Geophys. J. Int. (2010) **200**, 1–14

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 Research Group, 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] *Earthquakes and the Environment Program, NORSAR,Kjeller, Norway*

Received 2010 Month Day; in original form 2010 Month Day

SUMMARY
(Type abstract here)

Key words: Probability distributions; Earthquake source observations; Seismicity and tectonics; Wave scattering and diffraction

1 INTRODUCTION

Earthquake location is important for many applications. It assists our understanding of seismic activity by indicating regions of relatively higher seismicity such as major lithospheric plate boundaries (e.g., Sykes 1967; Isacks et al. 1968; Stein & Klosko 2002) or local active source zones (e.g., Gutenberg & Richter 1944; Giardini et al. 2002). Earthquake locations are required for magnitude determination (e.g., Richter 1935; Gutenberg 1945), computing moment tensors (e.g., Sipkin 2002) and seismological studies of the Earth's interior (e.g., Spencer & Gubbins 1980; Kennett et al. 1995; Curtis & Snieder 2002; Kennett et al. 2004). They are needed to understand strong motion and seismic attenuation (e.g., Toro et al. 1997; Campbell 2003) and model earthquake hazard or risk (e.g., Frankel et al. 2000; Stirling et al. 2002; Robinson et al. 2006). The accuracy required in earthquake location depends on the application. For example, imaging the structure of a fracture system from microseismicity requires greater detail than determining whether a $M_w = 7.5$ earthquake occurs offshore for tsunami warning purposes. This paper focuses on reducing the uncertainty on the locations of a cluster of events, particularly when they are recorded by a small number of stations.

Absolute location describes the location of an earthquake with respect to some 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. Errors in absolute location are typically of the order of several km because they are susceptible to uncertainty in the velocity structure along the entire path between the source and receiver. For example, Shearer (1999) states that location uncertainty in the ISC and PDE catalogues are generally around 25 km horizontally and at least 25 km in depth.[⋆] These locations are created using global seismic networks. Bondár et al. (2004) demonstrate that at the local scale, absolute locations are accurate to within 5 km with a 95% confidence level when local networks meet a number of station related criteria. Such errors are too large for many applications, particularly those focussed on the study of micro seismicity and imaging rupture surfaces from aftershock sequences.

Relative earthquake location involves locating a group of earthquakes with respect to one another and was first introduced by Douglas (1967) who developed the technique commonly known as joint hypocenter determination.[†] In principle, relative locations can be computed by differencing absolute locations. However, Pavlis (1992) shows that inadequate knowledge of velocity structure leads to systematic biases when relative positions are computed in this way. This is confirmed by Richards et al. (2006) who state that inaccuracies in the travel time model are one of two primary error sources in absolute earthquake location. The other error source is erroneous arrival picks. To reduce errors from the travel time model, relative location techniques compute locations directly from travel time differences between two waveforms (e.g., Ito 1985; Got et al. 1994; Nadeau & McEvilly 1997; Waldhauser et al. 1999). By doing so, they remove errors associated with unknown lateral velocity variations outside the local region, because such variations influence all waveforms in the same manner (Shearer 1999).

Reported uncertainties for event locations from relative techniques are around 15 to 75 m in local settings with good station coverage (e.g., Ito 1985; Got et al. 1994; Waldhauser et al. 1999; Waldhauser & Schaff 2008). Here, 'good coverage' implies multiple stations distributed across a broad range of azimuthal directions. Precise relative location techniques have been applied to a diverse range of problems. For example, the location of active fault planes (e.g., Deichmann & Garcia-Fernandez 1992; Got et al. 1994; Waldhauser et al. 1999; Waldhauser & Ellsworth 2002; Shearer et al. 2005), studying rupture mechanics (e.g., Rubin et al. 1999; Rubin 2002b), interpreting magma movement in volcanoes (e.g., Frèmont & Malone 1987) and monitoring pumping induced seismicity (e.g.,

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

[†] Douglas (1967) originally used the term joint epicentre determination. However, he was solving for hypocentre.

Lees 1998; Ake et al. 2005). Poupinet et al. (1984), Bokelmann & Harjes (2000) and Rubin (2002a) apply relative location techniques to identify earthquake doublets, events that have near identical locations. Poupinet et al. (1984) use these doublets to measure temporal variation in crustal velocity, Bokelmann & Harjes (2000) use them to identify systematic temporal variations in seismic anisotropy and Rubin (2002a) describes how they are useful for correcting time-dependent station delays.

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; the travel time difference computed between direct arrivals of two waveforms; or the time differences inferred from time-lagged cross correlation of relatively small windows around the body wave arrivals. In all three cases, the majority of the waveform is discarded. Furthermore, obtaining high accuracy with these techniques requires multiple stations with good azimuthal coverage. In this paper we demonstrate that it is possible to significantly reduce location uncertainty when few stations are available by using more of the waveform.

Coda refers to later arriving waves in the seismogram that arise from scattering (Aki 1969; Snieder 1999, 2006). Typically, the 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. Snieder & Vrijlandt (2005) demonstrate that the coda of two earthquakes can be used to estimate the separation between them. Their technique is based on the interference pattern between the coda waves and is known as coda wave interferometry (CWI). 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 (Robinson et al. 2007). This makes CWI particularly interesting for regions where station density is low such as intraplate areas. Our technique can be used on coda waves alone or in combination with travel times. We demonstrate that the use of coda waves improves the location of a cluster of events when the recording situation is poor and we show that it is possible to obtain accurate relative locations with as little as one station.

2 THEORY

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

$$\delta^2_{CWI} = g(\alpha,\beta)\sigma^2_\tau, \tag{1}$$

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

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

(Snieder & Vrijlandt 2005). The σ_τ in eq. (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 Robinson et al. (submitted) when relating these parameters and we apply a restricted time lag search when evaluating R_{max}. These extensions to the original technique of Snieder & Vrijlandt (2005) increase the range of applicability of CWI by 50%.

