Correcting for bias due to noise in coda wave interferometry

Huub Douma and Roel Snieder
Center for Wave Phenomena, Colorado School of Mines, Golden, CO 80401-1887, USA

ABSTRACT
Coda wave interferometry (CWI) utilizes multiply scattered waves to diagnose small changes in a medium by using the scattering medium as an interferometer. Since the medium is usually stationary over the duration of a seismic experiment, different (non-overlapping) time windows in the coda allow for independent estimates of the medium perturbation. If the seismograms are contaminated with noise, only those time windows can be used for which the amplitude of the coda is above the ambient noise level. This limits the usable number of independent time windows. Here, we show how bias due to noise in CWI can be accounted for, by deriving a correction factor for the cross-correlation coefficient. This correction factor allows more time windows further into the decaying coda to be used, and hence allows for a reduction of the error bars on the medium perturbation estimates. We demonstrate the validity of this correction factor by using data from a numerical experiment and field measurements. These experiments involve the displacement of point scatterers and a change in the source location, respectively. The application of our correction factor is not limited to CWI, but can be used to correct for bias induced by noise in any application that uses cross-correlation between different signals that are contaminated with noise.

Key words: Scattering, seismic coda, monitoring, noise, earthquake location

Introduction

Multiply scattered wavefields have been experimentally shown to be remarkably stable with respect to perturbations of the boundary conditions of experiments with multiply scattered waves (Derode *et al.*, 1995; Derode *et al.*, 1999). Due to this stability, the information carried by multiply scattered waves has been successfully used in an industrial context [e.g. Fink (1997)]. Coda wave interferometry (Snieder *et al.*, 2002; Snieder, 2002; Grêt *et al.*, 2004b; Grêt *et al.*, 2004c; Snieder, 2004a) uses multiply scattered waves to detect small changes by using the scattering medium as an interferometer. Since multiply scattered waves dominate the final portions of a seismogram, they are usually referred to as coda waves just as, in musical notation, the coda denotes the closing part of a musical piece. Hence the name *coda wave interferometry* (CWI). Since CWI uses multiply scattered waves, it is inherently more sensitive to changes in the medium than are techniques based on single scattering, as multiply scattered waves sense changes in the medium multiple times.

In parallel, but independently, diffusing acoustic wave spectroscopy (DAWS) (Page *et al.*, 2000; Cowan *et al.*, 2002) was developed as the classical equivalent of diffusing wave spectroscopy (DWS) (Maret & Wolf, 1987; Pine *et al.*, 1988; Yodh *et al.*, 1990; Weitz & Pine, 1993). In DWS light is used to study different aspects of strongly scattering media, whereas in DAWS classical waves are used to probe such media. DWS has been used in many applications such as, e.g., determining the aging of foams, particle sizing, and determining the motion of particles in fluidized suspensions on ångström length scales (Weitz & Pine, 1993). So far, DAWS has mainly been used to determine the relative mean square displacement of fluidized suspensions of particles (Page *et al.*, 1999; Cowan *et al.*, 2000; Page *et al.*, 2000; Cowan *et al.*, 2002). CWI has been succesfully used to measure the nonlinear dependence of seismic velocity in rocks on

temperature (Snieder *et al.*, 2002), to monitor volcanos (Grêt *et al.*, 2004a), and to estimate source displacement (Snieder & Vrijlandt, 2004). The ability to use CWI to determine the relative mean square displacement of point scatterers from noise-free waveforms was established using a numerical experiment by (Snieder *et al.*, 2002). Both CWI and DAWS use the amplitude information as well as the phase information of the multiply scattered wavefields, and are both based on a path summation approach to model the multiply scattered wavefields. Hence, both methods are in principle the same, but have been used for different applications.

CWI is based on a measure of cross-correlation between multiply scattered wavefields recorded before and after a medium has changed. The cross-correlation coefficients calculated for different (non-overlapping) time windows provide independent estimates of the medium perturbations. These independent estimates in turn allow for the calculation of error estimates of the perturbation; the larger the number of independent (unbiased) measurements, the smaller the error estimates.

When the coda is contaminated with noise, the number of independent time windows that can be used is limited to traveltimes where the ambient noise level is small compared to the amplitudes of the multiply scattered waves. To be able to use as many independent time windows as possible, and hence reduce the error bars on the inferred perturbation, it is important to correct the cross-correlation function for the bias caused by noise. In this paper, we show how this bias in CWI can be corrected for, by deriving a correction factor for the cross-correlation coefficient. We demonstrate its validity by using data from a numerical experiment involving the displacement of point scatterers, and field data involving the displacement of the source location.

The organization of this paper is as follows. We first review the principles of CWI and then derive the noise correction factor. Subsequently we show the validity of the correction factor using the numerical and field experiments mentioned above. We conclude with a short discussion of the results. In appendix A we derive a condition for the reliability of the correction factor. This condition can be used to determine the time windows in the coda where the correction factor is reliable.

Coda wave interferometry

Coda wave interferometry is based on a path summation approach, which is a generalization of the Neumann series solution of the Lippmann-Schwinger equation for scattering of classical waves [e.g. Snieder (1999)]. This path summation can be represented as

$$u_u(t) = \sum_T A_T(t)\,, \tag{1}$$

where T denotes the different trajectories the waves have traveled, and the function $A_T(t)$ denotes the contribution of trajectory T to the multiply scattered wavefield. The subscript in $u_u(t)$ denotes that the wavefield is recorded from the unperturbed medium. Here, for simplicity, we treat acoustic waves, but the theory has been generalized to elastic waves (Snieder, 2002).

Now suppose the medium has changed, and that the main contribution of the change to the wavefields is a perturbation of the traveltime of each trajectory T. [This is obviously not universally true for any type of perturbation of the medium, and limits the applicability of our subsequent results to certain types of perturbations. For the perturbations studied in this paper however, i.e., small displacement of the scatterers and change in the source location, this turns out to be an acceptable assumption.] Using this, we can write the multiply scattered wavefield in the perturbed medium, denoted by $u_p(t)$, as

$$u_p(t) = \sum_T A_T(t-\tau_T)\,, \tag{2}$$

with τ_T the change in the traveltime for trajectory T due to the perturbation in the medium.

