the pairwise functions are independent the joint likelihood becomes~~ In other words, knowing $|e1 - e2|$ and $|e2 - e3|$ does not mean that $|e1 - e3|$ is uniquely defined. Consequently, the likelihoods are weekly dependent, in that some likelihood-pairs share common events, an occurrence that becomes relatively less frequent as the number of events being located increases. For the purpose of this work we ignore this week dependance and assume independence

$$\begin{aligned} P(\widetilde{\delta}_{CWIN} | \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) \cancel{=} \approx & P(\widetilde{\delta}_{CWIN,12} | \mathbf{e}_1, \mathbf{e}_2) \\ & \times P(\widetilde{\delta}_{CWIN,13} | \mathbf{e}_1, \mathbf{e}_3) \times P(\widetilde{\delta}_{CWIN,23} | \mathbf{e}_2, \mathbf{e}_3). \end{aligned} \tag{10}$$

Similarly, the earthquake locations are independent and the joint prior becomes

$$P(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = P(\mathbf{e}_1) \times P(\mathbf{e}_2) \times P(\mathbf{e}_3). \tag{11}$$

Combining equations (10) and (11) gives the joint posterior function

$$
\begin{aligned}
P(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 | \widetilde{\boldsymbol{\delta}}_{CWIN}) &= c \prod_{i=1}^{3} P(\mathbf{e}_i) \\
&\times \prod_{i=1}^{2} \prod_{j=i+1}^{3} P(\widetilde{\delta}_{CWIN,ij} | \mathbf{e}_i, \mathbf{e}_j)
\end{aligned}
\tag{12}
$$

for three events.

A detailed understanding of the location of a single event (e.g. $\mathbf{e}_2$) is obtained by computing the marginal

$$
P(\mathbf{e}_2 | \delta_{CWIN}) = \int \int P(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 | \widetilde{\delta}_{CWIN}) d\mathbf{e}_1 d\mathbf{e}_3, \tag{13}
$$

where the intergral is taken over all plausible locations for $\mathbf{e}_1$ and $\mathbf{e}_3$. Alternatively, we can compute the marginal for a single event coordinate by integrating the posterior over all events and remaining coordinates for the chosen earthquake. Evaluation of the normalizing constant c in equation (12) involves finding the integral of the posterior function over all plausible locations. In many applications the constant of proportionality c can be ignored. For example, it is not required when seeking the combination of locations which maximise the posterior function, nor in Bayesian sampling algorithms such as Markov-chain Monte-Carlo techniques which only require evaluation of a function proportional to the PDF.

Extending to n events we get the posterior function

$$
\begin{aligned}
P(\mathbf{e}_1, ..., \mathbf{e}_n | \widetilde{\delta}_{CWIN}) &= c \prod_{i=1}^{n} P(\mathbf{e}_i) \\
&\times \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} P(\widetilde{\delta}_{CWIN,ij} | \mathbf{e}_i, \mathbf{e}_j).
\end{aligned}
\tag{14}
$$

When evaluating equation (14) over a range of locations it is necessary to compute and multiply many numbers close to zero. This is because the PDFs tend to zero as the locations get less likely (i.e. near the boundaries of the plausible region). Such calculations are prone to truncation errors and so we work with the negative logarithm

$$L(\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n) = -ln\left[P(\mathbf{e}_1, ..., \mathbf{e}_n|\widetilde{\delta}_{CWIN})\right] \tag{15}$$

or

$$\begin{aligned} L(\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n) = -ln\,[c] - \sum_{i=1}^{n} ln\,[P(\mathbf{e}_i)] \\ - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} ln\left[P(\widetilde{\delta}_{CWIN,ij}|\mathbf{e}_i, \mathbf{e}_j)\right]. \end{aligned} \tag{16}$$

The logarithm improves numerical stability by replacing products with summations. The negative facilitates the use of optimisation algorithms that are designed to minimise an objective function.

The event locations $\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n$ are defined by coordinates $\hat{x}$, $\hat{y}$ and $\hat{z}$ where the hat indicates use of a local coordinate system. We choose a local coordinate system which removes ambiguity associated with transformations of the coordinate system. It is necessary to do this because the distances between events are invariant for rotations, reflections and translations of the seismicity pattern and hence cannot be resolved from CWI alone. In defining this coordinate system we fix the first event at the origin

$$\mathbf{e}_1 = (0, 0, 0), \tag{17}$$

the second event on the positive $\hat{x}$-axis

$$\mathbf{e}_2 = (\hat{x}_2, 0, 0), \hat{x}_2 > 0 \tag{18}$$

the third on the $\hat{x} - \hat{y}$ plane

$$\mathbf{e}_3 = (\hat{x}_3, \hat{y}_3, 0), \hat{y}_3 > 0 \tag{19}$$

and the fourth to

$$\mathbf{e}_4 = (\hat{x}_4, \hat{y}_4, \hat{z}_4), \hat{z}_4 > 0. \tag{20}$$

This coordinate system reduces translational (equation 17) and rotational (equations 18 to 20) non-uniqueness without loss of generality. It is necessary to work with a local coordinate system when using ~~coda waves~~ coda-waves alone because the CWI technique constrains only event separation between earthquakes. The inclusion of ~~travel times in Combining Travel Time~~ arrival-times in the Combining Arrival-Time and CWI Constraints section allows us to move to a global reference system.

In summary, the posterior $P(\mathbf{e}_1, ..., \mathbf{e}_n|\widetilde{\delta}_{CWIN})$ and its negative logarithm L describe the joint probability of multiple event locations given the observed ~~coda waves~~coda-waves. The most likely set of locations is given by the minimum of L. In this paper we use the Polak-Ribiere technique (*Press et al.*, 1987), a conjugate gradient method, to minimize L. It uses the derivatives of L, derived in Appendix (Derivatives) , to guide the optimization procedure. Note that when optimizing equation 16 the values of $ln\,[c]$ and $ln\,[P(e_i)]$ can be ignored because they are constant ($ln\,[P(e_i)]$ is constant because we consider a uniform prior).

Synthetic experiments

We use synthetic examples in 2D and 3D with 50 earthquakes to test the performance of the optimization routine. In these examples the synthetic earthquakes are located randomly and CWI data generated according to the event separation. It is not necessary to generate synthetic waveforms and compute CWI estimates directly because we are testing the performance of the optimization routine only. The ability of CWI to estimate event separation has been demonstrated already (*Snieder and Vrijlandt*, 2005; *Robinson et al.*, 2007a, 2011). We undertake a complete coda wave location experiment, including the calculation of CWI separation estimates, for recorded earthquakes in Relocating Earthquakes on the Calaveras Fault and in Combining ~~Travel Time~~ Arrival-Time and CWI Constraints.

Examples 1 and 2 - 2D synthetic experiments

We design a 2D synthetic acoustic experiment (example 1) to test the performance of our CWI based relative location algorithm by randomly selecting $\hat{x}$- and $\hat{y}$-coordinates such that $-50 \leq \hat{x}, \hat{y} \leq 50\,\text{m}$. These are indicated with triangles in Figure 1. We assume a local velocity of ~~$\alpha = 3,300$~~ $\alpha = 3300\,\text{ms}^{-1}$ between all event pairs and a dominant frequency of 2.5 Hz to represent waveform data filtered between 1 and 5 Hz. The ~~CWI~~ purpose of these examples is to synthetically test the location algorithm. Hence, we do not need to synthetically generate waveforms and compute CWI separation estimates, Rather, we begin by synthetically generating the CWI separation estimates directly. *Robinson et al.* (2011) showed that the CWI data are defined by the dominant wavelength normalized positive bounded Gaussian PDF with statistics $\bar{\mu}_N$ and $\bar{\sigma}_N$. A hypothetical CWI mean is created by setting

$$\bar{\mu}_N = \mu_1\left(\widetilde{\delta}_t\right) \tag{21}$$

using equation (A7). This assumption ensures that the sample mean of hypothetical separation estimates is consistent with known CWI biases (*Robinson et al.*, 2011). In example 1 we use $\bar{\sigma}_N = 0.02$ between all event pairs. Application of our optimization procedure on the hypothetical CWI data yields the circles in Figure 1. The optimization does not lead to the exact solution due to the addition of noise ($\bar{\sigma}_N = 0.02$) on the hypothetical CWI data. The average coordinate error is 2.0 m (average location error ≈4 m) which is small compared to the noise of $\bar{\sigma}_N = 0.02$ which for $v_s = 3300\,\mathrm{ms}^{-1}$ and $f_{dom} = 2.5\,\mathrm{Hz}$ corresponds to roughly 25 m.