Robinson et al. (submitted) 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 representing the forward model between the desired separation $\widetilde{\delta}_t$ and the data $\widetilde{\delta}_{CWIN}$; and $P(\widetilde{\delta}_t)$ is the prior function 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 ($\delta = \delta_{CWIN}$ or δ_t) with respect to dominant wavelength λ_d. In this paper we consider a uniform prior which gives emphasis to the recorded data. The algorithm for computing the noisy likelihood $P(\widetilde{\delta}_{CWIN}|\widetilde{\delta}_t)$ is derived by Robinson et al. (submitted) and summarised in Appendix A. 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 $P(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN})$ and describes the location of multiple events with maximum identifying the most probable combination of locations. More importantly however, the PDF will 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 eq. (6) $\mathbf{d}$ represents observations that constrain the locations. They can be any combination of travel times, geodetic information or CWI separation estimations. 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 $P(\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3|\widetilde{\boldsymbol{\delta}}_{CWIN},\mathbf{t}_{DD})$ and $P(\widetilde{\boldsymbol{\delta}}_{CWIN},\mathbf{t}_{DD}|\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3)$ where $\mathbf{t}_{DD}$ represent travel time differences. In this derivation and in Sec. 2.1 we focus on the constraints imposed by coda waves, whereas in Sec. 4 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 likelihoods we have replaced the conditional term on separation $\widetilde{\delta}_t$ with the locations (e.g. $\mathbf{e}_1$ and $\mathbf{e}_2$). This can be done because knowledge of location

translates to separation.‡ Furthermore, since the pairwise functions are independent the joint likelihood becomes

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

Similarly, we note that earthquake locations are independent and the joint prior becomes

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

Combining eqs. (7) and (8) gives the joint posterior function

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

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

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

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 c in eq. (9) involves finding the value which normalises the posterior function to unit area over all plausible locations. In many applications the constant of proportionality c is not important. 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 require evaluation of a function proportional to the PDF.

By following the same argument we get the posterior function

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

for n events. When evaluating eq. (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 so we work with the negative logarithm

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

or

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

The logarithm improves numerical stability by replacing products with summations. The negative assists with optimisation, many algorithms for which 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, such as translations and rotations, that do not change the distance between events. In defining this coordinate system we fix the first event at the origin,

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

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

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

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

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

and the fourth to

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

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 Sec. 4 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 (Press et al. 1987) to optimise L. The Polak-Ribiere technique is a particular choice of conjugate gradient method. It uses the derivatives of L to guide the optimisation procedure. We derive the derivatives in Appendix B.

2.1 Synthetic experiments

We use synthetic examples in 2D with 50 earthquakes to test the performance of the optimisation routine.

2.2 2D synthetic experiments

Actual locations are created by randomly selecting $\hat{x}$- and $\hat{y}$-coordinates such that $-50 \leq \hat{x} \leq 50$ and $-50 \leq \hat{y} \leq 50$. These are indicated with triangles in Fig. 1. 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 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

$$\bar{\mu}_N = \mu_1\left(\widetilde{\delta}_t\right) \quad (18)$$

using eq. (A.7). This assumption ensures that the sample mean of hypothetical separation estimates is consistent with known CWI biases (Robinson et al. submitted). In example 1 we use $\bar{\sigma}_N = 0.02$ between all event pairs. Application of our optimisation procedure on the hypothetical CWI data yields the circles in Fig. 1. The optimisation does not lead to the correct 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

‡ Note that the reverse can not be said. That is, knowledge of separation between a single event pair does not translate to location but rather places a non-unique constraint on location.

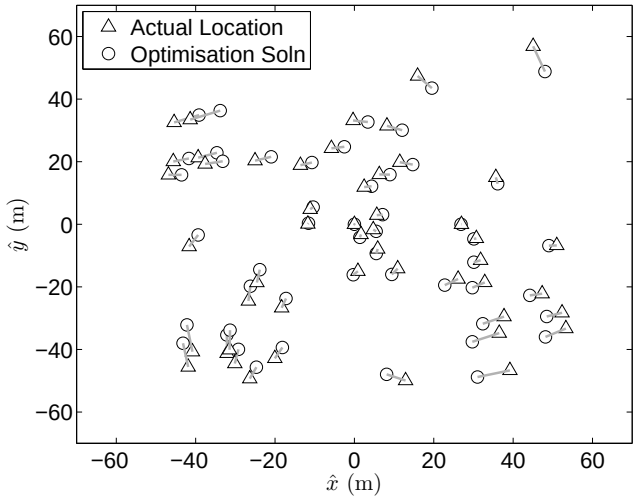

Figure 1. Example 1 - Synthetic relocation of 50 earthquakes in 2D using all constraints with noise $\bar{\sigma}_N = 0.02$. Actual and optimisation event locations are shown in blue and red circles, respectively.

$\bar{\sigma}_N = 0.02$ which for $v_s = 3300\,\mathrm{ms}^{-1}$ and $f_{dom} = 2.5\,\mathrm{Hz}$ corresponds to roughly 25 m.

Robinson et al. (submitted) demonstrates that the noise on CWI estimates is often larger than 0.02 and that it increases with event separation. Consequently, example 1 is overly 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 Fig4(b) of Robinson et al. submitted). Repeating the optimisation leads to the circles in Fig. 2 which have an average coordinate error of 2.8 m.

Conjugate gradient based optimisation 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 optimisation from 25 different randomly chosen starting positions. In this case we observe no differences in the solution for neither examples 1 or 2.

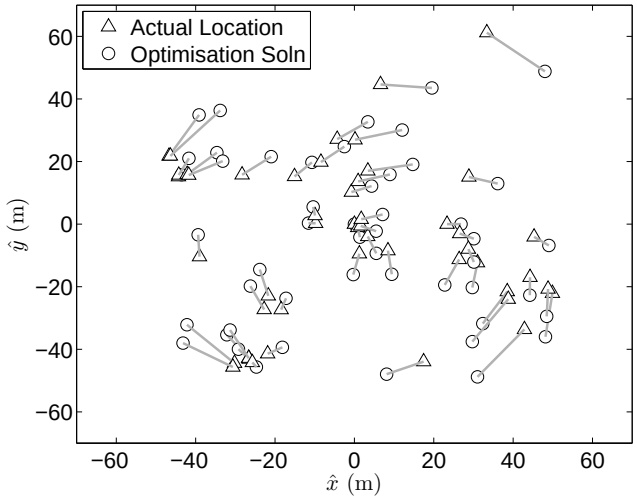

Figure 2. Example 3 - Synthetic relocation of 50 earthquakes in 2D using all constraints with noise $\bar{\sigma}_N = 2\sigma_{1to5Hz}(\delta_t)$. Actual and optimisation event locations are shown in blue and red circles, respectively.

We can draw three observations from the error structure in Figs. 1 and 2. Firstly, we note that the location errors depicted by grey bars in each figure increase between examples 1 and 2 with the introduction of larger noise. Secondly, the errors are larger for events at greater distances from the center. This is because events near the center of the cluster are constrained by links from all angles, whereas those on the outside are moderated by links from a limited number of directions. This observation is analogous to problems associated with poor azimuthal coverage in 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 appear to be 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 increases the strength of this constraint on later events becomes weaker 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 eqs. 14 to 17) we observe that in practice the effects of noise means that the rotational non-uniqueness reappears. This is because the 'easiest' way data noise can propagate is into the direction which is least constrained by the data. It is the same phenomena as in linear inversion where noise creates large spurious model changes in directions of the eigenvectors with the smallest singular values (e.g Aster et al. 2005). 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.

Despite these observations about the error structure however, we gain confidence in the optimisation 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.