To measure the change in waveform due to the medium perturbation, we define the correlation coefficient $r(t,t_s,t_w)$ as

$$r(t,t_s,t_w) \equiv \frac{(u_u,u_p)_{(t,t_s,t_w)}}{\sqrt{(u_u,u_u)_{(t,0,t_w)}\,(u_p,u_p)_{(t,0,t_w)}}}\,, \tag{3}$$

with

$$(u_u,u_p)_{(t,t_s,t_w)} \equiv \int_{t-t_w}^{t+t_w} u_u(t')u_p(t'+t_s)dt'\,, \tag{4}$$

where t_s is the time-shift in the cross-correlation, and $2t_w$ is the length of the time window. The maximum of this correlation coefficient occurs at time-shift $t_s = \langle\tau\rangle$, where $\langle\tau\rangle$ is the mean traveltime change given by (Snieder, 2002)

$$\langle\tau\rangle \equiv \frac{\sum_T A_T^2\tau_T}{\sum_T A_T^2}\,. \tag{5}$$

Using equations (1) and (2) in equation (3), Snieder *et al.* (2002) and Snieder (2002) show that the maximum value of this cross-correlation function, $r_{\max}^{(t,t_w)}$, is given by

$$r_{\max}^{(t,t_w)} = 1 - \frac{1}{2}\overline{\omega^2}\sigma_\tau^2\,, \tag{6}$$

where the frequency $\overline{\omega^2}$ is given by

$$\overline{\omega^2} \equiv \frac{\int_{t-t_w}^{t+t_w}\dot{u}_u^2(t')dt'}{\int_{t-t_w}^{t+t_w}u_u^2(t')dt'}\,. \tag{7}$$

Here, $\dot{u}$ denotes the time derivative of $u(t)$. The variance of the traveltime perturbation, σ_τ^2, is given by

$$\sigma_\tau^2 \equiv \frac{\sum_T A_T^2\,(\tau_T-\langle\tau\rangle)^2}{\sum_T A_T^2}\,. \tag{8}$$

In expression (8) and equation (5), the summations concern only trajectories with arrival times in the window with central window time t and window duration $2t_w$. In deriving equation (6), it is assumed that the double sums over trajectories T and T' with $T \neq T'$ are on average incoherent, and used a second-order Taylor expansion of the cross-correlation function in the quantity $\tau_T - \langle\tau\rangle$ (Snieder, 2002). The latter approximation is valid when the traveltime perturbation due to the medium perturbation is smaller than the dominant period of the multiply scattered wavefields. Snieder (2004b) showed that the double sums over trajectories T and T' with $T \neq T'$, are proportional to $1/\sqrt{t_w \Delta f}$, with Δf the bandwidth of the signal, so that their importance thus reduces for larger time windows and larger bandwidth data.

Correcting for the bias due to noise

So far, we assumed the waveforms $u_u(t)$ and $u_p(t)$ to be free of noise. To study the influence of noise on CWI, we derive a correction factor for the correlation coefficient when the waveforms $u_u(t)$ and $u_p(t)$ are contaminated with noise. We define

$$u'_u(t) = u_u(t) + n_u(t)\,, \quad u'_p(t) = u_p(t) + n_p(t)\,, \tag{9}$$

where $u'_u(t)$ and $u'_p(t)$ are the noise-contaminated signals, $u_u(t)$ and $u_p(t)$ the noise-free waveforms, and $n_u(t)$ and $n_p(t)$ the noise signals for the unperturbed and perturbed wavefields, respectively. Using the noise-contaminated signals, we define the noise-contaminated cross-correlation coefficient as

$$r'(t,t_s,t_w) = \frac{\left(u'_u, u'_p\right)_{(t,t_s,t_w)}}{\sqrt{\left(u'_u, u'_u\right)_{(t,0,t_w)}\left(u'_p, u'_p\right)_{(t,0,t_w)}}}\,. \tag{10}$$

Our aim is to derive a correction factor $c(t,t_s,t_w)$ such that

$$r(t,t_s,t_w) \approx c(t,t_s,t_w) r'(t,t_s,t_w)\,. \tag{11}$$

Throughout the remaining derivation, we assume that these noise signals are identically and independently distributed (i.i.d.), and wide-sense stationary with zero mean [i.e., the mean is zero for all times, and the autocorrelation depends only on the time shift t_s (Papoulis, 1991, p. 298)].

Using equation (9), we find

$$\begin{aligned}
\left(u'_u, u'_p\right)_{(t,t_s,t_w)} &= (u_u,u_p)_{(t,t_s,t_w)} + (u_u,n_p)_{(t,t_s,t_w)} \\
&\quad + (u_p,n_u)_{(t,t_s,t_w)} \\
&\quad + (n_u,n_p)_{(t,t_s,t_w)}\,, \qquad (12)\\
\left(u'_u, u'_u\right)_{(t,0,t_w)} &= (u_u,u_u)_{(t,0,t_w)} + 2\,(u_u,n_u)_{(t,0,t_w)} \\
&\quad + (n_u,n_u)_{(t,0,t_w)}\,, \qquad (13)\\
\left(u'_p, u'_p\right)_{(t,0,t_w)} &= (u_p,u_p)_{(t,0,t_w)} + 2\,(u_p,n_p)_{(t,0,t_w)} \\
&\quad + (n_p,n_p)_{(t,0,t_w)}\,. \qquad (14)
\end{aligned}$$

In equations (12)-(14), terms of the form $(u,n)_{(t,t_s,t_w)}$ appear, where $u = u(t)$ is a deterministic signal and $n = n(t)$ a realization of a stochastic process. Using that the noise signals have zero mean, we then find that

$$(u,n)_{(t,t_s,t_w)} \approx 0\,, \tag{15}$$

provided the window length t_w is at least several dominant periods of the noise-contaminated signal. For the terms $(n_u,n_p)_{(t,t_s,t_w)}$ and $(n_p,n_u)_{(t,t_s,t_w)}$ in equations (12)-(14), we use the independence of the noise realizations to find

$$(n_u,n_p)_{(t,t_s,t_w)} \approx 0\,, \tag{16}$$

with the same assumption for the window length t_w as used in equation (15). Approximations (15) and (16) become more accurate with increasing window lengths.

