Bulletin of the Seismological Society of America (2012) **200**, 1–**??**

Relocating a cluster of earthquakes using a single station

David J. Robinson[1,2], Malcolm Sambridge[1], Roel Snieder[3], Jürg Hauser[4]

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

[2] *Risk Analysis Methods, Geoscience Australia, GPO Box* 383 *Canberra ACT* 2601 *Australia, E-mail: david.robinson@ga.gov.au*

[3] *Center for Wave Phenomena and Department of Geophysics, Colorado School of Mines, Golden CO* 80401-1887*, USA*

[4] .....

Received 2012 Month Day; in original form 2012 Month Day

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 intereference 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 travel 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 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.

1 INTRODUCTION

Earthquake location is important for many applications. Locations are required for: magnitude determination (**??**); computing moment tensors (**?**); seismological studies of the Earth's interior (**????**); understanding strong motion and seismic attenuation (**??**) and modeling earthquake hazard or risk (**???**).

The accuracy required in earthquake location depends on the application. For example, imaging the structure of a fracture system from microseismicity requires greater detail 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, quality of signal and accuracy of the velocity model used in computing travel times. Uncertainties in absolute location are typically of the order of several kilometers because they are susceptible to uncertainty in the velocity structure along the entire path between the source and receiver. For example, **?**) states that location uncertainty in the ISC and PDE catalogues are generally around 25 km horizontally and at least 25 km in depth.* **?**) 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. 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 **?**) who developed the technique commonly known as joint hypocenter determination.† In principle, relative locations can be computed by differencing absolute locations. However, **?**) 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 compute locations directly from travel time differences between two waveforms (**????**). By doing so, they remove errors associated with velocity variations outside the local region, because such variations influence all waveforms in the same manner (**?**).

Reported location uncertainties from relative techniques are around 15 to 75 m in local settings with good station coverage (**????**). 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 (**?????**); studying rupture mechanics (**??**); interpreting magma movement in volcanoes (**?**); and monitoring pumping induced seismicity (**??**).

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 difference computed between picked arrivals of two waveforms; or 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.

* Here the depth uncertainties of 25 km assume the use of depth dependent phases such as pP. Without such phases the uncertainty is higher.

† **?**) originally used the term joint epicentre determination. However, he was solving for hypocentre.

Coda refers to later arriving waves in the seismogram that arise from scattering (**???**). Coda waves are ignored in most seismological applications due to the complexity involved in modeling their generation. In this paper we develop an approach for locating earthquakes using coda waves.

?) 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 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 (**?**). 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 travel times.

2 THEORY

?) 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 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,

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

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

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

(**?**). In this paper 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. **?**) explore the impact of a change in mechanism.

The σ_τ in equation (1) is related to R_{max}, the maximum of the cross correlation between the coda of the two waveforms, and hence can be computed directly from the recorded data. In this paper we use the autocorrelation approach of **?**) when relating these parameters and we apply a restricted time lag search when evaluating R_{max}. These extensions to the original technique of **?**) increase the range of applicability of CWI by 50% (i.e. from 300 to 450 m separation for 15 Hz filtered coda waves).

?) show that coda waves provide only 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

$$P(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN}) \propto P(\widetilde{\delta}_{CWIN}|\widetilde{\delta}_t) \times P(\widetilde{\delta}_t) \tag{4}$$

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 accounting for all a-priori information. The tilde above the separation parameters in eq. (4) indicates the use of a wavelength normalised separation parameter

$$\widetilde{\delta} = \frac{\delta}{\lambda_d}, \tag{5}$$

which measures separation (δ = δ_{CWIN} or δ_t) with respect to dominant wavelength λ_d. In this paper we consider a uniform prior over appropriate bounds which gives emphasis to the recorded data. The procedure for computing the noisy likelihood $P(\widetilde{\delta}_{CWIN}|\widetilde{\delta}_t)$ is derived by **?**) and summarised in Appendix **??**. 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) which links individual pairwise posteriors $P(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN})$ to describe the location of multiple events whose maximum corresponding 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

$$P(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3|\mathbf{d}) \propto P(\mathbf{d}|\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) \times P(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3), \tag{6}$$

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 (6) $\mathbf{d}$ represents observations that constrain the locations. They can be any combination of travel times, geodetic information or CWI separations. For example, if 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 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 differences. In this derivation and in Section 3 we focus on the constraints imposed by coda waves, whereas in Section **??** we demonstrate how CWI and travel 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 like-

lihoods 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.[‡] Furthermore, since the pairwise functions are independent the joint likelihood becomes

$$P(\widetilde{\delta}_{CWIN}|\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3) = P(\widetilde{\delta}_{CWIN,12}|\mathbf{e}_1,\mathbf{e}_2) \times P(\widetilde{\delta}_{CWIN,13}|\mathbf{e}_1,\mathbf{e}_3) \times P(\widetilde{\delta}_{CWIN,23}|\mathbf{e}_2,\mathbf{e}_3). \tag{7}$$

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

$$P(\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3) = P((\mathbf{e}_1) \times P(\mathbf{e}_2) \times P(\mathbf{e}_3). \tag{8}$$

Combining equations (7) and (8) gives the joint posterior function

$$P(\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3|\widetilde{\boldsymbol{\delta}}_{CWIN}) = c\prod_{i=1}^{3} P(\mathbf{e}_i) \times \prod_{i=1}^{2}\prod_{j=i+1}^{3} P(\widetilde{\delta}_{CWIN,ij}|\mathbf{e}_i,\mathbf{e}_j) \tag{9}$$

for three events.

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

$$P(\mathbf{e}_2|\delta_{CWIN}) = \int\int P(\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3|\widetilde{\delta}_{CWIN})d\mathbf{e}_1 d\mathbf{e}_3 \tag{10}$$

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 (9) 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

$$P(\mathbf{e}_1,...,\mathbf{e}_n|\widetilde{\delta}_{CWIN}) = c\prod_{i=1}^{n} P(\mathbf{e}_i) \times \prod_{i=1}^{n-1}\prod_{j=i+1}^{n} P(\widetilde{\delta}_{CWIN,ij}|\mathbf{e}_i,\mathbf{e}_j). \tag{11}$$

When evaluating equation (11) 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

$$L(\mathbf{e}_1,\mathbf{e}_2,...,\mathbf{e}_n) = -ln\left[P(\mathbf{e}_1,...,\mathbf{e}_n|\widetilde{\delta}_{CWIN})\right] \tag{12}$$

[‡] Note 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.

or

$$L(\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n) = -ln\,[c] - \sum_{i=1}^{n} ln\,[P(\mathbf{e}_i)] \\ - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} ln\left[P(\widetilde{\delta}_{CWIN,ij}|\mathbf{e}_i, \mathbf{e}_j)\right]. \tag{13}$$

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 $\hat{.}$ indicates use of a local coordinate system. We choose a local coordinate system which removes ambiguity associated with transformations of the coordinate system. In defining this coordinate system we fix the first event at the origin,

$$\mathbf{e}_1 = (0, 0, 0), \tag{14}$$

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

$$\mathbf{e}_2 = (\hat{x}_2, 0, 0), \hat{x}_2 > 0 \tag{15}$$

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

$$\mathbf{e}_3 = (\hat{x}_3, \hat{y}_3, 0), \hat{y}_3 > 0 \tag{16}$$

and the fourth to

$$\mathbf{e}_4 = (\hat{x}_4, \hat{y}_4, \hat{z}_4), \hat{z}_4 > 0. \tag{17}$$

This coordinate system reduces translational (constraint 14) and rotational (constraints 15 to 17) non-uniqueness without loss of generality. It is necessary to work with a local coordinate system when using coda waves alone because the CWI technique constrains only event separation between earthquakes. The inclusion of travel times in 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. The most likely set of locations is given by the minimum of L. In this paper we use the Polak-Ribiere technique (**?**), a conjugate gradient method, to optimize L. It uses the derivatives of L, derived in Appendix **??**, to guide the optimization procedure. Note that when optimizing equation 13 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.

