A probabilistic approach for estimating the separation between a pair of earthquakes directly from their coda waves

D. J. Robinson,[1,2] M. Sambridge,[1] and R. Snieder[3]

Received 2 June 2010; revised 15 October 2010; accepted 11 January 2011; published 19 April 2011.

[1] Coda wave interferometry (CWI) can be used to estimate the separation between a pair of earthquakes directly from the coda recorded at a single station. Existing CWI methodology leads to a single estimate of separation and provides no information on uncertainty. Here, the theory of coda wave interferometry is revisited and modifications introduced that extend the range of applicability by 50% (i.e., 300–450 m separation for 1–5 Hz filtered coda waves). Synthetic experiments suggest that coda wave separation estimates fluctuate around the actual separation and that they have an increased tendency to underestimate the actual separation as the distance between events increases. A Bayesian framework is used to build a probabilistic understanding of the coda wave constraints which accounts for both the fluctuations and bias. The resulting a posteriori function provides a conditional probability distribution of the actual separation given the coda wave constraints. It can be used in isolation, or in combination with other constraints such as travel times or geodetic data, and provides a method for combining data from multiple stations and events. Earthquakes on the Calaveras Fault, California, are used to demonstrate that CWI is relatively insensitive to the number of recording stations and leads to enhanced estimates of separation in situations where station geometry is unfavorable for traditional relative location techniques.

Citation: Robinson, D. J., M. Sambridge, and R. Snieder (2011), A probabilistic approach for estimating the separation between a pair of earthquakes directly from their coda waves, *J. Geophys. Res.*, *116*, B04309, doi:10.1029/2010JB007745.

1. Introduction

[2] The majority of seismological studies exploit early onset body waves to infer properties of the source (e.g., location and focal mechanism) and seismic velocity field. An alternative approach is to use the later arriving waves in the seismogram, or coda waves that arise from scattering [*Aki*, 1969; *Sato*, 1988; *Sato and Fehler*, 1998; *Snieder*, 1999, 2006]. These are ignored in most applications because they are difficult to model. Some notable exceptions include the work of *Aki* [1969], *Aki and Chouet* [1975], *Abubakirov and Gusev* [1990], and *Margerin et al.* [1999] who use coda to infer properties of earthquake sources and velocity field. In a seminal paper, *Aki* [1969] adopted a statistical treatment to describe coda generation in terms of single backscattering and used it to compute the seismic moment of local earthquakes. *Aki and Chouet* [1975] introduce an alternative explanation for coda via a diffusion process. They discuss links between source spectra, attenuation and coda using single backscattering and diffusion theories. More recent explanations of coda generation consider multiple scattering, an interpolation between the two extremes [*Hoshiba*, 1991; *Margerin et al.*, 2000]. An emerging field, known as coda wave interferometry (CWI), is based on the interference pattern between the coda of two events [*Snieder and Vrijlandt*, 2005; *Snieder*, 2006]. It is used to determine seismic velocity changes in laboratory specimens [*Roberts et al.*, 1992; *Snieder et al.*, 2002; *Grêt et al.*, 2006], volcanoes [*Ratdomopurbo and Poupinet*, 1995; *Grêt et al.*, 2005] and fault zones [*Poupinet et al.*, 1984]. In another variation of CWI, *Robinson et al.* [2007a] show how it can be extended to constrain differences in focal mechanism between pairs of earthquakes.

[3] In this paper we focus on CWI and source separation. *Snieder and Vrijlandt* [2005] demonstrate that coda can be used to estimate separation between the hypocenters of earthquakes with identical source properties. Unlike travel or delay time techniques, CWI does not require multiple stations. In fact, it is possible to obtain estimates of separation using a single station [*Robinson et al.*, 2007b]. This makes CWI particularly useful for regions where station density is low, such as intraplate areas. The limitation of the *Snieder and Vrijlandt* [2005] approach however, is that it leads to a single estimate of separation and neither accounts for the uncertainty (fluctuation of estimates around the true values), nor for its tendency to underestimate separation for earthquakes at greater distances. In this paper we extend the range of applicability of CWI theory to account for biases, and quantify its probabilistic nature. In doing so, we build a

[1]Research School of Earth Sciences, Australian National University, Canberra, ACT, Australia.
[2]Risk and Impact Analysis Group, Geoscience Australia, Canberra, ACT, Australia.
[3]Center for Wave Phenomena and Department of Geophysics, Colorado School of Mines, Golden, Colorado, USA.

0148-0227/11/2010JB007745

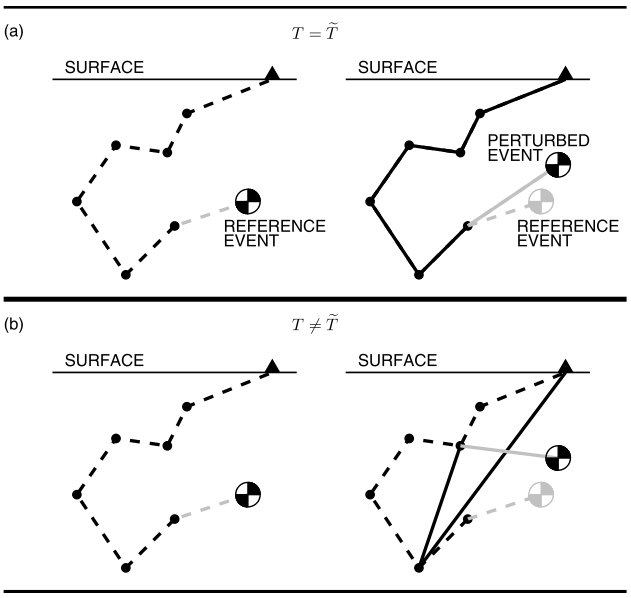

Figure 1. (a and b) Example scattering trajectories for reference (T, solid) and perturbed ($\widetilde{T}$, dashed) sources. Paths between source and first scatterer are grey. Later path components are black. In Figure 1a the only difference between T and $\widetilde{T}$ is the component of the trajectory between the source and first scatterer. To simplify notation we say that $T = \widetilde{T}$, despite the fact that they are not strictly identical. Figure 1b the paths T and $\widetilde{T}$ are clearly different. We denote such cases by $T \neq \widetilde{T}$.

computational framework for estimating the probability of true separation given coda wave estimates. With this probabilistic constraint it is possible to harness all available information from the coda and gain a detailed understanding of the uncertainty in actual separation. The method that we develop facilitates the combination of travel times and coda waves as well as the inclusion of data from different stations and/or earthquakes. We propose that these pairwise probability density functions can be useful in constraining the relative location for a cluster of earthquakes.

2. Coda Wave Interferometry Theory and Extensions

[4] Coda wave interferometry uses variations in coda to constrain separation between two events. The key idea is that when the source position is perturbed, some raypaths are longer and some are shorter [*Snieder*, 2006]. This is emphasized by defining the reference waveform $\mathbf{u}(t)$ recorded at an arbitrary station by

$$\mathbf{u}(t) = \sum_{T} \mathbf{A}_T(t) \tag{1}$$

and the perturbed waveform $\widetilde{\mathbf{u}}(t)$ resulting from the displaced source

$$\widetilde{\mathbf{u}}(t) = \sum_{\widetilde{T}} \mathbf{A}_{\widetilde{T}}(t - \tau_T), \tag{2}$$

where T (reference) and $\widetilde{T}$ (perturbed) are trajectories between source and station, $\mathbf{A}_T$ and $\mathbf{A}_{\widetilde{T}}$ are trajectory contributions to the reference and perturbed waveforms, and τ_T is the travel time perturbation between trajectories T and $\widetilde{T}$. If the sources do not share the same location, the trajectories T and $\widetilde{T}$ can never be identical. For small perturbations however, T and $\widetilde{T}$ may differ only in the component between source and first scatterer along the trajectory and we denote this by $T = \widetilde{T}$ (Figure 1a). We use $T \neq \widetilde{T}$ when the trajectories differ by more than the first path segment (Figure 1b).

2.1. Quantifying Similarity Between Two Waveforms

[5] Coda wave interferometry relies on cross correlation of waveforms computed for sliding windows across the coda. The parameter of interest, maximum cross correlation $R_{\max}$, is taken over all lag times in the sliding windows. This differs from existing relative location techniques which typically use cross correlation of early onset body waves to determine the lag (or delay time) leading to maximum phase correlation [e.g., *Waldhauser and Ellsworth*, 2000]. The normalized cross correlation used in CWI is

$$R^{(t,t_w)}(t_s) = \frac{\int_{t-t_w}^{t+t_w} u_i(t')\widetilde{u}_i(t'+t_s)dt'}{\left(\int_{t-t_w}^{t+t_w} u_i^2(t')dt' \int_{t-t_w+t_s}^{t+t_w+t_s} \widetilde{u}_i^2(t')dt'\right)^{\frac{1}{2}}}, \tag{3}$$

where t_s is lag time. It measures change between reference u_i and perturbed $\widetilde{u}_i$ displacement at station i over a time window of length $2t_w$ with center time t [*Snieder*, 2006].

[6] The numerator of equation (3) is the standard cross correlation

$$C_{\widetilde{uu}}^{(t,t_w)}(t_s) = \int_{t-t_w}^{t+t_w} u_i(t')\widetilde{u}_i(t'+t_s)dt' \tag{4}$$

and the denominator, $\sqrt{C_{uu}^{(t,t_w)}(0)C_{\widetilde{u}\widetilde{u}}^{(t,t_w)}(0)}$, is a normalization factor accounting for different magnitudes. Note that displacement u can be replaced with other wavefields such as velocity or acceleration. The maximum of $R^{(t,tw)}(t_s)$ over all t_s is written

$$R_{\max} = \max_{t_s}\left[R^{(t,t_w)}(t_s)\right] \tag{5}$$

and can be computed directly from the waveforms. In this paper we use $R_{\max}$ to constrain the separation between two events.

[7] Since random noise fluctuations reduce waveform similarity, its presence leads to a bias in $R_{\max}$. That is, $R_{\max}$ is always smaller for noisy waveforms. This bias is addressed by using the corrected normalized cross correlation [*Snieder and Vrijlandt*, 2005; *Douma and Snieder*, 2006]

$$R_{corr}^{(t,t_w)}(t_s) = \frac{C_{\widetilde{uu}}^{(t,t_w)}(t_s)}{\sqrt{\left(C_{uu}^{(t,t_w)}(0) - C_{nn}^{(t_w)}\right)\left(C_{\widetilde{u}\widetilde{u}}^{(t,t_w)}(0) - C_{\widetilde{n}\widetilde{n}}^{(t_w)}\right)}}, \tag{6}$$

where $C_{nn}^{(tw)}$ and $C_{\widetilde{n}\widetilde{n}}^{(tw)}$ are

$$C_{nn}^{(t_w)} = \frac{2t_w}{t_p}\int_{t_{start}}^{t_p} [u_i(t')]^2 dt' \tag{7}$$

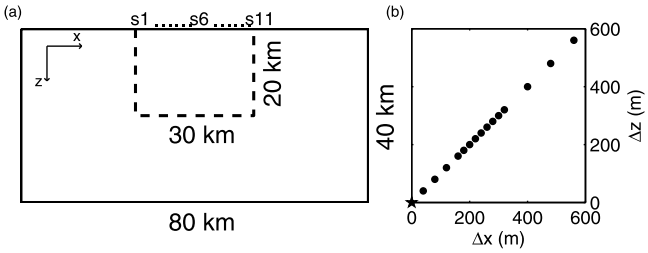

Figure 2. (a) Model domain with reflective boundaries as solid lines. The 11 recording stations are equally spaced along the surface and denoted s1 to s11. Reflections from the top represent the free-surface effect, whereas reflections from the bottom, left, and right do not reach stations within the time frame of the simulation for sources located in the dashed "region of interest." (b) Spatial relationship between the reference (star) and perturbed (circles) sources.

and

$$\underset{\sim\sim}{C}_{nn}^{(t_w)} = \frac{2t_w}{t_p}\int_{t_{start}}^{t_p}[\widetilde{u}_i(t')]^2 dt', \tag{8}$$

t_p is the P wave arrival, and t_{start} is the start of recording. Equations (7) and (8) measure the background noise energy in the reference and perturbed waveforms. The factor $2t_w/t_p$ ensures that noise energy is consistent with a time window width $2t_w$ and that it is constrained by all available data. We alternate between equations (3) and (6), our choice depending on whether studying synthetic noise free waveforms or real events.

2.2. Inferring Source Separation From Maximum Cross Correlation

[8] *Snieder* [2006] demonstrates that $R_{\max}$ is related to variance in the travel time perturbation σ_τ^2 by

$$R_{\max}^{(t,t_w)} = 1 - \frac{1}{2}\overline{\omega^2}\sigma_\tau^2, \tag{9}$$

where the square of dominant angular frequency

$$\overline{\omega^2} = \frac{\int_{t-t_w}^{t+t_w}\dot{u}_i^2(t')dt'}{\int_{t-t_w}^{t+t_w}u_i^2(t')dt'} \tag{10}$$

and $\dot{u}_i$ represents the time derivative of u_i. The derivation of equation (9) is provided in Appendix A. It is useful to note however, that the right-hand side arises from a Taylor series expansion of the normalized autocorrelation function about $t_s = 0$ s.