2.2.1 The impact of reduced linkage

A set of earthquakes and their corresponding coda wave estimates of separation can be thought of as a network. In this network the earthquakes are nodes and the constraints are branches or links which join them together. 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 less 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 optimisation with 25 randomly chosen starting locations.

In Figs. 3(a) and (b) we examine the performance of the optimisation statistically. In particular, we plot the maximum Δ_{max} and mean Δ_μ of the coordinate difference as a function of increas-

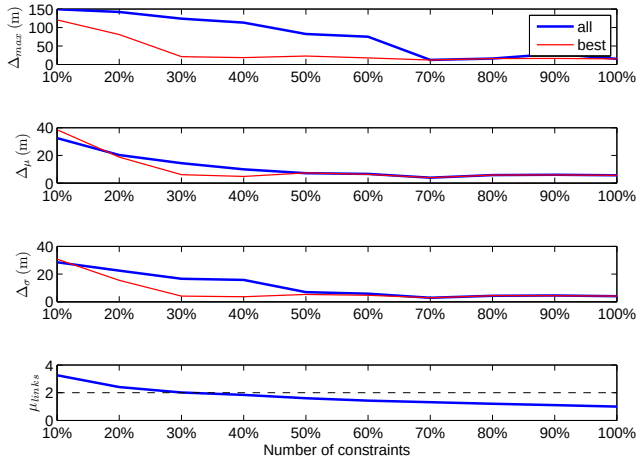

Figure 3. Statistical measures of error in the optimisation solutions for the 2D synthetic cases when all (blue) and best (red) results are considered. The statistics Δ_{max}, Δ_{μ} and Δ_{σ} are the maximum, mean and standard deviation of the coordinate error, respectively. The bottom subplot shows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

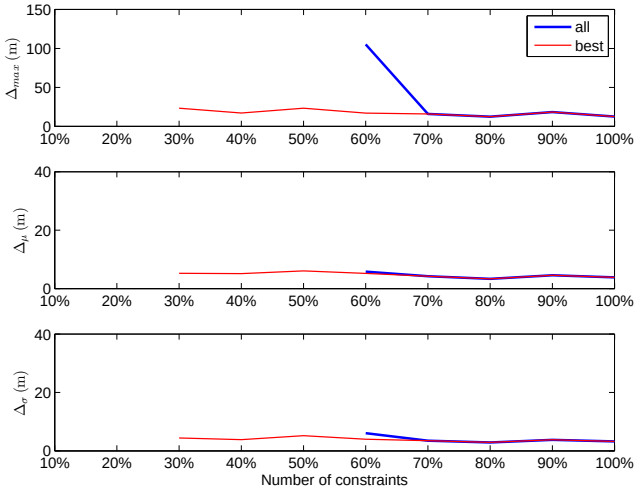

Figure 4. Statistical measures of error in the optimisation solutions for the 3D synthetic cases when all (blue) and best (red) results are considered. The statistics Δ_{max}, Δ_{μ} and Δ_{σ} are the maximum, mean and standard deviation of the coordinate error, respectively. The absence of the blue and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

ing linkage. We show the statistics for the 'best' optimisation solution (black) and for the solution space when all 25 optimisations are considered (grey). In this case the best solution is determined by the set of event locations which lead to the smallest value of L. We observe similar errors in the best solution for all cases when at least 30% of the branches are used. The errors increase drastically 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(c)). Since the average number of branches can be computed in advance it can be used as an indication of the potential stability of the inversion prior to optimisation. An earlier breakdown is observed when all 25 solutions are considered collectively. For example, the maximum coordinate error Δ_{max} is significantly higher than that for the best solution for all cases with 60% of the constraints or less. This observation suggests that the optimisation is susceptible to the presence of local minima and that a range of starting points should be considered.

Interestingly, we note that some of the optimisations chains fail to converge after 1200 iterations when the linkage is 60% or lower and all fail when then linkage is 20% or lower. Despite their failure to converge however, many of the solutions remain relatively close to the actual solutions. 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 taken care of during the iterative process where location information can spread to events which have no direct links. However, the lack of connection through the gradient could prevent convergence in extreme cases, or more likely slow the procedure down. This could explain why some examples do not converge after 1200 iterations. VanDecar & Snieder (1994) 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 optimisation by preconditioning.

2.2.2 3D synthetic examples with reduced linkage

In this section we expand the optimisation routine to 3D and repeat example 3 of the 2D case. We randomly pick a set of actual 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 eq. 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 optimisation using 10%, 20%, ..., 100% of the direct links. In each case we repeat the optimisation 25 times using randomly chosen starting locations. The results are summarised in Fig. 4.

As in the 2D example we observe a consistency of the optimisation procedure when the number of direct links used is high. That is, when 70% of the direct constraints are considered all optimisation results (grey) are consistent with the best solution (black). Furthermore, the best solution appears to constrain the event locations down to 30% of the direct links. There is however, one notable difference between the 3D and 2D results. In 2D the final iteration was typically close to the actual solution when the optimisation chain failed to converge after 1200 iterations. Contrastingly, in 3D the optimisation appears to converge to the correct solution or fail completely, leading to a set locations which in no way resemble the actual solution. This is depicted in Fig. 4 by the absence of the grey 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% or more of the links is encouraging for the potential of coda wave optimisation to constrain earthquake location more generally.

2.2.3 Summary of findings

In summary, the synthetic examples demonstrate the ability of coda wave data to constrain event location using optimisation. The optimisation 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 optimisation becomes less stable and the best solution from 25 random starting locations is required to find the optimal solution. Finally, as the number of links decrease below 30% the best solution also fails to converge to a useful result.

3 RELOCATING EARTHQUAKES ON THE CALAVERAS FAULT

In this section we consider real earthquakes and undertake a complete coda wave analysis beginning with waveform selection, filtering, pairwise application of CWI and optimisation.

- Introduce the Calaveras Fault and provide a brief background of why it is chosen here (Figure 5)
- Discuss Catalogue Locations - hypoDD locations and CWI locations (Figure A4)
  - mention difference between hypoDD inversions when 68 or 308 earthquakes considered but don't show figure (i.e. difference in coordinates for the two solutions is roughly 2 m)
  - provide details on CWI application - waveform selection, filtering etc.
- Discuss comparison between CWI and catalogue locations - first order performance of CWI and clustering
- Discuss comparison between CWI and hypoDD locations - second order result and streaks.