Robinson et al. (2011) demonstrates that the noise on CWI estimates is often larger than 0.02 and that it increases with event separation. Consequently, example 1 is simplistic because we fix $\bar{\sigma}_N = 0.02$ for all pairs. In example 2 we increase the uncertainty and introduce a distance dependance into the hypothetical $\bar{\sigma}_N$ by defining $\bar{\sigma}_N = \epsilon(\delta_t)$, where $\epsilon(\delta_t)$ is the half-width of the errorbars for a synthetic acoustic experiment with filtering between 1 and 5 Hz (see Fig. 4(b) of *Robinson et al.*, 2011). Repeating the optimization leads to the circles in Figure 2 which have an average coordinate error of 2.8 m (average location error ≈9 m).

Conjugate gradient based optimization techniques are susceptible to the presence of local minima. This is because they use the slope of the target function to explore the solution space. We explore the impact of local minima for our CWI location problem by beginning the optimization from 25 randomly chosen starting positions. We observe ~~no differences in the solution for either example~~ negligble difference in the solutions indicating that neither example is susceptible to local minima.

Three observations can be drawn from the error structure in Figures 1 and 2. Firstly, the location errors depicted by gray bars increase between examples 1 and 2 with the introduction of larger noise. Secondly, the errors are larger for events at greater distances from the center. This is because events near the center of the cluster are constrained by links from all angles, whereas those on the outside are moderated by links from a limited number of directions. This observation is analogous to problems associated with poor azimuthal coverage in triangulation ~~problems~~ such as individual earthquake location from limited ~~travel time~~ arrival-time data, or GPS positioning with few satellites. Our third observation is that the location errors form a pattern of circular rotation, despite our attempt to correct for rotational non-uniqueness with the local coordinate system.

The local coordinate system works by constraining the location of the first three earthquakes. Earthquake 1 is fixed at the origin, earthquake 2 on the positive $\hat{x}$-axis and earthquake 3 has $\hat{y} > 0$. As the number of events increase the strength of these constraints on later events weakens allowing small rotations of events with respect to each other. That is, even though the rotational freedom of the cluster is in ~~principal~~ principle removed by the constraints imposed on the events (see equations (17) to (19 - equation 20 is needed in 3D only) we observe that in practice the presence of noise allows the rotational non-uniqueness to reappear. This is because errors align themselves in directions least constrained by data. For the CWI technique this ammounts to rotations in 2D. The same ~~phenomena~~ phenomenon is observed in linear inversion where noise creates large spurious model changes in directions of the eigenvectors with the smallest singular values (*Aster et al.*, 2005). Fortunately, however, combining ~~coda waves~~ coda-waves with measurements of ~~travel times~~ arrival-times alleviates this problem and ~~facilitate~~ facilitates the removal of a local coordinate system altogether (see Combining ~~Travel Time~~ Arrival-Time and CWI Constraints). On balance, however, we gain confidence in the optimization procedure due to its stability for different starting locations and because of the small average coordinate errors of 2.0 m and 2.8 m for examples 1 and 2, respectively.

Example 3 - The impact of incomplete event pairs in 2D

Synthetic examples 1 and 2 use 100% direct linkage between event pairs. That is, there is a constraint between each earthquake and all other events. In reality, we might expect that the separation between some pairs will not be constrained by CWI data due to poor signal to noise ratio in the coda for common stations. Obviously, the fewer stations that record an event the more likely it is that links between it and other events will be broken. In such cases the probabilistic distance constraint between a pair of events may only exist indirectly through multiple pairs. In this section we consider the impact of reduced linkage between event pairs. In example 3, we repeat example 2 using 90%, 80%, ..., 10% of the links. That is, we randomly select 10% of the event pairs and remove the separation estimates between those pairs to create a data set with 90% linkage. Then, we randomly remove 20% of the links and so on. This experiment is designed to mimic a realistic recording situation where CWI estimates are not available for all event pairs due to station problems, poor signal-to-noise ratio or any number of other reasons. As with the above examples, we undertake the optimization with

25 randomly chosen starting locations.

~~Figures 3 (a) and (b)~~ Figure 3 illustrates the maximum Δ_{max} (top) and mean Δ_{μ} (middle) of the coordinate error as a function of percentage of earthquake pairs that are directly linked by a separation estimate. We show the statistics for the 'best' optimization solution (~~black~~thick) and for the solution space when all 25 optimizations are considered (~~gray~~thin). In the former case the best solution is determined by the set of event locations which lead to the smallest value of L. The error in the best solution is consistent when 30% or more of the branches are used. The errors increase when only 10% or 20% of the constraints are included. Interestingly, this breakdown around 20% to 30% coincides with the point where the average number of branches required to link an event pair reaches 2 (see Fig.3 (bottom)). Since the average number of branches can be computed in advance it can be used as an indication of ~~the~~ inversion stability prior to optimization. A higher breakdown is observed when all 25 solutions are considered collectively. For example, the maximum coordinate error Δ_{max} exceeds that for the best solution for linkage $\leq$60% confirming that the optimization is susceptible to local minima and that a range of starting points should be considered. Some optimizations fail to converge after 1200 iterations when the linkage is 60% or lower. All optimizations fail when the linkage is 20% or lower. ~~Despite their failure to converge, the locations at final iteration are close to the actual solution.~~

The derivatives used in the conjugate gradient method depend on events connected by CWI measurements. Consequently, earthquakes that are only connected via other events do not 'communicate' with each other directly. To some extent, this should be adressed during the iterative process where location information can spread to events which have no direct links. However, the lack of direct connection through the gradient could prevent convergence in extreme cases, or more likely slow the procedure down. This could explain why some examples do not converge after 1200 iterations. *VanDecar and Snieder* (1994) show that derivative based regularization acts slowly through iterative least-squares, because every cell in one iteration communicates only with its neighbours, and they demonstrate that this can be fixed with preconditioning in some cases. Their findings suggest that it may be possible to improve the convergence (stability and/or speed) of the CWI optimization by preconditioning.

Example 4 - The impact of incomplete event pairs in 3D

In Example 4 we expand the optimization routine to 3D by randomly picking a set of ~~actual~~ event locations for 50 earthquakes with $-50\,\mathrm{m} \leq \hat{x}, \hat{y}, \hat{z} \leq 50\,\mathrm{m}$. As in the 2D case we assume a local velocity of $\cancel{v = 3,300}$ $v = 3300\,\mathrm{ms}^{-1}$ between all event pairs and a dominant frequency of 2.5 Hz to represent waveform data filtered between 1 and 5 Hz. The hypothetical CWI mean is created using equation (21) which ensures consistency between the sample mean of hypothetical separation estimates and CWI biases. We use a standard deviation for the noisy CWI estimates of $\cancel{\bar{\sigma}_N = \epsilon}$ ~~and~~ $\bar{\sigma}_N = \epsilon(\delta_t)$ (where $\epsilon(\delta_t)$ is the same as that used in Examples 2 and 3) and perform the optimization using 10%, 20%, ..., 100% of the direct links. In each case we repeat the optimization 25 times using randomly chosen starting locations. The results are summarised in Figure 4.

When 70% of the direct constraints are considered all optimization results (~~gray~~thin) are consistent with the best solution obtained from all 25 starting locations (~~black~~thick). The best solution constrains the event locations down to 30% of the direct links. ~~There is one notable difference between the 3D and 2D results. In 2D the final iteration was close to the actual solution when the optimization failed to converge. Conversely, in 3D the optimization appears to converge to the correct solution or fail completely, leading to a set of locations at final iteration which do not resemble the actual solution.~~ This is depicted in Figure 4 by the absence of the ~~gray and black~~ thin and thick lines below 60% and 30% ~~of the constraints~~linkage, respectively. The ~~reason for this difference may be due to the increased number of degrees of freedom in 3D requiring a greater number of iterations to converge. Nevertheless, the~~ accurate convergence of the best solution for cases with 30% linkage or higher is encouraging for the potential of coda wave optimization to constrain earthquake location.