[9] *Snieder and Vrijlandt* [2005] show that σ_τ^2 is related to source separation δ_{CWI} by

$$\delta_{CWI}^2 = g(\alpha,\beta)\sigma_\tau^2, \tag{11}$$

with α and β representing the 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 same fault plane,

$$g(\alpha,\beta) = 7\frac{\left(\frac{2}{\alpha^6}+\frac{3}{\beta^6}\right)}{\left(\frac{6}{\alpha^8}+\frac{7}{\beta^8}\right)}, \tag{12}$$

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

$$g(\alpha,\beta) = 2\alpha^2. \tag{13}$$

2.3. Application to Synthetic Events in 2.5 Dimensions

[10] We employ a synthetic experiment to illustrate the nature of constraints provided by CWI and explore its range of applicability. Synthetic waveforms are computed by solving the acoustic (or scalar) wave equation

$$\frac{\partial^2\phi(x,z,t)}{\partial x^2} + \frac{\partial^2\phi(x,z,t)}{\partial z^2} = \frac{1}{\alpha^2}\frac{\partial^2\phi(x,z,t)}{\partial t^2} \tag{14}$$

using a finite difference solver with fourth-order accuracy in space and second-order in time. The model domain extends 80 km laterally, 40 km in depth and has reflective boundary conditions on all sides (Figure 2). To avoid unwanted reflections we define a "region of interest" with dimensions 20 km by 30 km. Reflections from the sides and bottom of the model domain arising from events within the region do not reenter it in the time frame of the experiment. Reflections from the top represent the free-surface effect. We use a grid spacing of $\Delta x = \Delta z = 20$ m and create a P wave velocity model by computing a realization of the Gaussian random medium with mean velocity $\mu_\alpha = 6$ km s^{-1}, correlation length $a = 1.2$ km and standard deviation $\sigma_\alpha = 1.5$ km s^{-1}. *Frankel and Clayton* [1986] and *Baig and Dahlen* [2004] provide techniques for creating Gaussian random media.

[11] Equation (14) is solved for a reference event buried 5 km below station S2 and illustrated at S6 in Figure 3a after 1–5 Hz bandpass filtering (grey). The event is a line source perpendicular to the model domain with Ricker wavelet source time function

$$s(t) = \left(1 - 2\pi^2 f_o^2 t^2\right)e^{-\pi^2 f_o^2 t^2} \tag{15}$$

with dominant frequency, $f_0 = 8$ Hz. We use a time step $\Delta t = \Delta x/2\alpha_{\max}$, where $\alpha_{\max}$ is the maximum velocity. This Δt satisfies the Courant, Friedrichs, Lewy stability condition [*Courant et al.*, 1928] which ensures the waves do not propagate more than half a cell within a single time step.

[12] We perturb the source by 226 m and repeat the calculation (black waveform in Figure 3a). The two waveforms have similar first arrivals (Figure 3b) and notably different coda (Figure 3c). We are interested in exploiting this difference to determine source separation. Figure 3d illustrates $R_{\max}$ using equations (3) and (5) with $2t_w = 0.75$ s. A high value of $R_{\max}$ is observed for early windows which is associated with the similarity of first arrivals. This is followed by lower estimates as we cross correlate the coda. Separation is computed using equations (11) and (13) and compared with the true value $\delta_t = 226$ m in Figure 3e. First arrival similarity leads to low CWI separations for early windows which we ignore. This is because early arrivals do not result from scattered waves leaving the source in all directions [*Snieder*

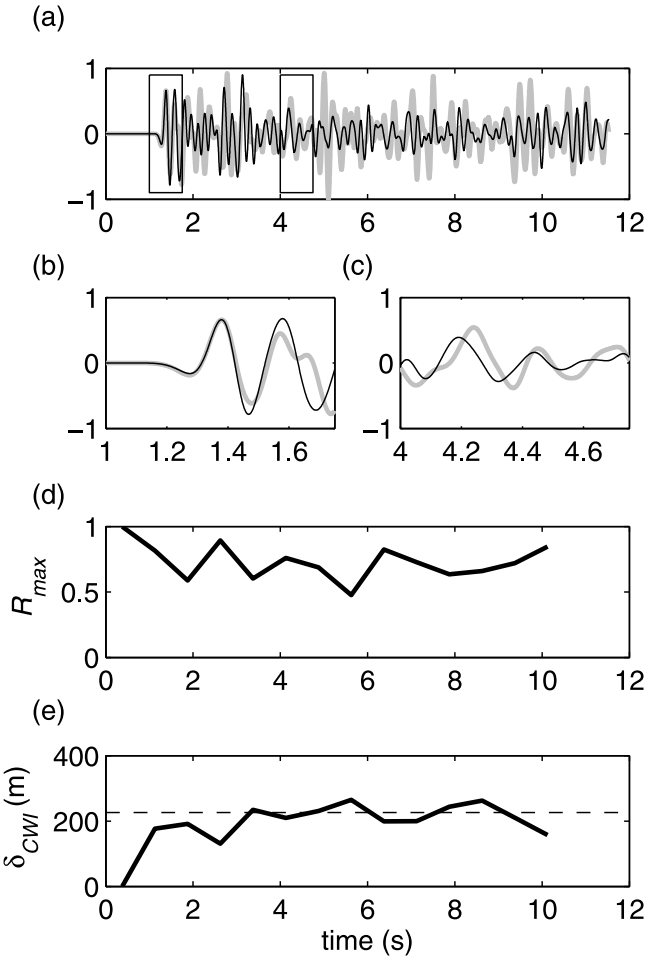

Figure 3. (a) Synthetic waveforms for reference (grey) and perturbed (black) events separated by 226 m at S6. (b) Left box around direct arrival and (c) right box surrounding example coda are shown in detail. (d) $R_{\max}$ as a function of sliding centroid for window width 0.75 s and (e) CWI (solid) and actual (dashed) separations.

and Vrijlandt, 2005]. Estimates after 2 s fluctuate around the true separation showing that CWI provides an accurate indication of separation for this pair. Fluctuations in $R_{\max}$ occur despite the absence of noise. Their size is inversely proportional to $\sqrt{\Delta f 2t_w}$ where $2t_w$ is the window size and Δf the signal bandwidth [*Snieder*, 2004, 2006]. Their presence implies that coda provide a soft (i.e., probabilistic) constraint on separation. We return to this point in section 3.

[13] To explore the range of applicability for CWI we consider perturbed sources at progressively increasing distances as shown in Figure 2b. The perturbed sources correspond to actual separations of 57, 113, 170, 226, 255, 283, 311, 339, 368, 396, 424, 453, 566, 679, and 792 m. Synthetic waveforms are generated at points *S*1 to *S*11 for all sources. We repeat the experiment at three depths by defining reference sources at $z \approx 3$ km (shallow), 5 km (mid) and 15 km (deep). In this paper we analyze the data collectively by grouping the CWI estimates. *Robinson* [2010] provides an independent analysis for different depths and stations and demonstrates that this aggregation is robust. When grouping $R_{\max}$ and δ_{CWI} estimates we ignore the first five sliding windows for each pair to remove the influence of first arrivals.

[14] Mean $R_{\max}$ is illustrated as a function of actual separation δ_t in Figure 4a (grey). We observe that it decreases monotonically up to $\delta_t = 300$ m, after which it is flat. This is a weakness of the present CWI theory. The associated δ_{CWI} are illustrated in the bottom panel. Here we show both the mean μ (solid grey) and the $\pm\sigma$ bounds (dashed grey). Mean δ_{CWI} flattens as a result of the leveling in $R_{\max}$ and once $\delta_t >$ 300 m, only provides information on the lower bound. We define the breakdown distance δ_b as the value of actual separation δ_t associated with the intersection between $\mu + \sigma$ and $\delta_t = \delta_{CWI}$ (circle in Figure 4b).

2.4. Extending Coda Wave Interferometry

[15] In this section we develop two extensions to the CWI separation theory. The first relates to our treatment of time lag when evaluating waveform similarity. The second pertains to how the variance of travel time perturbation is related to $R_{\max}$. Both extensions are designed to address the leveling of $R_{\max}$ around 0.7.

2.4.1. $R_{\max}$: Cycle Skipping and Time Lag

[16] When evaluating $R_{\max}$ above we considered the maximum over all time lags t_s. In general, as separation increases waveform similarity decreases and we expect a reduction in $R_{\max}$. Interestingly, Figure 4 demonstrates that $R_{\max}$ levels out around a value of 0.7. *Snieder and Vrijlandt* [2005] showed that for homogeneously distributed scatterers the mean travel time perturbation μ_τ is 0 and $R_{\max}$ is expected to occur near $t_s = 0$. The CWI technique can be extended to greater distances by exploiting this observation and reducing the time lag search to the vicinity of $t_s = 0$. Ideally we would compute $R_{\max}$ precisely at $t_s = 0$. This approach is susceptible to errors in waveform alignment so we seek the peak of the correlation function closest to $t_s = 0$, which is achieved in this paper by restricting the search to $-0.05 \le t_s \le 0.05$ s where 0.05 s is much smaller than one period for 1–5 Hz filtered waveforms.

[17] To understand why the reduced time lag search is useful we explore the relationship between $R_{\max}$ and the cross correlation function $C_{up}^{(t,tw)}(t_s)$. Figure 5 illustrates samples of $C_{up}^{(t,tw)}(t_s)$ using a 0.75 s wide time window for separations of 113, 283, and 368 m. Time lags $t_s = 0$ and $t_s \pm$ 0.05 s are shown by black solid and dashed lines, respectively. Peaks within the dashed lines represents $R_{\max}$ in the vicinity of $t_s = 0$ whereas those outside result from correlation of cycles that are at least one wavelength apart. We observe that for small separations (e.g., $\delta_t = 113$ m) the highest $C_{up}^{(t,tw)}(t_s)$ lies within $t_s = 0 \pm 0.05$ s. For $\delta_t = 283$ m the peak in $t_s = 0 \pm 0.05$ s is only marginally larger than its right counterpart and at $\delta_t = 368$ m the highest value is outside the dashed lines. In the latter case, considering the maximum over all time lags represents an overestimate of the desired $R_{\max}$. It is this correlation of out of phase cycles (or cycle skipping) which causes the flattening in Figure 4 for larger separations. A restriction of time lags when searching for $R_{\max}$ (equation (5)) ensures that only the required cycles are considered. Repeating the above experiment with this restriction leads to the black curve in Figure 4a. Note that unlike the original application of CWI (grey curve in Figure 4), $R_{\max}$ no longer flattens after $\delta_t = 300$ m suggest-

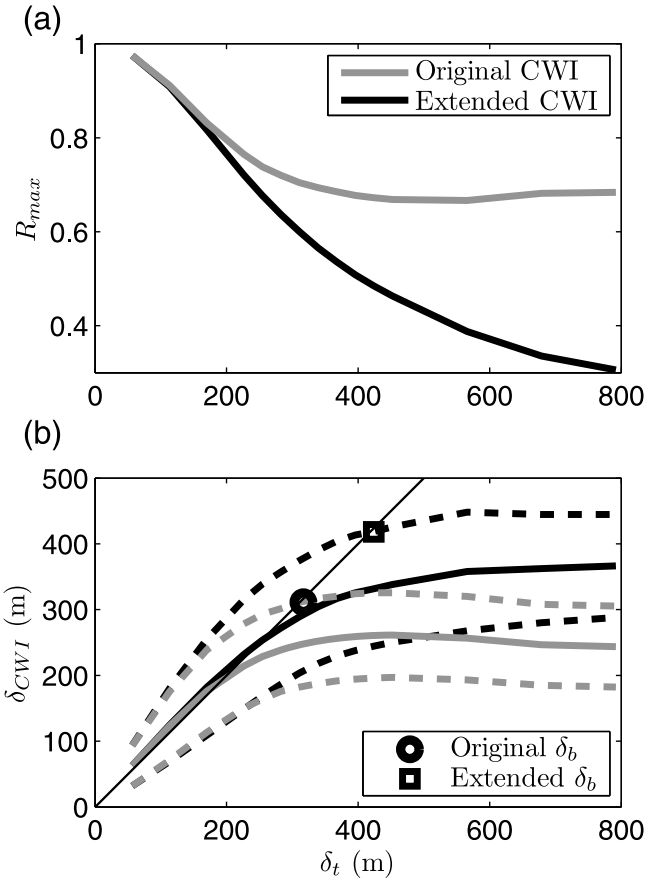

Figure 4. Comparison of (a) $R_{\max}$ and (b) δ_{CWI} using the original (sections 2.1 and 2.2) and extended (section 2.4) CWI theory. Thick solid lines represent the mean (μ), and dashed lines represent $\mu \pm \sigma$. A thin diagonal in Figure 4b indicates a one-to-one relationship between δ_{CWI} and δ_t. Breakdown distance δ_b is δ_t associated with illustrated symbols.

ing that the technique has the potential to work for greater separations.

