

1 **Relocating a Cluster of Earthquakes**
2 **Using a Single Seismic Station**

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

4 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 arise from scattering to form the later arriving components of a seismogram. Coda-wave interferometry is an emerging tool for constraining earthquake source properties from the interference pattern of coda-waves between nearby events. A new earthquake location algorithm is derived which relies on coda-wave based probabilistic estimates of earthquake separation. The algorithm can be used with coda-waves alone or in tandem with arrival-time data. Synthetic examples (2D and 3D) and real earthquakes on the Calaveras Fault, California are used to demonstrate the potential of coda-waves for locating poorly recorded earthquakes. It is demonstrated that 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*

27 *et al.*, 2006). The accuracy required in earthquake location depends on the
28 application. For example, imaging the structure of a fracture system from
29 microseismicity requires greater location accuracy than determining whether
30 a $M_w = 7.5$ earthquake occurs offshore for tsunami warning. This paper
31 focuses on reducing location uncertainty for a cluster of events when they
32 are recorded by a small number of stations.

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

- 49 1. there are 10 or more stations, all within 250 km,
- 50 2. an azimuthal gap of less than 110° ,

51 3. a secondary azimuthal gap of less than 160° , and

52 4. at least one station within 30 km.

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

55 Relative earthquake location involves locating a group of earthquakes
56 with respect to one another and was first introduced by *Douglas* (1967) who
57 developed the technique commonly known as joint hypocenter determination
58 (*Douglas* (1967) originally used the term joint epicentre determination. How-
59 ever, he was solving for hypocentre). In principle, relative locations can be
60 computed by differencing absolute locations. However, *Pavlis* (1992) shows
61 that inadequate knowledge of velocity structure leads to systematic biases
62 when relative positions are computed in this way. To reduce errors from
63 unknown velocity structure, relative location techniques typically compute
64 locations directly from arrival-time differences computed by time-lag cross
65 correlation of early-onset body waves (*Ito*, 1985; *Got et al.*, 1994; *Slunga*
66 *et al.*, 1995; *Nadeau and McEvilly*, 1997; *Waldhauser et al.*, 1999). Doing so
67 removes errors associated with velocity variations outside the local region,
68 because such variations influence all waveforms in a similar manner (*Shearer*,
69 1999).

70 Reported location uncertainties from relative techniques are around 15 to
71 75 m in local settings with good station coverage (*Ito*, 1985; *Got et al.*, 1994;
72 *Waldhauser et al.*, 1999; *Waldhauser and Schaff*, 2008). Here, ‘good cover-
73 age’ implies multiple stations distributed across a broad range of azimuthal
74 directions. Relative location techniques have been used to image active fault
75 planes (*Deichmann and Garcia-Fernandez*, 1992; *Got et al.*, 1994; *Wald-*

76 *hauser et al.*, 1999; *Waldhauser and Ellsworth*, 2002; *Shearer et al.*, 2005);
77 study rupture mechanics (*Rubin et al.*, 1999; *Rubin*, 2002); interpret magma
78 movement in volcanoes (*Frèmont and Malone*, 1987); and monitor pumping-
79 induced seismicity (*Lees*, 1998; *Ake et al.*, 2005).

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

90 Coda refers to later arriving waves in the seismogram that arise from
91 scattering (*Aki*, 1969; *Snieder*, 1999, 2006). Coda waves are ignored in most
92 seismological applications due to the complexity involved in constraining
93 complex heterogeneous velocity models in real settings. In this paper we
94 develop an approach for locating earthquakes using coda-waves. *Snieder and*
95 *Vrijlandt* (2005) demonstrate that the coda of two earthquakes can be used
96 to estimate the separation between them. Their technique, known as coda
97 wave interferometry (CWI), is based on the interference pattern between the
98 coda-waves. Unlike arrival-time based location techniques, CWI does not
99 require multiple stations or good azimuthal coverage. In fact, it is possible
100 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 alone or in combination with 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 arrival-time constraints.

Theory

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

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

where σ_τ is the standard deviation of the arrival-time perturbation between the 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,

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

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

$$122 \quad g(\alpha, \beta) = 2\alpha^2 \quad (3)$$

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

$$125 \quad \sigma_\tau^2 = \frac{\left(\frac{6}{\alpha^8} + \frac{7}{\beta^8}\right) \delta_{\parallel fault}^2 + 2\left(\frac{1}{\alpha^8} + \frac{2}{\beta^8}\right) \delta_{\perp fault}^2}{7\left(\frac{2}{\alpha^8} + \frac{3}{\beta^8}\right)}, \quad (4)$$

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

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

$$137 \quad R_{max}^{(t, t_w)} = 1 - \frac{1}{2} \overline{\omega^2} \sigma_\tau^2, \quad (5)$$

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

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

139 and \dot{u}_i represents the time derivative of u_i . In this paper we use the autocor-
 140 relation approach of *Robinson et al. (2011)* to relate the parameters directly

141 and we apply a restricted time lag search when evaluating R_{max} . These ex-
 142 tensions to the original technique of *Snieder and Vrijlandt* (2005) increase
 143 the range of applicability of CWI by 50% (i.e. from 300 to 450 m separation
 144 for 1 to 5 Hz filtered coda-waves).

145 *Robinson et al.* (2011) show that CWI leads to probabilistic constraints
 146 on source separation and introduce a Bayesian approach for describing the
 147 probability of true separation given the CWI data. Their approach is sum-
 148 marised by

$$149 \quad P(\tilde{\delta}_t | \tilde{\delta}_{CWI}) \propto P(\tilde{\delta}_{CWI} | \tilde{\delta}_t) \times P(\tilde{\delta}_t) \quad (7)$$

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

$$161 \quad \tilde{\delta} = \frac{\delta}{\lambda_d}, \quad (8)$$

162 which measures separation ($\delta = \delta_{CWI}$ or δ_t) with respect to dominant wave-
 163 length λ_d . In this paper we consider a uniform prior over appropriate bounds
 164 to ensure that the posterior function is dominated by the recorded data. The
 165 procedure for computing the likelihood $P(\tilde{\delta}_{CWI} | \tilde{\delta}_t)$ is derived by *Robinson*

166 *et al.* (2011) and summarised in Appendix (The Likelihood) . With these
 167 two pieces in place we can compute the posterior $P(\tilde{\delta}_t|\tilde{\delta}_{CWIN})$ (or PDF) for
 168 the separation between any pair of events directly from their coda waves.