3 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 (**???**). We undertake a complete coda wave location experiment, including the calculation of CWI separation estimates, for recorded earthquakes on the Calaveras Fault, California, in Section 4.

3.1 Examples 1 and 2 - 2D synthetic experiments

We design a 2D synthetic experiment (example 1) by randomly selecting $\hat{x}$- and $\hat{y}$-coordinates such that $-50 \leq \hat{x}, \hat{y} \leq 50\,\mathrm{m}$. These are indicated with triangles in Figure 1. We assume a local velocity of $v = 3,300\,\mathrm{ms}^{-1}$ between all event pairs and a dominant frequency of 2.5 Hz to represent waveform data filtered between 1 and 5 Hz. The CWI data are defined by the positive bounded Gaussian with statistics $\bar{\mu}_N$ and $\bar{\sigma}_N$. A hypothetical CWI mean is created by setting

$$\bar{\mu}_N = \mu_1\left(\widetilde{\delta_t}\right) \tag{18}$$

using equation (**??**). This assumption ensures that the sample mean of hypothetical separation estimates is consistent with known CWI biases (**?**). 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 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.

?) 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 $\bar{\sigma}_N = \sigma_{1to5Hz}(\delta_t)$, where $\sigma_{1to5Hz}(\delta_t)$ is the half-width of the errorbars for a synthetic acoustic experiment with filtering between 1 and 5 Hz (see Fig. 4(b) of **?**). Repeating the optimization leads to the circles in Figure 2 which have an average coordinate error of 2.8 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.

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.

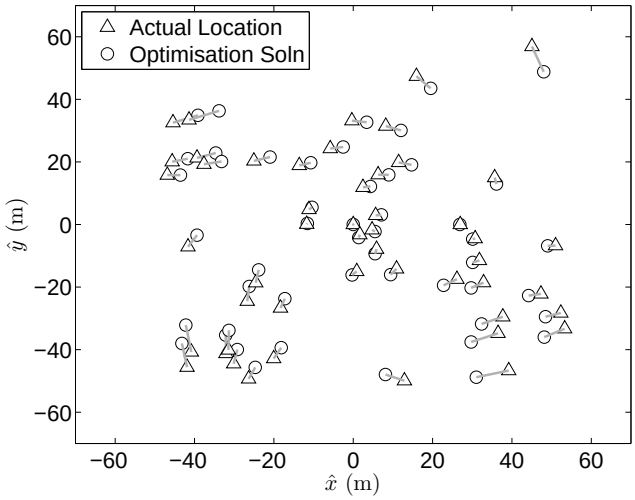

Figure 1. Example 1 - Synthetic relocation of 50 earthquakes in 2D using all constraints with noise $\bar{\sigma}_N = 0.02$. Actual and optimization event locations are shown in triangles and circles, respectively.

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 common triangulation problems such as individual earthquake location from limited travel time data, or GPS positioning with few satellites. Our third observation is that the location errors are circular, 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

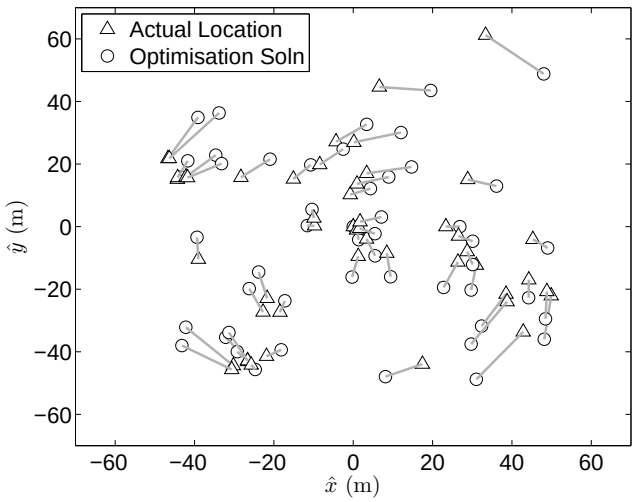

Figure 2. Example 2 - Synthetic relocation of 50 earthquakes in 2D using all constraints with noise $\bar{\sigma}_N = 2\sigma_{1to5Hz}(\delta_t)$. Actual and optimization event locations are shown in triangles and circles, respectively.

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 removed by the constraints imposed on the events (see equations (14) to (16)[¶]) we observe that in practice the presence of noise allows the rotational non-uniqueness to reappear. This is because the 'easiest' way data noise can propagate is into the direction which is least constrained by the data. The same phenomena is observed in linear inversion where noise creates large spurious model changes in directions of the eigenvectors with the smallest singular values (**?**). Unfortunately, there is no obvious way to overcome this issue when using coda waves alone since the constraints are based on separation. Fortunately however, combining coda waves with measurements of travel times alleviates this problem and facilitate the removal of a local coordinate system altogether (see Section **??**).

Despite our observations about the error structure 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.

3.2 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. As with the above examples, we undertake the optimization with 25 randomly chosen starting locations.

Figures 3(a) and (b) illustrates the maximum Δ_{max} (top) and mean Δ_{μ} (middle) of the coordinate error as a function of increasing linkage. We show the statistics for the 'best' optimization solution (black) and for the solution space when all 25 optimizations are considered (gray). In the latter 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 each 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.

[¶] equation 17 is needed in 3D only

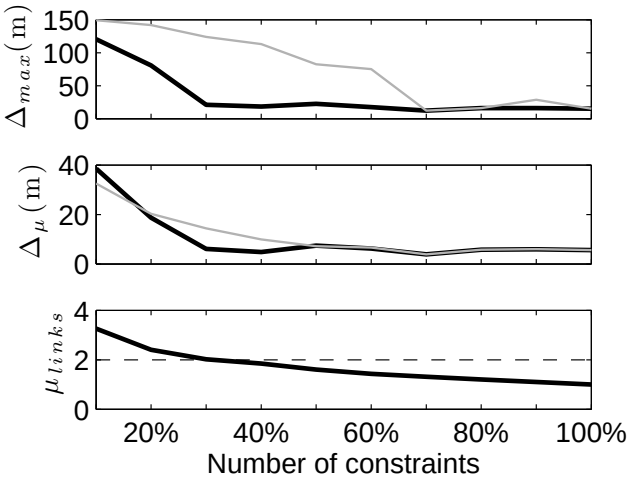

Figure 3. Example 3 - Statistical measures of error in the optimization solutions for the 2D synthetic cases when all (gray) and best (black) results are considered. The statistics Δ_{max} and Δ_{μ} are the maximum and mean coordinate error, respectively. The bottom subplot shows the average minimum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

Some optimizations fail to converge after 1200 iterations when the linkage is 60% or lower. All optimizations fail when then 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. **?**) show that gradient damping 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.

3.3 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 $v = 3300\,\mathrm{ms}^{-1}$ 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 (18) 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 $\bar{\sigma}_N = \sigma_{1to5Hz}$ 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.