3.1 Dependance on the number of stations

- Explain issues associated with poor station coverage - superimpose the three SWSZ (i.e. Australian) stations onto Figure 5
- Show setup for consideration of fewer stations (Table 1 and Figure 7)
- Show the results (Figures **??** to **??** and Figure A3)
- Discuss the better performance of CWI with fewer stations
- Question/Discuss why hypoDD continues to suggest streaks with 2 or 3 stations - there must be regularization

4 COMBINING TRAVEL TIME AND CWI CONSTRAINTS

- Pose the problem
- Introduce the idea of a travel time based separation PDF with mean event locations from hypoDD outputs and some measure of uncertainty (we will consider three levels of uncertainty as in paper1).
- Explain how to combine the CWI based posterior with these new travel time based separation PDFs
- Show the derivatives of the travel time based separation PDF
- Undertake the combined inversion show/discuss findings (Figure 11)

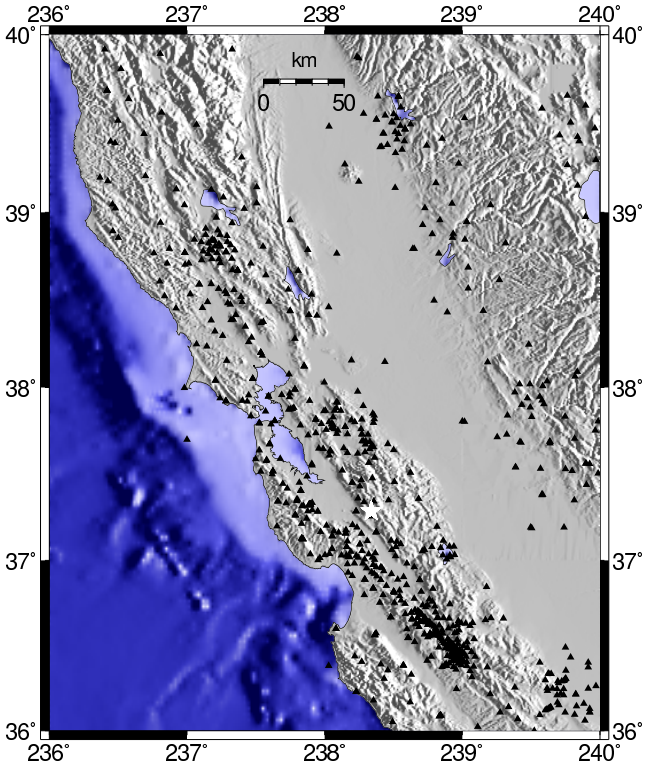

Figure 5. Elevation in California showing location of the 308 event Calaveras cluster (black star) and 805 seismic stations (red triangles).

4.1 Combining CWI and travel times when the travel times constrain a limited number of events

- Explain the idea of a temporary deployment of stations i.e. for aftershocks
- Discuss Omori formula and the M_s7.8 Hokkaido-Nansei-Oki earthquake (Utsu et al. 1995, Figure 12,)
- Show/Discuss results (Figure 13)

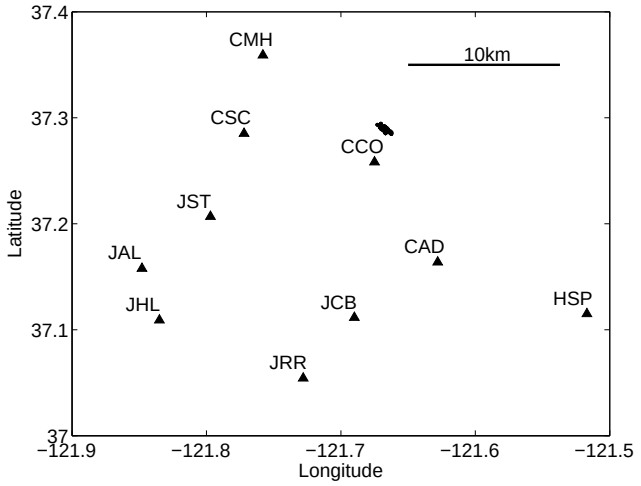

Figure 7. Location of the 10 stations (red triangles) used to relocate the Calaveras events. Stations are removed one at a time according to the order in Table 1 and the events relocated (see Figures **??** to **??**). The 304 Calaveras events are indicated with black circles.

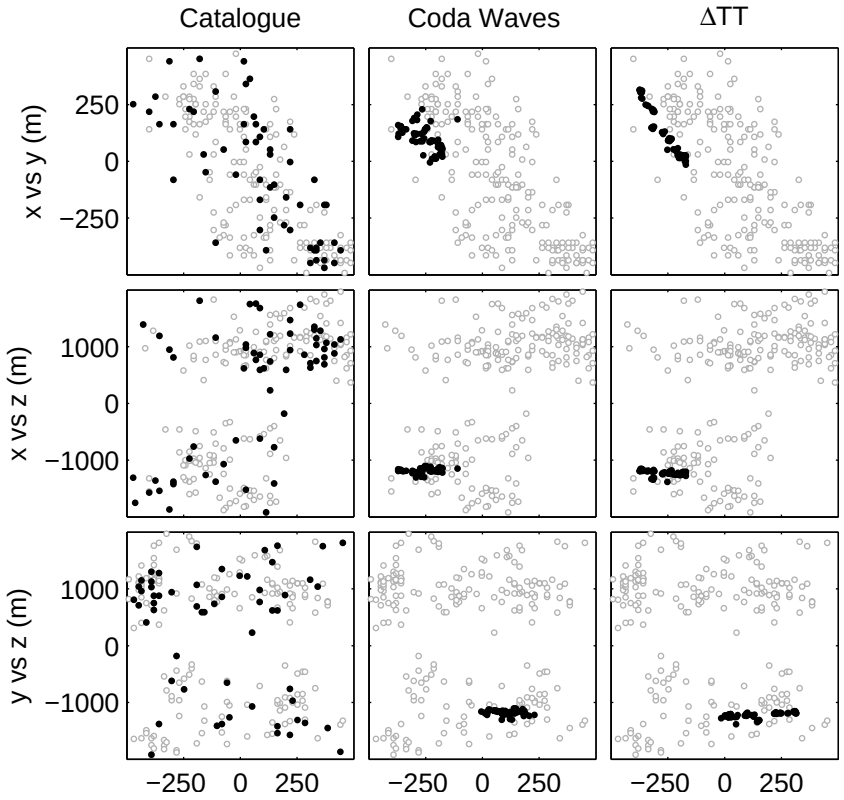

Figure 6. Comparison of earthquake hypocenters using three different methods: catalogue location (column 1), hypoDD with all 308 events (column 2) and CWI example 1 (column 3). Note that in the case of the CWI locations we consider only the 68 earthquakes in red, the blue events are shown for the purpose of orientation only.