2.4.2. $R_{\max}$ and σ_τ: Removal of Taylor Series Approximation

[18] Equation (9) relates $R_{\max}$ to the standard deviation of the travel time perturbation σ_τ. The key step in deriving this equation is an approximation of the autocorrelation function $C(t_s - \tau_T)$ by its second-order Taylor series expansion near $t_s - \tau_T = 0$ (see equation (A7)). This is depicted in Figure 6 which shows $C(t)$ (grey) and the second-order Taylor series expansion $\widetilde{C}(t)$ around the point $t = t_s - \tau_T = 0$ (black). In the original formulation σ_τ, and hence δ_{CWI}, is given by the intersection of the Taylor series approximation $\widetilde{C}(t)$ and the horizontal $R_{\max}$. However, the desired quantity is the intersection of the normalized autocorrelation function $C(t)$ with $R_{\max}$. The error in this approximation is therefore the width between the vertical dashed lines which widens with decreasing $R_{\max}$. For example, when $R_{\max} = 0.9$ the waveforms are similar and the error is roughly 5 m. However, as $R_{\max}$ approaches 0.4 the error tends to 40 m. In our second extension to the original CWI theory we remove the approximation for σ_τ (i.e., Taylor series expansion) and replace it with the intersection of $R_{\max}$ and the autocorrelation function $C(t)$ computed directly from the waveforms. Repeating our experiment with both extensions leads to the black curves in Figure 4 which demonstrate an increase in breakdown distance of 50% (300–450 m).

[19] We compute the intersection with an accuracy controlled by waveform sampling. Initially, it appears that there is a practical complication concerning the choice of waveform (i.e., $u_i(t)$ or $\widetilde{u}_i(t)$) when computing $C(t)$. Conveniently

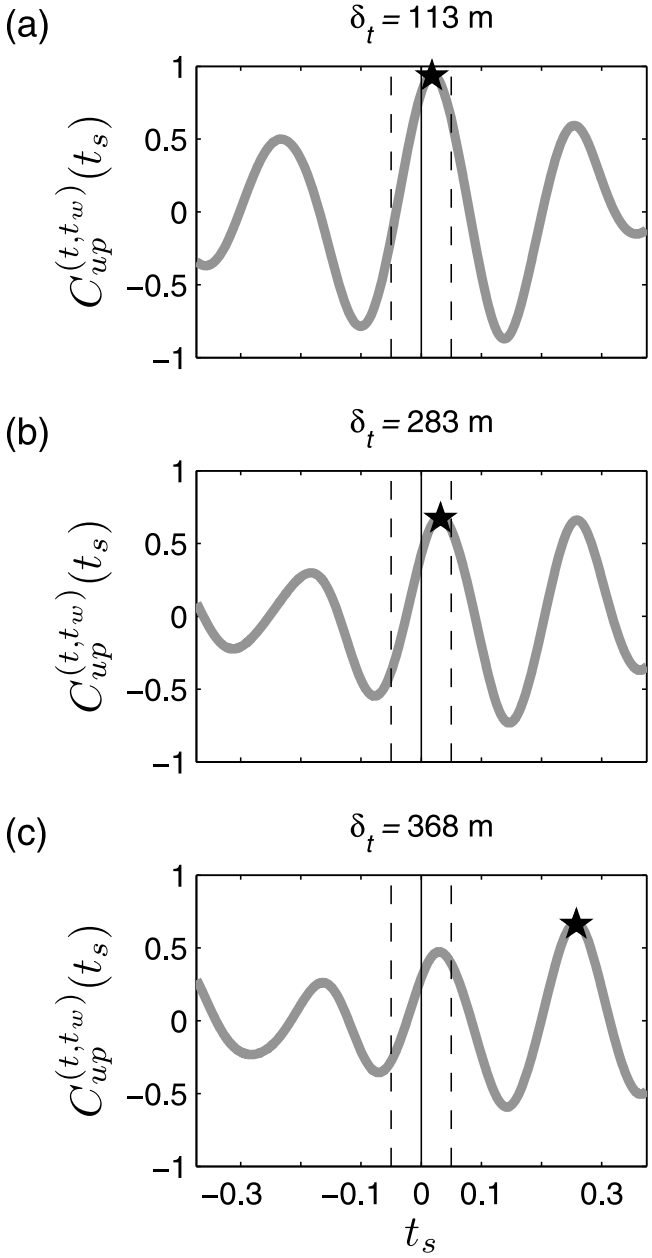

Figure 5. Cross correlation $C_{up}^{(t,tw)}(t_s)$ versus time lag t_s between two waveforms for a sample $2t_w = 0.75$ s window at three separations. Solid black lines depict $t_s = 0$, and dashed lines depict $t_s \pm 0.05$ s. Stars indicate the highest peak or $R_{\max}$ across all time lags. In the original application of CWI we search over all time lags in the window. In the extended version we restrict our search within the dashed lines. The latter means that we are less susceptible to cycle skipping.

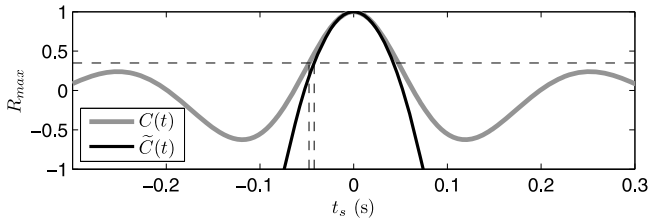

Figure 6. Taylor series approximation $\widetilde{C}(t)$ of the autocorrelation function $C(t)$. Width between the vertically oriented dashed lines indicates the error in σ_τ when $C(t)$ is approximated by $\widetilde{C}(t)$ at $R_{\max} = 0.35$.

however, the autocorrelations for $u_i(t)$ or $\widetilde{u}_i(t)$ are nearly identical under the assumption of similar power spectra (see Appendix A) and either waveform can be used with little difference.

2.5. Influence of Frequency

[20] Waveform frequency, in practice the dominant value, f_d, is influenced by the source time function, the velocity structure and the number of scattering events. We test the influence of the velocity structure by repeating our experiments at three depths in nine *P* wave velocity models with a broad range of properties. They are generated by computing realizations of Gaussian random media with mean velocity $\mu_\alpha = 6000$ m s^{-1}, correlation lengths a of 400, 800 and 1200 m and standard deviations σ_α of 500, 1000 and 1500 km s^{-1}. We observe minor variation amongst the media-depth experiments but no systematic trends, which suggests that CWI is not affected by the choice of Gaussian random medium or source-to-surface proximity over the ranges explored. Hereafter, we group the data for all media-depth combinations.

[21] To determine the influence of f_d on CWI we repeat the synthetic experiments with (a) 1–2 Hz, (b) 2–3 Hz, (c) 3–4 Hz, and (d) 4–5 Hz filtering and obtain breakdown distances δ_b of >800 m, 550 m, 400 m and 350 m, respectively. Recall that the 1–5 Hz bandwidth gave δ_b = 450 m. We conclude that CWI can be extended to distances exceeding 500 m as f_d tends toward 1–2 Hz.

[22] We seek a formulation of CWI which is uniform over a range of frequencies and hence simpler to use in practice. The dimensionless quantity

$$\widetilde{\delta} = \frac{\delta}{\lambda_d} \tag{16}$$

measures separation normalized by dominant wavelength λ_d, where δ can be the actual separation δ_t, CWI estimate δ_{CWI}, or breakdown distance δ_b which leads to $\widetilde{\delta}_t$, $\widetilde{\delta}_{CWI}$, and $\widetilde{\delta}_b$, respectively. The dominant wavelength is

$$\lambda_d = \frac{\beta}{f_d} = \frac{1}{1.65}\frac{\alpha}{f_d}, \tag{17}$$

and dominant frequency is given by

$$f_d = \sqrt{\frac{\overline{\omega^2}}{4\pi^2}}. \tag{18}$$

[23] We use β because shear waves dominate coda for earthquakes [*Aki*, 1992; *Snieder and Vrijlandt*, 2005]. The symbols in Figure 7 demonstrate the impact of applying equation (16) by illustrating the separation estimates as a function of actual separation for the four 1 Hz frequency bands. We observe consistent behavior across all bands and obtain $\widetilde{\delta}_b \approx 0.55$ for all frequencies suggesting that CWI can be applied in a unified formulation for all f_d.

[24] In summary, we have derived two modifications to the CWI theory which extend its range of applicability and have introduced a wavelength normalization parameter which allows CWI to be applied in a unified formulation across different frequencies. In the following we develop a probabilistic formulation which extends CWI further by changing the nature of CWI constraints from a single estimate to a probabilistic constraint.

3. A Bayesian Formulation of Coda Wave Interferometry

[25] From the previous discussion it is clear that coda wave data provide only probabilistic constraints on source separation (see Figures 4b and 7). This suggests that a Bayesian formulation of the information may be fruitful because it provides a convenient mechanism to explore probabilistic data [*Sivia and Skilling*, 2006]. A properly constructed Bayesian approach will account for the probabilistic nature of the constraints and address the known tendency of CWI to underestimate separation as it increases. Furthermore, it provides a framework within which it is possible to combine data from (1) different sources (stations, earthquakes etc.) and (2) different data types (coda waves, travel times, delay times, geodetic, etc.).

[26] In what follows we calculate the conditional probability of true separation given the CWI data. Our approach is based on the following statement of Bayes' rule [e.g., *Sivia and Skilling*, 2006],

$$P(\widetilde{\delta}_t|\widetilde{\delta}_{CWI}, I) \propto P(\widetilde{\delta}_{CWI}|\widetilde{\delta}_t, I) \times P(\widetilde{\delta}_t, I), \tag{19}$$

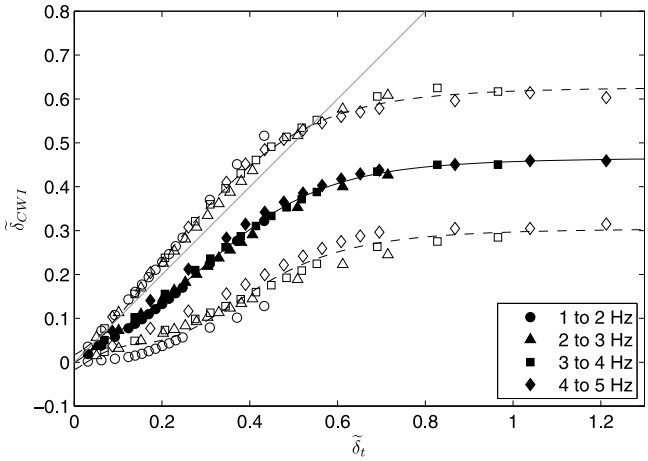

Figure 7. Least squares fits for $\mu_1(\widetilde{\delta}_t)$ and σ_1 $(\widetilde{\delta}_t)$ (see Table 1). Solid and open symbols represent μ_1 and $\mu_1 \pm \sigma_1$ for the CWI data after filtering to different bands. The one-to-one relationship between δ_{CWI} and δ_t is shown in grey.

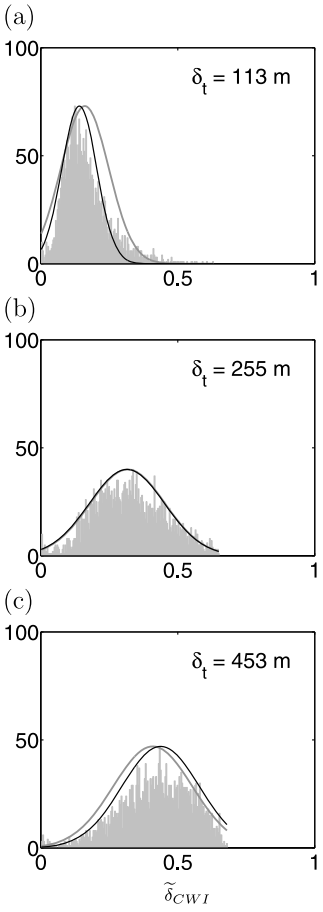

Figure 8. Histograms of $\widetilde{\delta}_{CWI}$ after 4–5 Hz bandpass filtering. Black line is the positive-bounded Gaussian likelihood with μ_1 and σ_1, which minimize the least squares difference between the observations. The Gaussian defined by sample mean and standard deviation of $\widetilde{\delta}_{CWI}$ data is shown in grey. Both PDFs are normalized to maximum bar height for convenience.

where I represents the underlying assumptions (e.g., similarity of source mechanism). For brevity we omit I in the notation. In equation (19) $P(\widetilde{\delta}_t|\widetilde{\delta}_{CWI})$ represents the probability density of the actual separation $\widetilde{\delta}_t$ given the CWI estimates $\widetilde{\delta}_{CWI}$ and is known as the posterior function or a posteriori. Our goal is to compute this function. The likelihood function $P(\widetilde{\delta}_{CWI}|\widetilde{\delta}_t)$ represents the probability of having observed the separation $\widetilde{\delta}_{CWI}$ (using CWI theory) if the true separation was $\widetilde{\delta}_t$. The remaining term $P(\widetilde{\delta}_t)$ is the prior on separation and can be used to incorporate existing information on separation. We consider only a uniform prior in this paper.