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

$$176 \quad P(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3|\mathbf{d}) \propto P(\mathbf{d}|\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) \times P(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3), \quad (9)$$

177 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, like-
 178 lihood and prior functions, respectively. In equation (9) \mathbf{d} represents obser-
 179 vations that constrain the locations. They can be any combination of arrival-
 180 times, geodetic information or CWI separations. For example, if coda-waves
 181 are used we have $P(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3|\tilde{\delta}_{CWIN})$ and $P(\tilde{\delta}_{CWIN}|\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, where $\tilde{\delta}_{CWIN}$
 182 are the wavelength normalised separation estimates. Alternatively, if we
 183 use CWI and arrival-time data we may write $P(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3|\tilde{\delta}_{CWIN}, \mathbf{\Delta}_{TT})$ and
 184 $P(\tilde{\delta}_{CWIN}, \mathbf{\Delta}_{TT}|\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ where $\mathbf{\Delta}_{TT}$ represents the arrival-time differences.
 185 In the following derivation and in the Synthetic Experiments and Relocat-
 186 ing Earthquakes on the Calaveras Fault section we focus on the constraints
 187 imposed by coda-waves, whereas in Combining Arrival-Time and CWI Con-
 188 straints we demonstrate how CWI and arrival-time data can be combined.

189 For three earthquakes we have likelihoods; $P(\tilde{\delta}_{CWIN,12}|\mathbf{e}_1, \mathbf{e}_2)$, $P(\tilde{\delta}_{CWIN,13}|\mathbf{e}_1, \mathbf{e}_3)$
 190 and $P(\tilde{\delta}_{CWIN,23}|\mathbf{e}_2, \mathbf{e}_3)$. In writing these likelihoods we have replaced the con-

ditional term on separation $\tilde{\delta}_t$ 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. In other words, knowing $|e_1 - e_2|$ and $|e_2 - e_3|$ does not mean that $|e_1 - e_3|$ 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

$$P(\tilde{\delta}_{CWIN}|\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) \approx P(\tilde{\delta}_{CWIN,12}|\mathbf{e}_1, \mathbf{e}_2) \times P(\tilde{\delta}_{CWIN,13}|\mathbf{e}_1, \mathbf{e}_3) \times P(\tilde{\delta}_{CWIN,23}|\mathbf{e}_2, \mathbf{e}_3). \quad (10)$$

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

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

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

$$P(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3|\tilde{\delta}_{CWIN}) = c \prod_{i=1}^3 P(\mathbf{e}_i) \times \prod_{i=1}^2 \prod_{j=i+1}^3 P(\tilde{\delta}_{CWIN,ij}|\mathbf{e}_i, \mathbf{e}_j) \quad (12)$$

for three events.

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

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

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

221 Extending to n events we get the posterior function

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

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

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

229 OR

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

231 The logarithm improves numerical stability by replacing products with sum-
 232 mations. The negative facilitates the use of optimisation algorithms that are
 233 designed to minimise an objective function.

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

$$241 \quad \mathbf{e}_1 = (0, 0, 0), \quad (17)$$

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

$$243 \quad \mathbf{e}_2 = (\hat{x}_2, 0, 0), \hat{x}_2 > 0 \quad (18)$$

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

$$245 \quad \mathbf{e}_3 = (\hat{x}_3, \hat{y}_3, 0), \hat{y}_3 > 0 \quad (19)$$

246 and the fourth to

$$247 \quad \mathbf{e}_4 = (\hat{x}_4, \hat{y}_4, \hat{z}_4), \hat{z}_4 > 0. \quad (20)$$

248 This coordinate system reduces translational (equation 17) and rotational
 249 (equations 18 to 20) non-uniqueness without loss of generality. It is neces-
 250 sary to work with a local coordinate system when using coda-waves alone
 251 because the CWI technique constrains only event separation between earth-
 252 quakes. The inclusion of arrival-times in the Combining Arrival-Time and
 253 CWI Constraints section allows us to move to a global reference system.

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

263 Synthetic experiments

264 We use synthetic examples in 2D and 3D with 50 earthquakes to test the
 265 performance of the optimization routine. In these examples the synthetic
 266 earthquakes are located randomly and CWI data generated according to the
 267 event separation. It is not necessary to generate synthetic waveforms and
 268 compute CWI estimates directly because we are testing the performance of
 269 the optimization routine only. The ability of CWI to estimate event separa-
 270 tion has been demonstrated already (*Snieder and Vrijlandt*, 2005; *Robinson*
 271 *et al.*, 2007a, 2011). We undertake a complete coda wave location experi-
 272 ment, including the calculation of CWI separation estimates, for recorded
 273 earthquakes in Relocating Earthquakes on the Calaveras Fault and in Com-
 274 bining Arrival-Time and CWI Constraints.

275 **Examples 1 and 2 - 2D synthetic experiments**

276 We design a 2D synthetic acoustic experiment (example 1) to test the perfor-
 277 mance of our CWI based relative location algorithm by randomly selecting
 278 \hat{x} - and \hat{y} -coordinates such that $-50 \leq \hat{x}, \hat{y} \leq 50$ m. These are indicated with
 279 triangles in Figure 1. We assume a local velocity of $\alpha = 3300 \text{ ms}^{-1}$ between
 280 all event pairs and a dominant frequency of 2.5 Hz to represent waveform data
 281 filtered between 1 and 5 Hz. The purpose of these examples is to synthetically
 282 test the location algorithm. Hence, we do not need to synthetically generate
 283 waveforms and compute CWI separation estimates, Rather, we begin by syn-
 284 thetically generating the CWI separation estimates directly. *Robinson et al.*
 285 (2011) showed that the CWI data are defined by the dominant wavelength
 286 normalized positive bounded Gaussian PDF with statistics $\bar{\mu}_N$ and $\bar{\sigma}_N$. A
 287 hypothetical CWI mean is created by setting

$$288 \quad \bar{\mu}_N = \mu_1 \left(\tilde{\delta}_t \right) \quad (21)$$

289 using equation (A7). This assumption ensures that the sample mean of hypo-
 290 thetical separation estimates is consistent with known CWI biases (*Robinson*
 291 *et al.*, 2011). In example 1 we use $\bar{\sigma}_N = 0.02$ between all event pairs. Ap-
 292 plication of our optimization procedure on the hypothetical CWI data yields
 293 the circles in Figure 1. The optimization does not lead to the exact solution
 294 due to the addition of noise ($\bar{\sigma}_N = 0.02$) on the hypothetical CWI data.
 295 The average coordinate error is 2.0 m (average location error ≈ 4 m) which
 296 is small compared to the noise of $\bar{\sigma}_N = 0.02$ which for $v_s = 3300 \text{ ms}^{-1}$ and
 297 $f_{dom} = 2.5 \text{ Hz}$ corresponds to roughly 25 m.