Summary of synthetic experiments

In summary, the synthetic examples demonstrate the ability of coda wave data to constrain relative event location using optimization. The optimization error is rotational in nature and influenced by the noise on CWI estimates with greater $\bar{\sigma}_N$ leading to larger errors in the solutions. ~~When~~ In 3D, when 70% or more of the direct branches are used the optimizer is stable with no observable difference in the solution for 25 randomly chosen starting locations. As the direct linkage reduces to 50% the optimization becomes less stable and the best solution from 25 random starting locations is required to find the optimal solution. All optimisations fail to converge as the number of links decrease below 30%.

Relocating Earthquakes on the Calaveras Fault

In this section we relocate 68 earthquakes from the Calaveras Fault, California. The 68 earthquakes are selected from the 308 earthquake Calaveras example released with the open source Double Difference algorithm or hypoDD (*Waldhauser and Ellsworth*, 2000; *Waldhauser*, 2001) [See also Data and Resources]. These events are chosen for four reasons. Firstly, they are recorded by a large number of stations (Fig. 5) and therefore lend themselves to accurate ~~travel time~~ arrival-time location. This makes them ideal for assessing the performance of a new location technique. Secondly, they are distributed with separations from near zero to hundreds of meters making them ideal for application of CWI. Thirdly, Calaveras earthquakes have been well researched with several studies having relocated events in the region (*Waldhauser*, 2001; *Schaff et al.*, 2002; *Waldhauser and Schaff*, 2008). Finally, the hypoDD locations for these 68 earthquakes align in a streak increasing the likelihood that they have near identical source mechanisms, a necessary assumption for the application of equation 2. The relocations in this paper are sorted into four examples as summarised in Table 1. Waveforms, cross correlations and separation estimates for example Calaveras event pairs are illustrated by *Robinson et al.* (2011) an are not repeated her for the sake of brevity.

Example 5 - comparison of CWI, catalogue and hypoDD locations

Figure 6 illustrates three sets of locations for the Calaveras earthquakes. The first column shows the original catalogue locations for all 308 earthquakes. That is, each event is located individually using all available ~~travel time arrivals~~ arrival-time data and a regional velocity model. The 68 earthquakes of interest in this study are differentiated in black. Catalogue locations suggest that the 68 earthquakes of interest are spatially widely distributed on the scale of Figure 6.

To apply CWI we download available waveforms from the Northern California Earthquake Data Center (See Data and Resources). Unsuitable waveforms are removed using the conditions summarised in Table 2. Remaining waveforms are filtered between 1 and 5 Hz and aligned to P arrivals at 0 s. CWI estimates are obtained from 5 s wide non-overlapping time windows between ~~$2.5 \leq t \leq 20$~~ $2.5 < t \leq 20$ s and used to create probabilistic constraints on event separation. We utilize the local coordinate system introduced in ~~Theory~~ the Theory Section and find the optimum relative locations using Polak-Ribiere optimization.

In this, and the following Calaveras examples, we allow the earthquakes to move freely in all three directions during the inversion despite using the in-fault separation estimates given by equation 2. We allow the events to move freely so that we can test the performance of our algorithm without assuming a-priori that the earthquakes are constrained on the same fault plane. We approximate the true event separation using the in-fault separation of equation 2 so that we can focus on developing a working algorithm and demonstrate capability without dealing with the complexity of in-fault ($\delta_{\parallel fault}$) and out-of-fault ($\delta^2_{\perp fault}$) displacement. Considering the more complicated formulation of equation 4 is left for future work. Another potential focus for future work involves refining our algorithm to simultaneously resolve event location and representative fault plane by restricting the events to align in a single (unknown a-priori) plane. That is, for cases where the earthquakes are believed a-priori to be in the same plan.

CWI locations for the 68 events are illustrated in column two of Figure 6. Catalogue locations (gray) are shown for the remaining 240 earthquakes and are included to ease comparison. The third column of Figure 6 illustrates the locations given by hypoDD with Singular Value Decomposition (SVD), absolute arrival times and cross correlation computed ~~travel time~~ arrival-time differences.

Absolute locations cannot be found by CWI alone. This is because of the ~~non-uniquesness~~ non-uniqueness associated with translation, rotation and reflection. For the sake of comparison, we arbitrarily choose a 'master' event and translate our relative locations to align with the hypoDD location for ~~the same~~ that event. This arbitrary translation does not change the relative locations. We return to this issue of relative versus absolute location in Example 7 by introducing a combined ~~travel time and coda wave~~ arrival-time and coda-wave inversion.

The spatial distribution of the CWI locations is clearly tighter than the catalogue locations of column 1. That is, CWI provides an independent indication of clustering for the 68 events and to first order, similar locations to those from hypoDD (column 3). There is a small second order difference between the CWI and hypoDD based locations. In particular, the lineation is less clear in the CWI locations (column 2) than the hypoDD locations (column 3). Our experience ~~suggest that the coda~~ suggests that the CWI locations are less supportive of the presence of streaks although a complete understanding of these differences is left for future work. Our attention now is devoted towards understanding how both techniques perform with fewer stations (Example 6) and exploring how CWI and ~~travel times~~ arrival-times can be combined (Examples 7 and 8).

Example 6 - Dependance on the number of stations

Accurate location of the Calaveras events is possible using arrival phases because of the excellent recording situation in California with many stations and strong azimuthal coverage (see Fig. 5). In contrast, a small number of stations and poor azimuthal coverage are common limitations when trying to locate intraplate clusters. For example, there are only four network seismic stations in the South West Seismic Zone of Western Australia, a region similar in size to that hosting 805 stations in Figure 5.

We explore the impact of poorer recording situations in example 6 by re-locating the 68 Calaveras events using hypoDD and ~~coda waves~~ coda-waves with a reduced number of stations. We begin with 10 stations and repeat the process removing one at a time until a single station remains. The 10 stations ~~considered are shown in Figure 7 and the~~ and there order of removal ~~explained in Table ??~~are shown in Figure 7.

CWI locations are illustrated in Figure 8 for the inversions with seven, five, four, three, two and one station. We observe a high level of consistency between these 6 inversions and the locations shown in Figure 6 (column 2) when all stations are considered. That is, the ~~coda wave~~ coda-wave approach is self-consistent regardless of the number of stations available, reinforcing our ~~hypothesis that coda waves~~ claim that coda-waves can constrain location in what would normally be regarded as a poor station network.

Figure 9 illustrates the hypoDD inversion results for seven, five and four stations. The ~~travel time~~ arrival-time problem is ill-posed for fewer than four stations so it is not possible to apply hypoDD with SVD for three or fewer stations. The hypoDD locations are ~~not~~ less self-consistent as the number of stations is reduced. We observe a general increase in scatter and a higher number of stray events outside the cluster when less stations are used with hypoDD. Even with seven stations the linear geometry of Figure 6 (column 3) is less evident.

As the number of stations ~~are reduced both the CWI and~~ is reduced neither the CWI nor hypoDD techniques are ~~not~~ able to re-locate all events. To use the coda waves we need at least one pairwise separation constraint to be formed from the available stations. This means that for every event there must be at least one station that records it, and at least one other earthquake, sufficiently well to apply CWI. Fortunately, we can make an assessment of this prior to starting the inversion. The top panel of Figure 10 demonstrates that when five or more stations are used, CWI can constrain the location of all 68 earthquakes. When less than five stations are used the ~~coda waves~~ coda-waves constrain a decreasing number of events until at one station it is only possible to locate 55 of the 68 events. The hypoDD algorithm also fails to locate all events as the number of stations is reduced. In the case of hypoDD an event can be identified as unconstrainable in one of two stages. Firstly, the data are analyzed to ensure that there ~~exists travel time~~ exist arrival-time differences for each event and at least one other earthquake. This is analogous to the situation for the ~~coda wave~~ coda-wave technique. The hypoDD program also has a secondary identification phase in which events that can not be located sufficiently well are rejected during the inversion. This process is related to the iterative removal of outliers described by *Waldhauser and Ellsworth* (2000). The top panel of Figure 10 shows that the number of events re-located by hypoDD fluctuates between 63 and 28 earthquakes for ten to four stations and it demonstrates that the number of events located by hypoDD is less than or equal to the number located by CWI.

The remaining panels of Figure 10 illustrate a statistical comparison of the CWI and hypoDD ~~reduced station locations~~ locations with a reduced number of stations to those using hypoDD with all available data. For the CWI inversions the mean and maximum coordinate difference is consistent regardless of the number of stations considered. In contrast, the hypoDD mean and maximum coordinate error fluctuate above those for CWI confirming that the hypoDD inversion is less stable than CWI with fewer stations.