3.1. Likelihood in Absence of Noise: Theoretical Likelihood

[27] Figure 8 illustrates $\widetilde{\delta}_{CWI}$ using histograms for separations 113, 255, and 453 m for the experiments of section 2.5. The probability density function which best describes these histograms is the likelihood function $P(\widetilde{\delta}_{CWI}|\widetilde{\delta}_t)$ for the given choice of $\widetilde{\delta}_t$. In a previous study *Robinson et al.* [2007b] used a Gaussian PDF to approximate the histograms. A Gaussian, however, has global support and does not reflect an ideal choice for separation which is positively bounded. Approximating the likelihood with a Gaussian is reasonable when $\delta_t \gg 0$ and the histograms do not have long tails. In the experiments of *Robinson et al.* [2007b] the smallest separation was 57 m and the Gaussian was acceptable. However, the normalized separations $\widetilde{\delta}_{CWI}$ and $\widetilde{\delta}_t$ are typically between 0 and 1 and can have values close to 0. For example, a true separation $\delta_t = 57$ m with $\beta = 3300$ m s^{-1} and $f_d = 2.5$ Hz leads to $\lambda_{dom} = 1320$ m and $\widetilde{\delta}_t = 0.04$. Similarly, $\delta_t = 20$ m gives $\widetilde{\delta}_t = 0.0152$. Since these values are close to zero, it follows that we should exercise care in our choice of likelihood when working with normalized separation.

[28] We seek a likelihood function $P(\widetilde{\delta}_{CWI}|\widetilde{\delta}_t)$ which satisfies $\widetilde{\delta}_{CWI}$, $\widetilde{\delta}_t > 0$ and which tends toward a Gaussian when $\widetilde{\delta}_t$ is sufficiently above zero. The positively bounded Gaussian

$$P\left(\widetilde{\delta}_{CWI}|\widetilde{\delta}_t\right) = \frac{1}{\left(1-\Phi_{\mu_1,\sigma_1}(0)\right)\sigma_1\sqrt{2\pi}} e^{\frac{-\left(\widetilde{\delta}_{CWI}-\mu_1\right)^2}{2\sigma_1^2}}, \widetilde{\delta}_{CWI} \geq 0 \quad (20)$$

satisfies these conditions, where $\Phi_{\mu,\sigma}(x)$ is the cumulative Gaussian distribution function

$$\Phi_{\mu_1,\sigma_1}(x) = \frac{1}{\sigma_1\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-(s-\mu_1)^2}{2\sigma_1^2}} ds \quad (21)$$

and the presence of $(1 - \Phi_{\mu,\sigma}(0))$ in the denominator ensures integration to unity

$$\int_0^{\infty} P\left(\widetilde{\delta}_{CWI}|\widetilde{\delta}_t\right) d\widetilde{\delta}_{CWI} = 1. \quad (22)$$

Parameters $\mu_1 = \mu_1(\widetilde{\delta}_t)$ and $\sigma_1 = \sigma_1(\widetilde{\delta}_t)$ are the expected value and standard deviation of the positive bounded Gaussian PDF at a given $\widetilde{\delta}_t$.

[29] We compute $\mu_1(\widetilde{\delta}_t)$ and $\sigma_1(\widetilde{\delta}_t)$ for the 2-D acoustic experiments using a least squares fitting procedure which minimizes the misfit between density of observed estimates and equation (20). Figure 8 shows the best fitting positive bounded Gaussians (black) for separations of 113, 255, and 453 m. The Gaussian (grey) defined by the sample mean and sample standard deviation are illustrated for comparison. As hypothesized, the positive bounded Gaussians fit the CWI data better for the smallest separation (δ_t = 113 m). This is followed by a region where the two PDFs are similar (δ_t = 255 m). Interestingly, after $\delta_t > 368$ m the difference becomes more evident again with the positive bounded Gaussians representing the best fit. This is due to the long tail of the histogram in this case. In general, the positive bounded Gaussian fits the data better across a broader range of separations. Consequently, we use equation (20) when working with normalized separation $\widetilde{\delta}_{CWI}$.

[30] In the above example we can assign the positive bounded Gaussian to a likelihood for given $\widetilde{\delta}_t$ because the true separations are known in advance. In practice however, the true separation is not known. Computing the likelihood function $P(\widetilde{\delta}_{CWI}|\widetilde{\delta}_t)$ for arbitrary $\widetilde{\delta}_t$ involves defining the functional dependance of μ_1 and σ_1 on $\widetilde{\delta}_t$ so that equation (20) can be computed for any separation. The desired functions $\mu_1 = \mu_1(\delta_t)$ and $\sigma_1 = \sigma_1(\delta_t)$ are created by empirically fitting the μ_1 and $\mu_1 + \sigma_1$ data for the 1–2 Hz, 2–3 Hz, 3–4 Hz, and

Table 1. Empirically Derived Functions for $\mu_1(\widetilde{\delta}_t)$ and $\sigma_1(\widetilde{\delta}_t)$

Variable	$\mu_1(\widetilde{\delta}_t) = a_1 \frac{a_2\widetilde{\delta}_t^{a_4} + a_3\widetilde{\delta}_t^{a_5}}{a_2\widetilde{\delta}_t^{a_4} + a_3\widetilde{\delta}_t^{a_5} + 1}$	$\sigma_1(\widetilde{\delta}_t) = c + a_1 \frac{a_2\widetilde{\delta}_t^{a_4} + a_3\widetilde{\delta}_t^{a_5}}{a_2\widetilde{\delta}_t^{a_4} + a_3\widetilde{\delta}_t^{a_5} + 1}$
a_1	0.4661	0.1441
a_2	48.9697	101.0376
a_3	2.4693	120.3864
a_4	4.2467	2.8430
a_5	1.1619	6.0823
c		0.017

4–5 Hz experiments (Figure 7). We restrict our treatment in this section to bandwidths $\Delta f = 1$ Hz and $2t_w = 0.75$ s. This is done to avoid complications associated with varying fluctuation levels, the size of which are inversely proportional to $\sqrt{\Delta f 2t_w}$ [*Snieder*, 2006]. A functional form for μ_1 and σ_1 is chosen such that (1) $\mu_1 = \mu_1(\widetilde{\delta}_t) \rightarrow 0$ and $\sigma_1 = \sigma_1(\widetilde{\delta}_t) \rightarrow 0.017$ as $\widetilde{\delta}_t \rightarrow 0$ (this choice for μ_1 ensures that expected CWI tend to zero; for σ_1, 0.017 guarantees that the fitted curve forecasts the same standard deviation at $\delta_t = 0$ that is observed at the smallest experimental $\widetilde{\delta}_t$) and (2) $\mu_1(\widetilde{\delta}_t)$ and $\sigma_1(\widetilde{\delta}_t)$ flatten as $\widetilde{\delta}_t$ increases beyond 0.6 as suggested by the synthetic data in Figure 7.

[31] A function of form

$$f(x) = c + \frac{p_1(x)}{p_2(x)}, \tag{23}$$

where p_1 and p_2 are polynomials of the same degree, has the potential to meet both requirements. After experimentation we identify

$$f(\widetilde{\delta}_t) = c + a_1 \frac{a_2\widetilde{\delta}_t^{a_4} + a_3\widetilde{\delta}_t^{a_5}}{a_2\widetilde{\delta}_t^{a_4} + a_3\widetilde{\delta}_t^{a_5} + 1} \tag{24}$$

as suitable where a_1, a_2, a_3, a_4 and a_5 are chosen using a nonlinear optimization algorithm which minimizes the residuals in a least squares sense. The fit of $\mu_1(\widetilde{\delta}_t)$ and $\sigma_1(\widetilde{\delta}_t)$ are illustrated in Figure 7 and the coefficients given in Table 1. With these functions we can define the likelihood for arbitrary δ_t. The coefficients in Table 1 apply to $\Delta f = 1$ Hz and $2t_w = 0.75$ s. This is somewhat limiting because some waveforms may warrant a CWI analysis with different bandwidth and/or sliding window size. In section 3.2 we explore the influence of noise and introduce a 'noisy likelihood' function which simultaneously accounts for changes in $\widetilde{\delta}_{CWI}$ fluctuations induced by different Δf and $2t_w$.

3.2. Likelihood in Presence of Noise

[32] Observational noise in seismic waveforms impacts CWI in two ways. Firstly, it biases $R_{\max}$ which we address by using the corrected normalized cross correlation of equation (6). Secondly, noise changes the size of δ_{CWI} fluctuations and consequently modifies the likelihood $P(\widetilde{\delta}_{CWI}|\widetilde{\delta}_t)$. To use CWI with real data we must extend our probabilistic approach to account for noisy waveforms. We redefine equation (19) as

$$P\left(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN}\right) \propto P\left(\widetilde{\delta}_{CWIN}|\widetilde{\delta}_t\right) \times P\left(\widetilde{\delta}_t\right), \tag{25}$$

where $\widetilde{\delta}_{CWIN}$ denotes the presence of noise in CWI estimates.

[33] We seek a functional form for the "noisy likelihood" $P(\delta_{CWIN}|\delta_t)$ which can be computed for arbitrary $\widetilde{\delta}_t$. In Appendix B we show that $P(X|Y) = \int P(X|Z,Y)P(Z|Y)dZ$ which becomes

$$P\left(\widetilde{\delta}_{CWIN}|\widetilde{\delta}_t\right) = \int P\left(\widetilde{\delta}_{CWIN}|\widetilde{\delta}_{CWI}, \widetilde{\delta}_t\right) \times P\left(\widetilde{\delta}_{CWI}|\delta_t\right) d\widetilde{\delta}_{CWI} \tag{26}$$

on setting $X = \widetilde{\delta}_{CWIN}$, $Y = \widetilde{\delta}_t$ and $Z = \widetilde{\delta}_{CWI}$. Assuming $\widetilde{\delta}_{CWIN}$ are distributed in the same manner as noise free $\widetilde{\delta}_{CWI}$, we have the positive bounded Gaussian

$$P\left(\widetilde{\delta}_{CWIN}|\widetilde{\delta}_t, \widetilde{\delta}_{CWI}\right) = \frac{1}{\left(1 - \Phi_{\overline{\mu}_N, \overline{\sigma}_N}(0)\right)\overline{\sigma}_N\sqrt{2\pi}} e^{\frac{-\left(\widetilde{\delta}_{CWIN} - \overline{\mu}_N\right)^2}{2\overline{\sigma}_N^2}}. \tag{27}$$

Parameters $\overline{\mu}_N$ and $\overline{\sigma}_N$ are obtained by minimizing the residuals between observations and equation (27) in a least squares sense.

[34] The noisy estimates $\widetilde{\delta}_{CWIN}$ in equation (27) can be represented by

$$\widetilde{\delta}_{CWIN} = \widetilde{\delta}_{CWI} + \epsilon, \tag{28}$$

where ϵ is the error or noise. We assume that noise is zero-centered (i.e., $E(\epsilon) = 0$) and set $\widetilde{\delta}_{CWIN} \approx \widetilde{\delta}_{CWI}$ which is true on average. Consequently, equation (27) reduces to

$$P\left(\widetilde{\delta}_{CWIN}|\widetilde{\delta}_t, \widetilde{\delta}_{CWI}\right) = \frac{1}{\left(1 - \Phi_{\overline{\mu}_N, \overline{\sigma}_N}(0)\right)\overline{\sigma}_N\sqrt{2\pi}} e^{\frac{-\left(\widetilde{\delta}_{CWI} - \overline{\mu}_N\right)^2}{2\overline{\sigma}_N^2}}. \tag{29}$$

For convenience we define

$$A\left(\widetilde{\delta}_t\right) = \frac{1}{\left(1 - \Phi_{\mu_1, \sigma_1}(0)\right)\sigma_1\sqrt{2\pi}}, \tag{30}$$

$$B\left(\widetilde{\delta}_t, \widetilde{\delta}_{CWI}\right) = e^{\frac{-\left(\widetilde{\delta}_{CWI} - \mu_1\right)^2}{2\sigma_1^2}}, \tag{31}$$

$$C(\overline{\mu}_N, \overline{\sigma}_N) = \frac{1}{\left(1 - \Phi_{\overline{\mu}_N, \overline{\sigma}_N}(0)\right)\sigma_N\sqrt{2\pi}}, \tag{32}$$

and

$$D\left(\widetilde{\delta}_{CWI}, \overline{\sigma}_N, \overline{\mu}_N\right) = e^{\frac{-\left(\widetilde{\delta}_{CWI} - \overline{\mu}_N\right)^2}{2\overline{\sigma}_N^2}}. \tag{33}$$

[35] With this new notation, the desired likelihood (equation (26)) becomes

$$P\left(\widetilde{\delta}_{CWIN}|\widetilde{\delta}_t\right) = A\left(\widetilde{\delta}_t\right)C(\overline{\mu}_N, \overline{\sigma}_N)\int_0^b B\left(\widetilde{\delta}_t, \widetilde{\delta}_{CWI}\right)D\left(\widetilde{\delta}_{CWI}, \overline{\sigma}_N, \overline{\mu}_N\right)d\widetilde{\delta}_{CWI} \tag{34}$$

which is computed numerically using the trapezoidal rule with upper integration limit $b = 1.2$. We choose 1.2 because our interest is in event pairs separated by less than one wavelength and because CWI should not be applied to events at greater distances.