Table 1. Stations considered when exploring the effect 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

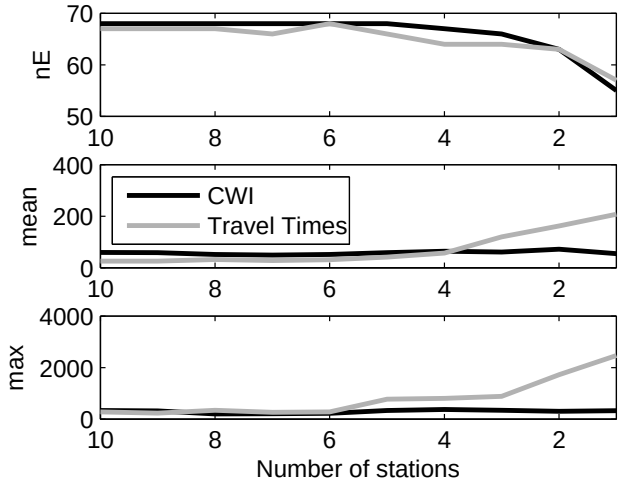

Figure 10. Statistics on coordinate differences for reduced station inversions. Differences are computed between the inversion results (CWI and hypoDD) and the complete hypoDD locations for all 308 events. The top subplot illustrates the number of constrainable events in the CWI and hypoDD inversions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

5 CONCLUSIONS

- Emphasise strengths of the CWI relocation (i.e. with small number of stations, linking temporary deployments with catalogue locations)
- Discuss potential applications e.g. intraplate/downhole/tunnel monitoring etc.

ACKNOWLEDGMENTS

(Text here)

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.

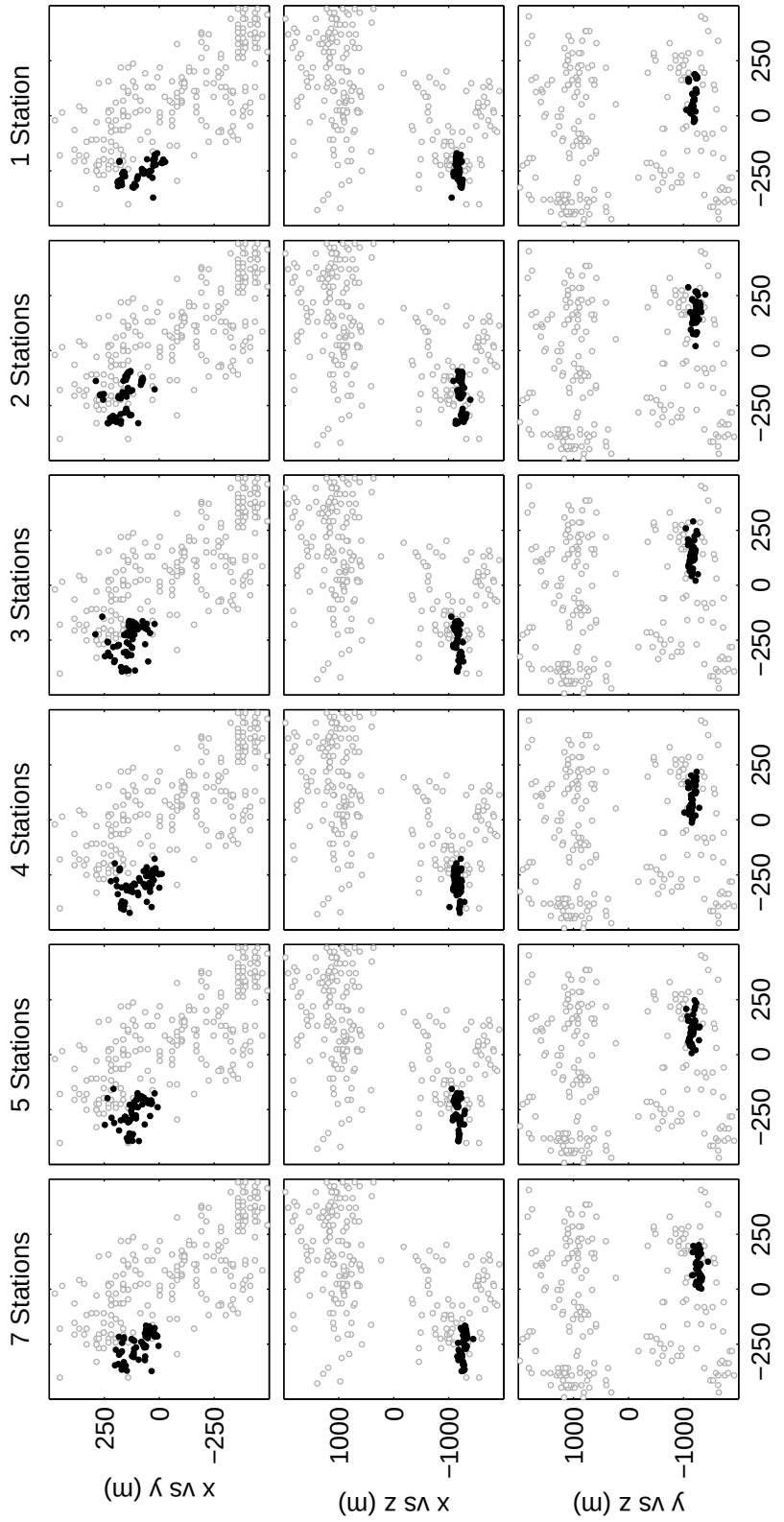
1 Station
2 Stations
3 Stations
4 Stations
5 Stations
7 Stations
x vs y (m)
x vs z (m)
y vs z (m)
250
0
−250
1000
0
−1000

Figure 8. ....