Using equations (15) and (16) in equations (12)-(14), it follows that

$$\begin{aligned}
\left(u'_u, u'_p\right)_{(t,t_s,t_w)} &\approx (u_u,u_p)_{(t,t_s,t_w)}\,, \qquad (17)\\
\left(u'_u, u'_u\right)_{(t,0,t_w)} &\approx (u_u,u_u)_{(t,0,t_w)} \\
&\quad + (n_u,n_u)_{(t,0,t_w)}\,, \qquad (18)\\
\left(u'_p, u'_p\right)_{(t,0,t_w)} &\approx (u_p,u_p)_{(t,0,t_w)} \\
&\quad + (n_p,n_p)_{(t,0,t_w)}\,. \qquad (19)
\end{aligned}$$

Using these approximations in equation (10), and substituting the resulting expression in equation (11), it follows that the correction factor $c(t,t_s,t_w)$ is approximately given by

$$c(t,t_s,t_w) \approx \left(1 - \frac{(n_u,n_u)_{(t,0,t_w)}}{\left(u'_u,u'_u\right)_{(t,0,t_w)}}\right)^{-\frac{1}{2}} \left(1 - \frac{(n_p,n_p)_{(t,0,t_w)}}{\left(u'_p,u'_p\right)_{(t,0,t_w)}}\right)^{-\frac{1}{2}} \tag{20}$$

This correction factor $c(t,t_s,t_w)$ depends on the unknown noise signals $n_u(t)$ and $n_p(t)$. In practice, assuming the noise is stationary, we can estimate these noise signals using the recorded wavefields $u_u(t)$ and $u_p(t)$ before $t = 0$ of the experiment. This gives us estimated noise signals $\tilde{n}_u(t)$ and $\tilde{n}_p(t)$. Replacing $n_u(t)$ and $n_p(t)$ in equation (20) with their estimated counterparts, is a good approximation only if the terms of the form $(u,n)_{(t,t_s,t_w)}$ and $(n_u,n_p)_{(t,t_s,t_w)}$ in equations (12)-(14) are small compared to the remaining terms. Hence, for this replacement to be valid, we want the following inequalities to hold:

$$\begin{aligned}
&|(u_u,n_p)_{(t,t_s,t_w)} + (u_p,n_u)_{(t,t_s,t_w)} + (n_u,n_p)_{(t,t_s,t_w)}| \\
&\qquad \ll |(u_u,u_p)_{(t,t_s,t_w)}|\,, \qquad (21)\\
&\qquad 2|(u_u,n_u)_{(t,0,t_w)}| \\
&\qquad \ll |(u_u,u_u)_{(t,0,t_w)} + (n_u,n_u)_{(t,0,t_w)}|\,, \qquad (22)\\
&\qquad 2|(u_p,n_p)_{(t,0,t_w)}| \\
&\qquad \ll |(u_p,u_p)_{(t,0,t_w)} + (n_p,n_p)_{(t,0,t_w)}|\,. \qquad (23)
\end{aligned}$$

In appendix A we rewrite these three inequalities into one inequality that can be evaluated using only the noise