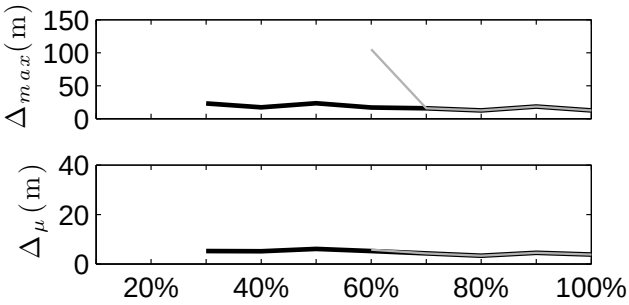

Figure 4. Example 4 - Statistical measures of error in the optimization solutions for the 3D synthetic cases when all (gray) and best (black) results are considered. The statistics Δ_{max} and Δ_{μ} are the maximum and mean coordinate error, respectively. The absence of the gray and black lines below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

When 70% of the direct constraints are considered all optimization results (gray) are consistent with the best solution (black). 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 lines below 60% and 30% of the constraints, 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.

3.3.1 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 influenced by the noise on CWI estimates with greater $\bar{\sigma}_N$ leading to larger errors in the solutions. When 70% or more of the direct branches are used the optimiser 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. As the number of links decrease below 30% the best solution fails to converge.

4 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 (**??**)[See also Data and Resources Section]. These events are chosen for three reasons. Firstly, they are recorded by a large number of stations (Fig. 5) and therefore lend themselves to accurate

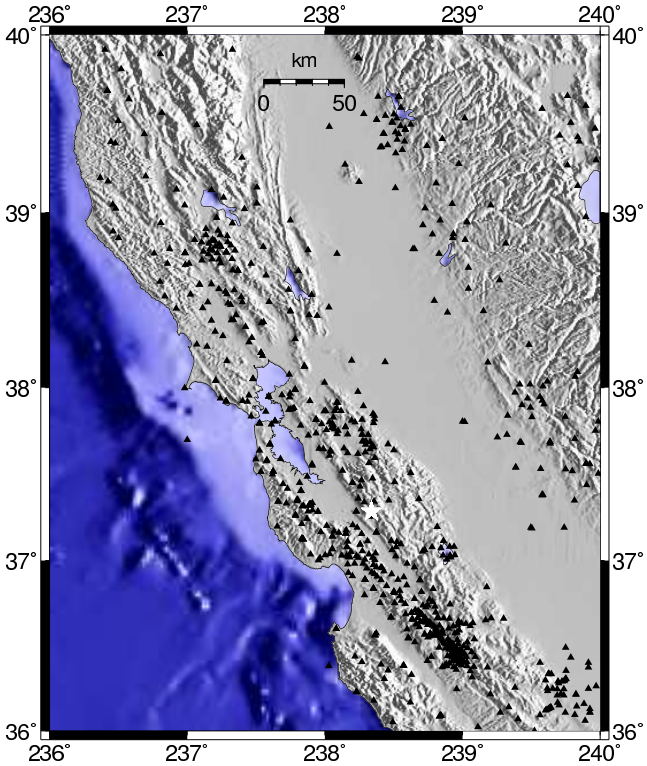

Figure 5. Elevation in California showing location of the Calaveras cluster (white star) and 805 seismic stations (black triangles).

travel 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 meters making them ideal for application of CWI. Finally, Calaveras earthquakes have been well researched with several studies having relocated events in the region (**???**). The relocations in this paper are sorted into four examples summarised in Table 1.

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

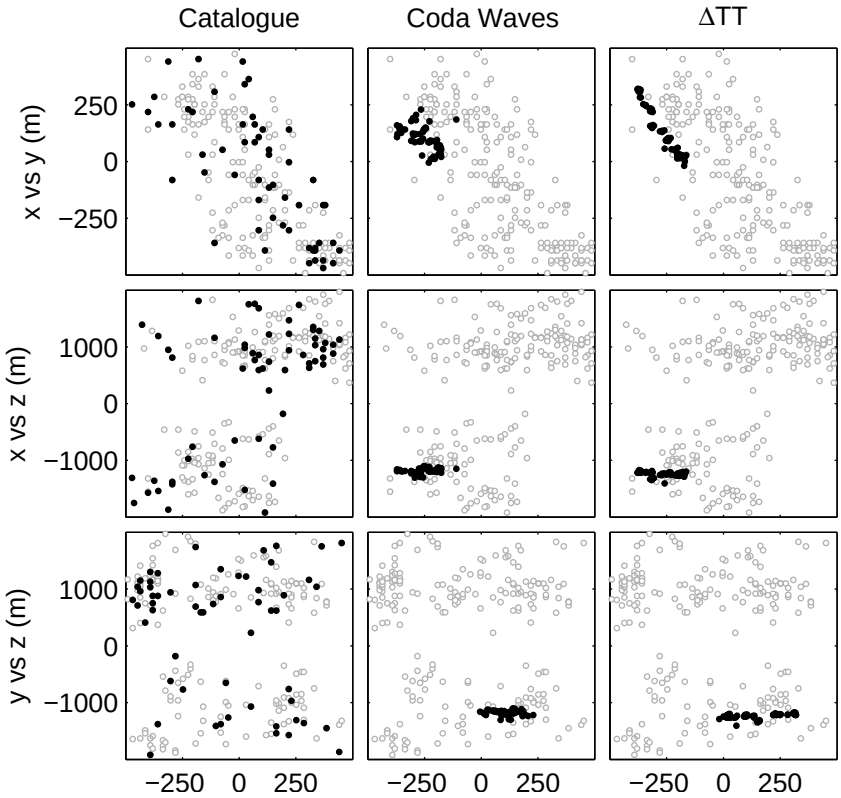

Figure 6. Example 5 - Comparison of relative earthquake locations using three different methods: catalogue location (column 1), CWI (column 2) and hypoDD (column 3). Note that in the case of the hypoDD and CWI inversions we consider only the 68 earthquakes in black, the gray catalogue locations for the remaining 240 (308-68) earthquakes are shown for the purpose of orientation only.

4.1 Example 5 - comparison of CWI, catalogue and hypoDD locations

Figure **??** 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 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 distributed spatially throughout all events.

To apply CWI we download available waveforms from the Northern California Earthquake Data Center (See Data and Resources Section). Unsuitable waveforms are removed using the conditions summarised in Table 5 of **?**). 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 $2.5 \leq t \leq 20$ s and used to create probabilistic constraints on event separation. We introduce the local coordinate system of Section 2 and find the optimum relative locations using Polak-Ribiere optimization.

CWI locations for the 68 events are illustrated in column two of Figure **??**. Catalogue locations (gray) are shown for the remaining 240

earthquakes and are included to ease comparison. The third column of Figure **??** illustrates the locations given by hypoDD with Singular Value Decomposition (SVD), absolute arrival times and cross correlation computed travel time differences.

Global locations cannot be found by CWI alone. 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 event. This arbitrary translation does not change the relative locations. We return to this issue of relative versus actual location in Example 7 by introducing a combined travel 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 results suggest that the coda 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 can be combined (Examples 7 and 8).

4.2 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 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 **??** and the order of removal explained in Table **??**.

CWI locations are illustrated in Figure **??** for the inversions with 7, 5, 4, 3, 2 and 1 station. We observe a high level of consistency between these 6 inversions and the locations shown in Figure **??** (column 2) when all stations are considered. That is, the coda wave approach is self-consistent regardless of the number of stations available, reinforcing our hypothesis that coda waves can constrain location in what would normally be regarded as a poor station network.

Figure **??** illustrates the hypoDD inversion results for seven, five and four stations. The travel 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 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 **??** (column 3) is less evident.