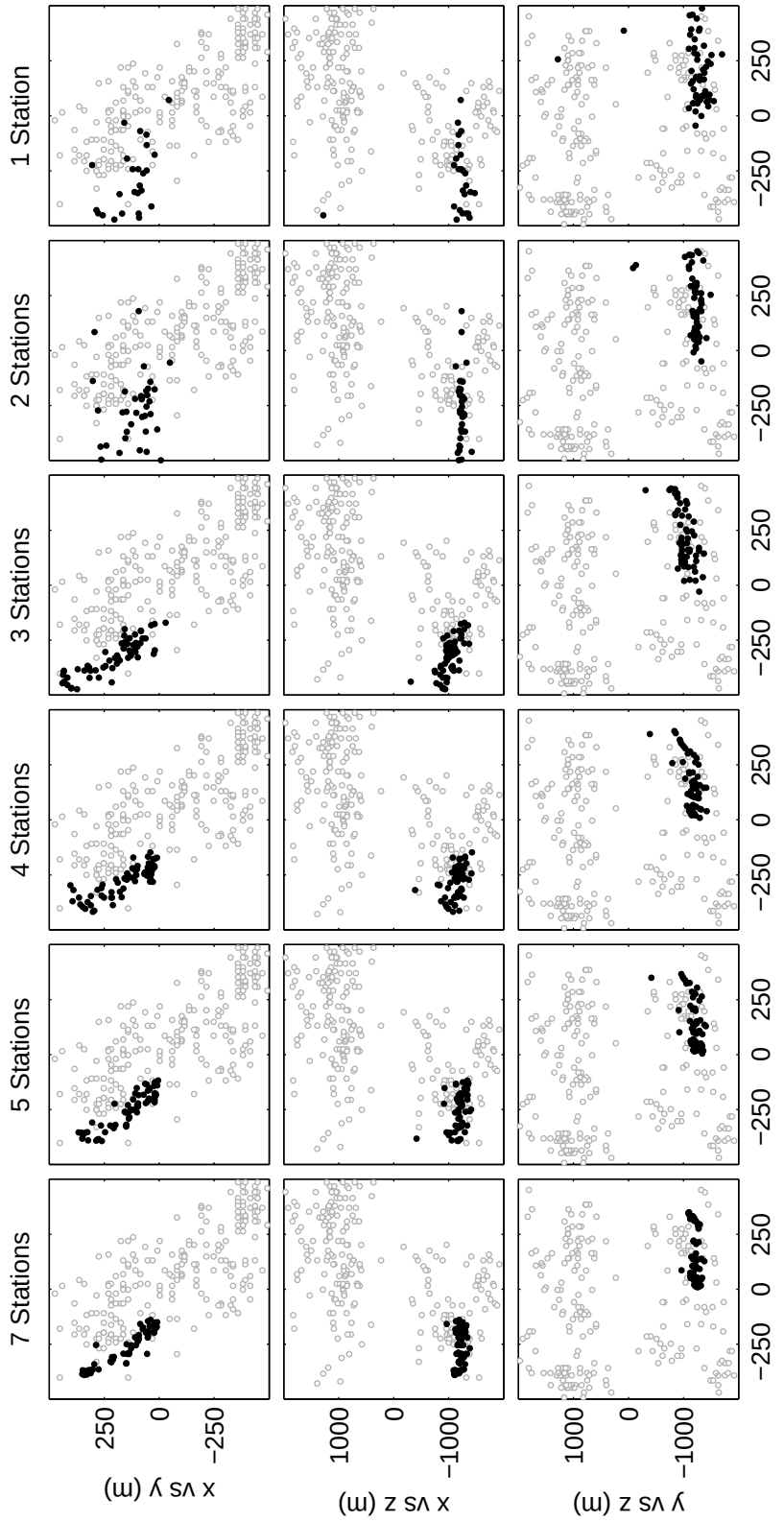
1 Station
2 Stations
3 Stations
4 Stations
5 Stations
7 Stations
x vs y (m)
x vs z (m)
y vs z (m)
250
0
−250
1000
0
−1000

Figure 9. ....

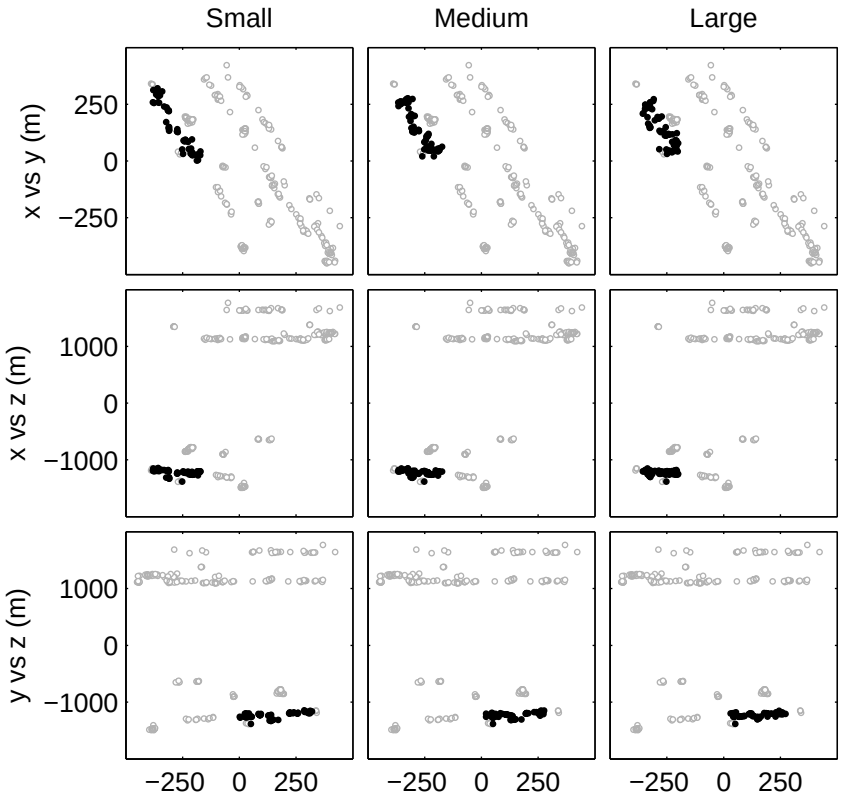

Figure 11. Locations of the 68 Calaveras earthquakes combining all available coda wave and travel time constraints with three different levels of uncertainty on the travel time PDFs. Column 1 sets $\sigma_x = \sigma_y = 19.5$ m and $\sigma_z = 15$ m (Waldhauser & Schaff 2008, after), Column 2 uses $\sigma_x = \sigma_y = 3 \times 19.5$ m and $\sigma_z = 3 \times 15$ m and column 3 considers $\sigma_x = \sigma_y = 152$ m and $\sigma_z = 232$ m (Shearer 1997, after).

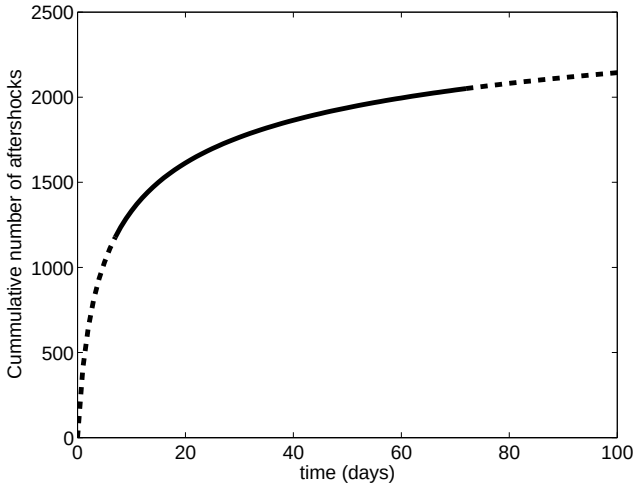

Figure 12. Cumulative number of aftershocks for the Hokkaido-Nansei-Oki, Japan $M_s = 7.8$ earthquake of 12 July 1993 according to the best fitting modified Omori Formula (Utsu et al. 1995). The leftmost dashed, middle solid and rightmost dashed signify aftershocks occurring before, during and after the deployment of a temporary array installed 7 days after the main shock and left for 65 days.

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.

Bokelmann, G. H. R. & Harjes, H., 2000. Evidence for temporal variation of seismic velocity within the upper continental crust, *Journal of Geophysical Research*, **105**(B10), 23,879–23,894.

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.

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.