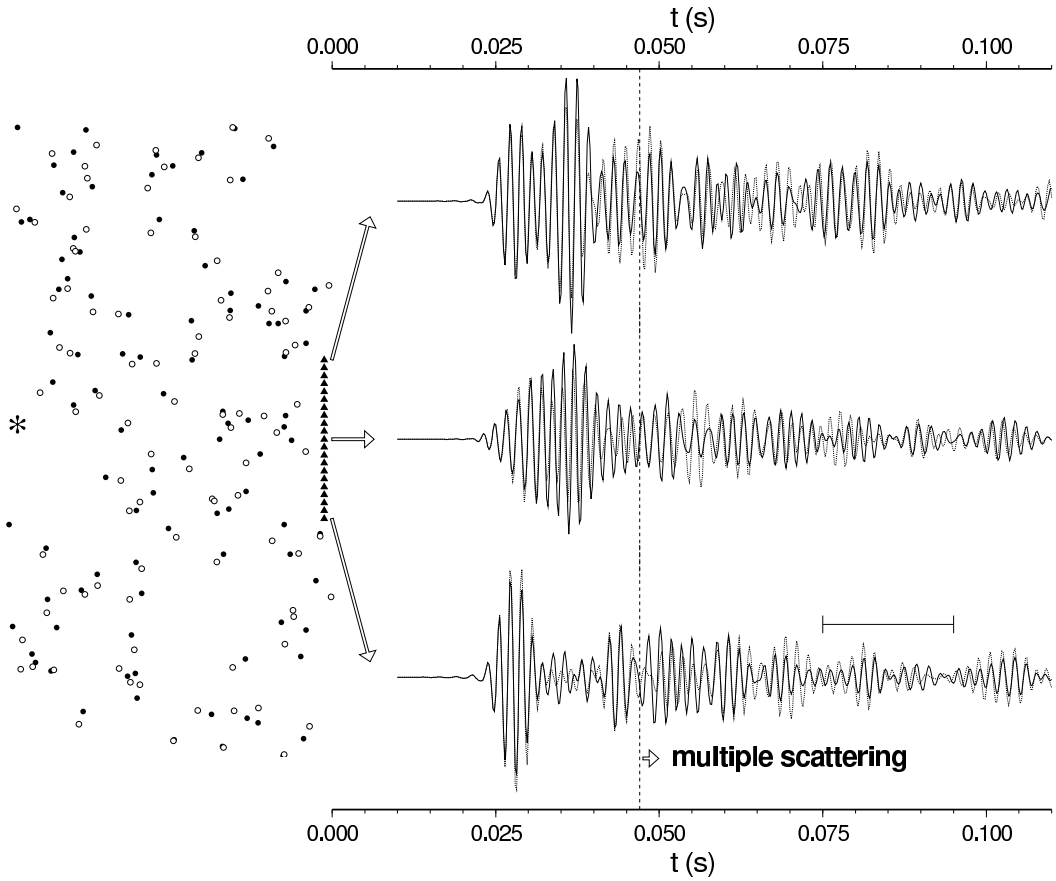

Figure 1. Unperturbed (filled circles) and perturbed (open circles) locations of the point scatterers in the medium. For display purposes the differences between the unperturbed and perturbed locations are magnified by a factor 10. The star denotes the source and the triangles the receivers. The modeled seismograms are shown for three receivers. The drawn seismograms are related to the unperturbed scatterer locations, and the dotted seismograms to the perturbedlocations. For times later than $t = 4.7 \times 10^{-2}$s (marked by the dotted line), the waves have scattered more than four times. The horizontal bar indicates the window length for the time-windowed cross-correlations used to calculate the RMS displacement δ shown in figure 2.

contaminated signals $u'_u(t)$ and $u'_p(t)$, and an estimated noise signal $n_0(t)$. The resulting inequality [equation (A5)] then determines if a certain time window has a reliable correction factor associated with it, when the unknown noise signals $n_u(t)$ and $n_p(t)$ in equation (20) are replaced with $n_0(t)$. Replacing $n_u(t)$ and $n_p(t)$ with $n_0(t)$ in equation (20), the correction factor is given by

$$c(t, t_s, t_w) \approx \left(1 - \frac{(n_0, n_0)_{(t,0,t_w)}}{(u'_u, u'_u)_{(t,0,t_w)}}\right)^{-\frac{1}{2}} \left(1 - \frac{(n_0, n_0)_{(t,0,t_w)}}{(u'_p, u'_p)_{(t,0,t_w)}}\right)^{-\frac{1}{2}} \quad (24)$$

and is considered reliable if inequality (A5) is satisfied with an appropriate value of γ $[O(10^{-1})]$.

To derive condition (A5), we assume that the noise is stationary and that $n_u(t)$ and $n_p(t)$ have about the same noise levels, (i.e., the same variance). If the noise levels of $n_u(t)$ and $n_p(t)$ are substantially different, separate noise estimates of $n_u(t)$ and $n_p(t)$ can be used in equation (20) to calculate the correction factor. In this case condition (A5) does not apply, and a new condition could be derived. To avoid belaboring the point, we refrain from such a treatment.

Displacement of the scatterers

The problem of inferring the average displacement of scatterers in a strongly scattering medium from the multiply scattered wavefields, has been used to study fluidized particle suspensions (Weitz & Pine, 1993; Heckmeier & Maret, 1997; Page *et al.*, 1999; Cowan *et al.*, 2000; Page *et al.*, 2000; Cowan *et al.*, 2002). In geophysics, this problem may be relevant when a strongly scattering region in the earth is strained, causing the scattering heterogeneities to move. In such a situation the displacement of the scatterers is not expected to be random, but will be correlated among scatterers.

Here, we present a numerical experiment with point scatterers in a homogeneous background model, where we randomly perturb the scatterer locations and use CWI to infer their root-mean-square (RMS) displacement. Although this experiment is not directly related to a changing strain in the earth, it serves the purpose of testing the workings of our correction factor.

Snieder & Scales (1998) showed that for independent perturbations of the scatterer positions and isotropic scattering, the variance of the path length L is given by

$$\sigma_L^2 = 2n\delta^2 \,, \tag{25}$$

where n is the number of scatterers in the path, and δ is the RMS displacement of the scatterers in the direction of either coordinate axis (horizontal and vertical for two dimensions). Note that Snieder and Scales assume all directions of random displacement to be equally likely. As a result, the RMS displacement is the same in each direction (horizontal and vertical for 2D), i.e., for 2D the true RMS displacement would be $\sqrt{2}\delta$. Using that the number of scatterers is on average given by $n = vt/l^*$, with l^* the transport mean free path (Lagendijk & van Tiggelen, 1996) and v the velocity, and using $L = vt$, it follows that the variance of the traveltime perturbations is given by

$$\sigma_\tau^2 = \frac{2\delta^2 t}{vl^*} \,. \tag{26}$$

Inserting this in equation (6), it follows that the RMS displacment δ can be found from

$$\delta = \sqrt{\left(1 - r_{\max}^{(t,t_w)}\right) \frac{vl^*}{\overline{\omega^2} t}} \,. \tag{27}$$

Since the perturbations of the scatterer locations are assumed to be independent, and since the scattering is assumed to be isotropic, the mean traveltime perturbation $\langle \tau \rangle = 0$. This means that the maximum of the cross-correlation function $r_{\max}^{(t,t_w)}$ occurs at zero lag, i.e., $t_s = 0$.

Figure 1 shows the setup of our numerical experiment to test the inference of the scatterer displacements from the seismic coda using equation (27). This experiment was also outlined by Snieder *et al.* (2002). One hundred point scatterers (solid dots) are contained in an area of 40×80 m^2, and the waveforms are calculated using a numerical implementation (Groenenboom & Snieder, 1995) of Foldy's method (Foldy, 1945). The resulting seismograms are shown by the solid lines for three locations on the edge of the area, and the source location is indicated by the asterisk. In these calculations the scattering amplitude was set to $-4i$, in order to get the maximum possible scattering strength as constrained by the optical theorem [(Groenenboom & Snieder, 1995); in their notation we used $\gamma = 4$]. The background velocity equalled 1500 m/s and the source spectrum $S(\omega) = e^{-\omega^2/\omega_0^2}$, with ω the angular frequency, $\omega_0 = 2\pi f_0$, and $f_0 = 600$ Hz. The frequency band used was $400 - 800$ Hz, with a resulting dominant frequency of about 500 Hz due to tapering on either side of the spectrum. Since in our experiment we have isotropic scattering, the transport mean free path equals the mean free path, i.e., $l^* = l$. The mean free path in our experiment was measured to be $l = 17.6$ m, which, using $n = vt/l$, can be used to infer that after $t = 4.7 \times 10^{-2}$ s the waves are on average scattered more than four times. This time is indicated in Figure 1 by the dotted vertical line. The perturbed scatterer locations are indicated by the open circles in Figure 1. For display purposes the displacements are magnified by a factor 10. The actual RMS displacement $\delta_{\rm true} = 8 \times 10^{-2}$ m in both the horizontal and vertical direction. *This displacement equals just* 1/38 *of the dominant wavelength* (the dominant wavelength $\lambda = 3$ m). The resulting waveforms calculated using the displaced scatterers are shown for three receivers by the dotted lines.