As the number of stations are reduced both the CWI and hypoDD techniques are not able to re-locate all events. To use the coda waves

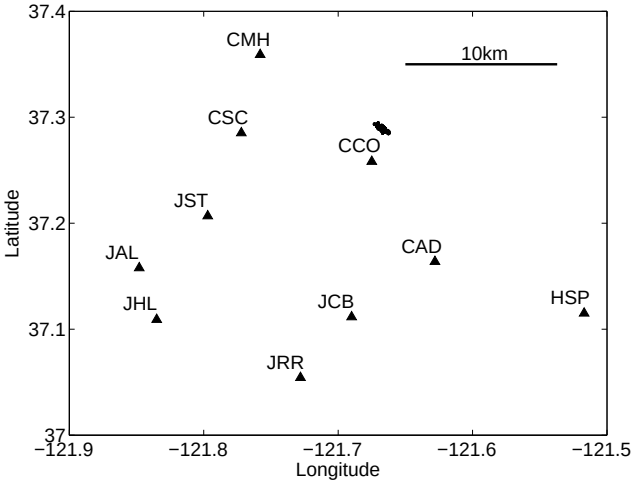

Figure 7. Location of the 10 stations (triangles) used to relocate the Calaveras events in Example 6 to 8. Stations are removed one at a time according to the order in Table **??** and the events relocated. The events are indicated with black circles.

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 row of Figure **??** 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 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 struggles to locate all events as the number of stations is reduced. In the case of hypoDD an event can

Table 2. Stations considered when exploring the impact of reduced station coverage.

Number of Stations	Station Names
10	CCO, JCB, JST, CMH, HSP, JAL, CSC, JST, CAD, JHL, JRR
9	CCO, JCB, JST, CMH, HSP, JAL, CSC, JST, CAD, JHL
8	CCO, JCB, JST, CMH, HSP, JAL, CSC, JST, CAD
7	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

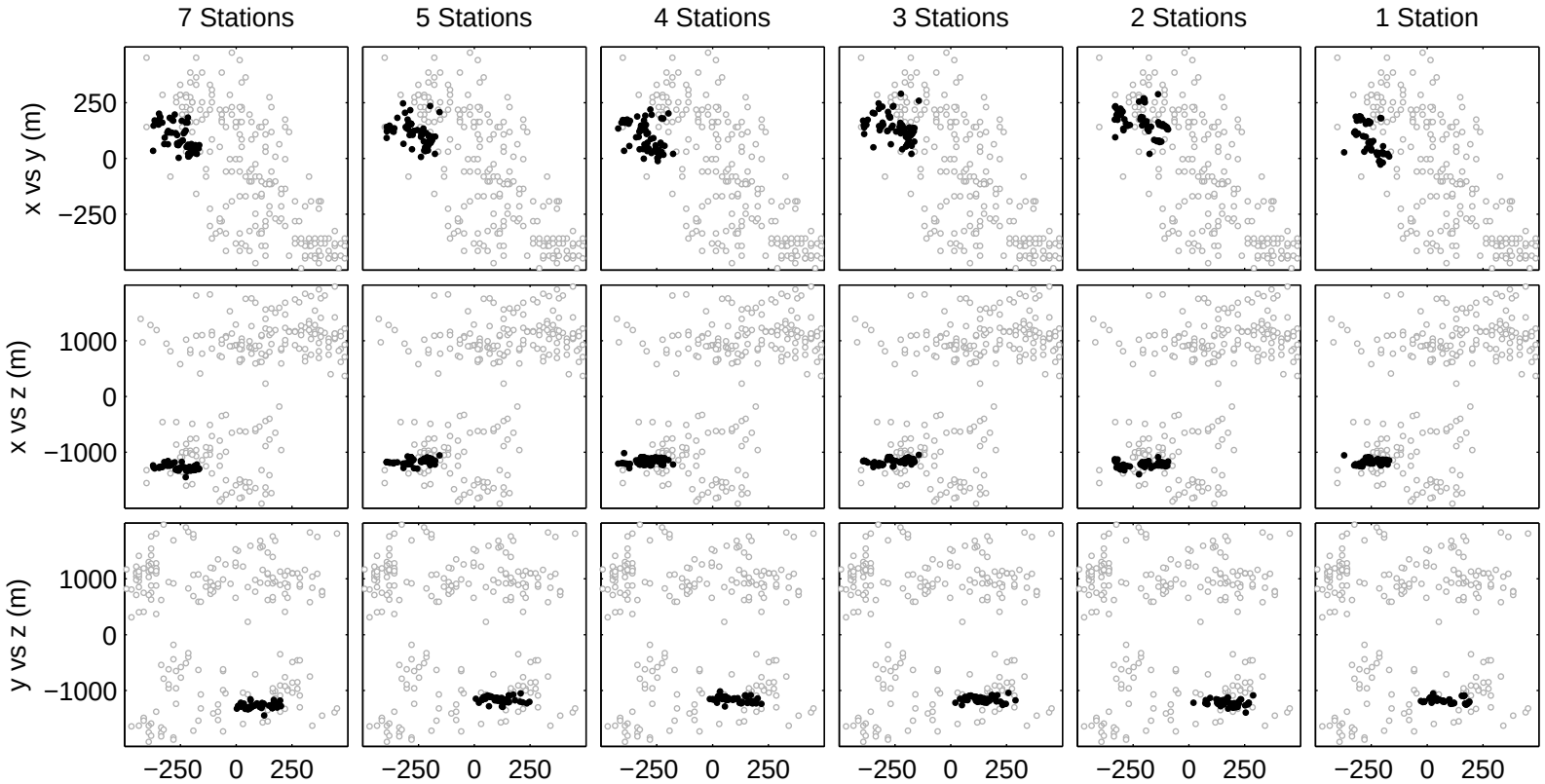

Figure 8. Example 6 - CWI relative locations with reduced stations.

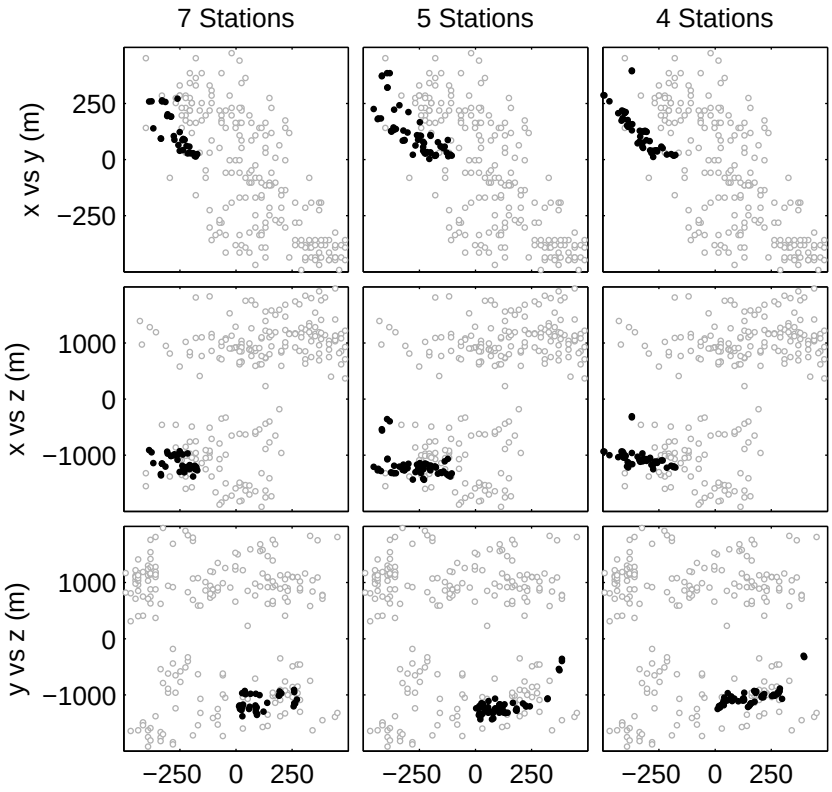

Figure 9. Example 6 - HypoDD (SVD) relative locations with reduced stations.

be identified as unconstrainable in one of two stages. Firstly, the data are analysed to ensure that there exists travel time differences for each event and at least one other earthquake. This is analogous to the situation for the coda wave technique. The hypoDD program also has a secondary identification phase in which events that can not be located sufficiently are rejected during the inversion. This process is presumably related to the iterative removal of outliers described by **?**). The top row of Figure **??** shows that the number of events re-located by hypoDD fluctuates between 63 and 28 earthquakes for 10 to 4 stations and it demonstrates that the number of events located by hypoDD is less than the number located by CWI in all cases.