Giardini, D., Gruenthal, G., Shedlock, K., & Zhang, P., 2002. The GSHAP global seismic hazard map, in *International Handbook of Earthquake Engineering Seismology*, vol. A, chap. 74, pp. 1233–1239, eds Lee, W. H., Kanamori, H., Jennings, P. C., & Kisslinger, C., Academic Press, London.

Got, J.-L., Frèchet, J., & Klein, F. W., 1994. Deep fault plane geometry inferred from multiplet relative relocation beneath the south flank of

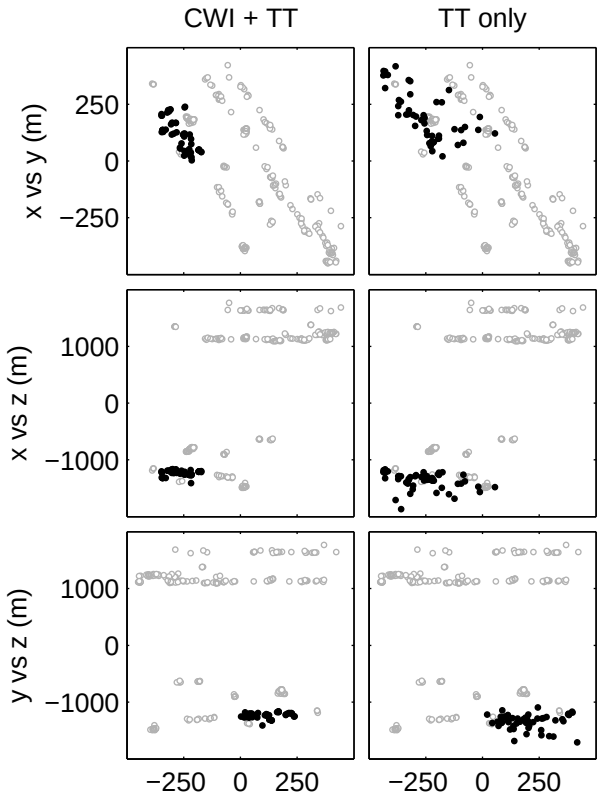

Figure 13. Left - Combined travel time and coda wave inversions using travel time constraints on 22 (or 32%) of the events and coda waves from station CCO only. All travel time based uncertainties are assigned $\sigma_x = \sigma_y = 3 \times 19.5\,\mathrm{m}$ and $\sigma_z = 3 \times 15\,\mathrm{m}$. Right - travel time relocation only (i.e. no CWI) using .....

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.

Gutenberg, B. & Richter, C. F., 1944. Frequency of earthquakes in California, *Bulletin of Seismological Society of America*, **34**, 185–188.

Isacks, B., Oliver, J., & Sykes, L. R., 1968. Seismology and the new Global tectonics, *Journal of Geophysical Research*, **73**(18), 5855–5899.

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.

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.

Poupinet, G., Ellsworth, W. L., & Frechet, J., 1984. Monitoring velocity variations in the crust using earthquake doublets: An application to the Calaveras Fault, California, *Journal of Geophysical Research*, **89**(B7), 5719–5731.

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

Richards, P. G., Waldhauser, F., Schaff, D., & Kim, W.-Y., 2006. The applicability of modern methods of earthquake location, *Pure and Applied Geophysics*, **163**, 351–372.

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., 2007. Constraints on coda wave interferometry estimates of source separation: The 2.5d acoustic case, *Exploration Geophysics*, **38**(3), 189–199.

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

Rubin, A., 2002a. Using repeating earthquakes to correct high-precision earthquake catalogs for time-dependant station delays, *Bulletin of the Seismological Society of America*, **92**(5), 1647–1659.

Rubin, A. M., 2002b. 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.

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., 1997. Improving local earthquake locations using the L1 norm and waveform cross correlation: Application to the Whittier Narrows, California, aftershock sequence, *Journal of Geophysical Research*, **102**(B4), 8269–8283.

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.

Stein, S. & Klosko, E., 2002. Earthquake mechanism and plate tectonics, in *International Handbook of Earthquake Engineering Seismology*, vol. A, chap. 7, pp. 69–78, eds Lee, W. H., Kanamori, H., Jennings, P. C., & Kisslinger, C., Academic Press, London.

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.

Sykes, L., 1967. Mechanism of earthquake and nature of faulting on the Mid-Oceanic Ridges, *Journal of Geophysical Research*, **72**(8), 2131–2153.

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. & 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 eq. (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,

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

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

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

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

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

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

(Robinson et al. submitted). The parameters μ_1 and σ_1 used in eq. (A.2) are defined by the expresions

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

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 A1. The parameters $\bar{\mu}_N$ and $\bar{\sigma}_N$ used in eq. (A.4) are obtained by finding the values which minimise the difference in a least squares sense between the noisy CWI estimates $\widetilde{\delta}_{CWIN}$ computed from the waveforms and the positively bounded Gaussian density function

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

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

Table A1. Coefficients for eqs. (A.7) and (A.8).

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

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 optimisation procedure towards the values of $(\mathbf{e}_1,\mathbf{e}_2,...,\mathbf{e}_N)$ which minimise L.

The equations for the derivatives are convoluted so we build them gradually. We start with an expression for δ_t, the wavelength normalised 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 B.2 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 eq. (A.7) gives

$$\frac{\partial\mu_1}{\partial\widetilde{\delta}_t} = a_1\frac{a_2a_4\widetilde{\delta}_t^{a_4-1} + a_3a_5\widetilde{\delta}_t^{a_5-1}}{\left(a_2\widetilde{\delta}_t^{a_4} + a_3\widetilde{\delta}_t^{a_5}+1\right)^2}. \tag{B.6}$$

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 B.6 with different constants $a_1,a_2,...,a_5$ (see table A1).

The cumulative Gaussian distribution function A.6 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

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

and for the logarithm we have

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

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

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

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 in the following sections by demonstrating that the optimisation procedure converges to the correct solution for a number of synthetic problems in 2 and 3 dimensions.

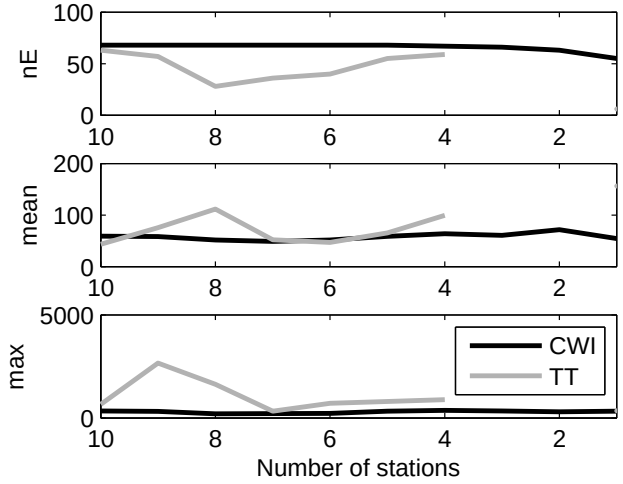

Figure A3. Statistics on coordinate differences for reduced station inversions. Differences are computed between the inversion results (CWI and hypoDD) and the complete hypoDD locations for all 308 events. The top subplot illustrates the number of constrainable events in the CWI and hypoDD inversions as a function of the stations considered.