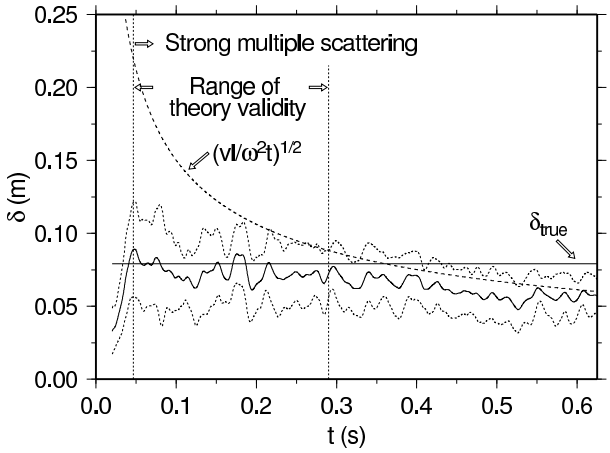

Figure 2. Inferred RMS displacement δ as a function of the center window time t (solid line) plus or minus one standard deviation (dotted lines). The true value of δ is indicated by the horizontal solid line. The range of validity of CWI is indicated by the vertical lines. The half-window time t_w used in calculating the time windowed correlation coefficient equals 0.01s.

Figure 2 shows the inferred value of δ [using equation (27)] as a function of the central window time t, where the estimated values for δ from all 21 receivers were averaged. [The receiver spacing was chosen such that the calculated multiply scattered waveforms were uncorrelated, meaning they can be treated as independent.] The dotted lines show the average inferred value of δ as a function of time plus or minus one standard deviation. The half-window duration t_w was 10^{-2} s, resulting in a window length of 10 dominant periods, and $\sqrt{\overline{\omega^2}} = 3.66 \times 10^3$ rad/s. The vertical dotted lines indicated the range of validity of equation (27). For early times (i.e., $t < 4.7 \times 10^{-2}$ s) the relation $n = vt/l$ for the number of scatterings used in the derivation

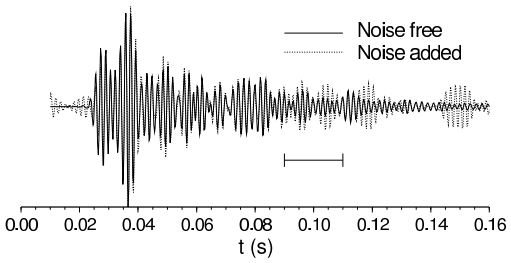

Figure 3. An example signal without noise (drawn) and with noise (dotted). The horizontal bar indicates the window length for the time-windowed cross-correlations used to calculate the RMS displacement δ shown in figure 4.

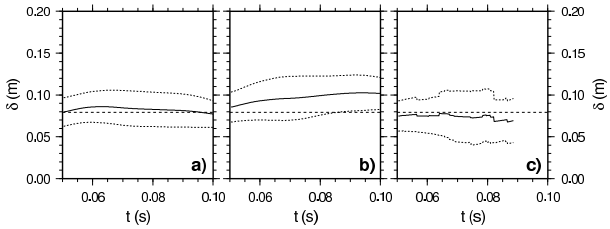

Figure 4. Inferred value of δ as a function of the central window time t, obtained using equation (27) with the noise-free signals (a), with the noise-contaminated signals and no application of the correction factor (b), and with the noise-contaminated signals and application of the correction factor using equation (24) (c).

of equation (27) is not valid, and for late times (i.e., $t > 2.9\times10^{-1}$s), the second-order Taylor approximation of the auto-correlation of the source signal is inaccurate by more than 15%. This latter time is indicated by the rightmost dotted vertical line in Figure 2. Within the range of validity of equation (27), the true displacement is recovered within the range given by the average RMS displacement plus or minus one standard deviation. For late times, the correlation coefficient is close to zero and, according to equation (27), the inferred value of δ is then given by $\sqrt{vl/\left(\overline{\omega^2}t\right)}$. This function is indicated by the dashed line in Figure 2 and indeed agrees well with the inferred value of δ for late times. For these times the inferred value of δ is of course no longer a good estimate of the true RMS displacement.