The remaining rows of Figure **??** illustrate a statistical comparison of the CWI and hypoDD reduced station locations 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.

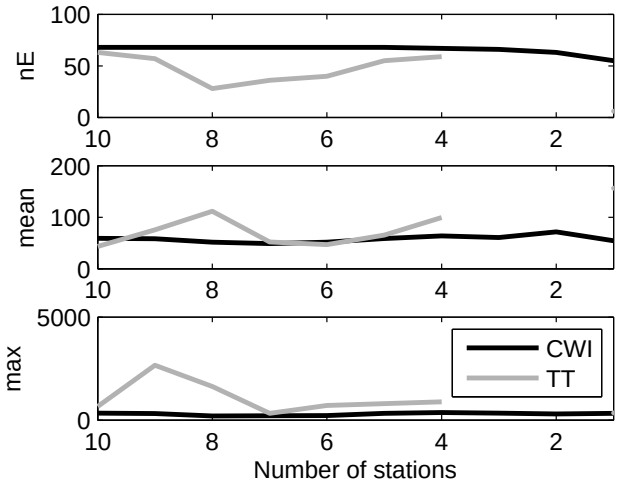

Figure 10. Example 6 - Statistics on coordinate differences for reduced station inversions. Differences are computed between the inversion results (CWI and hypoDD) and the complete hypoDD locations for all 308 events. The top subplot illustrates the number of constrainable events,nE in the CWI and hypoDD inversions as a function of the stations considered.

5 THEORY FOR COMBINING TRAVEL 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 travel time data.

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

$$P(\mathbf{e}_p|\Delta_{TT})\frac{1}{(2\pi)^{\frac{3}{2}}\sqrt{|\Sigma|}} \times \exp\left(-\frac{1}{2}\left([\mathbf{e}_p-\mu_{\mathbf{e}_p}]^T\Sigma^{-1}[\mathbf{e}_p-\mu_{\mathbf{e}_p}]\right)\right), \tag{19}$$

where

$$\mathbf{e}_p = (x_p, y_p, z_p)^T \tag{20}$$

is the location of event p,

$$\mu_{\mathbf{e}_p} = (\mu_{x_p}, \mu_{y_p}, \mu_{z_p})^T \tag{21}$$

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

$$\Sigma = \begin{pmatrix} \sigma^2_{x_p} & 0 & 0 \\ 0 & \sigma^2_{y_p} & 0 \\ 0 & 0 & \sigma^2_{z_p} \end{pmatrix} \tag{22}$$

is the covariance matrix. In this paper 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 (**?**).

We pose the location problem using the negative log likelihood

$$\begin{aligned} L(\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_1, \mathbf{e}_n) = &- \textstyle\sum_{i=1}^{n} ln\left[P(\mathbf{e}_i|\Delta_{TT})\right] \\ &- \textstyle\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} ln\left[P(\delta_{CWIN}|\mathbf{e}_i, \mathbf{e}_j)\right], \end{aligned} \tag{23}$$

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

$$\sum_{i=1}^{n} ln\left[P(\mathbf{e}_i|\Delta_{TT})\right] \tag{24}$$

incorporates the travel time constraints and

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} ln\left[P(\delta_{CWIN}|\mathbf{e}_i, \mathbf{e}_j)\right] \tag{25}$$

the coda waves.

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

$$\begin{aligned} \frac{\partial L}{\partial x_p} = &- \frac{\partial ln\left[P(\mathbf{e}_p|t_{DD})\right]}{\partial x_p} - \textstyle\sum_{i=p+1}^{N} \frac{\partial \ln\left[P(\delta_{CWIN}|\mathbf{e}_p, \mathbf{e}_i)\right]}{\partial x_p} \\ &- \textstyle\sum_{j=1}^{p-1} \frac{\partial \ln\left[P(\delta_{CWIN}|\mathbf{e}_j, \mathbf{e}_p)\right]}{\partial x_p} \end{aligned} \tag{26}$$

where

$$\sum_{i=p+1}^{N} \frac{\partial \ln\left[P(\delta_{CWIN}|\mathbf{e}_p, \mathbf{e}_i)\right]}{\partial x_p} \tag{27}$$

and

$$\sum_{j=1}^{p-1} \frac{\partial \ln\left[P(\delta_{CWIN}|\mathbf{e}_j, \mathbf{e}_p)\right]}{\partial x_p} \tag{28}$$

are defined in Appendix **??** and

$$\begin{aligned} \frac{\partial ln\left[P(\mathbf{e}_p|t_{DD})\right]}{\partial x_p} = &-\tfrac{1}{2}[1,0,0]^T \Sigma^{-1}[\mathbf{e}_p - \mu_{\mathbf{e}_p}] \\ &-\tfrac{1}{2}[\mathbf{e}_p - \mu_{\mathbf{e}_p}]^T \Sigma^{-1}[1,0,0]. \end{aligned} \tag{29}$$

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

$$\begin{aligned} \frac{\partial ln\left[P(\mathbf{e}_p|t_{DD})\right]}{\partial y_p} = &-\tfrac{1}{2}[0,1,0]^T \Sigma^{-1}[\mathbf{e}_p - \mu_{\mathbf{e}_p}] \\ &-\tfrac{1}{2}[\mathbf{e}_p - \mu_{\mathbf{e}_p}]^T \Sigma^{-1}[0,1,0] \end{aligned} \tag{30}$$

and

$$\frac{\partial ln\left[P(\mathbf{e}_p|t_{DD})\right]}{\partial z_p} = -\tfrac{1}{2}[0,0,1]^T\Sigma^{-1}[\mathbf{e}_p - \mu_{\mathbf{e}_p}] \\ -\tfrac{1}{2}[\mathbf{e}_p - \mu_{\mathbf{e}_p}]^T\Sigma^{-1}[0,0,1]. \tag{31}$$

Combining the travel 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 constraints in the form of equation (**??**) resolve the inherent non-uniqueness 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 to consider multiple randomly chosen starting locations.

5.1 Example 7 - Combining travel time and CWI constraints

Figure **??** illustrates the earthquake locations obtained when we combine the travel time and coda wave data using all data (left) and five stations (right). The linear features observed in the original hypoDD inversions (see Fig. **??**) are evident in both cases. However, the coda waves introduce a scatter around these streaks. That is, the locations in figure **??** result from a trade-off between hypoDD's desire to place the events on linear features and the 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.