Combining ~~Travel Time~~ Arrival-Time and CWI Constraints

In Examples 5 and 6 we compare the location of the Calaveras earthquakes using ~~coda wave~~ coda-wave and arrival time based constraints independently. Since the arrival time (direct or difference) and ~~coda wave~~ coda-wave data come from different sections of the waveform they provide independent constraints on the locations. In this section we devise a location algorithm which incorporates both CWI and ~~travel time~~ arrival-time data.

We do not propose a new technique for earthquake location using ~~travel time~~ arrival-time differences. Rather, we exploit the information created by hypoDD with SVD to define a probability density (or posterior) function

$$
\begin{aligned}
P(\mathbf{e}_p|\Delta_{TT}) & \frac{1}{(2\pi)^{3/2}\sqrt{|\Sigma|}} \\
& \times \exp\left(-\frac{1}{2}\left([\mathbf{e}_p - \mu_{\mathbf{e}_p}]^T \Sigma^{-1} [\mathbf{e}_p - \mu_{\mathbf{e}_p}]\right)\right),
\end{aligned} \tag{22}
$$

where

$$
\mathbf{e}_p = (x_p, y_p, z_p)^T \tag{23}
$$

is the location of event p,

$$
\mu_{\mathbf{e}_p} = (\mu_{x_p}, \mu_{y_p}, \mu_{z_p})^T \tag{24}
$$

is the most likely location as determined using the ~~travel time~~ arrival-time data, and

$$
\Sigma = \begin{pmatrix} \sigma_{x_p}^2 & 0 & 0 \\ 0 & \sigma_{y_p}^2 & 0 \\ 0 & 0 & \sigma_{z_p}^2 \end{pmatrix} \tag{25}
$$

is the covariance matrix. ~~In this paper we~~ We define the mean location $\mu_{\mathbf{e}_p}$ and covariance matrix by the hypoDD optimum solution and its uncertainties. It is important to note that hypoDD must be used with SVD to obtain useful estimates of σ_{x_p}, σ_{y_p} and σ_{z_p} because the errors reported by conjugate gradient methods (LSQR) are grossly underestimated in hypoDD (*Waldhauser*, 2001).

We pose the location problem using the negative log likelihood

$$\begin{aligned} L(\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_1, \mathbf{e}_n) = - \textstyle\sum_{i=1}^{n} ln \left[P(\mathbf{e}_i | \Delta_{TT}) \right] \\ - \textstyle\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} ln \left[P(\delta_{CWIN} | \mathbf{e}_i, \mathbf{e}_j) \right], \end{aligned} \tag{26}$$

where $(\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n)$ is the joint location,

$$\sum_{i=1}^{n} ln \left[P(\mathbf{e}_i | \Delta_{TT}) \right] \tag{27}$$

incorporates the ~~travel time~~ arrival-time constraints and

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} ln \left[P(\delta_{CWIN} | \mathbf{e}_i, \mathbf{e}_j) \right] \tag{28}$$

the ~~coda waves~~coda-waves.

We must differentiate L to use the Polak-Ribiere conjugate gradient technique of *Press et al.* (1987). The derivative of $L(\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n)$ with respect to x_p is given by

$$\begin{aligned} \frac{\partial L}{\partial x_p} = - \frac{\partial ln[P(\mathbf{e}_p | t_{DD})]}{\partial x_p} - \textstyle\sum_{i=p+1}^{N} \frac{\partial \ln[P(\delta_{CWIN} | \mathbf{e}_p, \mathbf{e}_i)]}{\partial x_p} \\ - \textstyle\sum_{j=1}^{p-1} \frac{\partial \ln[P(\delta_{CWIN} | \mathbf{e}_j, \mathbf{e}_p)]}{\partial x_p} \end{aligned} \tag{29}$$

where

$$\sum_{i=p+1}^{N} \frac{\partial \ln \left[P(\delta_{CWIN} | \mathbf{e}_p, \mathbf{e}_i) \right]}{\partial x_p} \tag{30}$$

and

$$\sum_{j=1}^{p-1} \frac{\partial \ln\left[P(\delta_{CWIN}|\mathbf{e}_j, \mathbf{e}_p)\right]}{\partial x_p} \tag{31}$$

are defined in Appendix and

$$\begin{aligned} \tfrac{\partial ln[P(\mathbf{e}_p|t_{DD})]}{\partial x_p} &= -\tfrac{1}{2}[1,0,0]^T \Sigma^{-1}[\mathbf{e}_p - \mu_{\mathbf{e}_p}] \\ &\quad -\tfrac{1}{2}[\mathbf{e}_p - \mu_{\mathbf{e}_p}]^T \Sigma^{-1}[1,0,0]. \end{aligned} \tag{32}$$

Similarly, for the derivatives with respect to y_p and z_p we have

$$\begin{aligned} \tfrac{\partial ln[P(\mathbf{e}_p|t_{DD})]}{\partial y_p} &= -\tfrac{1}{2}[0,1,0]^T \Sigma^{-1}[\mathbf{e}_p - \mu_{\mathbf{e}_p}] \\ &\quad -\tfrac{1}{2}[\mathbf{e}_p - \mu_{\mathbf{e}_p}]^T \Sigma^{-1}[0,1,0] \end{aligned} \tag{33}$$

and

$$\begin{aligned} \tfrac{\partial ln[P(\mathbf{e}_p|t_{DD})]}{\partial z_p} &= -\tfrac{1}{2}[0,0,1]^T \Sigma^{-1}[\mathbf{e}_p - \mu_{\mathbf{e}_p}] \\ &\quad -\tfrac{1}{2}[\mathbf{e}_p - \mu_{\mathbf{e}_p}]^T \Sigma^{-1}[0,0,1]. \end{aligned} \tag{34}$$

Combining the ~~travel time and coda wave~~ arrival-time and coda-wave data offers two advantages. Firstly, it combines independent constraints on the event locations offering further confidence in the resulting solution. Secondly, the ~~travel time~~ arrival-time constraints in the form of equation (27) resolve the inherent non-uniqueness of the CWI inversion that is associated with translation, rotation and reflection around a global coordinate system. This means that it is no longer necessary to use a local coordinate system and we can solve directly for location with respect to a global reference. Collectively, these advantages improve the behavior of the Polak-Ribiere optimization leading to faster and more stable convergence. Consequently, we no longer ~~have~~ need to consider multiple randomly chosen starting locations.

Example 7 - Combining ~~travel time~~ arrival-time and CWI constraints

Figure 11 illustrates the earthquake locations obtained when we combine the ~~travel time and coda wave~~ arrival-time and coda-wave data using all data (left) and five stations (right). The linear features observed in the original hypoDD inversions (see Fig. 6) are evident in both cases. However, the ~~coda waves~~ coda-waves introduce a scatter around these streaks. That is, the locations in figure 11 result from a trade-off between hypoDD's desire to place the events on linear features and the ~~coda waves~~ coda-waves voracity to push them away from streaks. When all stations are used the hypoDD constraints are strong and little off-streak scatter is introduced. As we reduce hypoDD's leverage by decreasing the number of stations to five, we observe an increase in off-streak scatter resulting from the enhanced influence of the coda.

Example 8 - Combining CWI and ~~travel times~~ arrival-times when the ~~travel times~~ arrival-times constrain a limited number of events

In intraplate regions such as Australia it is common to deploy temporary seismometers to monitor aftershocks for significant events (*Bowman et al.*, 1990; *Leonard*, 2002). Traditionally, these deployments facilitate a higher accuracy of location for events occuring during the deployment period. Using our combined inversion it is possible to re-locate all events by employing the detailed ~~travel time~~ arrival-time data when the temporary network is in-situ and using ~~coda waves~~ coda-waves from network stations when the deployment is absent. The hypothesis, to be tested in this section, is that conducting such a combined inversion will improve the location accuracy of events outside the deployment period.

An estimate of the cummulative number of aftershocks $N(t)$ after t days ~~is given~~ can be modeled by the modified Omori formula

$$N(t) = K\frac{c^{1-p} + (t+c)^{1-p}}{p-1} \tag{35}$$

(*Utsu et al.*, 1995). The empirically derived constants, K, C and p vary between tectonic settings. For example, using recorded aftershocks with $M \geq 3.2$ of the Hokkaido-Nansei-Oki, Japan $M_s = 7.8$ earthquake of 12 July 1993, *Utsu et al.* (1995) obtained maximum likelihood estimates for K, p and c of 906.5, 1.256 and 1.433, respectively. With these empirically derived values an array deployed within 4 days and left for 150 days will record roughly one half of the aftershocks occurring within the first 1000 days. That is,

$$\frac{N(150+4) - N(4)}{N(1000)} = \frac{2257 - 934}{2626} \approx 0.5. \tag{36}$$

This ~~idea~~ is illustrated in Figure 12 which shows the best fitting Omori Formula separated into segments before~~(gray), during (black) and after (gray)~~ , during and after the pseudo temporary deployment.