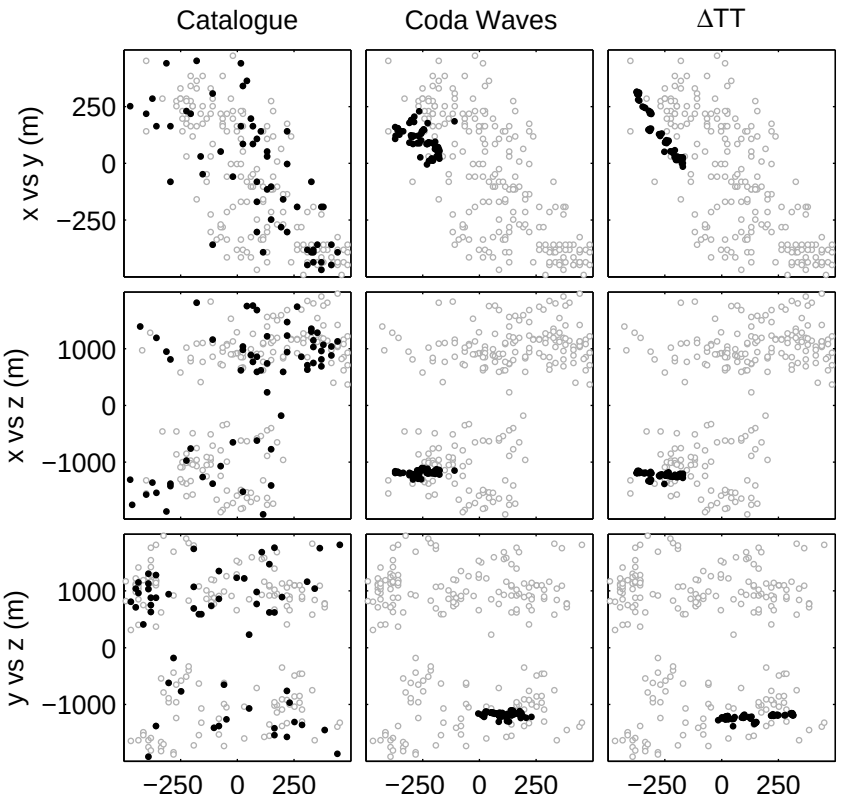

Figure A1. Comparison of earthquake hypocenters using three different methods: catalogue location (column 1), hypoDD with all 308 events (column 2) and CWI example 1 (column 3). Note that in the case of the CWI locations we consider only the 68 earthquakes in red, the blue events are shown for the purpose of orientation only.

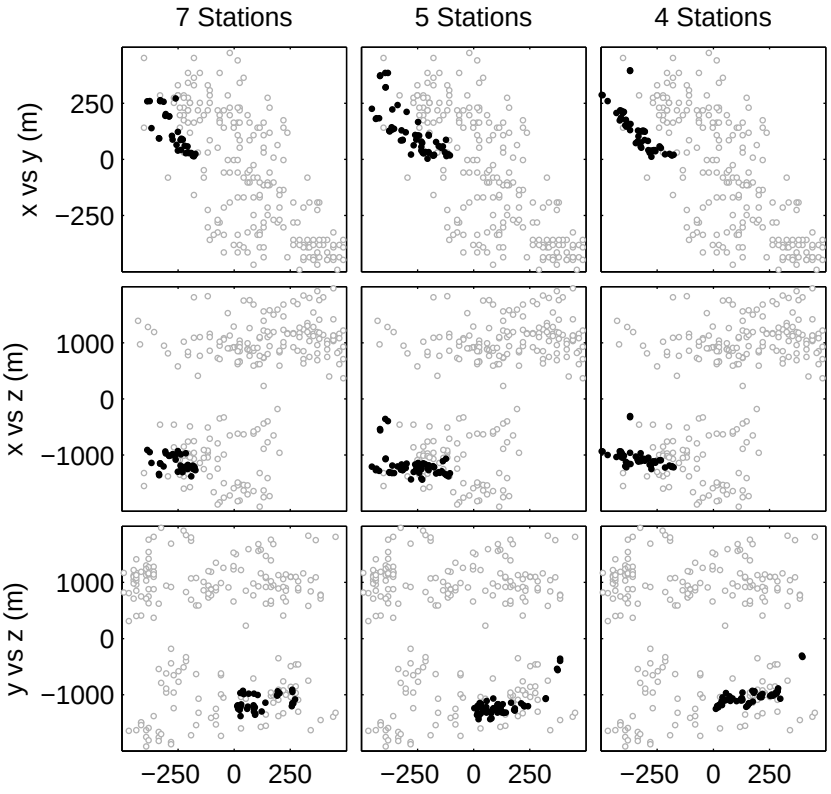

Figure A2. HypoDD (SVD) with reduced stations.

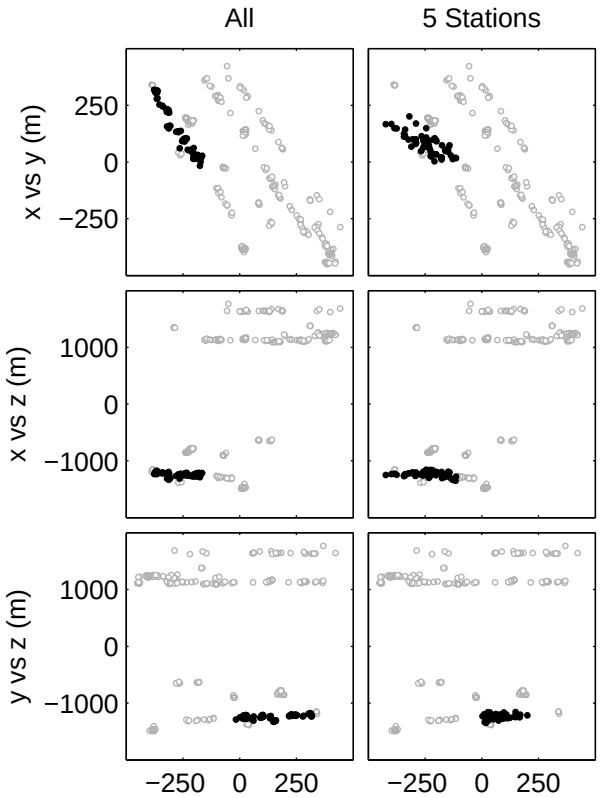

Figure A4. Combined HypoDD (SVD) and CWI: all stations and 5 stations.