5.2 Example 8 - Combining CWI and travel times when the travel 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 (**??**). 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 data when the temporary network is in-situ and using 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 by the modified Omori formula

$$N(t) = K\frac{c^{1-p} + (t+c)^{1-p}}{p-1} \tag{32}$$

(**?**). 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, **?**) obtained maximum likelihood estimates for K, p and c of

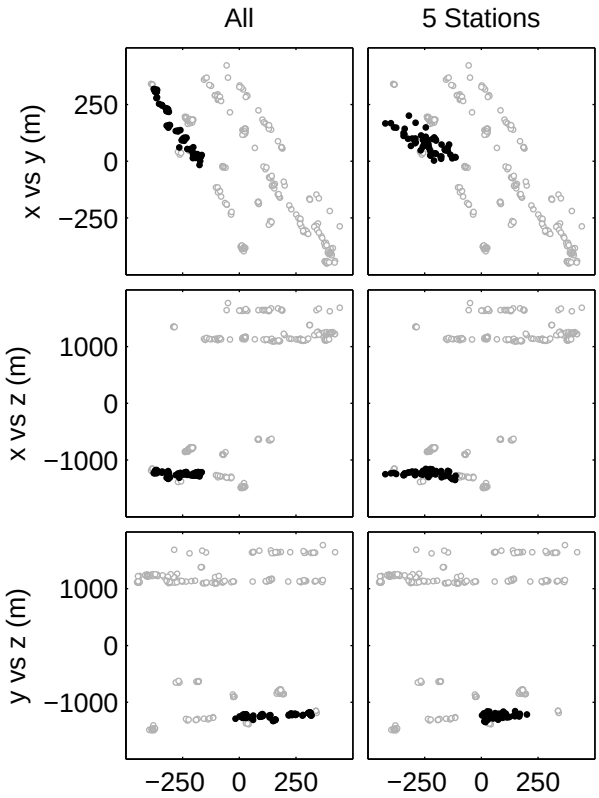

Figure 11. Example 7 - Combined HypoDD (SVD) and CWI relative locations using data form all stations (left) and 5 stations (right).

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,

$$\frac{N(150+4)-N(4)}{N(1000)} = \frac{2257-934}{2626} \approx 0.5. \tag{33}$$

This idea is illustrated in Figure **??** which shows the best fitting Omori Formula separated into segments before (gray), during (black) and after (gray) 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 constraints on half (34) of the earthquakes and incorporate coda wave data from a single station for all 68 earthquakes. The combined inversion is shown in column 1 of Figure **??**. The inversion result is similar to the combined inversion when all travel time data is incorporated (see Fig. **??**). The slight increase in scatter observed here can be explained by the events with no travel time constraints and the tendency of the coda to push events away from streaks.

Remarkably, the combined coda wave and travel time inversion locates all 68 earthquakes to an accuracy similar to the inversions with all data. In contrast when travel time data is used alone it is only possible to locate the 34 events recorded by the pseudo temporary deployment.

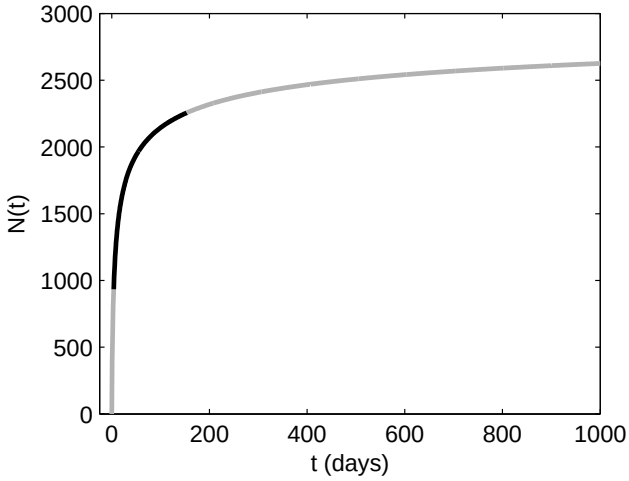

Figure 12. Cumulative number of aftershocks for the Hokkaido-Nansei-Oki, Japan $M_s = 7.8$ earthquake of 12 July 1993 using equation (**??**). The leftmost gray, middle black and rightmost gray lines signify aftershocks occurring before, during and after the deployment of a pseudo temporary array installed 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.

This ability of 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.

6 DISCUSSION AND CONCLUSIONS

Coda wave interferometry is an emerging technique for constraining earthquake location. The technique relies on the interference between coda waves of closely located events and is hence useful for studying earthquake clusters and/or aftershock sequences. Coda wave constraints are independent of travel 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 **??** and **??**.

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 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 improved our understanding of intraplate seismicity and earthquake hazard.

Another potential application of CWI is in the area of hydraulic fracturing such as hot rock geothermal projects and/or petroleum reservoir engineering. Monitoring pumping induced micro earthquakes is a key step in understanding the migration of fluids in such reservoirs.

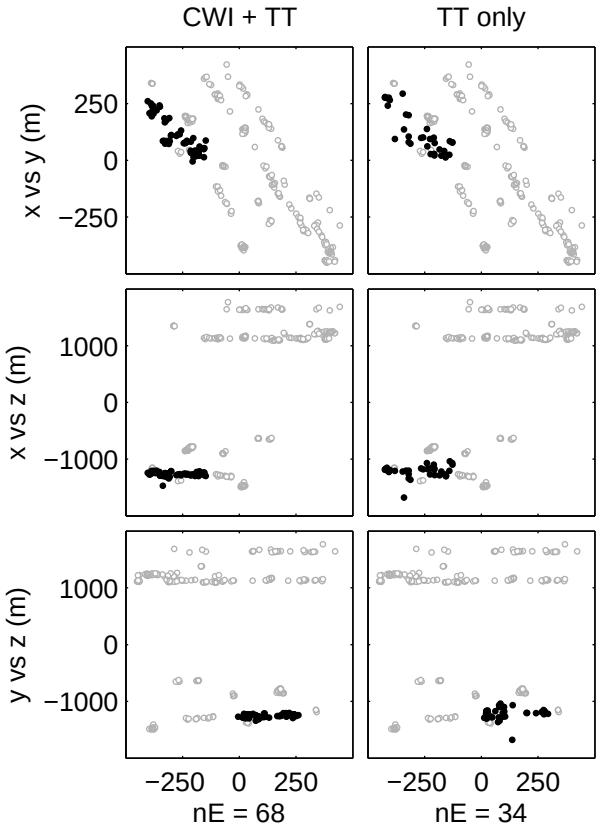

Figure 13. Example 8 - Mimicking the deployment of a temporary network by ignoring data from all but station CCO for 50% (or 34) of the events. Relative locations are shown for the combined CWI and travel time inversion (left) and the inversion with travel times only (right) Only by combining the data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

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.

7 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. 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.

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 well as comments and advice received from Felix Waldhauser on the this and earlier research on the application of CWI to earthquake location.

REFERENCES

Ake, J., O'Connell, D., & Block, L., 2005. Deep-injection and closely monitored induced seismicity at Paradox Valley, Colorado, *Bulletin of the Seismological Society of America*, **95**(2), 664–683.

Aki, K., 1969. Analysis of the seismic coda of local earthquakes as scattered waves, *Journal of Geophysical Research*, **74**(2), 615–631.

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

Bondár, I., Myers, S. C., Engdahl, E. R., & Bergman, E. A., 2004. Epicentre accuracy based on seismic network criteria, *Geophysical Journal International*, **156**, 483–496.

Bowman, J. R., Gibson, G., & Jones, T., 1990. Aftershocks of the 1988 January 22 Tennant Creek, Australia intraplate earthquakes: evidence for a complex thrust-fault geometry, *Geophysical Journal International*, **100**, 87–97.

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

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

Deichmann, N. & Garcia-Fernandez, M., 1992. Rupture geometry from high-precision relative hypocentre locations of microearthquake clusters, *Geophysical Journal International*, **110**, 501–517.

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

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

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

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

Gutenberg, B., 1945. Amplitudes of surface waves and magnitudes of shallow earthquakes, *Bulletin of the Seismological Society of America*, **35**, 3–12.

Ito, A., 1985. High resolution relative hypocenters of similar earthquakes by cross-spectral analysis method, *Journal of Physics of the Earth*, **33**, 279–294.

Kennett, B. L. N., Engdahl, E. R., & Buland, R., 1995. Constraints on seismic velocities in the Earth from traveltimes, *Geophysical Journal International*, **122**, 108–124.