With this idea of a temporary deployment in mind we have another attempt at relocating the Calaveras earthquakes. In Example 8 we consider the ~~travel time~~ arrival-time constraints on half (34) of the earthquakes and incorporate ~~coda wave~~ coda-wave data from a single station for all 68 earthquakes. The combined inversion is shown in column 1 of Figure 13. The inversion result is similar to the combined inversion when all ~~travel time data is~~ arrival-time data are incorporated (see Fig. 11). The slight increase in scatter observed here can be explained by the events with no ~~travel time~~ arrival-time constraints and the tendency of the coda to push events away from streaks.

Remarkably, the combined ~~coda wave and travel time~~ coda-wave and arrival-time inversion locates all 68 earthquakes to an accuracy similar to the inversions with all data. In contrast when ~~travel time data is~~ arrival-time data are used alone it is only possible to locate the 34 events recorded by the pseudo temporary deployment. This ability of ~~coda waves~~ coda-waves to constrain the location of events recorded by a single station creates new opportunities for understanding earthquakes in regions with limited station coverage.

Discussion and Conclusions

~~Coda wave~~ Coda-wave interferometry is an emerging technique for constraining earthquake location. The technique relies on the interference between ~~coda waves~~ coda-waves of closely located events and is hence useful for studying earthquake clusters and/or aftershock sequences. ~~Coda wave~~ Coda-wave constraints are independent of ~~travel times~~ arrival-times and can be used in isolation or combination with early onset body waves. The strength of coda is that it is possible to constrain earthquake location from a single station, an outcome demonstrated most clearly by Figures 8 and 13.

~~Coda wave~~ Coda-wave interferometry offers a new technique for understanding earthquakes in intraplate areas with sparse networks and poor azimuthal coverage. In particular, the ability to combine ~~coda wave constraints with travel times~~ coda-wave constraints with arrival-times makes it possible to link well constrained events from a temporary deployment with those recorded outside the deployment period. All that is required to achieve this is at least one network station which has recorded sufficient events from both periods. CWI facilitates the location of poorly recorded events to an accuracy approaching those recorded during the temporary deployment and therefore opens new avenues for imaging intraplate fault structures and improving our understanding of intraplate seismicity and earthquake hazard. Importantly, this analysis can be conducted for any historical aftershock sequence or earthquake swarm recorded by a temporay deployment. Our technique is, in that sense, related to the retorspective sesimological observation technique of *Curtis et al.* (2012) that utilizes interferometry to obtain seismic signals on newly installed sensors regardless of whether the event occurs before, during or after the physical installation of the sensor.

Another potential application of CWI is in the area of hydraulic fracturing such as hot rock geothermal projects, petroleum reservoir engineering, tight gas extraction, CO_2 geosequestration and/or underground brine injection. Monitoring pumping-induced micro earthquakes is a key step in understanding the migration of fluids in such reservoirs. There is a trade-off in the ability of surface deployed networks to locate events which are small and/or deep. Downhole seismic monitoring is likely to play increasingly important roles in deep reservoir projects. CWI creates new possibilities to monitor pumping induced micro earthquakes from fewer boreholes and hence dramatically reduce the costs of reservoir monitoring at large depths. It may also be possible to utilize coda for understanding hazard in tunneled mining operations where the location of deep tunnels prohibits azimuthal coverage of induced events.

Data and Resources

We thank the Northern California Earthquake Data Center (NCEDC) for providing the Calaveras data and the Northern California Seismic Network (NCSN), U.S. Geological Survey, Menlo Park and Berkeley Seismological Laboratory, University of California, Berkeley for contributing it to the NCEDC. The waveforms can be downloaded from
http://www.ncedc.org/ncedc/access.html (last accessed August 2012). We also acknowledge Felix Waldhauser and William Ellsworth, the authors of the openly available Double Difference location algorithm, hypoDD which can be downloaded from
http://www.ldeo.columbia.edu/ felixw/hypoDD.html (last accessed August 2012). The International Sesimological Centre can be found at
http://www.isc.ac.uk/ (last accessed December 2012). The National Earthquake Information Center catalogue can be accessed from
http://earthquake.usgs.gov/earthquakes/eqarchives/epic/ (last accessed December 2012).

Acknowledgments

Geoscience Australia, the Research School of Earth Sciences at The Australian National University, and the Center for Wave Phenomena at the Colorado School of Mines, are acknowledged for supporting this research. The paper is published with permission of the CEO Geoscience Australia. Work was conducted as part of an Australian Research Council Discovery Project (DP0665111). This paper has benefited significantly from: reviews by Trevor Allen and Clive Collins at Geoscience Australia~~; as well as~~ , the BSSA editorial team, Andrew Curtis from the School of GeoSciences at the University of Edinburgh and an anonymous reviewer. The authors have also benefited from comments and advice received from Felix Waldhauser on this and earlier research on the application of CWI to earthquake location.

References

Ake, J., D. O'Connell, and L. Block (2005), Deep-injection and closely monitored induced seismicity at Paradox Valley, Colorado, *Bull. Seism. Soc. Am.*, *95*(2), 664–683.

Aki, K. (1969), Analysis of the seismic coda of local earthquakes as scattered waves, *J. Geophys. Res.*, *74*(2), 615–631.

Aster, R. C., B. Borchers, and C. H. Thurber (2005), *Parameter estimation and inverse problems*, *International Geophysics Series*, vol. 90, Elsevier Academic Press, USA.

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

Bowman, J. R., G. Gibson, and T. Jones (1990), Aftershocks of the 1988 January 22 Tennant Creek, Australia intraplate earthquakes: evidence for a complex thrust-fault geometry, *Geophys. J. Int.*, *100*, 87–97.

Campbell, K. W. (2003), Strong motion attenuation, in *International Handbook of Earthquake and Engineering Seismology*, vol. B, edited by W. H. K. Lee, H. Kanamori, P. C. Jennings, and C. Kisslinger, chap. 60, pp. 1003–1012, Academic Press, London.

Curtis, A., and R. Snieder (2002), Probing the Earth's interior with seismic tomography, in *International Handbook of Earthquake Engineering Seismology*, vol. A, edited by W. H. Lee, H. Kanamori, P. C. Jennings, and C. Kisslinger, chap. 52, pp. 861–874, Academic Press, London.

Curtis, A., Behr, Y., Entwistle, E., Galetti, E., Townend J. Bannister, S. (2012), The benefit of hindsight in observational science: Retrospective seismological observations, *Earth Planet. Sci. Lett.*, *345–348*, 212–220.

Deichmann, N., and M. Garcia-Fernandez (1992), Rupture geometry from high-precision relative hypocentre locations of microearthquake clusters, *Geophys. J. Int.*, *110*, 501–517.

Douglas, A. (1967), Joint epicentre determination, *Nature*, *215*, 47–48.

Frankel, A. D., C. S. Mueller, T. P. Barnhard, E. V. Leyendecker, R. L. Wesson, S. C. Harmsen, F. W. Klein, D. M. Perkins, N. C. Dickman, S. L.

Hanson, and M. G. Hopper (2000), USGS National seismic hazard maps, *Earthquake Spectra*, *16*(1), 1–19.

Frèmont, M.-J., and S. D. Malone (1987), High precision relative locations of earthquakes at Mount St. Helens, *J. Geophys. Res.*, *92*(B10), 10,223–10,236.

Got, J.-L., J. Frèchet, and F. W. Klein (1994), Deep fault plane geometry inferred from multiplet relative relocation beneath the south flank of Kilauea, *J. Geophys. Res.*, *99*(B8), 15,375–15,386.

Gutenberg, B. (1945), Amplitudes of surface waves and magnitudes of shallow earthquakes, *Bull. Seism. Soc. Am.*, *35*, 3–12.

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

Kennett, B. L. N., E. R. Engdahl, and R. Buland (1995), Constraints on seismic velocities in the Earth from traveltimes, *Geophys. J. Int.*, *122*, 108–124.