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

307 Conjugate gradient based optimization techniques are susceptible to the
 308 presence of local minima. This is because they use the slope of the target
 309 function to explore the solution space. We explore the impact of local min-
 310 ima for our CWI location problem by beginning the optimization from 25
 311 randomly chosen starting positions. We observe negligible difference in the
 312 solutions indicating that neither example is susceptible to local minima.

313 Three observations can be drawn from the error structure in Figures 1
 314 and 2. Firstly, the location errors depicted by gray bars increase between
 315 examples 1 and 2 with the introduction of larger noise. Secondly, the errors
 316 are larger for events at greater distances from the center. This is because
 317 events near the center of the cluster are constrained by links from all angles,
 318 whereas those on the outside are moderated by links from a limited number
 319 of directions. This observation is analogous to problems associated with poor
 320 azimuthal coverage in triangulation such as individual earthquake location
 321 from limited arrival-time data, or GPS positioning with few satellites. Our
 322 third observation is that the location errors form a pattern of circular rota-

tion, 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 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 amounts to rotations in 2D. The same 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 with measurements of arrival-times alleviates this problem and facilitates the removal of a local coordinate system altogether (see Combining 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.

347 In reality, we might expect that the separation between some pairs will not
 348 be constrained by CWI data due to poor signal to noise ratio in the coda
 349 for common stations. Obviously, the fewer stations that record an event
 350 the more likely it is that links between it and other events will be broken.
 351 In such cases the probabilistic distance constraint between a pair of events
 352 may only exist indirectly through multiple pairs. In this section we consider
 353 the impact of reduced linkage between event pairs. In example 3, we repeat
 354 example 2 using 90%, 80%, ..., 10% of the links. That is, we randomly select
 355 10% of the event pairs and remove the separation estimates between those
 356 pairs to create a data set with 90% linkage. Then, we randomly remove
 357 20% of the links and so on. This experiment is designed to mimic a realistic
 358 recording situation where CWI estimates are not available for all event pairs
 359 due to station problems, poor signal-to-noise ratio or any number of other
 360 reasons. As with the above examples, we undertake the optimization with
 361 25 randomly chosen starting locations.

362 Figure 3 illustrates the maximum Δ_{max} (top) and mean Δ_{μ} (middle) of
 363 the coordinate error as a function of percentage of earthquake pairs that
 364 are directly linked by a separation estimate. We show the statistics for the
 365 ‘best’ optimization solution (thick) and for the solution space when all 25
 366 optimizations are considered (thin). In the former case the best solution is
 367 determined by the set of event locations which lead to the smallest value of L .
 368 The error in the best solution is consistent when 30% or more of the branches
 369 are used. The errors increase when only 10% or 20% of the constraints are
 370 included. Interestingly, this breakdown around 20% to 30% coincides with
 371 the point where the average number of branches required to link an event

372 pair reaches 2 (see Fig.3 (bottom)). Since the average number of branches
 373 can be computed in advance it can be used as an indication of inversion
 374 stability prior to optimization. A higher breakdown is observed when all 25
 375 solutions are considered collectively. For example, the maximum coordinate
 376 error Δ_{max} exceeds that for the best solution for linkage $\leq 60\%$ confirming
 377 that the optimization is susceptible to local minima and that a range of
 378 starting points should be considered. Some optimizations fail to converge
 379 after 1200 iterations when the linkage is 60% or lower. All optimizations fail
 380 when the linkage is 20% or lower.

381 The derivatives used in the conjugate gradient method depend on events
 382 connected by CWI measurements. Consequently, earthquakes that are only
 383 connected via other events do not ‘communicate’ with each other directly. To
 384 some extent, this should be addressed during the iterative process where loca-
 385 tion information can spread to events which have no direct links. However,
 386 the lack of direct connection through the gradient could prevent convergence
 387 in extreme cases, or more likely slow the procedure down. This could explain
 388 why some examples do not converge after 1200 iterations. *VanDecar and*
 389 *Snieder* (1994) show that derivative based regularization acts slowly through
 390 iterative least-squares, because every cell in one iteration communicates only
 391 with its neighbours, and they demonstrate that this can be fixed with pre-
 392 conditioning in some cases. Their findings suggest that it may be possible to
 393 improve the convergence (stability and/or speed) of the CWI optimization
 394 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 event locations for 50 earthquakes with $-50 \text{ m} \leq \hat{x}, \hat{y}, \hat{z} \leq 50 \text{ m}$. As in the 2D case we assume a local velocity of $v = 3300 \text{ ms}^{-1}$ between all event pairs and a dominant frequency of 2.5 Hz to represent waveform data filtered between 1 and 5 Hz. The 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 $\bar{\sigma}_N = \epsilon(\delta_t)$ (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 (thin) are consistent with the best solution obtained from all 25 starting locations (thick). The best solution constrains the event locations down to 30% of the direct links. This is depicted in Figure 4 by the absence of the thin and thick lines below 60% and 30% linkage, respectively. 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

418 with greater $\bar{\sigma}_N$ leading to larger errors in the solutions. In 3D, when 70% or
 419 more of the direct branches are used the optimizer is stable with no observ-
 420 able difference in the solution for 25 randomly chosen starting locations. As
 421 the direct linkage reduces to 50% the optimization becomes less stable and
 422 the best solution from 25 random starting locations is required to find the
 423 optimal solution. All optimisations fail to converge as the number of links
 424 decrease below 30%.