Kennett, B. L. N., Fishwick, S., & Heintz, M., 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, *Bulletin of the Seismological Society of America*, **88**(5), 1127–1143.

Leonard, M., 2002. The Burakin WA earthquake sequence Sept 2000 – June 2002, in *Total Risk Management in the Privatised Era*, vol. 10th of **Australian Earthquake Engineering Society Conference**, pp. 22(1)–22(5), AEES, University of Adelaide.

Nadeau, R. M. & McEvilly, T. V., 1997. Seismological studies at Parkfield V: Characteristic microearthquake sequences as fault–zone drilling targets, *Bulletin of the Seismological Society of America*, **87**(6), 1463–1472.

Pavlis, G. L., 1992. Appraising relative earthquake location errors, *Bulletin of the Seismological Society of America*, **82**(2), 836–859.

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

Richter, C. F., 1935. An instrumental earthquake magnitude scale, *Bulletin of the Seismological Society of America*, **25**(1), 1–32.

Robinson, D., Dhu, T., & Schneider, J., 2006. Practical probabilistic seismic risk analysis: A demonstration of capability, *Seismological Research Letters*, **77**(4), 452–458.

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

Robinson, D. J., Snieder, R., & Sambridge, M., 2007b. Using coda wave interferometry for estimating the variation in source mechanism between double couple events, *Journal of Geophysical Research*, **112**, B12302, doi:10.1029/2007JB004925.

Robinson, D. J., Sambridge, M., & Snieder, R., 2011. A probabilistic approach for estimating the separation between a pair of earthquakes directly from their coda waves, *Journal of Geophysical Research*, **B04309**, 1–17.

Rubin, A. M., 2002. Aftershocks of microearthquakes as probes of the mechanics of rupture, *Journal of Geophysical Research*, **107**(B7,2142), 10.1029/2001JB000496.

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

Schaff, D. P., Bokelmann, G. H. R., & Beroza, G. C., 2002. High-resolution image of Calaveras Fault seismicity, *Journal of Geophysical Research*, **107**(B9), 2186, doi:10.1029/2001JB000633.

Shearer, P., Hauksson, E., & Lin, G., 2005. Southern California hypocenter relocation with waveform cross-correlation, Part 2: Results using source-specific station terms and cluster analysis, *Bulletin of the Seismological Society of America*, **95**(3), 904–915. doi:10.1785/0120040168.

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, chap. 50, pp.

823–825, eds Lee, W. H., Kanamori, H., Jennings, P. C., & Kisslinger, C., Academic Press, London.

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

Snieder, R., 2006. The theory of coda wave interferometry, *Pure and Applied Geophysics*, **163**, 455–473.

Snieder, R. & Vrijlandt, M., 2005. Constraining the source separation with coda wave interferometry: Theory and application to earthquake doublets in the Hayward Fault, California, *Journal of Geophysical Research*, **110**(B04301), doi:10.1029/2004JB003317.

Spencer, C. & Gubbins, D., 1980. Travel-time inversion for simultaneous earthquake location and velocity structure determination in laterally varying media, *Geophysical Journal of the Royal Astronomical Society*, **63**, 95–116.

Stirling, M. W., McVerry, G. H., & Berrryman, K. R., 2002. A new seismic hazard model for New Zealand, *Bulletin of the Seismological Society of America*, **92**(5), 1878–1903.

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

Utsu, T., Ogata, Y., & Matsu'ura, R. S., 1995. The Centenary of the Omori Formula for a decay law of aftershock activity, *Journal of Physics of the Earth*, **43**, 1–33.

VanDecar, J. C. & Snieder, R., 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. & Ellsworth, W. L., 2000. A double–difference earthquake location algorithm: method and application to the northern Hayward Fault, California, *Bulletin of the Seismological Society of America*, **90**(6), 1353–1368.

Waldhauser, F. & Ellsworth, W. L., 2002. Fault structure and mechanics of the Hayward Fault, California, from double–difference earthquake locations, *Journal of Geophysical Research*, **107**(B3), 10.1029/2000JB000084.

Waldhauser, F. & Schaff, D. P., 2008. Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods, *Journal of Geophysical Research*, **133**, B08311, doi10.1029/2007JB005479.

Waldhauser, F., Ellsworth, W. L., & Cole, A., 1999. Slip-parallel lineations on the Northern Hayward Fault, California, *Geophysical Research Letters*, **26**(23), 3525–3528.

APPENDIX A: THE NOISY LIKELIHOOD

The noisy likelihood $P(\widetilde{\delta}_{CWIN}|\widetilde{\delta}_t)$ used in equation (4) is given by

$$P(\widetilde{\delta}_{CWIN}|\widetilde{\delta}_t) = A(\widetilde{\delta}_t)C(\bar{\mu}_N, \bar{\sigma}_N) \times \int_0^\infty B(\widetilde{\delta}_t, \widetilde{\delta}_{CWI})D(\widetilde{\delta}_{CWI}, \bar{\sigma}_N, \bar{\mu}_N)d\widetilde{\delta}_{CWI} \tag{A.1}$$

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

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

Table A1. Coefficients for equations. (**??**) and (**??**).

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

$$B(\widetilde{\delta}_t, \widetilde{\delta}_{CWI}) = e^{\frac{-(\widetilde{\delta}_{CWI}-\mu_1)^2}{2\sigma_1^2}}, \tag{A.3}$$

$$C(\bar{\mu}_N, \bar{\sigma}_N) = \frac{1}{(1-\Phi_{\bar{\mu}_N,\bar{\sigma}_N}(0))\sigma_N\sqrt{2\pi}}, \tag{A.4}$$

$$D(\widetilde{\delta}_{CWI}, \bar{\sigma}_N, \bar{\mu}_N) = e^{\frac{-(\widetilde{\delta}_{CWI}-\bar{\mu}_N)^2}{2\bar{\sigma}_N^2}} \tag{A.5}$$

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

$$\Phi_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{\frac{-(s-\mu)^2}{2\sigma^2}}\,ds \tag{A.6}$$

(**?**). The parameters μ_1 and σ_1 used in equation (**??**) are defined by the expressions

$$\mu_1(\widetilde{\delta}_t) = a_1\frac{a_2\widetilde{\delta}_t^{a_4} + a_3\widetilde{\delta}_t^{a_5}}{a_2\widetilde{\delta}_t^{a_4} + a_3\widetilde{\delta}_t^{a_5} + 1} \tag{A.7}$$

and

$$\sigma_1(\widetilde{\delta}_t) = c + a_1\frac{a_2\widetilde{\delta}_t^{a_4} + a_3\widetilde{\delta}_t^{a_5}}{a_2\widetilde{\delta}_t^{a_4} + a_3\widetilde{\delta}_t^{a_5} + 1} \tag{A.8}$$

with coefficients a_1 to a_5 and c defined in Table **??**. The parameters $\bar{\mu}_N$ and $\bar{\sigma}_N$ used in equation (**??**) 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

$$\begin{aligned} &P(\widetilde{\delta}_{CWIN}|\widetilde{\delta}_t, \widetilde{\delta}_{CWI}) \\ &\qquad = \frac{1}{\left(1-\Phi_{\bar{\mu}_N,\bar{\sigma}_N}(0)\right)\bar{\sigma}_N\sqrt{2\pi}} e^{\frac{-(\widetilde{\delta}_{CWIN}-\bar{\mu}_N)^2}{2\bar{\sigma}_N^2}} \end{aligned} \tag{A.9}$$