Kennett, B. L. N., S. Fishwick, and M. Heintz (2004), Lithospheric structure in the Australian region - a synthesis of surface wave and body wave studies, *Exploration Geophysics*, *35*, 242–250.

Lees, J. M. (1998), Multiplet analysis at Coso Geothermal, *Bull. Seism. Soc. Am.*, *88*(5), 1127–1143.

Leonard, M. (2002), The Burakin WA earthquake sequence Sept 2000 – June 2002, in *Total Risk Management in the Privatised Era*, *Australian Earth-*

quake Engineering Society Conference, vol. 10th, edited by M. Griffith, D. Love, P. McBean, A. McDougall, and B. Butler, pp. 22(1)–22(5), AEES, University of Adelaide.

Nadeau, R. M., and T. V. McEvilly (1997), Seismological studies at Parkfield V: Characteristic microearthquake sequences as fault–zone drilling targets, *Bull. Seism. Soc. Am.*, *87*(6), 1463–1472.

Pavlis, G. L. (1992), Appraising relative earthquake location errors, *Bull. Seism. Soc. Am.*, *82*(2), 836–859.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling (1987), *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, USA.

Richter, C. F. (1935), An instrumental earthquake magnitude scale, *Bull. Seism. Soc. Am.*, *25*(1), 1–32.

Robinson, D., T. Dhu, and J. Schneider (2006), Practical probabilistic seismic risk analysis: A demonstration of capability, *Seism. Res. Let.*, *77*(4), 452–458.

Robinson, D. J., M. Sambridge, and R. Snieder (2007a), Constraints on coda wave interferometry estimates of source separation: The 2.5d acoustic case, *Exploration Geophysics*, *38*(3), 189–199.

Robinson, D. J., R. Snieder, and M. Sambridge (2007b), Using coda wave interferometry for estimating the variation in source mechanism between double couple events, *J. Geophys. Res.*, *112*, b12302, doi:10.1029/2007JB004925.

Robinson, D. J., M. Sambridge, and R. Snieder (2011), A probabilistic approach for estimating the separation between a pair of earthquakes directly from their coda waves, *J. Geophys. Res.*, *B04309*, 1–17.

Rubin, A. M. (2002), Aftershocks of microearthquakes as probes of the mechanics of rupture, *J. Geophys. Res.*, *107*(B7,2142), 10.1029/2001JB000,496.

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

Schaff, D. P., G. H. R. Bokelmann, and G. C. Beroza (2002), High-resolution image of Calaveras Fault seismicity, *J. Geophys. Res.*, *107*(B9), 2186, doi:10.1029/2001JB000,633.

Shearer, P., E. Hauksson, and G. Lin (2005), Southern California hypocenter relocation with waveform cross-correlation, Part 2: Results using source-specific station terms and cluster analysis, *Bull. Seism. Soc. Am.*, *95*(3), 904–915. doi:10.1785/0120040,168.

Shearer, P. M. (1999), *Introduction to Seismology*, Cambridge University Press, USA, 260pp.

Sipkin, S. A. (2002), USGS earthquake moment tensor catalog, in *International Handbook of Earthquake Engineering Seismology*, vol. A, edited by W. H. Lee, H. Kanamori, P. C. Jennings, and C. Kisslinger, chap. 50, pp. 823–825, Academic Press, London.

Slunga, R., Rögnvaldsson, S. Th. and Bödvarsson, R. (1995), Absolute and relative locations of similar events with application to microearthquakes in southern Iceland, *Geophys. J. Int.*, *123*(2), 409–419.

Snieder, R. (1999), Imaging and averaging in complex media, in *Diffuse waves in complex media*, *NATO Science Series C*, vol. 531, edited by J. P. Fouque, pp. 405–454, Kluwer Academic Publishers.

Snieder, R. (2006), The theory of coda wave interferometry, *Pure Appl. Geophys.*, *163*, 455–473.

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.

Spencer, C., and D. Gubbins (1980), Travel-time inversion for simultaneous earthquake location and velocity structure determination in laterally varying media, *Geophys. J. R. Astr. Soc.*, *63*, 95–116.

Stirling, M. W., G. H. McVerry, and K. R. Berrryman (2002), A new seismic hazard model for New Zealand, *Bull. Seism. Soc. Am.*, *92*(5), 1878–1903.

Toro, G. R., N. A. Abrahamson, and J. F. Schneider (1997), Model of strong ground motions from earthquakes in Central and Eastern North America: Best estimates and uncertainties, *Seism. Res. Let.*, *68*(1), 41–57.

Utsu, T., Y. Ogata, and R. S. Matsu'ura (1995), The Centenary of the Omori Formula for a decay law of aftershock activity, *J. Phys. Earth*, *43*, 1–33.

VanDecar, J. C., and R. Snieder (1994), Obtaining smooth solutions to large linear inverse problems, *Geophysics*, *59*, 818–829.

Waldhauser, F. (2001), hypoDD – a program to compute double-difference hypocenter locations (hypoDD version 1.0 - 03/2001), *Open file report 01-113*, United States Geological Survey, Menlo Park, California.

Waldhauser, F., and W. L. Ellsworth (2000), A double–difference earthquake location algorithm: method and application to the northern Hayward Fault, California, *Bull. Seism. Soc. Am.*, *90*(6), 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), 10.1029/2000JB000,084.

Waldhauser, F., and D. P. Schaff (2008), Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods, *J. Geophys. Res.*, *133*, B08311, doi10.1029/2007JB005479.

Waldhauser, F., W. L. Ellsworth, and A. Cole (1999), Slip-parallel lineations on the Northern Hayward Fault, California, *Geophys. Res. Lett.*, *26*(23), 3525–3528.

905 Authors' Affiliations, Adresses

David Robinson: Earth Monitoring and Hazard Group
Geoscience Australia
GPO Box *383*
Canberra ACT *2601* Australia
E-mail: david.robinson@ga.gov.au

Also at: Research School of Earth Sciences
Australian National University
Canberra ACT *0200*, Australia
E-mail: david.robinson@anu.edu.au

906 Malcolm Sambridge: Research School of Earth Sciences
Australian National University
Canberra ACT *0200*, Australia

Roel Snieder: Center for Wave Phenomena and Department of Geophysics
Colorado School of Mines
Golden CO *80401-1887*, USA

Juerg Hauser: CSIRO Earth Science and Resource Engineering
Dick Perry Av
Kensington WA *6151* Australia

Table 1: Location examples for the 68 Calaveras earthquakes.

Example 5	Comparison of CWI, catalogue and hypoDD locations (using all available data).
Example 6	Exploration of station dependance for CWI and hypoDD (using a subset of data).
Example 7	Combined use of CWI and ~~travel time~~ arrival-time data with all and a reduced number of stations.
Example 8	Combined use of CWI and ~~travel time~~ arrival-time data when ~~travel times~~ arrival-times constrain only 50% of the events.

~~Stations considered when exploring the impact of reduced station coverage. Number of Station NamesStations 10 CCO, JCB, JST, CMH, HSP, JAL, CSC, JST, CAD, JHL, JRR9 CCO, JCB, JST, CMH, HSP, JAL, CSC, JST, CAD, JHL8 CCO, JCB, JST, CMH, HSP, JAL, CSC, JST, CAD7 CCO, JCB, JST, CMH, HSP, JAL, CSC 6 CCO, JCB, JST, CMH, HSP, JAL 5 CCO, JCB, JST, CMH, HSP 4 CCO, JCB, JST, CMH 3 CCO, JCB, JST 2 CCO, JCB 1 CCO~~

Table 2: Conditions used to identify unsuitable waveforms before applying CWI (Originally published as Table 5 *Robinson et al.*, 2011)

	condition
1	waveform is clearly corrupted
2	waveform indicates recording of more then one event
3	signal to noise ratio is obviously low
4	there is insufficient coda recorded after the first arrivals
5	there is insufficient recording before the arrivals (needed for accurate noise energy estimate)

Figure 1: Example 1 - Synthetic relocation of 50 earthquakes in 2D using all constraints with noise $\bar{\sigma}_N = 0.02$. Actual and optimization event locations are identified by triangles and circles, respectively.

Figure 2: Example 2 - Synthetic relocation of 50 earthquakes in 2D using all constraints with noise $\bar{\sigma}_N = 2\epsilon(\delta_t)$. Actual and optimization event locations are identified by triangles and circles, respectively.