425 Relocating Earthquakes on the Calaveras 426 Fault

427 In this section we relocate 68 earthquakes from the Calaveras Fault, Cali-
 428 fornia. The 68 earthquakes are selected from the 308 earthquake Calaveras
 429 example released with the open source Double Difference algorithm or hy-
 430 poDD (*Waldhauser and Ellsworth, 2000; Waldhauser, 2001*) [See also Data
 431 and Resources]. These events are chosen for four reasons. Firstly, they are
 432 recorded by a large number of stations (Fig. 5) and therefore lend themselves
 433 to accurate arrival-time location. This makes them ideal for assessing the per-
 434 formance of a new location technique. Secondly, they are distributed with
 435 separations from near zero to hundreds of meters making them ideal for ap-
 436 plication of CWI. Thirdly, Calaveras earthquakes have been well researched
 437 with several studies having relocated events in the region (*Waldhauser, 2001;*
 438 *Schaff et al., 2002; Waldhauser and Schaff, 2008*). Finally, the hypoDD lo-
 439 cations for these 68 earthquakes align in a streak increasing the likelihood

440 that they have near identical source mechanisms, a necessary assumption
 441 for the application of equation 2. The relocations in this paper are sorted
 442 into four examples as summarised in Table 1. Waveforms, cross correlations
 443 and separation estimates for example Calaveras event pairs are illustrated by
 444 *Robinson et al.* (2011) and are not repeated here for the sake of brevity.

445 **Example 5 - comparison of CWI, catalogue and hypoDD** 446 **locations**

447 Figure 6 illustrates three sets of locations for the Calaveras earthquakes. The
 448 first column shows the original catalogue locations for all 308 earthquakes.
 449 That is, each event is located individually using all available arrival-time
 450 data and a regional velocity model. The 68 earthquakes of interest in this
 451 study are differentiated in black. Catalogue locations suggest that the 68
 452 earthquakes of interest are spatially widely distributed on the scale of Figure
 453 6.

454 To apply CWI we download available waveforms from the Northern Cali-
 455 fornia Earthquake Data Center (See Data and Resources). Unsuitable wave-
 456 forms are removed using the conditions summarised in Table 2. Remaining
 457 waveforms are filtered between 1 and 5 Hz and aligned to P arrivals at 0 s.
 458 CWI estimates are obtained from 5 s wide non-overlapping time windows be-
 459 tween $2.5 < t \leq 20$ s and used to create probabilistic constraints on event
 460 separation. We utilize the local coordinate system introduced in the Theory
 461 Section and find the optimum relative locations using Polak-Ribiere opti-
 462 mization.

463 In this, and the following Calaveras examples, we allow the earthquakes

464 to move freely in all three directions during the inversion despite using the
 465 in-fault separation estimates given by equation 2. We allow the events to
 466 move freely so that we can test the performance of our algorithm without
 467 assuming a-priori that the earthquakes are constrained on the same fault
 468 plane. We approximate the true event separation using the in-fault separa-
 469 tion of equation 2 so that we can focus on developing a working algorithm
 470 and demonstrate capability without dealing with the complexity of in-fault
 471 ($\delta_{\parallel fault}$) and out-of-fault ($\delta_{\perp fault}^2$) displacement. Considering the more com-
 472 plicated formulation of equation 4 is left for future work. Another potential
 473 focus for future work involves refining our algorithm to simultaneously re-
 474 solve event location and representative fault plane by restricting the events
 475 to align in a single (unknown a-priori) plane. That is, for cases where the
 476 earthquakes are believed a-priori to be in the same plan.

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

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

489 The spatial distribution of the CWI locations is clearly tighter than the
 490 catalogue locations of column 1. That is, CWI provides an independent in-
 491 dication of clustering for the 68 events and to first order, similar locations
 492 to those from hypoDD (column 3). There is a small second order difference
 493 between the CWI and hypoDD based locations. In particular, the lineation
 494 is less clear in the CWI locations (column 2) than the hypoDD locations
 495 (column 3). Our experience suggests that the CWI locations are less sup-
 496 portive of the presence of streaks although a complete understanding of these
 497 differences is left for future work. Our attention now is devoted towards un-
 498 derstanding how both techniques perform with fewer stations (Example 6)
 499 and exploring how CWI and arrival-times can be combined (Examples 7 and
 500 8).

501 **Example 6 - Dependence on the number of stations**

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

509 We explore the impact of poorer recording situations in example 6 by
 510 re-locating the 68 Calaveras events using hypoDD and coda-waves with a
 511 reduced number of stations. We begin with 10 stations and repeat the process
 512 removing one at a time until a single station remains. The 10 stations and

513 there order of removal are shown in Figure 7.

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

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

529 As the number of stations is reduced neither the CWI nor hypoDD tech-
530 niques are able to re-locate all events. To use the coda waves we need at
531 least one pairwise separation constraint to be formed from the available sta-
532 tions. This means that for every event there must be at least one station
533 that records it, and at least one other earthquake, sufficiently well to ap-
534 ply CWI. Fortunately, we can make an assessment of this prior to starting
535 the inversion. The top panel of Figure 10 demonstrates that when five or
536 more stations are used, CWI can constrain the location of all 68 earthquakes.
537 When less than five stations are used the coda-waves constrain a decreasing

538 number of events until at one station it is only possible to locate 55 of the 68
 539 events. The hypoDD algorithm also fails to locate all events as the number
 540 of stations is reduced. In the case of hypoDD an event can be identified
 541 as unconstrainable in one of two stages. Firstly, the data are analyzed to
 542 ensure that there exist arrival-time differences for each event and at least
 543 one other earthquake. This is analogous to the situation for the coda-wave
 544 technique. The hypoDD program also has a secondary identification phase
 545 in which events that can not be located sufficiently well are rejected during
 546 the inversion. This process is related to the iterative removal of outliers de-
 547 scribed by *Waldhauser and Ellsworth* (2000). The top panel of Figure 10
 548 shows that the number of events re-located by hypoDD fluctuates between
 549 63 and 28 earthquakes for ten to four stations and it demonstrates that the
 550 number of events located by hypoDD is less than or equal to the number
 551 located by CWI.