As mentioned in the introduction, when the wavefields are contaminated with noise, the number of independent time windows that can be used is limited to traveltimes where the ambient noise level is small compared to the amplitudes of the multiply scattered waves. In order to be able to use as many independent time windows as possible, and hence reduce the error bars on the inferred perturbation, it is important to correct the cross-correlation function for the bias due to the noise. To test the correction factor in equation (24), we added band-limited noise to the waveforms for all 21 receivers from our numerical experiment. The bandwidth of the noise was the same as that of the noise-free signals (i.e., 400 − 800 Hz). Figure 3 shows a waveform with and without the added noise. Using the noise-contaminated waveforms, we again calculated the inferred values of δ both with and without the correction factor; see Figure 4c and 4b respectively. For reference, Figure 4a shows the inferred value of δ when the noise-free signals were used. Figure 4b shows that the noise induces a bias in the estimated value of δ; the presence of noise reduces the correlation between the unperturbed and perturbed waveforms, and causes the inferred values of δ, calculated using equation (27), to be larger. For early times, the true value of δ is embedded within the average value of δ plus or minus one standard deviation, but the estimated average value of δ is too high, especially for later times, where the lower amplitude values of the coda result in lower signal-to-noise ratios. Figure 4c shows that the correction factor from equation (24) accurately accounts for the bias due to the noise. The noise estimates were obtained using the signal before the main first arrival in the seismograms. In the calculation of the results shown in Figure 4c, we used condition (A5) for each receiver, with $\gamma = 0.125$ to select the time windows used for the inversion of δ [note that this is a different γ than that used by Groenenboom & Snieder (1995)]. As a result, the number of receivers that had usable time windows for a given central window time t, varies for different t. This causes the jagged appearance of the average inferred value of δ (and the standard deviation). We used a time window only when at least seven receivers (i.e., 30% of the receivers) satisfied condition (A5) at the central window time t. After $t \approx 9 \times 10^{-2}$ s fewer than seven time windows satisfied condition (A5) with $\gamma = 0.125$, and hence the correction factor was judged unreliable. As a result, the noise-corrected estimates of the RMS displacement are not shown for times larger than $t \approx 9 \times 10^{-2}$ s.

Source separation

Snieder & Vrijlandt (2004) used CWI to estimate the distance between seismic events having the same source mechanism, that are recorded at a single station. They derive the imprint of a change in source location on the variance of the traveltime differences, and then use equation (6) to infer this change from the maximum of the cross-correlation function, i.e., $r^{t,t_w}_{\max}$. For two double-couple sources with a source separation in the fault plane, they show that the relation between the source

displacement Δs and the variance of the traveltimes σ_τ^2 is given by

$$\sigma_\tau^2 = \frac{\left(\frac{6}{\alpha^8} + \frac{7}{\beta^8}\right)}{7\left(\frac{2}{\alpha^6} + \frac{3}{\beta^6}\right)}(\Delta s)^2 \,, \tag{28}$$

where α and β are the P and S wave velocities, respectively. Different non-overlapping time-windows provide independent estimates of the source separation. These independent estimates in turn allow for the calculation of error estimates of the source separation.

Figure 5a shows two seismograms (events 242003 and 242020) from earthquakes on the Hayward fault, California (Waldhauser & Ellsworth, 2000), recorded at station CSP of the Northern California Seismic Network. The recorded signal before the arrival of the P wave shows that the noise level is considerable. Figure 5b shows the maximum of the time-windowed cross-correlation function without the correction factor (thin line) and with the correction factor (thick line) applied, and Figure 5c shows the inferred values of the source separation using equation (28). Here the estimate of the noise, i.e., $n_0(t)$, was obtained from the waveforms before the first arrivals. The half-window duration t_w used is 5 s (the full window length is indicated by the horizontal bar in Figure 5a), and the P and S wave velocities used to calculate the source displacement are $\alpha = 5750$ m/s and $\beta = 3320$ m/s, respectively. Note that we used overlapping time windows, since we plot $r_{\max}^{(t,t_w)}$ simply as a continuous function of the central window time t. Of course, non-overlapping windows could be used to ensure independent estimates of the source separation.

Figure 5b shows that the corrected values of $r_{\max}^{(t,t_w)}$ maintain a fairly constant level for times t late in the coda, whereas the uncorrected values decrease earlier in the coda because the noise decreases the similarity between both waveforms. As a result, the inferred values of the source separation using equation (28) are more or less constant for larger traveltimes when the corrected values of $r_{\max}^{(t,t_w)}$ are used (Figure 5c). This indicates that the correction factor $c(t, t_s, t_w)$ given by equation (24), accurately corrects for the influence of the noise on the cross-correlation function. For very large times (say $t > 40$ s) the corrected values are more variable because the correction factor becomes unreliable. We purposely showed the times where the corrected values become variable, to indicate the level of variation caused by an unreliable correction factor. Of course, the time where the correction factor becomes unreliable could have been estimated using condition (A5) with an appropriate value of γ [$O(10^{-1})$].

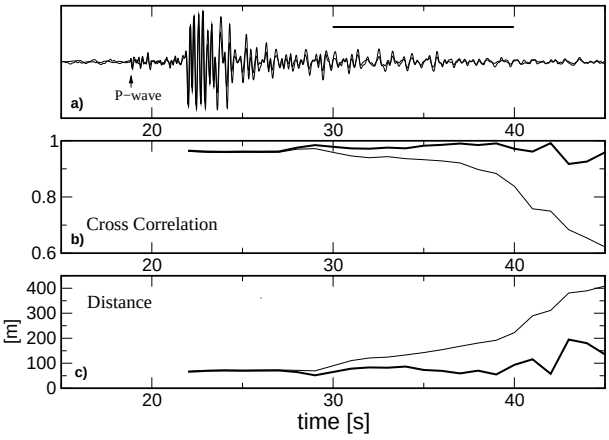

Figure 5. Two seismograms from two earthquakes on the Hayward fault, California, recorded at station CSP of the Northern California Seismic Network (a), their cross-correlation maximum $r_{max}^{(t,t_w)}$ (b), uncorrected (thin line) and corrected (thick line), and the inferred source displacement (c) using both the uncorrected (thin line) and corrected (thick line) values of $r_{max}^{(t,t_w)}$ shown in (b). The horizontal line in (a) indicates the window length used in the cross-correlation.

Conclusion

We have derived a factor that corrects for the influence of noise on the cross-correlation function, and have shown its accuracy using both numerical and field data. The application of this correction factor is shown in the context of CWI, for the inference of the RMS displacement of scatterers and a displacement of the source, from multiply scattered wavefields. For the displacement of the scatterers, we showed that in the presence of noise, a displacement of *only* 1/38 *from the dominant wavelength* can be succesfully retrieved from the cross-correlation between the unperturbed and perturbed signals. This shows the power of CWI when compared to methods that use singly-scattered waves only.