[36] As stated above, noise reduces waveform similarity and increases fluctuations. The noisy likelihood derived herein accounts for changes in the fluctuations and can simultaneously be used to correct for variations in Δf and $2t_w$

Table 2. Properties of the Six Velocity Models Used for the Synthetic Tests[a]

Model	μ_α (m s^{-1})	σ_α (m s^{-1})	a (m)	$v_\alpha(d)$ (m s^{-1})
1	1500	6000	1200	NA
2	500	6000	400	NA
3	180	6000	1200	NA
4	180	6000	400	NA
5	180	6000	400	$v_\alpha(d) = \begin{cases} -2000 & \text{if } d \le 1000 \text{ m} \\ 0 & \text{if } d > 1000 \text{ m,} \end{cases}$
6	180	6000	400	$v_\alpha(d) = \begin{cases} 2d - 2000 & \text{if } d \le 1000 \text{ m} \\ 0 & \text{if } d > 1000 \text{ m,} \end{cases}$

[a]Parameters μ_α, σ_α, and a are mean P wave velocity, standard deviation of P wave velocity, and correlation length, respectively. Models 1–4 are random realizations of the exponential media. A 1 km thick depth d dependant velocity layer $v_\alpha(d)$ is incorporated at the top of model 4 to create models 5 and 6. In 5 the layer is a constant 2000 m s^{-1} slower than μ_α for 4. In 6 a linear gradient is introduced such that the velocity is 2000 m s^{-1} slower at the surface and equivalent at d = 1 km.

from 1 Hz and 0.75 s. This is because noise, variation in $2t_w$ and changes in Δf all modify the size of $\widetilde{\delta}_{CWI}$ fluctuations.

4. Applications

[37] With the theoretical and noisy likelihoods in place, we now have all the pieces needed to apply the probabilistic formulation. In sections 4.1 and 4.2 we illustrate the performance of our probabilistic CWI framework using both synthetic tests and real data.

4.1. Synthetic Examples

[38] The theoretical likelihood discussed in section 3.1 forms the basis of the probabilistic inference from CWI data. Its derivation relies on analysis of 2-D acoustic experiments in Gaussian random media. Our aim however, is to apply CWI to real earthquakes whose seismic waves travel through media with different statistical properties. It is therefore important to investigate the suitability of the technique in media with different properties. In this section we perform further 2-D acoustic experiments using exponential random media.

[39] The primary difference between Gaussian and exponential random media is the fall off in spectral amplitude of velocity perturbations [*Frankel and Clayton*, 1986]. A Gaussian medium varies smoothly, whereas exponential media exhibit a greater level of roughness at smaller length scales. The exponential model corresponds to a von Karman random medium with Hurst parameter $\mu = 0.5$ [*Hong and Kennett*, 2005]. Von Karman media more accurately depict characteristics of the Earth than Gaussian random media due to improved representation of small-scale roughness [e.g., *Kamei et al.*, 2005]. Consequently, the synthetic experiments of this section use media that more closely represent the environment in which seismic waves propagate. Six new velocity models are introduced and the ability of CWI to determine event separation retested. Models 1 to 4 are created by taking realizations from exponential random media with statistical properties defined in Table 2. The remaining models, 5 and 6, are created by adding a layered velocity model $v_\alpha(d)$ to model 4.

[40] Waveforms are simulated for events separated by 57, 170 and 368 m and filtered between 1 and 5 Hz. We compute CWI estimates and treat them as noisy because a 4 Hz bandwidth differs from 1 Hz used in determining the theoretical likelihood. Computed noisy likelihoods $P(\widetilde{\delta}_{CWIN}|\widetilde{\delta}_t)$ are used with uniform priors to determine posteriors $P(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN})$ (Figure 9). These are indistinguishable for the six new models and provide the most information for the

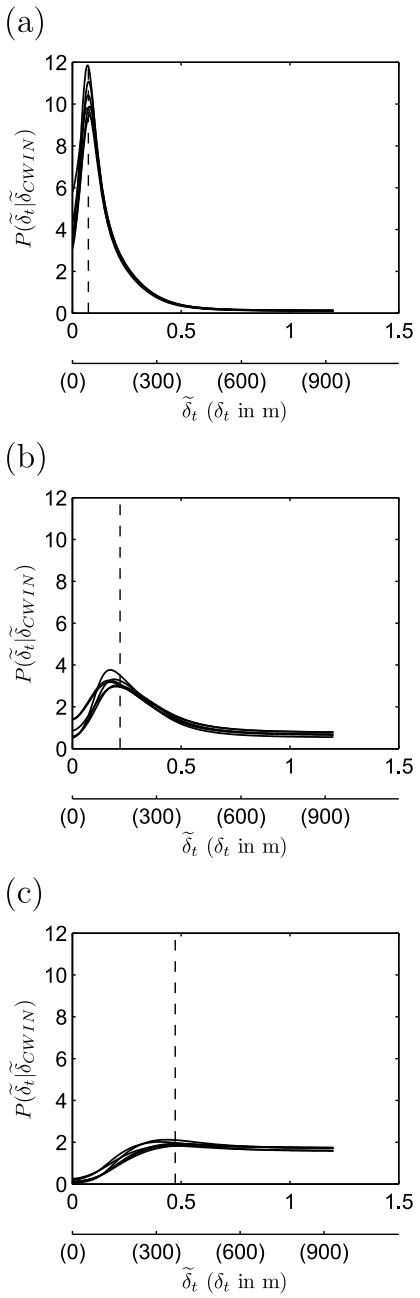

Figure 9. Posteriors $P(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN})$ for synthetic experiments in models 1 to 5 at true separations (a) 57, (b) 170, and (c) 368 m (dashed lines). The posteriors for different media are identical for practical purposes and plotted black. Two scales are provided for convenience: the unitless wavelength normalized separation $\widetilde{\delta}_t$ (top) and actual separation in meters (bottom).

Table 3. Six Calaveras Pairs Used for Comparing CWI and Travel Time Separations[a]

Pair	Event IDs	δ_{DD}(m)	$\mu_{\delta_{CWI}} \pm \sigma_{\delta_{CWI}}$ (m)
1	101362, 62514	4	[0,30]
2	103138, 62520	70	[25,107]
3	103138, 207043	130	[77,132]
4	101362, 93248	150	[36,182]
5	108512, 292015	247	[69,309]
6	103138, 292015	341	[114,233]

[a]Event IDs are those given by the Northern California Earthquake Data Center (NCEDC). The Euclidean distance between optimum travel time locations (δ_{DD}) and bounds on the CWI estimates ($\mu_{\delta_{CWI}} \pm \sigma_{\delta_{CWI}}$) are also provided. Details for individual events are provided in Table 4.

smallest separation. That is, the highest and tightest peak is observed for $\delta_t = 57$ m. It is aligned with known separation confirming the technique's accuracy for this pair. When δ_t increases to 170 m, the constraint is less precise becoming broader but remaining aligned with true separation. This implies that our approach remains accurate with uncertainty around δ_t increasing. For $\delta_t = 368$ m a prominent peak is no longer visible and CWI provides only a lower bound on separation. We can draw three conclusions from these experiments. Firstly, the probabilistic technique is not biased by exclusive use of Gaussian random media during its derivation. Secondly, presence of a 1 km thick slower layer near the surface does not influence CWI performance. Thirdly, the noisy likelihood permits use of different frequency bands Δf and/or sliding window widths $2t_w$.

4.2. Calaveras Fault

[41] The Calaveras fault is one of the most active structures in northern California [*Schaff et al.*, 2002] and has hosted several moderate sized events. *Oppenheimer et al.* [1990] list nine earthquakes between 1943 and 1988 with $4.9 \leq M_s \leq 6.2$ which were located on the Calaveras and a further 16 with similar magnitudes that occurred nearby between 1858 and 1911. The fault also hosts many smaller earthquakes [e.g., *Schaff et al.*, 2002] and is ideal for exploring the performance of CWI for three reasons. Firstly, the presence of repeating events leads to separations ranging from near zero to hundreds of meters. Consequently, it is possible to test the performance of CWI across a broad range of separations. Secondly, the large number of stations and good azimuthal coverage in Northern California represents a recording situation which exceeds that found in most tectonic settings. Therefore, we can be confident that existing locations are well constrained and provide a good basis for comparison. Finally, Calaveras earthquakes have been well researched with several studies having relocated events in the region [e.g., *Waldhauser*, 2001; *Schaff et al.*, 2002; *Waldhauser and Schaff*, 2008].

4.2.1. Application of CWI

[42] We apply CWI to six earthquake pairs on the Calaveras Fault and compare our CWI septation PDFs with results from the Double Difference algorithm or hypoDD [*Waldhauser and Ellsworth*, 2000; *Waldhauser*, 2001]. Details of the six pairs are summarized in Table 3 (see also Table 4) along with our calculated CWI separations and the Euclidean distance between optimum hypoDD locations δ_{DD}. These δ_{DD} are calculated using cross correlation computed delay times with waveforms from all available stations during simultaneous relocation of the 308 Calaveras earthquakes provided with the open source version of hypoDD. Note that the pairs are at increasing separations and that CWI and delay time separations are consistent up to 250 m, after which $\delta_{CWI} < \delta_{DD}$.

Table 4. Details of Events Used for the Pairs in Table 3 Showing NCEDC Event IDs, Depth, Date, Time, and Duration Magnitude M_d

Event ID	Depth (km)	Date	Time	M_d
101362	3.638	3 June 1987	06:58.51	1.7
62514	3.641	31 December 1985	15:50.52	1.7
62520	3.676	31 December 1985	18:03.03	1.6
207043	3.526	22 February 1991	15:51.34	2.3
93248	3.512	26 January 1987	13:47.46	2.3
108512	3.418	16 November 1987	23:19.9	2.0
292015	3.594	9 June 1992	12:59.24	2.9

[43] Before applying CWI we remove unsuitable waveforms using the conditions in Table 5. In general, the larger the magnitudes and smaller the source-to-station distance, the more likely it is that the required signal-to-noise ratio is satisfied. When considering events in this region we obtain useable pairs from stations CCO (3.5 km), JCB (19 km), JST (14 km), CMH (11 km) and HSP (24 km) for at least 50% of possible pairings. Similarly, stations JAL (21 km), CSC (9.1 km), CAD (14 km), JHL (24 km), JRR (26 km) lead to useable data for 20–50% of possible pairings. Bracketed distances are measured along the surface of the reference ellipsoid between station and mean epicenter of the cluster using the *Vincenty* [1975] inverse formula.

[44] Waveforms are filtered between 1 and 5 Hz before applying CWI and aligned to a *P* arrival at 0 s. The traces fluctuate wildly at the start (before −2.5 s) and end (after 20 s) in Figure 10. Fluctuations before −2.5 s arise from cross correlations of noise before the waves arrive and are ignored. Fluctuations at the end appear when coda strength falls to a level comparable with noise. These are also ignored. The low separation estimate immediately before $t = 0$ s is associated with the direct *P* arrivals and occurs because waves leaving the source from all directions have not had sufficient time to be backscattered toward the stations. Recall that we observed a similar effect in the synthetic experiments. In the following we consider independent CWI estimates (i.e., from non-overlapping time windows) between $2.5 \leq t \leq 20$ s. We note that in this range the δ_{CWI} traces for different stations and channels are consistent and that any individual station could be used to reliably estimate the separation.

4.2.2. Application of HypoDD

[45] Table 3 compares a range of CWI estimates with δ_{DD}. A complete comparison should include uncertainty in the hypoDD solutions. The 95% confidence intervals for the

Table 5. Conditions Used to Identify Unsuitable Waveforms Before Applying CWI

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

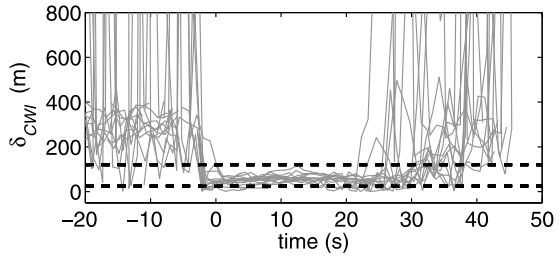

Figure 10. CWI separations (grey) as a function of sliding window centroid for pair 2 using different channels and stations. Black dashed lines represent the 95% confidence interval on separation from delay time locations with $\sigma_x = \sigma_y =$ 19.5 m and $\sigma_z =$ 15 m.

delay time based separation shown in Figure 10 are obtained by creating a separation PDF $P(\delta_t|t_{DD})$ between two earthquakes whose locations are defined by 3-D Gaussian density functions with mean given by optimum delay time locations and uncertainties σ_x, σ_y and σ_z. We derive $P(\delta_t|t_{DD})$ in Appendix C and use hypoDD with Singular Value Decomposition (SVD) to obtain formal estimates of σ_x, σ_y and σ_z from the least squares procedure. Note that when computing $P(\delta_t|t_{DD})$ we incorporate only the cross correlation measure of delay time in hypoDD and ignore the travel time picks. This is because the cross correlation technique provides the best estimate of delay time. Moreover, we consider only the two events that constitute the pair and ignore all other earthquakes in the relocation.