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

Combining Arrival-Time and CWI Constraints

In Examples 5 and 6 we compare the location of the Calaveras earthquakes using coda-wave and arrival time based constraints independently. Since the arrival time (direct or difference) and 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 arrival-time data.

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

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

where

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

is the location of event p ,

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

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

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

is the covariance matrix. We define the mean location $\mu_{\mathbf{e}_p}$ and covariance matrix by the hypoDD optimum solution and its uncertainties. It is important

579 to note that hypoDD must be used with SVD to obtain useful estimates of
 580 σ_{x_p} , σ_{y_p} and σ_{z_p} because the errors reported by conjugate gradient methods
 581 (LSQR) are grossly underestimated in hypoDD (*Waldhauser, 2001*).

582 We pose the location problem using the negative log likelihood

$$583 \quad L(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_1, \mathbf{e}_n) = - \sum_{i=1}^n \ln [P(\mathbf{e}_i | \Delta_{TT})] \quad (26)$$

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

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

$$585 \quad \sum_{i=1}^n \ln [P(\mathbf{e}_i | \Delta_{TT})] \quad (27)$$

586 incorporates the arrival-time constraints and

$$587 \quad \sum_{i=1}^{n-1} \sum_{j=i+1}^n \ln [P(\delta_{CWIN} | \mathbf{e}_i, \mathbf{e}_j)] \quad (28)$$

588 the coda-waves.

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

$$592 \quad \frac{\partial L}{\partial x_p} = - \frac{\partial \ln [P(\mathbf{e}_p | t_{DD})]}{\partial x_p} - \sum_{i=p+1}^N \frac{\partial \ln [P(\delta_{CWIN} | \mathbf{e}_p, \mathbf{e}_i)]}{\partial x_p} \quad (29)$$

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

593 where

$$594 \quad \sum_{i=p+1}^N \frac{\partial \ln [P(\delta_{CWIN} | \mathbf{e}_p, \mathbf{e}_i)]}{\partial x_p} \quad (30)$$

595 and

$$596 \quad \sum_{j=1}^{p-1} \frac{\partial \ln [P(\delta_{CWIN} | \mathbf{e}_j, \mathbf{e}_p)]}{\partial x_p} \quad (31)$$

are defined in Appendix and

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

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

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

and

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

Combining the 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 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 need to consider multiple randomly chosen starting locations.

Example 7 - Combining arrival-time and CWI constraints

Figure 11 illustrates the earthquake locations obtained when we combine the arrival-time and coda-wave data using all data (left) and five stations

617 (right). The linear features observed in the original hypoDD inversions (see
 618 Fig. 6) are evident in both cases. However, the coda-waves introduce a
 619 scatter around these streaks. That is, the locations in figure 11 result from a
 620 trade-off between hypoDD's desire to place the events on linear features and
 621 the coda-waves voracity to push them away from streaks. When all stations
 622 are used the hypoDD constraints are strong and little off-streak scatter is
 623 introduced. As we reduce hypoDD's leverage by decreasing the number of
 624 stations to five, we observe an increase in off-streak scatter resulting from
 625 the enhanced influence of the coda.

626 **Example 8 - Combining CWI and arrival-times when** 627 **the arrival-times constrain a limited number of events**

628 In intraplate regions such as Australia it is common to deploy temporary
 629 seismometers to monitor aftershocks for significant events (*Bowman et al.*,
 630 1990; *Leonard*, 2002). Traditionally, these deployments facilitate a higher
 631 accuracy of location for events occurring during the deployment period. Using
 632 our combined inversion it is possible to re-locate all events by employing
 633 the detailed arrival-time data when the temporary network is in-situ and
 634 using coda-waves from network stations when the deployment is absent. The
 635 hypothesis, to be tested in this section, is that conducting such a combined
 636 inversion will improve the location accuracy of events outside the deployment
 637 period.

638 An estimate of the cumulative number of aftershocks $N(t)$ after t days

639 can be modeled by the modified Omori formula

$$640 \quad N(t) = K \frac{c^{1-p} + (t+c)^{1-p}}{p-1} \quad (35)$$

641 (*Utsu et al.*, 1995). The empirically derived constants, K , C and p vary
642 between tectonic settings. For example, using recorded aftershocks with
643 $M \geq 3.2$ of the Hokkaido-Nansei-Oki, Japan $M_s = 7.8$ earthquake of 12
644 July 1993, *Utsu et al.* (1995) obtained maximum likelihood estimates for
645 K , p and c of 906.5, 1.256 and 1.433, respectively. With these empirically
646 derived values an array deployed within 4 days and left for 150 days will
647 record roughly one half of the aftershocks occurring within the first 1000
648 days. That is,

$$649 \quad \frac{N(150+4) - N(4)}{N(1000)} = \frac{2257 - 934}{2626} \approx 0.5. \quad (36)$$

650 This is illustrated in Figure 12 which shows the best fitting Omori Formula
651 separated into segments before, during and after the pseudo temporary de-
652 ployment.

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

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

669 Discussion and Conclusions

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

678 Coda-wave interferometry offers a new technique for understanding earth-
679 quakes in intraplate areas with sparse networks and poor azimuthal coverage.
680 In particular, the ability to combine coda-wave constraints with arrival-times
681 makes it possible to link well constrained events from a temporary deploy-
682 ment with those recorded outside the deployment period. All that is re-
683 quired to achieve this is at least one network station which has recorded

684 sufficient events from both periods. CWI facilitates the location of poorly
685 recorded events to an accuracy approaching those recorded during the tem-
686 porary deployment and therefore opens new avenues for imaging intraplate
687 fault structures and improving our understanding of intraplate seismicity
688 and earthquake hazard. Importantly, this analysis can be conducted for any
689 historical aftershock sequence or earthquake swarm recorded by a temporary
690 deployment. Our technique is, in that sense, related to the retrospective
691 seismological observation technique of *Curtis et al.* (2012) that utilizes inter-
692 ferometry to obtain seismic signals on newly installed sensors regardless of
693 whether the event occurs before, during or after the physical installation of
694 the sensor.