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

APPENDIX B: DERIVATIVES

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

$$\frac{\partial L}{\partial \hat{x}_1}, \frac{\partial L}{\partial \hat{y}_1}, \frac{\partial L}{\partial \hat{z}_1}, \frac{\partial L}{\partial \hat{x}_2}, \frac{\partial L}{\partial \hat{y}_2}, \frac{\partial L}{\partial \hat{z}_2}, ..., \frac{\partial L}{\partial \hat{x}_N}, \frac{\partial L}{\partial \hat{y}_N}, \frac{\partial L}{\partial \hat{z}_N} \tag{B.1}$$

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 $\mathbf{e}_q = (\hat{x}_q, \hat{y}_q, \hat{z}_q)$

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

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

$$\begin{aligned}
&\frac{\partial \widetilde{\delta_t}}{\partial \hat{x}_p} = \frac{f_{dom}^2(\hat{x}_p - \hat{x}_q)}{v_s^2 \widetilde{\delta_t}}, \frac{\partial \widetilde{\delta_t}}{\partial \hat{y}_p} = \frac{f_{dom}^2(\hat{y}_p - \hat{y}_q)}{v_s^2 \widetilde{\delta_t}}, \\
&\frac{\partial \widetilde{\delta_t}}{\partial \hat{z}_p} = \frac{f_{dom}^2(\hat{z}_p - \hat{z}_q)}{v_s^2 \widetilde{\delta_t}}, \frac{\partial \widetilde{\delta_t}}{\partial \hat{x}_q} = \frac{f_{dom}^2(\hat{x}_q - \hat{x}_p)}{v_s^2 \widetilde{\delta_t}}, \\
&\frac{\partial \widetilde{\delta_t}}{\partial \hat{y}_q} = \frac{f_{dom}^2(\hat{y}_q - \hat{y}_p)}{v_s^2 \widetilde{\delta_t}}, \frac{\partial \widetilde{\delta_t}}{\partial \hat{z}_q} = \frac{f_{dom}^2(\hat{z}_q - \hat{z}_p)}{v_s^2 \widetilde{\delta_t}}.
\end{aligned} \tag{B.3}$$

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

$$\frac{\partial \widetilde{\delta_t}}{\partial \hat{x}_q} = -\frac{\partial \widetilde{\delta_t}}{\partial \hat{x}_p}. \tag{B.4}$$

The chain rule gives

$$\frac{\partial \mu_1}{\partial \hat{x}_p} = \frac{\partial \mu_1}{\partial \widetilde{\delta_t}} \frac{\partial \widetilde{\delta_t}}{\partial \hat{x}_p} \tag{B.5}$$

where differentiating equation (**??**) gives

$$\frac{\partial \mu_1}{\partial \widetilde{\delta_t}} = a_1 \frac{a_2 a_4 \widetilde{\delta_t}^{a_4 - 1} + a_3 a_5 \widetilde{\delta_t}^{a_5 - 1}}{\left(a_2 \widetilde{\delta_t}^{a_4} + a_3 \widetilde{\delta_t}^{a_5} + 1\right)^2}. \tag{B.6}$$

Similarly, we have

$$\frac{\partial \sigma_1}{\partial \hat{x}_p} = \frac{\partial \sigma_1}{\partial \widetilde{\delta_t}} \frac{\partial \widetilde{\delta_t}}{\partial \hat{x}_p} \tag{B.7}$$

where $\frac{\partial \sigma_1}{\partial \widetilde{\delta_t}}$ has the identical form as **??** with different constants $a_1, a_2, ..., a_5$ (see table **??**).

The cumulative Gaussian distribution function **??** is

$$\Phi_{\mu_1, \sigma_1}(0) = \frac{1}{\sigma_1 \sqrt{2\pi}} \int_{-\infty}^{0} e^{\frac{-(s-\mu_1)^2}{2\sigma_1^2}} ds \tag{B.8}$$

which has derivative

$$\frac{\partial \Phi_{\mu_1, \sigma_1}(0)}{\partial \hat{x}_p} = \frac{\sigma_1 \int_{-\infty}^{0} \frac{\partial g}{\partial \hat{x}_p} e^g ds - \frac{\partial \sigma_1}{\partial \hat{x}_p} \int_{-\infty}^{0} e^g ds}{\sigma_1^2 \sqrt{2\pi}} \tag{B.9}$$

where

$$g = \frac{-(s-\mu_1)^2}{2\sigma_1^2} \tag{B.10}$$

and

$$\frac{\partial g}{\partial \hat{x}_p} = \frac{4\sigma_1^2(s-\mu_1)\frac{\partial \mu_1}{\partial \hat{x}_p} + 4\sigma_1 \frac{\partial \sigma_1}{\partial \hat{x}_p}(s-\mu_1)^2}{4\sigma_1^4}. \tag{B.11}$$

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

$$\frac{\partial A}{\partial \hat{x}_p} = -\frac{-\frac{\partial \Phi_{\mu_1,\sigma_1}(0)}{\partial \hat{x}_p}\sigma_1 + (1-\Phi_{\mu_1,\sigma_1}(0))\,\frac{\partial \sigma_1}{\partial \hat{x}_p}}{(1-\Phi_{\mu_1,\sigma_1}(0))^2\,\sigma_1^2\sqrt{2\pi}} \tag{B.12}$$

and

$$\frac{\partial B}{\partial \hat{x}_p} = e^h \frac{\partial h}{\partial \hat{x}_p} \tag{B.13}$$

where

$$h = \frac{-(\delta_{CWI}-\mu_1)^2}{2\sigma_1^2} \tag{B.14}$$

and

$$\frac{\partial h}{\partial \hat{x}_p} = \frac{4\sigma_1^2(\delta_{CWI}-\mu_1)\frac{\partial \mu_1}{\partial \hat{x}_p} + 4(\delta_{CWI}-\mu_1)^2\sigma_1 \frac{\partial \sigma_1}{\partial \hat{x}_p}}{4\sigma_1^4}. \tag{B.15}$$

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

$$\begin{aligned} \tfrac{\partial P(\delta_{CWIN}|\widetilde{\delta}_t)}{\partial \hat{x}_p} &= \tfrac{\partial A(\widetilde{\delta}_t)}{\partial \hat{x}_p} C(\bar{\mu}_N, \bar{\sigma}_N) \\ &\times \textstyle\int_0^\infty B(\widetilde{\delta}_t, \widetilde{\delta}_{CWI}) D(\widetilde{\delta}_{CWI}, \bar{\sigma}_N, \bar{\mu}_N) d\widetilde{\delta}_{CWI} \\ &+ A(\widetilde{\delta}_t) C(\bar{\mu}_N, \bar{\sigma}_N) \\ &\times \textstyle\int_0^\infty \tfrac{\partial B(\widetilde{\delta}_t, \widetilde{\delta}_{CWI})}{\partial \hat{x}_p} D(\widetilde{\delta}_{CWI}, \bar{\sigma}_N, \bar{\mu}_N) d\widetilde{\delta}_{CWI} \end{aligned} \tag{B.16}$$

and for the logarithm we have

$$\frac{\partial \ln\left[P(\delta_{CWIN}|\delta_t)\right]}{\partial \hat{x}_p} = \frac{1}{P(\delta_{CWIN}|\delta_t)} \frac{\partial P(\delta_{CWIN}|\delta_t)}{\partial \hat{x}_p}. \tag{B.17}$$

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

$$\begin{aligned} \tfrac{\partial L(E_1, E_2, \ldots, E_n)}{\partial \hat{x}_p} = \quad & -\textstyle\sum_{i=p+1}^{N} \tfrac{\partial \ln\left[P(\delta_{CWIN}|E_p, E_i)\right]}{\partial \hat{x}_p} \\ & + \textstyle\sum_{j=1}^{p-1} \tfrac{\partial \ln\left[P(\delta_{CWIN}|E_j, E_p)\right]}{\partial \hat{x}_p} \end{aligned} \tag{B.18}$$

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.