4.2.3. Comparison of CWI and HypoDD

[46] There are a number of differences between the hypoDD and CWI approaches. When applying hypoDD in this paper we utilize cross correlation derived measures of the delay time between recordings for waveform packages around the early onset body waves. The CWI technique however, is based on the strength of cross correlation between the coda waves of each event. We show later, how the independence of the two data sources makes CWI a compliment to delay time based location procedures such as hypoDD.

[47] A second difference between CWI and hypoDD relates to the solution. The delay time techniques are able to

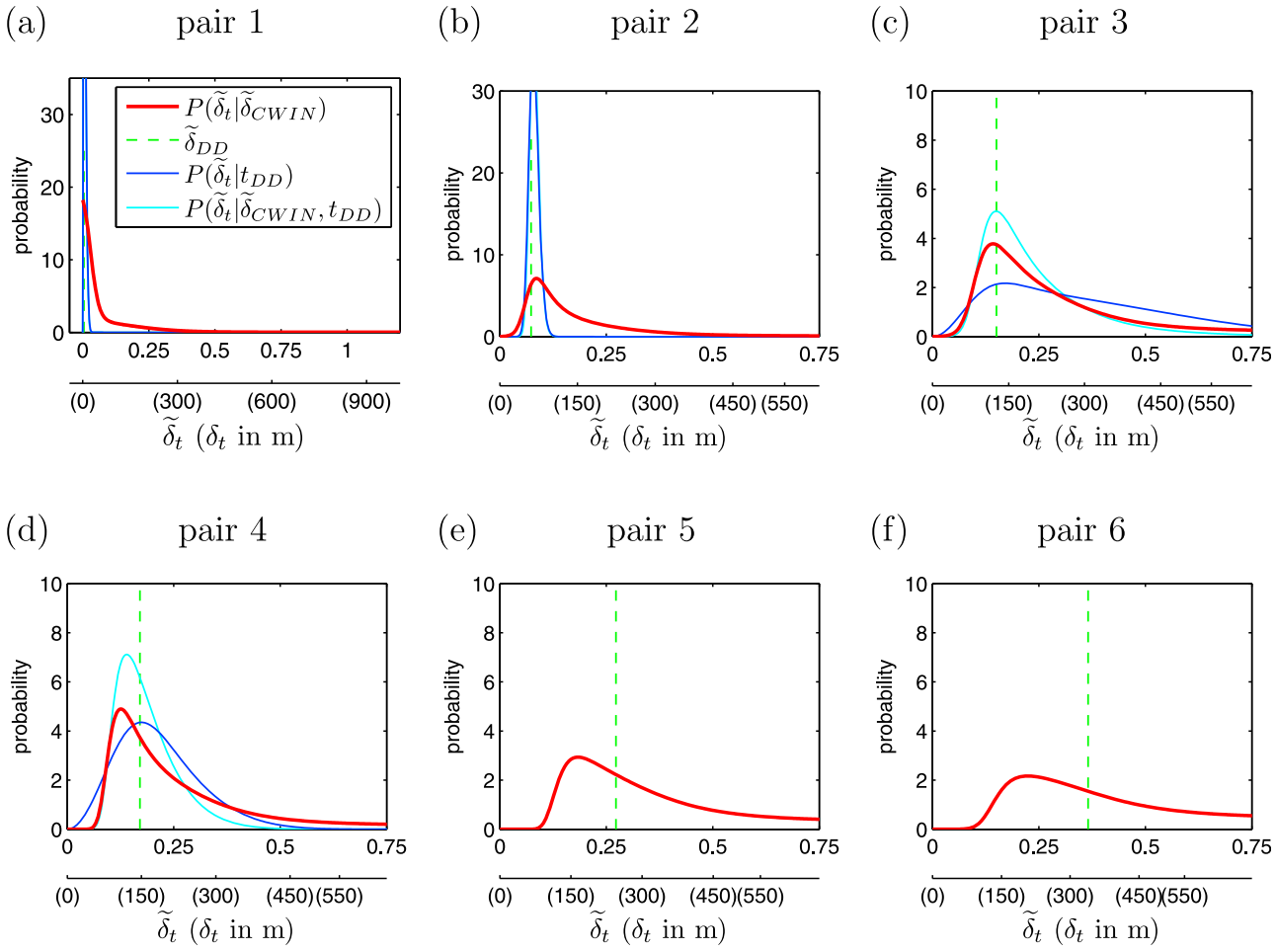

Figure 11. (a–e) Posterior functions for Calaveras pairs 1–6 (see Table 3) for example 1. $P(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN})$ and $P(\widetilde{\delta}_t|t_{DD})$ are the posteriors when coda waves and travel times are considered individually. $P(\widetilde{\delta}_t|t_{DD})$ are computed using hypoDD with all available stations for each event pair, $\widetilde{\delta}_{DD}$ is the Euclidean distance between optimum travel time locations using all 308 events and all available stations, and $P(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN}, t_{DD})$ is the posterior when all data is combined. $P(\widetilde{\delta}_t|t_{DD})$ and $P(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN}, t_{DD})$ are omitted for pairs 5 and 6 because hypoDD is unable to solve for the locations of these events because of insufficient data.

Table 6. Standard Deviation of Relative Locations for the Event Pairs in Example 1 as Determined Using HypoDD With SVD and Cross Correlation Computed Phase Differences[a]

Event IDs	σ_x	σ_y	σ_z
	Pair 1		
62514	1.3	1.1	5.3
101362	1.3	1.2	5.2
	Pair 2		
62520	5.2	5.1	12.0
103138	5.2	5.1	12.0
	Pair 3		
103138	47.9	40.6	185.3
207043	47.9	40.6	185.3
	Pair 4		
93248	45.8	44.4	92.4
101362	45.8	44.4	92.4

[a]Recall that all available stations are used in Example 1. Information is omitted for pairs 5 and 6 because hypoDD is unable to obtain a solution. Results are measured in meters.

solve for the displacement vector between events and are therefore well suited for relative and/or absolute location. To achieve this, they require at least four stations to record each event. In many cases a network of four stations is in itself insufficient however, due to problems with the waveforms (e.g., poor signal to noise) or station downtime. For the events, in this paper it was not possible to identify four stations which adequately recorded all six pairs. In contrast, the CWI technique solves only for event separation and is therefore unable to uniquely locate the events when a single pair is used.

[48] In sections 4.2.4–4.2.6 we compare our CWI posteriors for each pair to those obtained from hypoDD as well as the posterior when hypoDD and CWI data are combined. We repeat the comparison using all available data (example 1), with waveforms recorded from 6 stations (example 2) and with a single station (example 3). The hypoDD based δ_{DD} is included for reference in all examples. Recall that δ_{DD} is computed using all available data during a simultaneous inversion of multiple events and is believed to represent the best single estimate of separation from hypoDD.

4.2.4. Example 1: All Available Stations

[49] We compute posteriors $P(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN})$ for the six pairs and illustrate them in Figure 11. Unlike the 2-D acoustic

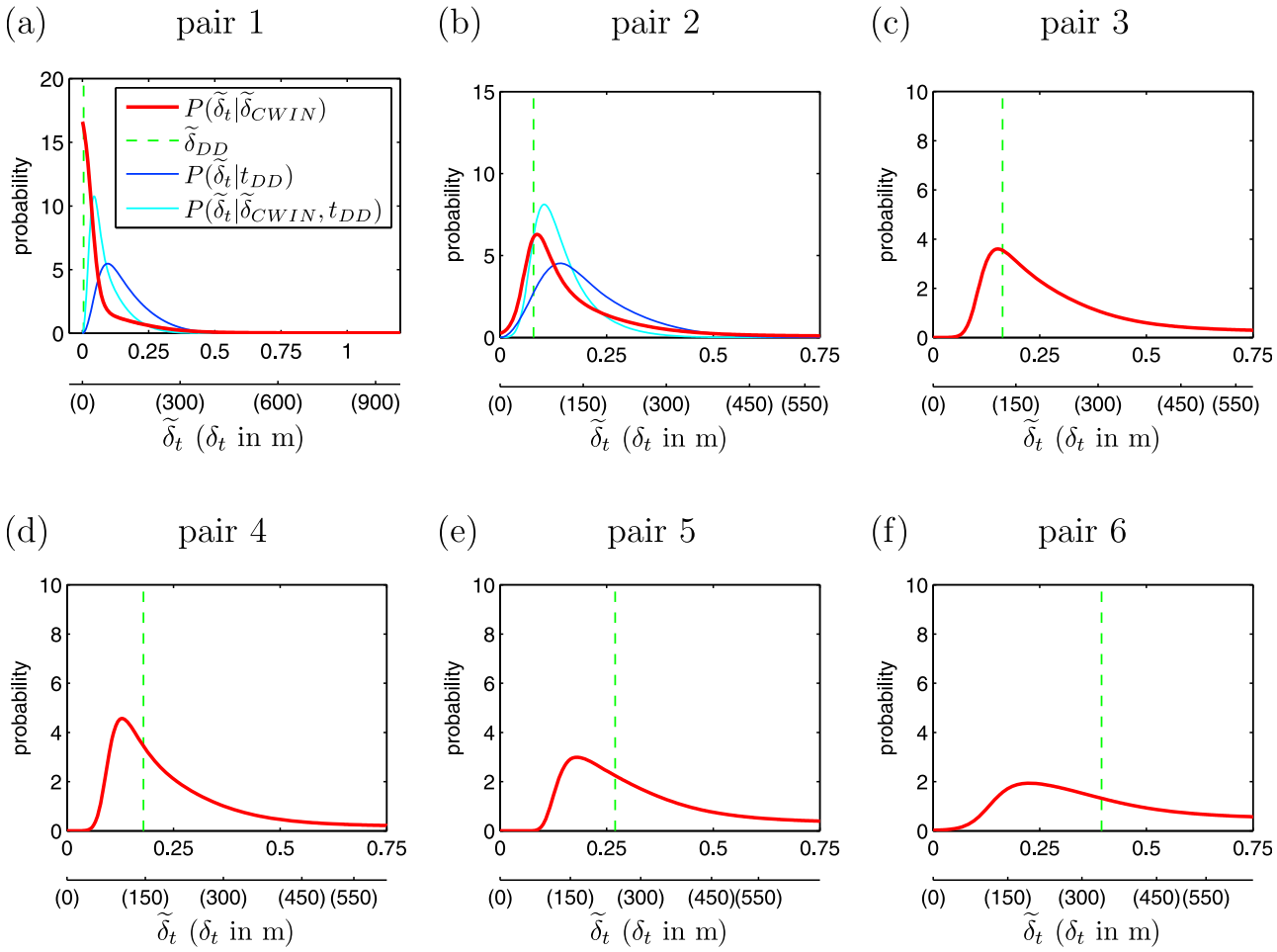

Figure 12. (a–e) Posteriors for Calaveras example 2 pairs using data from only six stations. $P(\widetilde{\delta}_t|t_{DD})$ and $P(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN}, t_{DD})$ are omitted for pairs 3–6 because hypoDD is unable to solve for the locations of these events. Remaining details as in Figure 11.

Table 7. Standard Deviation of Relative Locations for the Event Pairs in Example 2 as Determined Using HypoDD With SVD and Cross Correlation Computed Phase Differences[a]

Event IDs	σ_x	σ_y	σ_z
	Pair 1		
62514	19.3	38.4	27.9
101362	19.3	31.9	122.6
	Pair 2		
62520	30.7	31.2	104.8
103138	30.7	31.2	104.8

[a]Recall that only data from six stations (NCCCO, NCJCB, NCJST, NCCMH, NCHSP, and NCJAL) are used in Example 2. Information is omitted for pairs 3–6 because hypoDD is unable to obtain a solution. Results are measured in meters.

experiments, the exact locations of these earthquakes are unknown. Therefore, we compare the CWI posteriors with $\widetilde{\delta}_{DD}$ and $P(\widetilde{\delta}_t|t_{DD})$. Arguably, $\widetilde{\delta}_{DD}$ represents the best single separation estimate given by hypoDD. However, as stated above it is a single estimate and unlike $P(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN})$ provides no bounds on uncertainty for the separation. We compute $P(\widetilde{\delta}_t|t_{DD})$ using the hypoDD derived standard deviations σ_x, σ_y, and σ_z shown in Table 6.

[50] Figure 11 demonstrates that both hypoDD and CWI are accurate for events separated by small distances (e.g., pairs 1 to 2 with separations of less than 100 m). The hypoDD based conditional probability $P(\widetilde{\delta}_t|t_{DD})$ is higher and narrower than the CWI based $P(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN})$ however, suggesting that hypoDD provides the tighter constraint on separation for these events. The combined posterior given by

$$P(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN}, t_{DD}) \propto P(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN}) \times P(\widetilde{\delta}_t|t_{DD}) \tag{35}$$

is only marginally different from $P(\widetilde{\delta}_t|t_{DD})$ suggesting that there is little benefit in combining the CWI and delay time based data for these pairs. The individual CWI and hypoDD posteriors can be combined because they are derived from different sections of the waveform and are hence independent. Note that $P(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN}, tt_{DD})$ is renormalized over the region of interest ($0 \le \widetilde{\delta}_t \le 1.2$).