695 Another potential application of CWI is in the area of hydraulic fracturing
696 such as hot rock geothermal projects, petroleum reservoir engineering, tight
697 gas extraction, CO₂ geosequestration and/or underground brine injection.
698 Monitoring pumping-induced micro earthquakes is a key step in understand-
699 ing the migration of fluids in such reservoirs. There is a trade-off in the
700 ability of surface deployed networks to locate events which are small and/or
701 deep. Downhole seismic monitoring is likely to play increasingly important
702 roles in deep reservoir projects. CWI creates new possibilities to monitor
703 pumping induced micro earthquakes from fewer boreholes and hence dra-
704 matically reduce the costs of reservoir monitoring at large depths. It may
705 also be possible to utilize coda for understanding hazard in tunneled mining
706 operations where the location of deep tunnels prohibits azimuthal coverage
707 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 Seismological 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

730 (DP0665111). This paper has benefited significantly from: reviews by Trevor
731 Allen and Clive Collins at Geoscience Australia, the BSSA editorial team,
732 Andrew Curtis from the School of GeoSciences at the University of Edin-
733 burgh and an anonymous reviewer. The authors have also benefited from
734 comments and advice received from Felix Waldhauser on this and earlier
735 research on the application of CWI to earthquake location.

736 References

- 737 Ake, J., D. O’Connell, and L. Block (2005), Deep-injection and closely mon-
738 itored induced seismicity at Paradox Valley, Colorado, *Bull. Seism. Soc.*
739 *Am.*, *95*(2), 664–683.
- 740 Aki, K. (1969), Analysis of the seismic coda of local earthquakes as scattered
741 waves, *J. Geophys. Res.*, *74*(2), 615–631.
- 742 Aster, R. C., B. Borchers, and C. H. Thurber (2005), *Parameter estimation*
743 *and inverse problems*, *International Geophysics Series*, vol. 90, Elsevier
744 Academic Press, USA.
- 745 Bondár, I., S. C. Myers, E. R. Engdahl, and E. A. Bergman (2004), Epicentre
746 accuracy based on seismic network criteria, *Geophys. J. Int.*, *156*, 483–496.
- 747 Bowman, J. R., G. Gibson, and T. Jones (1990), Aftershocks of the 1988
748 January 22 Tennant Creek, Australia intraplate earthquakes: evidence for
749 a complex thrust-fault geometry, *Geophys. J. Int.*, *100*, 87–97.
- 750 Campbell, K. W. (2003), Strong motion attenuation, in *International Hand-*

751 *book of Earthquake and Engineering Seismology*, vol. B, edited by W. H. K.
 752 Lee, H. Kanamori, P. C. Jennings, and C. Kisslinger, chap. 60, pp. 1003–
 753 1012, Academic Press, London.

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

758 Curtis, A., Behr, Y., Entwistle, E., Galetti, E., Townend J. Bannister, S.
 759 (2012), The benefit of hindsight in observational science: Retrospective
 760 seismological observations, *Earth Planet. Sci. Lett.*, *345–348*, 212–220.

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

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

765 Frankel, A. D., C. S. Mueller, T. P. Barnhard, E. V. Leyendecker, R. L.
 766 Wesson, S. C. Harmsen, F. W. Klein, D. M. Perkins, N. C. Dickman, S. L.
 767 Hanson, and M. G. Hopper (2000), USGS National seismic hazard maps,
 768 *Earthquake Spectra*, *16*(1), 1–19.

769 Frèmont, M.-J., and S. D. Malone (1987), High precision relative locations
 770 of earthquakes at Mount St. Helens, *J. Geophys. Res.*, *92*(B10), 10,223–
 771 10,236.

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

775 Gutenberg, B. (1945), Amplitudes of surface waves and magnitudes of shal-
776 low earthquakes, *Bull. Seism. Soc. Am.*, *35*, 3–12.

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

779 Kennett, B. L. N., E. R. Engdahl, and R. Buland (1995), Constraints on
780 seismic velocities in the Earth from traveltimes, *Geophys. J. Int.*, *122*,
781 108–124.

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

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

787 Leonard, M. (2002), The Burakin WA earthquake sequence Sept 2000 – June
788 2002, in *Total Risk Management in the Privatised Era, Australian Earth-*
789 *quake Engineering Society Conference*, vol. 10th, edited by M. Griffith,
790 D. Love, P. McBean, A. McDougall, and B. Butler, pp. 22(1)–22(5), AEES,
791 University of Adelaide.

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

- 795 Pavlis, G. L. (1992), Appraising relative earthquake location errors, *Bull.*
796 *Seism. Soc. Am.*, 82(2), 836–859.
- 797 Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling (1987),
798 *Numerical Recipes: The Art of Scientific Computing*, Cambridge Univer-
799 sity Press, USA.
- 800 Richter, C. F. (1935), An instrumental earthquake magnitude scale, *Bull.*
801 *Seism. Soc. Am.*, 25(1), 1–32.
- 802 Robinson, D., T. Dhu, and J. Schneider (2006), Practical probabilistic seismic
803 risk analysis: A demonstration of capability, *Seism. Res. Let.*, 77(4), 452–
804 458.
- 805 Robinson, D. J., M. Sambridge, and R. Snieder (2007a), Constraints on coda
806 wave interferometry estimates of source separation: The 2.5d acoustic case,
807 *Exploration Geophysics*, 38(3), 189–199.
- 808 Robinson, D. J., R. Snieder, and M. Sambridge (2007b), Using coda
809 wave interferometry for estimating the variation in source mecha-
810 nism between double couple events, *J. Geophys. Res.*, 112, b12302,
811 doi:10.1029/2007JB004925.
- 812 Robinson, D. J., M. Sambridge, and R. Snieder (2011), A probabilistic ap-
813 proach for estimating the separation between a pair of earthquakes directly
814 from their coda waves, *J. Geophys. Res.*, B04309, 1–17.
- 815 Rubin, A. M. (2002), Aftershocks of microearthquakes as probes
816 of the mechanics of rupture, *J. Geophys. Res.*, 107(B7,2142),
817 10.1029/2001JB000496.