Since for both the scatterer displacement and source separation cases, the perturbation is independent of the traveltimes of the multiply scattered waves, using non-overlapping time windows to estimate the perturbations provides a consistency check of the method, and allows the calculation of error estimates. In this context, our correction factor is relevant, as it increases the number of usable time windows and hence allows for a reduction of the error estimates. In addition the correction factor adjusts for bias in the cross-correlation induced by the noise. Since our factor depends on an estimate of the noise level in the data, we present a condition that allows determination of the reliability of the correction. This condition can be verified using only noise contaminated signals and an estimate of the noise level in the data. Using this condition, the time windows used in the

windowed cross-correlation can be judged to be reliable or not.

The use of the proposed correction factor is of course not limited to CWI. Any application that uses cross-correlations between different and noisy signals, and needs to correct for bias induced by noise, can benefit from the correction factor presented here.

ACKNOWLEDGMENTS

We thank Mark Vrijlandt for making Figure 5, and Ken Larner for his critical review. In addition, we thank Western Geophysical (now WesternGeco), in particular Peter Nuttal, for funding a three-month visit of H.D. to the Center for Wave Phenomena at the Colorado School of Mines in the summer of 2000, that formed the basis for this work. This work was supported by the NSF (grant EAR-0106668) and by the sponsors of the Consortium Project on Seismic Inverse Methods for Complex Structures at the Center for Wave Phenomena.

REFERENCES

Cowan, M.L., Page, J.H., & Weitz, D.A. 2000. Velocity fluctuations in fluidized suspensions probed by ultrasonic correlation spectroscopy. *Phys. Rev. Lett.*, **85**, 453–456.

Cowan, M.L., Jones, I.P., Page, J.H., & Weitz, D.A. 2002. Diffusing Acoustic Wave Spectroscopy. *Phys. Rev. E*, **65**, 066605–1/11.

Derode, A., Roux, P., & Fink, M. 1995. Robust acoustic time reversal with high-order multiple scattering. *Phys. Rev. Lett.*, **75**, 4206–4209.

Derode, A., Tourin, A., & Fink, M. 1999. Ultrasonic pulse compression with one-bit time reversal through multiple scattering. *J. Appl. Phys.*, **85**, 6343–6352.

Fink, M. 1997. Time-reversed Acoustics. *Physics Today*, **50**, 34–40.

Foldy, L.O. 1945. The Multiple scattering of waves. *Phys. Rev.*, **67**, 107–119.

Grêt, A., Snieder, R., Aster, R.C., & Kyle, P.R. 2004a. Monitoring rapid temporal changes in a volcano with coda wave interferometry. *Submitted to Geophys. Res. Lett.*

Grêt, A., Snieder, R., & Ozbay, U. 2004b. Monitoring stress change in an underground mining environment with coda wave interferometry. *Submitted to the J. of Mining Applications.*

Grêt, A., Snieder, R., & Scales, J. 2004c. Time-lapse monitoring of rock properties with coda wave interferometry. *Submitted to J. Geophys. Res.*

Groenenboom, J., & Snieder, R.K. 1995. Attenuation, dispersion, and anisotropy by multiple scattering of transmitted waves through distributions of scatterers. *J. Acoust. Soc. Am.*, **98**(6), 1–11.

Heckmeier, M., & Maret, G. 1997. Dark Speckle Imaging of Colloidal Suspensions in Multiple Light Scattering Media. *Progr. Colloid. Polym. Sci.*, **104**, 12–16.

Lagendijk, A., & van Tiggelen, B.A. 1996. Resonant Multiple Scattering of Light. *Phys. Rep.*, **270**, 143–215.

Maret, G., & Wolf, G.E. 1987. Effect of Brownian Motion of Scatterers. *Z. Phys. B*, **65**, 409–413.

Page, J.H., Jones, I.P., Schriemer, H.P., Cowan, M.L., Sheng, P., & Weitz, D.A. 1999. Diffusive transport of acoustic waves in strongly scattering media. *Physica B*, **263-264**, 37.

Page, J.H., Cowan, M.L., & Weitz, D.A. 2000. Diffusing acoustic wave spectroscopy of fluidized suspensions. *Physica B*, **279**, 130–133.

Papoulis, A. 1991. *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill Inc.

Pine, D.J., Weitz, D.A., Chaikin, P.M., & Herbolzheimer, E. 1988. Diffusing Wave Spectroscopy. *Phys. Rev. Lett*, **60**, 1134–1137.

Snieder, R. 2004a. Coda wave interferometry. *Pages 54–56 of: 2004 McGraw-Hill yearbook of science & technology.* New York: McGraw-Hill.

Snieder, R. 2004b. Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase. *Phys. Rev. E*, **69**, 046610.

Snieder, R., & Vrijlandt, M. 2004. Constraining Relative Source Locations with Coda Wave Interferometry: Theory and Application to Earthquake Doublets in the Hayward Fault, California. *submitted to J. Geophys. Res.*

Snieder, R.K. 1999. Imaging and averaging in complex media. *Pages 405–454 of:* Fouque, J.P. (ed), *Diffuse waves in complex media.* Kluwer, Dordrecht.

Snieder, R.K. 2002. Coda wave interferometry and the equilibration of energy in elastic media. *Phys. Rev. E*, **66**, 046615–1,8.

Snieder, R.K., & Scales, J.A. 1998. Time reversed imaging as a diagnostic of wave and particle chaos. *Phys. Rev. E*, **58**, 5668–5675.

Snieder, R.K., Grêt, A., Douma, H., & Scales, J. 2002. Coda Wave Interferometry for estimating nonlinear behavior in seismic velocity. *Science*, **295**(22 March), 2253–2255.

Waldhauser, F., & Ellsworth, W. 2000. A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, California. *Bull. Seism. Soc. Am.*, **90**, 1353–1368.

Weitz, D.A., & Pine, D.J. 1993. Diffusing Wave Spectroscopy. *Page 652 of:* Brown, W. (ed), *Dynamic Light Scattering.* Clarendon Press, Oxford.