[51] Both the CWI and hypoDD posteriors become broader with lower peaks as the separation increases (i.e., pairs 3 and 4 versus pairs 1 and 2). This is because the ability of both techniques to resolve separation reduces with increasing distance. Figure 11 demonstrates that the falloff for hypoDD performance (see also Table 6) is more rapid than the CWI technique. That is, the CWI based $P(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN})$ is now marginally narrower and higher for pairs 3 and 4. The best estimate of separation for these pairs is given by the combined posterior $P(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN}, t_{DD})$. For pair 4 ($\delta_t \approx 150$ m) in Figure 11, $P(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN})$ slightly underestimates $\widetilde{\delta}_{DD}$. In this case the problem is due to a poor fit of the likelihood resulting in an underestimate of $\overline{\sigma}_N$. Fortunately, this fitting problem can be identified easily before computing the posterior. This observation suggests that care should be taken when fitting $\widetilde{\delta}_{CWIN}$ to obtain $\overline{\mu}_N$ and $\overline{\sigma}_N$, particularly if CWI is the only constraint on event separation. For pairs 5 and 6 with separations exceeding 200 m it is no longer possible to obtain a hypoDD solution using SVD with cross correlation computed phase delay because the data is insufficient to constrain the problem.

4.2.5. Example 2: Six Stations

[52] In Example 2 we simulate a less ideal recording situation by selecting only data from 6 stations (NCCCO, NCJCB, NCJST, NCCMH, NCHSP, and NCJAL) and repeating Example 1. A comparison of Figures 11 and 12 demonstrate little change in the performance of CWI for each of the six pairs. In contrast, the change in hypoDD estimates of separation are more significant. For example, the hypoDD based $P(\widetilde{\delta}_t|t_{DD})$ for pairs 1 and 2 are now lower and broader than those computed using all available stations. This broadening is directly associated with an increase in the uncertainty on the relative locations (as seen by comparing Tables 6 and 7). Consequently, the CWI technique leads to the best estimate of separation for these pairs. Furthermore, hypoDD is unable to resolve locations for the events in pairs 3 to 6.

4.2.6. Example 3: One Station

[53] In Example 3, we consider the extreme case of a single station (NCCCO) only and repeat the above experiment. For the sake of brevity, we consider only pairs 1 to 3. Obviously,

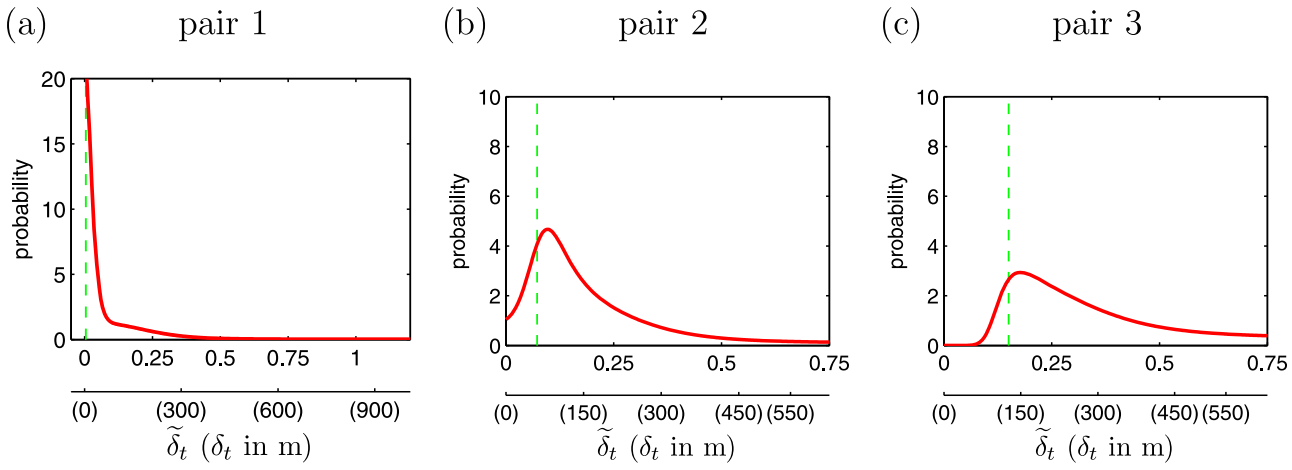

Figure 13. (a–c) Posteriors for Calaveras example 3 pairs with $P(\widetilde{\delta}_t|t_{DD})$ assuming hypoDD locations using only one station. $P(\widetilde{\delta}_t|t_{DD})$ and $P(\widetilde{\delta}_t|\widetilde{\delta}_{CWIN}, t_{DD})$ are omitted for all cases because hypoDD is unable to solve for the locations of these events. Remaining details as in Figure 11.

it is not possible to obtain any separation estimates from hypoDD because the location problem is unsolvable. In contrast, Figure 13 illustrates no noticeable change in the performance of CWI, therefore confirming our hypothesis that CWI remains useful for events recorded by a single station.

5. Conclusions

[54] Seismic coda are ignored in most applications because it is difficult to deterministically model the complexity of scattered waves. However, coda contains information that has left the source in all directions and been scattered toward the recording station. Consequently, coda from a single station contain information that has traditionally only been accessible from a number of spatially distributed stations. Coda wave interferometry (CWI) is a relatively new technique which allows an estimation of event separation directly from the coda.

[55] In this paper we lay the statistical foundation for the use of coda waves in earthquake location problems by extending CWI theory and expanding the algorithms available for analyzing event pairs. In particular, we extend the range of applicability by 50% (i.e., 300 to 450 m for 1 to 5 Hz waveforms) and introduce a wavelength normalized measure of separation which facilitates a uniform treatment of CWI across all frequencies. We formulate a probabilistic theoretical likelihood function which explains the CWI data and provides a mechanism for producing probabilistic constraints on event separation. This facilitates the treatment of uncertainties in CWI estimates and puts the probabilistic constraints on a quantitative footing allowing the combination of different stations and events. Moreover, this property of the posterior functions leads to a convenient approach for combining different types of data as demonstrated in this paper with the combined CWI (coda waves) and hypoDD (delay time) examples.

[56] We extend our probabilistic approach to waveforms containing noise and apply it to earthquake pairs on the Calaveras Fault, California. The CWI results are compared to separations derived from hypoDD, an existing algorithm for relative earthquake location [*Waldhauser and Ellsworth*, 2000]. The CWI technique leads to poorer estimates of separation for earthquakes recorded by a large number of stations and separated at small distances (<70 m). As the separation increases to ≈150 m there is less difference between the two techniques. In contrast, CWI outperforms hypoDD at all distances in poorer recording situations (e.g., Examples 2 and 3 with 6 and 1 station(s), respectively).

[57] It is important to note that coda waves are also sensitive to temporal variation in velocity [e.g., *Poupinet et al.*, 1984; *Ratdomopurbo and Poupinet*, 1995; *Grêt et al.*, 2005] or differences in source mechanism [*Robinson et al.*, 2007a]. Care should be taken when applying CWI that the observed signals are not corrupted by other variations. For example, it may be useful to limit the application of CWI in fault zones to moderate sized events observed between large earthquakes. This is because larger events can create significant variation in near fault velocity structure.

[58] Employing coda waves in earthquake location is a topic of ongoing research. We argue that the techniques introduced in this paper will be useful in constraining earthquake locations and aftershock sequences, particularly when only a small number of stations are available. For example, in intraplate regions such as Australia where only 60% of the earthquakes are located within an uncertainty of 10 km or less [*Leonard*, 2008]. This is because it is possible to obtain a detailed understanding about the separation between two earthquakes from the coda waves recorded at a single station. Moreover, the ability to combine coda waves and information from early onset body waves allows us to utilize more of the recorded data. Another feature of this process is that it is possible to combine events recorded by different networks provided one or more stations have recorded all events. This makes it possible to combine permanent network monitored events with those recorded using temporary arrays. Consequently, we suggest that it will become possible to locate earthquakes with the accuracy of delay times when a well designed temporary array is in place and combine them with earthquakes recorded by a sparse permanent network whose locations can be controlled by the coda waves. These remain directions for further studies.

Appendix A

[59] In this appendix we provide a derivation of equation (9) (see *Snieder* [2006] for more details). We begin by substituting equations (1) and (2) into equation (4) which gives

$$C_{\widetilde{uu}}^{(t,t_w)}(t_s) = \int_{t-t_w}^{t+t_w} \sum\nolimits_T \mathbf{A}_T(t') \sum\nolimits_{\widetilde{T}} \mathbf{A}_{\widetilde{T}}(t'+t_s-\tau_T)dt'$$
$$= \sum\nolimits_T \sum\nolimits_{\widetilde{T}} \int_{t-t_w}^{t+t_w} \mathbf{A}_T(t')\mathbf{A}_{\widetilde{T}}(t'+t_s-\tau_T)dt'. \quad \text{(A1)}$$

This becomes

$$C_{\widetilde{uu}}^{(t,t_w)}(t_s) = \sum_T \int_{t-t_w}^{t+t_w} \mathbf{A}_T(t')\mathbf{A}_T(t'+t_s-\tau_T)dt' \quad \text{(A2)}$$

if we assume that only the paths $T = \widetilde{T}$ interfere constructively and that paths $T \neq \widetilde{T}$ cancel on average. Note that the integral

$$C_T(t_s-\tau_T) = \int_{t-t_w}^{t+t_w} \mathbf{A}_T(t')\mathbf{A}_T(t'+t_s-\tau_T)dt' \quad \text{(A3)}$$

is the cross correlation of waves that have traveled along trajectory $T = \widetilde{T}$. We follow a similar treatment for the denominator of equation (4) and the normalized cross correlation becomes

$$R^{(t,t_w)}(t_s) = \frac{\sum_T C_T(t_s-\tau_T)}{\sum_T C_T(0)}. \quad \text{(A4)}$$

[60] *Snieder* [2006] explains that $C_T(t)$ can be related to the auto correlation function $C(t)$ by

$$C_T(t) = I_t C(t), \quad \text{(A5)}$$

where $C(t)$ is normalized to a maximum of 1 at $t = 0$ (i.e., $C(0) = 1$) and I_T accounts for the intensity of the wave propagating along T. The assumption necessary in deriving equation (A5) is that the power spectra of the reference and

perturbed waveforms are comparable, a requirement that can be satisfied by filtering. Thus, equation (A4) becomes

$$R^{(t,t_w)}(t_s) = \frac{\sum_T I_T C(t_s - \tau_T)}{\sum_T I_T}. \quad \text{(A6)}$$

[61] Replacing $C(t)$ with its second-order Taylor series expansion near $t = 0$ gives

$$C(t) = C(t=0) + t\dot{C}(t=0) + \frac{1}{2}t^2\ddot{C}(t=0), \quad \text{(A7)}$$

where $\dot{C}$ and $\ddot{C}$ represent the first- and second-order derivatives of $C(t)$ with respect to t, respectively. Recognizing that $C(0) = 1$ and that the autocorrelation function is even we have

$$C(t) = 1 + \frac{1}{2}t^2\ddot{C}(t=0). \quad \text{(A8)}$$

Snieder [2006] shows that

$$\ddot{C}(t=0) = -\overline{\omega^2}, \quad \text{(A9)}$$

where $\overline{\omega^2}$ is given by equation (10).

[62] Substituting equations (A9) and (A8) into equation (A6) gives

$$R^{(t,t_w)}(t_s) = 1 - \frac{1}{2}\overline{\omega^2}\frac{\sum_T (t_s - \tau_T)^2 I_T}{\sum_T I_T}. \quad \text{(A10)}$$

It can be shown by differentiating equation (A10) with respect to t_s that $R^{(t,tw)}(t_s)$ attains its maximum when

$$t_s = \frac{\sum_T \tau_T I_T}{\sum_T I_T}, \quad \text{(A11)}$$

which *Snieder* [2006] denotes $\langle\tau\rangle$ in recognition that it is the average travel time perturbation for all trajectories weighted by the intensity of each trajectory. We can now write

$$R_{\max} = 1 - \frac{1}{2}\overline{\omega^2}\frac{\sum_T (\tau_T - \langle\tau\rangle)^2 I_T}{\sum_T I_T}, \quad \text{(A12)}$$

which becomes equation (9) in recognition that the standard deviation of the travel time perturbation σ_τ is given by

$$\sigma_\tau^2 = \langle(\tau_T - \langle\tau\rangle)^2\rangle = \frac{\sum_T (\tau_T - \langle\tau\rangle)^2 I_T}{\sum_T I_T}. \quad \text{(A13)}$$

Appendix B

[63] Here, we prove a probabilistic relationship which is required for our Bayesian formulation. For a joint probability density $P(X, Y, Z)$ over random variables X, Y, and Z we can define a marginal probability density for any subset. For example, the marginal probability density for (X, Y) is given by

$$P(X,Y) = \int P(X,Y,Z)dZ. \quad \text{(B1)}$$

[64] Furthermore, the product rule of probability tells us that

$$P(X,Y) = P(X|Y)P(Y) \quad \text{(B2)}$$

for two variables, or similarly,

$$P(X,Y,Z) = P(X,Z|Y)P(Y) \quad \text{(B3)}$$

and

$$P(X,Z|Y) = P(X|Y,Z)P(Z,Y) \quad \text{(B4)}$$

for three variables. Substituting equations (B2) and (B3) into the left and right sides of equation (B1), gives

$$P(X|Y)P(Y) = \int P(X,Z|Y)P(Y)dZ \quad \text{(B5)}$$

which simplifies to

$$P(X|Y) = \int P(X,Z|Y)dZ \quad \text{(B6)}$$

and, ultimately

$$P(X|Y) = \int P(X|Z,Y)P(Z|Y)dZ \quad \text{(B7)}$$

upon using equation (B4).