- 818 Rubin, A. M., D. Gillard, and J.-L. Got (1999), Streaks of microearthquakes
819 along creeping faults, *Nature*, *400*, 635–641.
- 820 Schaff, D. P., G. H. R. Bokelmann, and G. C. Beroza (2002), High-resolution
821 image of Calaveras Fault seismicity, *J. Geophys. Res.*, *107*(B9), 2186,
822 doi:10.1029/2001JB000633.
- 823 Shearer, P., E. Hauksson, and G. Lin (2005), Southern California hypocenter
824 relocation with waveform cross-correlation, Part 2: Results using source-
825 specific station terms and cluster analysis, *Bull. Seism. Soc. Am.*, *95*(3),
826 904–915. doi:10.1785/0120040168.
- 827 Shearer, P. M. (1999), *Introduction to Seismology*, Cambridge University
828 Press, USA, 260pp.
- 829 Sipkin, S. A. (2002), USGS earthquake moment tensor catalog, in *Interna-*
830 *tional Handbook of Earthquake Engineering Seismology*, vol. A, edited by
831 W. H. Lee, H. Kanamori, P. C. Jennings, and C. Kisslinger, chap. 50, pp.
832 823–825, Academic Press, London.
- 833 Slunga, R., Rögnvaldsson, S. Th. and Bödvarsson, R. (1995), Absolute and
834 relative locations of similar events with application to microearthquakes
835 in southern Iceland, *Geophys. J. Int.*, *123*(2), 409–419.
- 836 Snieder, R. (1999), Imaging and averaging in complex media, in *Diffuse waves*
837 *in complex media*, *NATO Science Series C*, vol. 531, edited by J. P. Fouque,
838 pp. 405–454, Kluwer Academic Publishers.
- 839 Snieder, R. (2006), The theory of coda wave interferometry, *Pure Appl. Geo-*
840 *phys.*, *163*, 455–473.

- 841 Snieder, R., and M. Vrijlandt (2005), Constraining the source separation
842 with coda wave interferometry: Theory and application to earthquake
843 doublets in the Hayward Fault, California, *J. Geophys. Res.*, *110*(B04301),
844 doi:10.1029/2004JB003317.
- 845 Spencer, C., and D. Gubbins (1980), Travel-time inversion for simultane-
846 ous earthquake location and velocity structure determination in laterally
847 varying media, *Geophys. J. R. Astr. Soc.*, *63*, 95–116.
- 848 Stirling, M. W., G. H. McVerry, and K. R. Berryman (2002), A new seismic
849 hazard model for New Zealand, *Bull. Seism. Soc. Am.*, *92*(5), 1878–1903.
- 850 Toro, G. R., N. A. Abrahamson, and J. F. Schneider (1997), Model of strong
851 ground motions from earthquakes in Central and Eastern North America:
852 Best estimates and uncertainties, *Seism. Res. Let.*, *68*(1), 41–57.
- 853 Utsu, T., Y. Ogata, and R. S. Matsu’ura (1995), The Centenary of the Omori
854 Formula for a decay law of aftershock activity, *J. Phys. Earth*, *43*, 1–33.
- 855 VanDecar, J. C., and R. Snieder (1994), Obtaining smooth solutions to large
856 linear inverse problems, *Geophysics*, *59*, 818–829.
- 857 Waldhauser, F. (2001), hypoDD – a program to compute double-difference
858 hypocenter locations (hypoDD version 1.0 - 03/2001), *Open file report 01-*
859 *113*, United States Geological Survey, Menlo Park, California.
- 860 Waldhauser, F., and W. L. Ellsworth (2000), A double-difference earthquake
861 location algorithm: method and application to the northern Hayward
862 Fault, California, *Bull. Seism. Soc. Am.*, *90*(6), 1353–1368.

- 863 Waldhauser, F., and W. L. Ellsworth (2002), Fault structure and mechan-
864 ics of the Hayward Fault, California, from double-difference earthquake
865 locations, *J. Geophys. Res.*, *107*(B3), 10.1029/2000JB000,084.
- 866 Waldhauser, F., and D. P. Schaff (2008), Large-scale relocation of
867 two decades of Northern California seismicity using cross-correlation
868 and double-difference methods, *J. Geophys. Res.*, *133*, B08311,
869 doi10.1029/2007JB005479.
- 870 Waldhauser, F., W. L. Ellsworth, and A. Cole (1999), Slip-parallel lineations
871 on the Northern Hayward Fault, California, *Geophys. Res. Lett.*, *26*(23),
872 3525–3528.

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

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 Engineer-
 ing
 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 arrival-time data with all and a reduced number of stations.
Example 8	Combined use of CWI and arrival-time data when arrival-times constrain only 50% of the events.

Table 2: Conditions used to identify unsuitable waveforms before applying

CWI*

	condition
1	waveform is clearly corrupted
2	waveform indicates recording of more than 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)

875

* Originally published as Table 5 in *Robinson et al.* (2011)

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 defined by the 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 from all stations (left) and 5 stations (right). Axes as defined in Figure 6.

Figure 12: 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 arrival-time inversion (left) and the inversion with 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 $P(\tilde{\delta}_{CWIN}|\tilde{\delta}_t)$ used in equation (7) is given by

$$P(\tilde{\delta}_{CWIN}|\tilde{\delta}_t) = A(\tilde{\delta}_t)C(\bar{\mu}_N, \bar{\sigma}_N) \times \int_0^\infty B(\tilde{\delta}_t, \tilde{\delta}_{CWI})D(\tilde{\delta}_{CWI}, \bar{\sigma}_N, \bar{\mu}_N)d\tilde{\delta}_{CWI} \quad (\text{A1})$$

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

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

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

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

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

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

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

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

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

and

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

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

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

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 $\tilde{\delta}_{CWIN}$ computed from the waveforms and the positively bounded Gaussian density function

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

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

Derivatives

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

$$\frac{\partial L}{\partial \hat{x}_1}, \frac{\partial L}{\partial \hat{y}_1}, \frac{\partial L}{\partial \hat{z}_1}, \frac{\partial L}{\partial \hat{x}_2}, \frac{\partial L}{\partial \hat{y}_2}, \frac{\partial L}{\partial \hat{z}_2}, \dots, \frac{\partial L}{\partial \hat{x}_N}, \frac{\partial L}{\partial \hat{y}_N}, \frac{\partial L}{\partial \hat{z}_N} \quad (\text{A10})$$