Yodh, A.G., Kaplan, P.D., & Pine, D.J. 1990. Pulsed diffusing-wave spectroscopy: High resolution through nonlinear optical gating. *Phys. Rev. B*, **42**, 4744–4747.

APPENDIX A: A CONDITION TO ESTIMATE THE RELIABILITY OF THE CROSS-CORRELATION CORRECTION FACTOR

Equation (20) for the correction factor of the cross-correlation coefficient, depends on the unknown noise functions $n_u(t)$ and $n_p(t)$. In practice, we don't know these noise functions, and can at best estimate the average noise levels. Since in practice we often only have one estimate of the noise level, as opposed to some ensemble average, we want the correction factor to be reliable when the unknown noise signals in equation (20) are

replaced by a single estimate of the noise. If inequalities (21)-(23) are satisfied, the correction factor will only weakly depend on the estimate of the noise signals. Here we rewrite these inequalities into one inequality that can be verified using the noise contaminated signals and an estimate (or single realization) of the noise $n_0(t)$.

To write inequalities (21)-(23) as a single one, we first add the left- and right-hand sides of equations (21)-(23), while multiplying equation (21) by two for convenience in the further derivation. This gives

$$\begin{aligned} 2\left(|(u_u,n_p)_{(t,t_s,t_w)}+(u_p,n_u)_{(t,t_s,t_w)}+(n_u,n_p)_{(t,t_s,t_w)}|+|(u_u,n_u)_{(t,0,t_w)}|+|(u_p,n_p)_{(t,0,t_w)}|\right) \\ \ll \\ 2|(u_u,u_p)_{(t,t_s,t_w)}|+(u_u,u_u)_{(t,0,t_w)}+(n_u,n_u)_{(t,0,t_w)}+(u_p,u_p)_{(t,0,t_w)}+(n_p,n_p)_{(t,0,t_w)}\,, \end{aligned} \tag{A1}$$

where we have used that zero-lag auto-correlations are positive definite. If conditions (21)-(23) hold, equations (17)-(19) from the main text are good approximations. We can use these approximations, together with the linearity of $(\cdot\,,\cdot)_{(t,t_s,t_w)}$ and equation (9), to approximate inequality (A1) as

$$\begin{aligned} 2|(u'_u,n_p)_{(t,t_s,t_w)}+(u'_p,n_u)_{(t,t_s,t_w)}-(n_u,n_p)_{(t,t_s,t_w)}| \\ +2|(u'_u,n_u)_{(t,0,t_w)}-(n_u,n_u)_{(t,0,t_w)}|+2|(u'_p,n_p)_{(t,0,t_w)}-(n_p,n_p)_{(t,0,t_w)}| \\ \ll \\ 2|(u'_u,u'_p)_{(t,t_s,t_w)}|+(u'_u,u'_u)_{(t,0,t_w)}+(u'_p,u'_p)_{(t,0,t_w)}\,. \end{aligned} \tag{A2}$$

Since in CWI we assume that the travel-time perturbations in the time window $[t-t_w, t+t_w]$ are small, we expect a positive cross-correlation between the signals $u'_u(t)$ and $u'_p(t)$ for times where CWI is valid. Using this, we can write condition (A2) as

$$\begin{aligned} |(u'_u+u'_p,n_0)_{(t,t_s,t_w)}|+|(u'_u,n_0)_{(t,0,t_w)}-(n_0,n_0)_{(t,0,t_w)}|+|(u'_p,n_0)_{(t,0,t_w)}-(n_0,n_0)_{(t,0,t_w)}| \\ \ll 2([u'_u+u'_p]/2,[u'_u+u'_p]/2)_{(t,0,t_w)} \end{aligned} \tag{A3}$$

where we substituted for both noise signals $n_u(t)$ and $n_p(t)$ the estimated noise signal $n_0(t)$, and we assumed $\left|(n_u,n_p)_{(t,t_s,t_w)}\right| \ll \left|(u'_u,n_p)_{(t,t_s,t_w)}+(u'_p,n_u)_{(t,t_s,t_w)}\right|$ to eliminate the $(n_u,n_p)_{(t,t_s,t_w)}$ term. The latter approximation is more appropriate for larger signal-to-noise ratios, larger time windows, and uncorrelated noise realizations $n_u(t)$ and $n_p(t)$. Note that substituting a single noise signal for both unknown noise signals $n_u(t)$ and $n_p(t)$ is appropriate only, if both noise signals have similar noise levels. Dividing both sides of condition (A3) by $(n_0,n_0)_{(t,0,t_w)}$ (which is positive definite), and defining the ratio

$$\Gamma(t,t_w) \equiv \frac{\left([u'_u+u'_p]/2,[u'_u+u'_p]/2\right)_{(t,0,t_w)}}{(n_0,n_0)_{(t,0,t_w)}}\,, \tag{A4}$$

inequality (A3) leads to

$$\frac{1}{2}\left(\left|\frac{(u'_u+u'_p,n_0)_{(t,t_s,t_w)}}{(n_0,n_0)_{(t,0,t_w)}}\right|+\left|\frac{(u'_u,n_0)_{(t,0,t_w)}}{(n_0,n_0)_{(t,0,t_w)}}-1\right|+\left|\frac{(u'_p,n_0)_{(t,0,t_w)}}{(n_0,n_0)_{(t,0,t_w)}}-1\right|\right)/\Gamma(t,t_w) \le \gamma\,, \tag{A5}$$

where γ is $\mathrm{O}(10^{-1})$. Here $\sqrt{\Gamma(t,t_w)}$ can be interpreted as the average signal-to-noise ratio.

Condition (A5) is satisfied only for time windows that have a large average signal-to-noise ratio. The l.h.s of condition (A5) can be evaluated using the noise contaminated signals $u'_u(t)$ and $u_p(t)$, and an estimated noise signal $n_0(t)$. This condition can thus be used as a selection criterion to determine which time windows have reliable correction factors associated with them, when an estimated noise signal $n_0(t)$ is used (i.e., when equation (24) is used to calculate the correction factor). Note that γ can be interpreted as the inverse of the signal-to-noise ratio.