Figure 3: Example 3 - Statistical measures of error in the solutions for the 2D synthetic cases when all and best optimization results are considered. The statistics Δ_{max} and Δ_μ are the maximum and mean coordinate error, respectively. The bottom subplot shows the average minimum number of branches required to link the 2450 pairs.

Figure 4: Example 4 - Statistical measures of error in the optimization solutions for the 3D synthetic cases when all and best results are considered. The statistics Δ_{max} and Δ_μ are the maximum and mean coordinate error, respectively. The absence of the lines below 60% and 30% indicates a breakdown in the solutions when all or best optimization result(s) are considered, respectively.

Figure 5: Map showing location of the Calaveras cluster (star) and 805 seismic stations (triangles).

Figure 6: Example 5 - Comparison of relative earthquake locations using three different methods: catalogue location (column 1), CWI (column 2) and hypoDD (column 3). Note that in the case of the hypoDD and CWI inversions we consider only the 68 earthquakes in black, the gray catalogue locations for the remaining 240 (308-68) earthquakes are shown for the purpose of orientation only. In this and subsequent similar figures (Figures 8, 9, 11 and 13) x is defined as positive towards the east, y is positive towards the north and z is positive down.

Figure 7: Location of the 10 stations (triangles) used to relocate the Calaveras events in Examples 6 to 8. Stations are removed one at a time according to the order ~~in Table ?? and~~ defined by the ~~events relocated~~bracketed numbers. That is, JRR is the first station to be removed, JHL is the second and so on. Events are indicated with circles.

Figure 8: Example 6 - CWI relative locations with reduced stations. Axes as defined in Figure 6.

Figure 9: Example 6 - HypoDD (SVD) relative locations with reduced stations. Axes as defined in Figure 6.

Figure 10: Example 6 - Number of constrainable events nE in the CWI and hypoDD inversions as a function of the number of stations considered (top). Mean (middle) and maximum (bottom) of the difference computed between the reduced station inversion results (CWI and hypoDD) and the complete hypoDD locations for all 308 events.

Figure 11: Example 7 - Combined HypoDD (SVD) and CWI relative locations using data form all stations (left) and 5 stations (right). Axes as defined in Figure 6.

Figure 12: ~~Cumulative~~ Modeled cumulative number of aftershocks for the Hokkaido-Nansei-Oki, Japan $M_s = 7.8$ earthquake of 12 July 1993 using equation (35). The leftmost, middle and rightmost lines signify aftershocks occurring before, during and after the deployment of a pseudo temporary array installed 4 days after the main shock and left for 150 days. A temporary deployment of this kind will record roughly 50% of the aftershocks in the 1000 days following the mainshock.

Figure 13: Example 8 - Mimicking the deployment of a temporary network by ignoring data from all but station CCO for 50% (or 34) of the 68 events. Relative locations are shown for the combined CWI and ~~travel time~~ arrival-time inversion (left) and the inversion with ~~travel times~~ arrival-times only (right). Only by combining the data is it possible to locate all 68 events. Furthermore, combining the data leads to a solution more consistent with Figure 6. Axes as defined in Figure 6.

Appendix

The Likelihood

The likelihood $P(\widetilde{\delta}_{CWIN}|\widetilde{\delta}_t)$ used in equation (7) is given by

$$\begin{aligned} P(\widetilde{\delta}_{CWIN}|\widetilde{\delta}_t) = {} & A(\widetilde{\delta}_t)C(\bar{\mu}_N, \bar{\sigma}_N) \\ & \times \int_0^\infty B(\widetilde{\delta}_t, \widetilde{\delta}_{CWI})D(\widetilde{\delta}_{CWI}, \bar{\sigma}_N, \bar{\mu}_N)d\widetilde{\delta}_{CWI} \end{aligned} \tag{A1}$$

where $\widetilde{\delta}_{CWI}$ is an estimate of CWI separation in the absence of noise,

$$A(\widetilde{\delta}_t) = \frac{1}{(1 - \Phi_{\mu_1,\sigma_1}(0))\sigma_1\sqrt{2\pi}}, \tag{A2}$$

$$B(\widetilde{\delta}_t, \widetilde{\delta}_{CWI}) = e^{\frac{-(\widetilde{\delta}_{CWI} - \mu_1)^2}{2\sigma_1^2}}, \tag{A3}$$

$$C(\bar{\mu}_N, \bar{\sigma}_N) = \frac{1}{(1 - \Phi_{\bar{\mu}_N,\bar{\sigma}_N}(0))\sigma_N\sqrt{2\pi}}, \tag{A4}$$

$$D(\widetilde{\delta}_{CWI}, \bar{\sigma}_N, \bar{\mu}_N) = e^{\frac{-(\widetilde{\delta}_{CWI} - \bar{\mu}_N)^2}{2\bar{\sigma}_N^2}} \tag{A5}$$

and $\Phi_{\mu,\sigma}(x)$ is the cumulative Gaussian distribution function

$$\Phi_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-(s-\mu)^2}{2\sigma^2}} ds \tag{A6}$$

(*Robinson et al.*, 2011). The parameters μ_1 and σ_1 used in equation (A2) are defined by the expressions

$$\mu_1(\widetilde{\delta}_t) = a_1 \frac{a_2\widetilde{\delta}_t^{a_4} + a_3\widetilde{\delta}_t^{a_5}}{a_2\widetilde{\delta}_t^{a_4} + a_3\widetilde{\delta}_t^{a_5} + 1} \tag{A7}$$

and

$$\sigma_1(\widetilde{\delta}_t) = c + a_1 \frac{a_2\widetilde{\delta}_t^{a_4} + a_3\widetilde{\delta}_t^{a_5}}{a_2\widetilde{\delta}_t^{a_4} + a_3\widetilde{\delta}_t^{a_5} + 1} \tag{A8}$$

Table A1: Coefficients for equations (A7) and (A8).

$\mu_1(\widetilde{\delta}_t)$	$\sigma_1(\widetilde{\delta}_t)$
$a1 = 0.4661$	$a1 = 0.1441$
$a2 = 48.9697$	$a2 = 101.0376$
$a3 = 2.4693$	$a3 = 120.3864$
$a4 = 4.2467$	$a4 = 2.8430$
$a5 = 1.1619$	$a5 = 6.0823$
	$c = 0.017$

with coefficients a_1 to a_5 and c defined in Table A1. The parameters $\bar{\mu}_N$ and $\bar{\sigma}_N$ used in equation (A4) are obtained by finding the values which minimize the difference in a least squares sense between the noisy CWI estimates $\widetilde{\delta}_{CWIN}$ computed from the waveforms and the positively bounded Gaussian density function

$$P(\widetilde{\delta}_{CWIN}|\widetilde{\delta}_t, \widetilde{\delta}_{CWI}) = \frac{1}{\left(1-\Phi_{\bar{\mu}_N,\bar{\sigma}_N}(0)\right)\bar{\sigma}_N\sqrt{2\pi}} e^{\frac{-(\widetilde{\delta}_{CWIN}-\bar{\mu}_N)^2}{2\bar{\sigma}_N^2}} \tag{A9}$$

with $\widetilde{\delta}_{CWIN} \geq 0$.

Derivatives

The derivatives of $L(\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_N)$

$$\frac{\partial L}{\partial \hat{x}_1}, \frac{\partial L}{\partial \hat{y}_1}, \frac{\partial L}{\partial \hat{z}_1}, \frac{\partial L}{\partial \hat{x}_2}, \frac{\partial L}{\partial \hat{y}_2}, \frac{\partial L}{\partial \hat{z}_2}, ..., \frac{\partial L}{\partial \hat{x}_N}, \frac{\partial L}{\partial \hat{y}_N}, \frac{\partial L}{\partial \hat{z}_N} \tag{A10}$$

are required by the Polak-Ribiere algorithm. These are used to guide the optimization procedure towards the values of $(\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_N)$ which minimize

L.