Appendix C

[65] Here, we derive a PDF for the separation R between two events $\mathbf{e}_1$ and $\mathbf{e}_2$ with locations defined by 3-D multivariate Gaussians. We denote the PDF for the location of $\mathbf{e}_1$ as follows

$$f(X_1,Y_1,Z_1) = \frac{1}{(2\pi)^{\frac{3}{2}}\sqrt{|\Sigma|}}\exp\left(-\frac{1}{2}\left([\mathbf{e}_1 - \mu_{\mathbf{e}_1}]\Sigma^{-1}[\mathbf{e}_1 - \mu_{\mathbf{e}_1}]^T\right)\right), \quad \text{(C1)}$$

where $\mathbf{e}_1 = (x_1, y_1, z_1)$ is the location of the event, $\mu_{\mathbf{e}1} = (\mu_{x1}, \mu_{y1}, \mu_{z1})$ is the mean location, and

$$\Sigma = \begin{pmatrix} \sigma_{x_1}^2 & 0 & 0 \\ 0 & \sigma_{y_1}^2 & 0 \\ 0 & 0 & \sigma_{z_1}^2 \end{pmatrix} \quad \text{(C2)}$$

is the covariance matrix. The omission of off-diagonal terms in equation (C2) implies that individual coordinates of the location are not correlated. That is, the marginals of each coordinate for $\mathbf{e}_1$ can be written as Gaussians of form $X_1 \sim N(\mu_{x_1}, \sigma_{x_1})$, $Y_1 \sim N(\mu_{y_1}, \sigma_{y_1})$ and $Z_1 \sim N(\mu_{z_1}, \sigma_{z_1})$ with a similar set of equations for $\mathbf{e}_2$.

[66] The separation between $\mathbf{e}_1$ and $\mathbf{e}_2$ is given by R, where

$$R^2 = X^2 + Y^2 + Z^2, \quad \text{(C3)}$$

and $X = X_1 - X_2$, $Y = Y_1 - Y_2$ and $Z = Z_1 - Z_2$. The PDF for X, the difference between two Gaussian distributed random

variables, is also Gaussian and can be denoted $X \sim N(\mu_x, \sigma_x)$, with $\mu_x = \mu_{x_1} - \mu_{x_2}$ and $\sigma_x^2 = {\sigma_{x_1}}^2 + {\sigma_{x_2}}^2$ (see *Sivia and Skilling* [2006] for derivation). Similar expressions can be written for Y and Z. Since, X, Y and Z are independent of one another, we can build the multivariate Gaussian

$$f(X,Y,Z) = \frac{1}{(2\pi)^{\frac{3}{2}}\sqrt{|\Sigma|}} \cdot \exp\left[-\frac{1}{2}\left(\begin{bmatrix} X-\mu_x \\ Y-\mu_y \\ Z-\mu_z \end{bmatrix} \Sigma^{-1} \begin{bmatrix} X-\mu_x, Y-\mu_y, Z-\mu_z \end{bmatrix}\right)\right], \tag{C4}$$

where

$$\Sigma = \frac{1}{|\Sigma|}\begin{pmatrix} \sigma_x^2 & 0 & \\ 0 & \sigma_y^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{pmatrix} \tag{C5}$$

leads to

$$|\Sigma| = \sigma_x^2\sigma_y^2\sigma_z^2, \tag{C6}$$

and

$$\Sigma^{-1} = \begin{pmatrix} \sigma_z^2\sigma_y^2 & 0 & \\ 0 & \sigma_z^2\sigma_x^2 & 0 \\ 0 & 0 & \sigma_x^2\sigma_y^2 \end{pmatrix}. \tag{C7}$$

Substituting equations (C6) and (C7) into (C4) gives

$$f(X,Y,Z) = \frac{1}{(2\pi)^{\frac{3}{2}}\sigma_x\sigma_y\sigma_z} \cdot \exp\left[\frac{-\left(\sigma_y^2\sigma_z^2(X-\mu_x)^2 + \sigma_x^2\sigma_z^2(Y-\mu_y)^2 + \sigma_x^2\sigma_y^2(Z-\mu_z)^2\right)}{2\sigma_x^2\sigma_y^2\sigma_z^2}\right]. \tag{C8}$$

[67] To obtain the PDF for R we perform a transformation to spherical coordinates using

$$f(R,\theta,\phi) = \left|\frac{\partial(x,y,z)}{\partial(r,\theta,\phi)}\right| f(X,Y,Z), \tag{C9}$$

where $x = r\sin(\theta)\cos(\phi)$, $y = r\sin(\theta)\sin(\phi)$, $z = r\cos(\theta)$, $r \geq 0$, $0 < \theta \leq \pi$, $0 < \phi \leq 2\pi$ and

$$\left|\frac{\partial(x,y,z)}{\partial(r,\theta,\phi)}\right| = r^2\sin\theta. \tag{C10}$$

Evaluation of equation (C9) gives

$$f(R,\theta,\phi) = \frac{1}{(2\pi)^{\frac{3}{2}}\sigma_x\sigma_y\sigma_z} r^2\sin\theta \cdot \exp[-A(B(r,\theta,\phi) + C(r,\theta,\phi) + D(r,\theta)], \tag{C11}$$

where

$$A = -\frac{1}{2\sigma_x^2\sigma_y^2\sigma_z^2}, \tag{C12}$$

$$B(r,\theta,\phi) = \sigma_y^2\sigma_z^2(r\sin\theta\cos\phi - \mu_x), \tag{C13}$$

$$C(r,\theta,\phi) = \sigma_x^2\sigma_z^2(r\sin\theta\sin\phi - \mu_y), \tag{C14}$$

and

$$D(r,\theta) = \sigma_x^2\sigma_y^2(r\cos\theta - \mu_z). \tag{C15}$$

Finally, the desired PDF $f(R)$ is given by

$$f(R) = \int_0^{2\pi}\int_0^{\pi} f(R,\theta,\phi)\mathrm{d}\theta\mathrm{d}\phi, \tag{C16}$$

which can be computed numerically.

[68] **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 of Geoscience Australia. Work was conducted as part of an Australian Research Council Discovery Project (DP0665111). Heiner Igel is thanked for providing the original MATLAB 2-D acoustic wave equation solver which we recoded in Fortran for improved speed. 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. We also acknowledge Felix Waldhauser and William Ellsworth, the authors of the openly available Double Difference location algorithm, hypoDD. This paper has benefited significantly from reviews by David Burbidge, Marthijn de Kool, and Jane Sexton at Geoscience Australia, two anonymous reviewers, and the Associate Editor, Felix Waldhauser.

References

Abubakirov, I. R., and A. A. Gusev (1990), Estimation of scattering properties of lithosphere of Kamchatka based on Monte-Carlo simulation of record envelope of a near earthquake, *Phys. Earth Planet. Inter.*, *64*, 52–67.

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

Aki, K. (1992), Scattering conversions from P to S versus S to P, *Bull. Seismol. Soc. Am.*, *82*, 1969–1972.

Aki, K., and B. Chouet (1975), Origin of coda waves: Source, attenuation, and scattering effects, *J. Geophys. Res.*, *80*(23), 3322–3342.

Baig, A. M., and F. A. Dahlen (2004), Statistics of traveltimes and amplitudes in random media, *Geophys. J. Int.*, *158*, 187–210.

Courant, R., K. Friedrichs, and H. Lewy (1928), Über die partiellen Differenzengleichungen der mathematischen Physik, *Math. Ann.*, *100*, 32–74.

Douma, H., and R. Snieder (2006), Correcting for bias due to noise in coda wave interferometry, *Geophys. J. Int.*, *164*, 99–108.

Frankel, A., and R. W. Clayton (1986), Finite difference simulations of seismic scattering: Implications for the propagation of short-period seismic waves in the crust and models of crustal heterogeneity, *J. Geophys. Res.*, *91*(B6), 6465–6489.

Grêt, A., R. Snieder, R. C. Aster, and P. R. Kyle (2005), Monitoring rapid temporal change in a volcano with coda wave interferometry, *Geophys. Res. Lett.*, *32*, L06304, doi:10.1029/2004GL021143.

Grêt, A., R. Snieder, and J. Scales (2006), Time-lapse monitoring of rock properties with coda wave interferometry, *J. Geophys. Res.*, *111*, B03305, doi:10.1029/2004JB003354.

Hong, T. K., and B. L. N. Kennett (2005), Stochastic features of scattering, *Phys. Earth Planet. Inter.*, *148*, 131–148.

Hoshiba, M. (1991), Simulation of multiple-scattered coda wave excitation based on the energy conservation law, *Phys. Earth Planet. Inter.*, *67*, 123–136.

Kamei, R., M. Hato, and T. Matsuoka (2005), Random heterogeneous model with bimodal velocity distribution for Methane Hydrate exploration, *Explor. Geophys.*, *36*, 41–49.

Leonard, M. (2008), One hundred years of earthquake recording in Australia, *Bull. Seismol. Soc. Am.*, *98*(3), 1458–1470, doi:10.1785/0120050193.

Margerin, L., M. Campillo, N. M. Shapiro, and B. Van Tiggelen (1999), Residence time of diffuse waves in the crust as a physical interpretation of coda Q: Application to seismograms recorded in Mexico, *Geophys. J. Int.*, *138*, 343–352.

Margerin, L., M. Campillo, and B. Van Tiggelen (2000), Monte carlo simulation of multiple scattering of elastic waves, *J. Geophys. Res.*, *105*(B4), 7873–7892.

Oppenheimer, D. H., W. H. Bakun, and A. G. Lindh (1990), Slip partioning of the Calaveras Fault, California, and prospects for future earthquakes, *J. Geophys. Res.*, *95*, 8483–8498.

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

Ratdomopurbo, A., and G. Poupinet (1995), Monitoring a temporal change of seismic velocity in a volcano: Application to the 1992 eruption of Mt. Merapi (Indonesia), *Geophys. Res. Lett.*, *22*(7), 775–778.

Roberts, P. M., W. S. Phillips, and M. C. Fehler (1992), Development of the active doublet method for measuring small velocity and attenuation changes in solids, *J. Acoust. Soc. Am.*, *91*(6), 3291–3302.

Robinson, D. J. (2010), Studies on earthquake location and source determination using coda waves, Ph.D. thesis, Aust. Natl. Univ., Canberra.

Robinson, D. J., R. Snieder, and M. Sambridge (2007a), Using coda wave interferometry for estimating the variation in source mechanism between double couple events, *J. Geophys. Res.*, *112*, B12302, doi:10.1029/2007JB004925.

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

Sato, H. (1988), Temporal change in scattering and attenuation associated with the earthquake occurence—A review of recent studies on coda waves, *Pure Appl. Geophys.*, *126*, 465–497.

Sato, H., and M. C. Fehler (1998), *Seismic Wave Propagation and Scattering in the Heterogeneous Earth*, Springer, New York.

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

Sivia, D. S., and J. Skilling (2006), *Data Analysis: A Bayesian Tutorial*, 2nd ed., Oxford Univ. Press, New York.

Snieder, R. (1999), Imaging and averaging in complex media, in *Diffuse Waves in Complex Media*, *NATO ASI Ser. C*, vol. 531, edited by J. P. Fouque, pp. 405–454, Kluwer Acad., Dordrecht, Netherlands.

Snieder, R. (2004), Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase, *Phys. Rev. E*, *69*, 046610, doi:10.1103/PhysRevE.69.046610.

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

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

Snieder, R., A. Grêt, H. Douma, and J. Scales (2002), Coda wave interferometry for estimating nonlinear behaviour in seismic velocity, *Science*, *295*, 2253–2255, doi:10.1126/science.1070015.

Vincenty, T. (1975), Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations, *Surv. Rev.*, *23*(176), 88–93.

Waldhauser, F. (2001), HypoDD: A program to compute double-difference hypocenter locations (hypoDD version 1.0, 03/2001), *U.S. Geol. Surv. Open File Rep., 01-113*.

Waldhauser, F., and W. L. Ellsworth (2000), A double-difference earthquake location algorithm: Method and application to the northern Hayward Fault, California, *Bull. Seismol. Soc. Am.*, *90*(6), 1353–1368.

Waldhauser, F., and D. P. Schaff (2008), Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods, *J. Geophys. Res.*, *133*, B08311, doi:10.1029/2007JB005479.

D. J. Robinson, Risk and Impact Analysis Group, Geoscience Australia, GPO Box 378, Canberra, ACT 2601, Australia. (david.robinson@ga.gov.au)

M. Sambridge, Research School of Earth Sciences, Australian National University, Bldg. 61, Mills Road, Canberra, ACT 0200, Australia. (malcolm.sambridge@anu.edu.au)

R. Snieder, Department of Geophysics, Colorado School of Mines, 1500 Illinois St., Golden, CO 80401, USA. (rsnieder@mines.edu)