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

907 L .

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

$$911 \quad \delta_t = \frac{f_{dom}}{v_s} \sqrt{(\hat{x}_p - \hat{x}_q)^2 + (\hat{y}_p - \hat{y}_q)^2 + (\hat{z}_p - \hat{z}_q)^2}, \quad (\text{A11})$$

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

$$914 \quad \begin{aligned} \frac{\partial \tilde{\delta}_t}{\partial \hat{x}_p} &= \frac{f_{dom}^2 (\hat{x}_p - \hat{x}_q)}{v_s^2 \tilde{\delta}_t}, \quad \frac{\partial \tilde{\delta}_t}{\partial \hat{y}_p} = \frac{f_{dom}^2 (\hat{y}_p - \hat{y}_q)}{v_s^2 \tilde{\delta}_t}, \\ \frac{\partial \tilde{\delta}_t}{\partial \hat{z}_p} &= \frac{f_{dom}^2 (\hat{z}_p - \hat{z}_q)}{v_s^2 \tilde{\delta}_t}, \quad \frac{\partial \tilde{\delta}_t}{\partial \hat{x}_q} = \frac{f_{dom}^2 (\hat{x}_q - \hat{x}_p)}{v_s^2 \tilde{\delta}_t}, \\ \frac{\partial \tilde{\delta}_t}{\partial \hat{y}_q} &= \frac{f_{dom}^2 (\hat{y}_q - \hat{y}_p)}{v_s^2 \tilde{\delta}_t}, \quad \frac{\partial \tilde{\delta}_t}{\partial \hat{z}_q} = \frac{f_{dom}^2 (\hat{z}_q - \hat{z}_p)}{v_s^2 \tilde{\delta}_t}. \end{aligned} \quad (\text{A12})$$

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

$$918 \quad \frac{\partial \tilde{\delta}_t}{\partial \hat{x}_q} = -\frac{\partial \tilde{\delta}_t}{\partial \hat{x}_p}. \quad (\text{A13})$$

919 The chain rule gives

$$920 \quad \frac{\partial \mu_1}{\partial \hat{x}_p} = \frac{\partial \mu_1}{\partial \tilde{\delta}_t} \frac{\partial \tilde{\delta}_t}{\partial \hat{x}_p} \quad (\text{A14})$$

921 where differentiating equation (A7) gives

$$922 \quad \frac{\partial \mu_1}{\partial \tilde{\delta}_t} = a_1 \frac{a_2 a_4 \tilde{\delta}_t^{a_4-1} + a_3 a_5 \tilde{\delta}_t^{a_5-1}}{\left(a_2 \tilde{\delta}_t^{a_4} + a_3 \tilde{\delta}_t^{a_5} + 1 \right)^2}. \quad (\text{A15})$$

923 Similarly, we have

$$924 \quad \frac{\partial \sigma_1}{\partial \hat{x}_p} = \frac{\partial \sigma_1}{\partial \tilde{\delta}_t} \frac{\partial \tilde{\delta}_t}{\partial \hat{x}_p} \quad (\text{A16})$$

925 where $\frac{\partial \sigma_1}{\partial \delta_t}$ has the identical form as A15 with different constants a_1, a_2, \dots, a_5
 926 (see table A1).

927 The cumulative Gaussian distribution function A6 is

$$928 \quad \Phi_{\mu_1, \sigma_1}(0) = \frac{1}{\sigma_1 \sqrt{2\pi}} \int_{-\infty}^0 e^{-\frac{(s-\mu_1)^2}{2\sigma_1^2}} ds \quad (\text{A17})$$

929 which has derivative

$$930 \quad \frac{\partial \Phi_{\mu_1, \sigma_1}(0)}{\partial \hat{x}_p} = \frac{\sigma_1 \int_{-\infty}^0 \frac{\partial g}{\partial \hat{x}_p} e^g ds - \frac{\partial \sigma_1}{\partial \hat{x}_p} \int_{-\infty}^0 e^g ds}{\sigma_1^2 \sqrt{2\pi}}, \quad (\text{A18})$$

931 where

$$932 \quad g = \frac{-(s - \mu_1)^2}{2\sigma_1^2} \quad (\text{A19})$$

933 and

$$934 \quad \frac{\partial g}{\partial \hat{x}_p} = \frac{4\sigma_1^2(s - \mu_1) \frac{\partial \mu_1}{\partial \hat{x}_p} + 4\sigma_1 \frac{\partial \sigma_1}{\partial \hat{x}_p} (s - \mu_1)^2}{4\sigma_1^4}. \quad (\text{A20})$$

935 Now, we have all the pieces to compute the derivatives of $A = A(\delta_t)$ and
 936 $B = B(\delta_t, \delta_{CWI})$ as follows

$$937 \quad \frac{\partial A}{\partial \hat{x}_p} = -\frac{-\frac{\partial \Phi_{\mu_1, \sigma_1}(0)}{\partial \hat{x}_p} \sigma_1 + (1 - \Phi_{\mu_1, \sigma_1}(0)) \frac{\partial \sigma_1}{\partial \hat{x}_p}}{(1 - \Phi_{\mu_1, \sigma_1}(0))^2 \sigma_1^2 \sqrt{2\pi}} \quad (\text{A21})$$

938 and

$$939 \quad \frac{\partial B}{\partial \hat{x}_p} = e^h \frac{\partial h}{\partial \hat{x}_p}, \quad (\text{A22})$$

940 where

$$941 \quad h = \frac{-(\delta_{CWI} - \mu_1)^2}{2\sigma_1^2} \quad (\text{A23})$$

942 and

$$943 \quad \frac{\partial h}{\partial \hat{x}_p} = \frac{4\sigma_1^2(\delta_{CWI} - \mu_1) \frac{\partial \mu_1}{\partial \hat{x}_p} + 4(\delta_{CWI} - \mu_1)^2 \sigma_1 \frac{\partial \sigma_1}{\partial \hat{x}_p}}{4\sigma_1^4}. \quad (\text{A24})$$

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

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

946 and for the logarithm we have

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

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

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

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