The equations for the derivatives are convoluted so we build them gradually. We start with an expression for δ_t, the wavelength normalized separation between two events $\mathbf{e}_p = (\hat{x}_p, \hat{y}_p, \hat{z}_p)$ and $\mathbf{e}_q = (\hat{x}_q, \hat{y}_q, \hat{z}_q)$

$$\delta_t = \frac{f_{dom}}{v_s}\sqrt{(\hat{x}_p - \hat{x}_q)^2 + (\hat{y}_p - \hat{y}_q)^2 + (\hat{z}_p - \hat{z}_q)^2}, \tag{A11}$$

where f_{dom} is the dominant frequency of the waveforms and v_s is the velocity between the events. Expression A11 has derivatives

$$\begin{aligned} \frac{\partial\widetilde{\delta}_t}{\partial\hat{x}_p} &= \frac{f_{dom}^2(\hat{x}_p - \hat{x}_q)}{v_s^2\widetilde{\delta}_t}, \frac{\partial\widetilde{\delta}_t}{\partial\hat{y}_p} = \frac{f_{dom}^2(\hat{y}_p - \hat{y}_q)}{v_s^2\widetilde{\delta}_t}, \\ \frac{\partial\widetilde{\delta}_t}{\partial\hat{z}_p} &= \frac{f_{dom}^2(\hat{z}_p - \hat{z}_q)}{v_s^2\widetilde{\delta}_t}, \frac{\partial\widetilde{\delta}_t}{\partial\hat{x}_q} = \frac{f_{dom}^2(\hat{x}_q - \hat{x}_p)}{v_s^2\widetilde{\delta}_t}, \\ \frac{\partial\widetilde{\delta}_t}{\partial\hat{y}_q} &= \frac{f_{dom}^2(\hat{y}_q - \hat{y}_p)}{v_s^2\widetilde{\delta}_t}, \frac{\partial\widetilde{\delta}_t}{\partial\hat{z}_q} = \frac{f_{dom}^2(\hat{z}_q - \hat{z}_p)}{v_s^2\widetilde{\delta}_t}. \end{aligned} \tag{A12}$$

For brevity we focus the following derivation in terms of $\hat{x}_p$. The remaining terms for $\mathbf{e}_p$ (i.e. $\hat{y}_p$ and $\hat{z}_p$) can be computed by following the same procedure. The derivatives for $\mathbf{e}_q$ can be attained by exploiting the symmetry

$$\frac{\partial\widetilde{\delta}_t}{\partial\hat{x}_q} = -\frac{\partial\widetilde{\delta}_t}{\partial\hat{x}_p}. \tag{A13}$$

The chain rule gives

$$\frac{\partial\mu_1}{\partial\hat{x}_p} = \frac{\partial\mu_1}{\partial\widetilde{\delta}_t}\frac{\partial\widetilde{\delta}_t}{\partial\hat{x}_p} \tag{A14}$$

where differentiating equation (A7) gives

$$\frac{\partial\mu_1}{\partial\widetilde{\delta}_t} = a_1\frac{a_2a_4\widetilde{\delta}_t^{a_4-1} + a_3a_5\widetilde{\delta}_t^{a_5-1}}{\left(a_2\widetilde{\delta}_t^{a_4} + a_3\widetilde{\delta}_t^{a_5} + 1\right)^2}. \tag{A15}$$

Similarly, we have

$$\frac{\partial\sigma_1}{\partial\hat{x}_p} = \frac{\partial\sigma_1}{\partial\widetilde{\delta}_t}\frac{\partial\widetilde{\delta}_t}{\partial\hat{x}_p} \tag{A16}$$

where $\frac{\partial \sigma_1}{\partial \tilde{\delta}_t}$ has the identical form as A15 with different constants $a_1, a_2, ..., a_5$ (see table A1).

The cumulative Gaussian distribution function A6 is

$$\Phi_{\mu_1,\sigma_1}(0) = \frac{1}{\sigma_1\sqrt{2\pi}} \int_{-\infty}^{0} e^{\frac{-(s-\mu_1)^2}{2\sigma_1^2}} ds \tag{A17}$$

which has derivative

$$\frac{\partial \Phi_{\mu_1,\sigma_1}(0)}{\partial \hat{x}_p} = \frac{\sigma_1 \int_{-\infty}^{0} \frac{\partial g}{\partial \hat{x}_p} e^g ds - \frac{\partial \sigma_1}{\partial \hat{x}_p} \int_{-\infty}^{0} e^g ds}{\sigma_1^2 \sqrt{2\pi}}, \tag{A18}$$

where

$$g = \frac{-(s-\mu_1)^2}{2\sigma_1^2} \tag{A19}$$

and

$$\frac{\partial g}{\partial \hat{x}_p} = \frac{4\sigma_1^2 (s-\mu_1) \frac{\partial \mu_1}{\partial \hat{x}_p} + 4\sigma_1 \frac{\partial \sigma_1}{\partial \hat{x}_p} (s-\mu_1)^2}{4\sigma_1^4}. \tag{A20}$$

Now, we have all the pieces to compute the derivatives of $A = A(\delta_t)$ and $B = B(\delta_t, \delta_{CWI})$ as follows

$$\frac{\partial A}{\partial \hat{x}_p} = -\frac{-\frac{\partial \Phi_{\mu_1,\sigma_1}(0)}{\partial \hat{x}_p} \sigma_1 + (1 - \Phi_{\mu_1,\sigma_1}(0)) \frac{\partial \sigma_1}{\partial \hat{x}_p}}{(1 - \Phi_{\mu_1,\sigma_1}(0))^2 \sigma_1^2 \sqrt{2\pi}} \tag{A21}$$

and

$$\frac{\partial B}{\partial \hat{x}_p} = e^h \frac{\partial h}{\partial \hat{x}_p}, \tag{A22}$$

where

$$h = \frac{-(\delta_{CWI} - \mu_1)^2}{2\sigma_1^2} \tag{A23}$$

and

$$\frac{\partial h}{\partial \hat{x}_p} = \frac{4\sigma_1^2 (\delta_{CWI} - \mu_1) \frac{\partial \mu_1}{\partial \hat{x}_p} + 4(\delta_{CWI} - \mu_1)^2 \sigma_1 \frac{\partial \sigma_1}{\partial \hat{x}_p}}{4\sigma_1^4}. \tag{A24}$$

Finally, we can differentiate the likelihood for an individual event pair

$$\begin{aligned}\frac{\partial P(\delta_{CWIN}|\widetilde{\delta}_t)}{\partial \hat{x}_p} = & \frac{\partial A(\widetilde{\delta}_t)}{\partial \hat{x}_p} C(\bar{\mu}_N, \bar{\sigma}_N) \\ & \times \int_0^\infty B(\widetilde{\delta}_t, \widetilde{\delta}_{CWI}) D(\widetilde{\delta}_{CWI}, \bar{\sigma}_N, \bar{\mu}_N) d\widetilde{\delta}_{CWI} \\ & + A(\widetilde{\delta}_t) C(\bar{\mu}_N, \bar{\sigma}_N) \\ & \times \int_0^\infty \frac{\partial B(\widetilde{\delta}_t, \widetilde{\delta}_{CWI})}{\partial \hat{x}_p} D(\widetilde{\delta}_{CWI}, \bar{\sigma}_N, \bar{\mu}_N) d\widetilde{\delta}_{CWI}\end{aligned} \tag{A25}$$

and for the logarithm we have

$$\frac{\partial \ln\left[P(\delta_{CWIN}|\delta_t)\right]}{\partial \hat{x}_p} = \frac{1}{P(\delta_{CWIN}|\delta_t)} \frac{\partial P(\delta_{CWIN}|\delta_t)}{\partial \hat{x}_p}. \tag{A26}$$

Thus, it follows that the derivative of L with respect to $\hat{x}_p$ is given by

$$\begin{aligned}\frac{\partial L(E_1, E_2, \ldots, E_n)}{\partial \hat{x}_p} = & -\sum_{i=p+1}^{N} \frac{\partial \ln[P(\delta_{CWIN}|E_p, E_i)]}{\partial \hat{x}_p} \\ & + \sum_{j=1}^{p-1} \frac{\partial \ln[P(\delta_{CWIN}|E_j, E_p)]}{\partial \hat{x}_p}\end{aligned} \tag{A27}$$

for a uniform prior. The change of sign in the middle (i.e. to addition) accounts for the change in order of the events under the conditional. Its inclusion here assumes the correct use of $\partial\widetilde{\delta}_t/\partial\hat{x}_p$ or $\partial\widetilde{\delta}_t/\partial\hat{x}_q$ when evaluating the left and right hand terms of the summation. The derivatives shown in this section appear complicated but are in practice trivial to compute numerically. Confidence in their accuracy is enhanced by demonstrating that the optimization procedure converges to the correct solution for a number of synthetic problems in 2 and 3 dimensions.